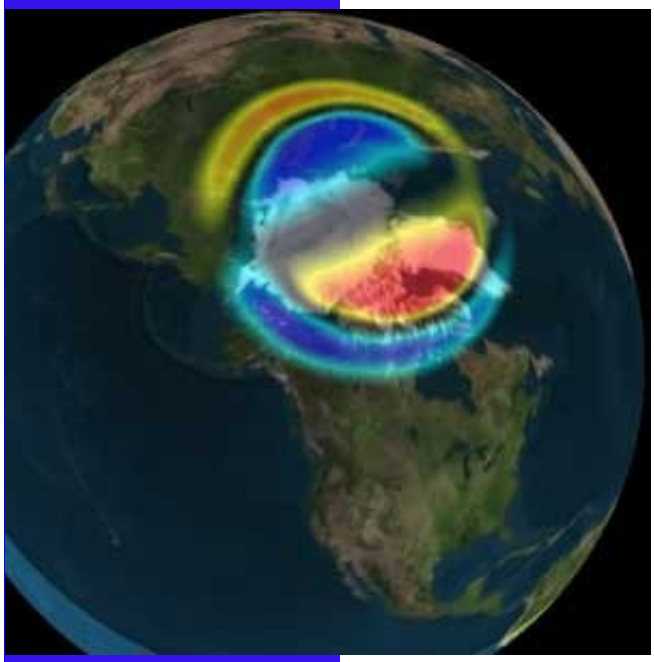
MODELLING FOR ENGINEERING AND HUMAN BEHAVIOUR 2012

Instituto de Matemática Multidisciplinar



BOOK OF EXTENDED ABSTRACTS

L. Jódar, L. Acedo, J. C. Cortés and F. Pedroche, Editors
Instituto Universitario de Matemática Multidisciplinar

im² Instituto de Matemática Multidisciplinar

GENERALITAT VALENCIANA

UNIVERSIDAD POLITECNICA DE VALENCIA

MODELLING FOR ENGINEERING, & HUMAN BEHAVIOUR 2012

Instituto Universitario de Matemática Multidisciplinar
Universidad Politécnica de Valencia
Valencia 46022, SPAIN

Edited by
Lucas Jódar, Luis Acedo, Juan Carlos Cortés and F. Pedroche
Instituto Universitario de Matemática Multidisciplinar
I.S.B.N.: 978-84-695-6701-2

CONTENTS

An implementation of Encke's method for the spacecraft flybys of the Earth

L. Acedo*

Instituto de Matemática Multidisciplinar,

Universitat Politècnica de València,

Edificio 8G, Piso 2, 46022 Valencia, España.

November 30, 2012

1 Introduction

Nowadays, spacecraft missions provide a very stringent test to our understanding of gravitation on the Solar System scale. On the contrary to natural planet and satellites, spacecraft's mass, thermal properties, geometry, etc... are very well-known because of its design. Thanks to the careful monitoring of these missions new effects on spacecraft navigation have been discovered. For example, analysis of the Doppler data for the Pioneer missions to Jupiter and Saturn has revealed an anomalous acceleration $a_{\rm Pioneer} = (8.74 \pm 1.33) \times 10^{-10}$ m/s^2 directed towards the Sun [1]. Many possible conventional and unconventional explanations have been proposed, and for a long time there was no no consensus on the origin of this phenomenon [2, 3]. However, a very recent study of the whole dataset for the Pioneer 10 and Pioneer 11 orbits strongly suggested that this sunward acceleration is the consequence of anisotropic emission of on-board heat [4] and this has been finally proven beyond any doubt [5].

Another anomalous behaviour of spacecraft have been reported recently by Anderson et al. These authors have analyzed the data for six Earth flybys of five deep-space missions[6]: Galileo, NEAR, Cassini, Rosetta and Messenger that took place between December 1990 and September 2005. Flybys are a common maneuver in spacecraft missions which allows the spacecraft to gain or lose of heliocentric energy with the purpose of reaching their objective [7]. An analysis of the data for these flybys have shown X-band Doppler residuals that are interpreted in terms of a change of the hyperbolic excess velocity, V_∞, of a few mm/s. Anderson et al. have proposed the phenomenological formula:

$$\frac{\Delta V_\infty}{V_\infty} = K(\cos\delta_i - \cos\delta_o) \ , \tag{1}$$

*e-mail: luiacrod@imm.upv.es

where δ_i, δ_o are the declinations for the incoming and outgoing osculating velocity vectors and K is a constant. The value of K seems to be close to $2\omega_E R_E/c$, where ω_E is the angular rotational velocity of the Earth, R_E is the Earth radius and c is the speed of light. Although this formula works reasonably well for the six flybys studied in the paper, the proposal for the relation of K with the Earth's tangential velocity at the Equator is a daring hypothesis, taking into account that the flybys of other planets with different rotational velocities and radius have not been considered.

A revival of interest in celestial mechanics is the consequence of this important advances in high-precision astrodynamics. Celestial mechanics has been an area of active research since the times of Newton and in the modern era of spacecraft navigation it continues to be of paramount interest for both physicists, mathematicians and engineers working in this interdisciplinary field. Trying to unveil the nature of these anomalies is quite challenging for several reasons: it is not known whether these come from standard physics effects (as the thermodynamic origin of the Pioneer anomaly) or there is some unknown underlying physical phenomena beyond standard physics that requires the formulationi of a new theoretical paradigm.

However, an important lesson to be learn is that observational, computational, engineering and physical sources of error should be studied very carefully before declaring that new physics is necessary to account for these effects. In this spirit we propose an implementation of Encke's method to the calculation of flyby orbits that could be used in orbit determination an analysis. In Sec. 2 we develop of the lunar-solar perturbations of the flyby trajectory. Also we consider the displacement of the Sun and the Moon in the sky during the duration of the flyby. Conclusions are given in 3.

2 Orbit parameters and perturbation theory

Positions of the planets, the Sun and the spacecraft on the sky are usually expressed in terms of the declination angle, δ, which is the angle of the line of sight of the object with the equatorial celestial plane. Similarly, the right ascension angle, α, is the angle between the projection of the position vector of the object upon the celestial equator and the first point of Aries (the point where the Sun crosses the Celestial equator at the Vernal equinox). In the following we will use the celestial polar angle $\theta = \pi/2 - \delta$ instead of the declination.

In order to analyze the flyby orbit and its subsequent perturbations, it is highly convenient to define a system of coordinates anchored to that orbit. As such a system we choose a unit vector along the periapsis direction corresponding to the point of closest approach, $\hat{s}$, a second unit vector pointing along the direction of the inclination vector of the orbit, $\hat{w}$, and a third one perpendicular to those two, $\hat{n}$. This third unit vector is defined in such a way that the scalar product with the initial radiovector of the spacecraft, $r_{\rm in}$, is positive.

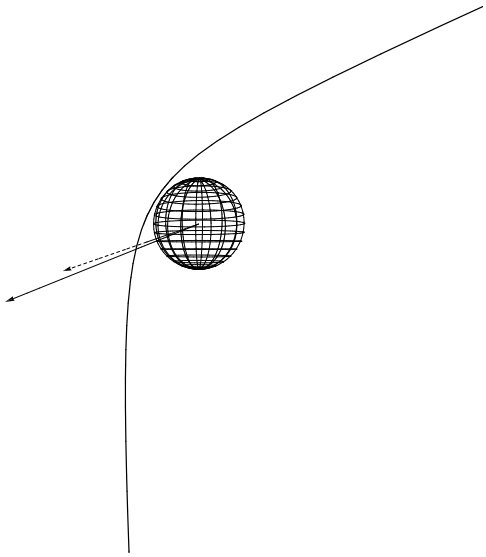


Figure 1: Plot of the NEAR flyby orbit (January 23, 1998). The solid vector points towards the Sun, the dashed vector points towards the location of the Moon at the instant of the closest approach.

These vectors are given as follows:

$$\hat{s} = \cos\theta_p\ \hat{k} + \sin\theta_p \cos\alpha_p\ \hat{\imath} + \sin\theta_p \sin\alpha_p\ \hat{\jmath} \tag{2}$$

$$\hat{w} = \cos I\ \hat{k} + \sin I \cos\alpha_I\ \hat{\imath} + \sin I \sin\alpha_I\ \hat{\jmath} \tag{3}$$

$$\hat{n} = \pm\hat{w} \times \hat{s}\ , \tag{4}$$

where θ_p, α_p are the celestial polar angle and right ascension of the periapsis, I and α_I are the inclination and the right ascension of the inclination vector and the sign in the last expression for $\hat{n}$ depends on the orientation of the orbit. The orthogonal system $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$ is, obviously, the celestial coordinate system.

The NEAR flyby orbit (January 23, 1998) is plotted in Fig. 1. In this case the parameters were (all angles in degrees): $\theta_p = 57$, $\alpha_p = 280.43$, $I = 108.0$, $\alpha_I = \alpha_p + \arccos(-\cot I \cot\theta_p) = 358.24$. The incoming direction is given by $\theta_i = 69.24$ and $\alpha_i = 81.17$. In this particular case, it can be shown that $\hat{n} = \hat{w} \times \hat{s}$. Instead of using time or the true anomaly (the angle formed by the radiovector of the spacecraft and the periapsis vector) to parametrize the orbit, we can use, more conveniently, the eccentric anomaly defined as follows:

$$\cosh H = \frac{\epsilon + \cos\nu}{1 + \epsilon\cos\nu}\ , \tag{5}$$

ν being the true anomaly and $\epsilon > 1$ the eccentricity of the hyperbolic orbit. The time of flight can be given in terms of the eccentric anomaly by

$$t = T(\epsilon \sinh H - H)\ , \tag{6}$$

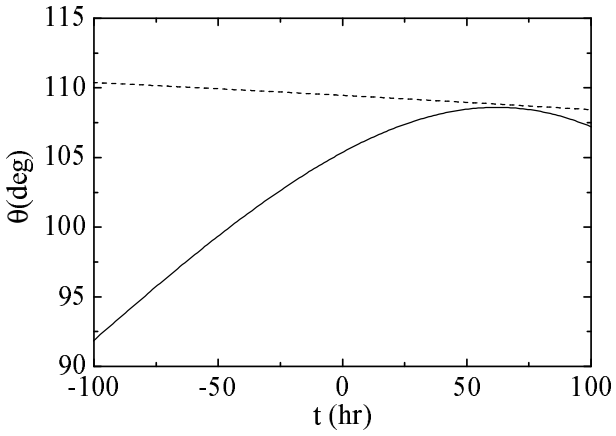


Figure 2: Polar celestial angle of the Sun (dotted line) and the Moon (solid line) before and after 100 hours of the closest approach to Earth of the NEAR spacecraft (January 23, 1998).

where the time-scale $T = \sqrt{(-a)^3/\mu}$, a is the semi-major axis of the orbit and μ is the product of the gravitational constant and the mass of the Earth, $\mu = 398600.4\,\mathrm{km}^3/\mathrm{s}^2$. The equations for the radiovector and the velocity of the spacecraft in the ideal hyperbolic orbit are then given by

$$\mathbf{r}(H) = a(\cosh H - \epsilon)\hat{s} - a\sqrt{\epsilon^2 - 1}\sinh H\hat{w} \tag{7}$$

$$\mathbf{v}(H) = \frac{a}{T(\epsilon\cosh H - 1)}(\sinh H\hat{s} - \sqrt{\epsilon^2 - 1}\cosh H\hat{w})\ , \tag{8}$$

In order to determine the perturbations of the orbit of the spacecraft in the geocentrical system of reference caused by other bodies (Sun or Moon) the tidal force generated by the difference of forces exerted upon the Earth and the spacecraft must be calculated. In general, this tidal force is given as follows:

$$\mathbf{F}_{\mathrm{tidal}} = -\frac{\mathbf{R}}{R^3} + \frac{\mathbf{R} - \mathbf{r}}{(r^2 + R^2 - 2\mathbf{r}\cdot\mathbf{R})^{3/2}}\ , \tag{9}$$

where $\mathbf{R}$ is the radiovector from the center of the Earth towards the perturbing body and R its modulus. We must take into account that $\mathbf{R}$ has a significant change during the flyby maneuver which it is considered to last about 200 hours. In Figs. 2 and 3 we have plotted the variation of the polar celestial angle and the right ascension for the Sun and the Moon during the time span of the NEAR flyby. The unit radiovectors of the Sun and the Moon in the geocentrical coordinate system corresponding to the spacecraft orbit are defined as $\hat{R}_S = \alpha(H)\hat{s} + \beta(H)\hat{w} + \gamma(H)\hat{n}$ and $\hat{R}_M = \eta(H)\hat{s} + \chi(H)\hat{w} + \kappa(H)\hat{n}$. From Eqs. (7) and (9) we can now give the components of the tidal force generated by the Sun in the

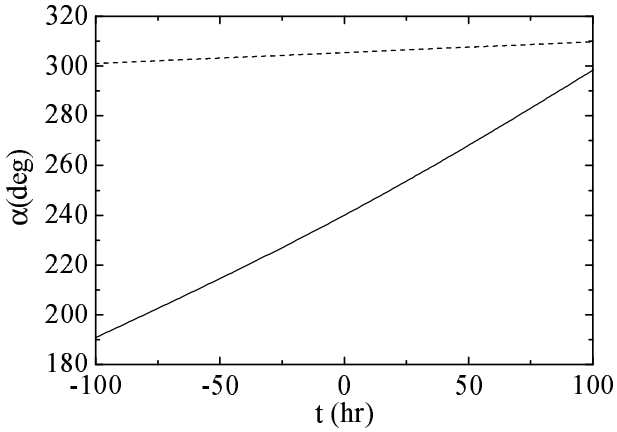


Figure 3: The same as Fig. 2 but for the right ascension of the Sun (dotted line) and the Moon (solid line).

orthogonal system of reference $\hat{s}$, $\hat{w}$, $\hat{n}$ as follows

$$\begin{aligned}
\mathbf{F}_S(H) &= \mu_S \left(-\frac{\alpha(H)}{R_S(H)^2} + \frac{a(\epsilon - \cosh H) + \alpha(H) R_S(H)}{\rho(H)^{3/2}} \right) \hat{s} \\
&+ \mu_S \left(-\frac{\beta(H)}{R_S(H)^2} + \frac{a(\sqrt{\epsilon^2 - 1} \sinh H + \beta(H) R_S(H)}{\rho(H)^{3/2}} \right) \hat{w} \\
&+ \mu_S \left(-\frac{\gamma(H)}{R_S(H)^2} + \frac{\gamma(H) R_S(H)}{\rho(H)^{3/2}} \right) \hat{n} \, ,
\end{aligned} \tag{10}$$

where $R_S(H)$ is the distance between the Sun and the Earth, $\mu_S = 1.3271244 \times 10^{11}$ km^3/s^2 is the mass of the Sun times the gravitational constant and $\rho(H)$ is the square of the distance from the spacecraft to the Sun:

$$\rho(H) = a^2(\epsilon \cosh H - 1)^2 + R_S(H)^2 - 2a\alpha(H) R_S(H)(\cosh H - \epsilon) + 2a\sqrt{\epsilon^2 - 1}\beta(H) R_S(H) \sinh H \, . \tag{11}$$

Once the osculating orbit parameters (the ideal hyperbolic orbit corresponding to the velocity and position at the periapsis) are known, the perturbation tidal force generated by the Sun gravitational field is only a function of the eccentric anomaly, H. A similar expression can be written for the tidal force exerted by the Moon $\mathbf{F}_M(H)$ (in this case, $\mu_M = 4902.8$ km^3/s^2).

Using the relation among the time of flight, t, and the eccentric anomaly, H, in Eq. (6) we can now write the perturbations in the velocity and position of the spacecraft as integrals over H as follows:

$$\Delta\mathbf{v}(H) = T \int_0^H du(\epsilon \cosh u - 1) \left(\mathbf{F}_S(u) + \mathbf{F}_M(u) \right) \tag{12}$$

$$\Delta\mathbf{r}(H) = T^2 \int_0^H du(\epsilon \cosh u - 1) \int_0^u dv(\epsilon \cosh v - 1) \left(\mathbf{F}_S(v) + \mathbf{F}_M(v) \right) \, . \tag{13}$$

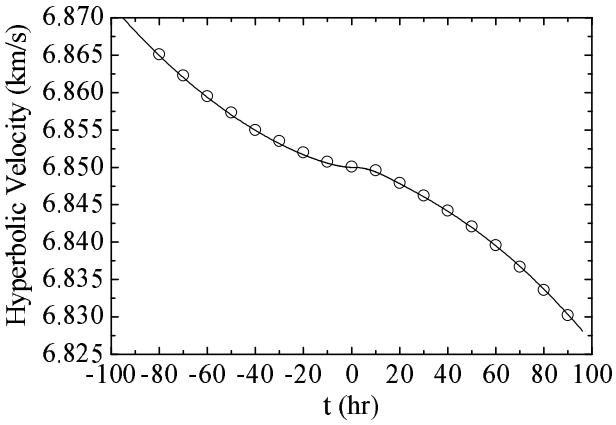


Figure 4: The prediction for the osculating hyperbolic asymptotic velocity (solid line) compared with observations for the NEAR flyby (circles) as a function of time. The instant of time corresponding to the closest approach was taken as $t = 0$.

In the case of the NEAR flyby we have an orbital eccentricity $\epsilon = 1.81352$, the minimum altitude over the Earth geoid at periapsis is $h_p = 539$ km which, from Eq. (7), implies a semi-major axis $a = (h_p + R_E)/(1 - \epsilon) = -8494.97$ km, where $R_E = 6371$ km is the mean radius of Earth. Consequently, the time scale, T, appearing in Eq. (6) is $T = 1240.13$ sec.

Finally, we can measure the effect of the perturbation in terms of the asymptotic hyperbolic velocity of the osculating orbit at every instant, t. The asymptotic hyperbolic velocity of the osculating orbit at every point of the real trajectory is given by

$$V_\infty^2(H) = |\mathbf{v}(H) + \Delta\mathbf{v}(H)|^2 - 2\mu_E / |\mathbf{r}(H) + \Delta\mathbf{r}(H)| \ , \tag{14}$$

where $|\ldots|$ denote the vector modulus. In Fig. (4) we compare the prediction of Eq. (14) for the hyperbolic velocity at a function of time by performing the integrals in Eq. (12) numerically with the observational results [6]. The agreement is very good and this proves the effectiveness of the perturbation approach in its first order approximation.

3 Conclusions

Flyby trajectories are conventionally used in astrodynamics as a way to gain or lose energy in order to reach far distant objectives. This is the so-called slingshot maneuver [7]

In this paper we have pursued a conventional model for slingshot trajectories based upon a detailed implementation of classical perturbation theory. The nature of flyby maneuvers require the monitoring of trajectories from the Earth and, consequently, they are naturally followed in a geocentrical coordinate system. From the point of view of this system, any other celestial body will be the source of a tidal force due to the different positions of the Earth and the spacecraft relative to the third body. Moreover, celestial bodies indeed change their positions in the sky during the flyby maneuver which lasts several days and

this change must be taken into account. The most important contributions to perturbations in the case of Earth flybys come from the Sun and the Moon.

We have checked that this numerical method provide a very precise account of the perturbed trajectory and could be used in studies of high-precision astrodynamics as those unveiling the recently discovered anomalies. Work along this line is in progress and will be published elsewhere.

Acknowledgements

The author gratefully acknowledges R. M. Shoucri for many useful discussions and a critical reading of the manuscript. The NASA's Jet Propulsion Laboratory is also acknowledged for their HORIZONs web-based system which was used to compute the ephemeris used in this work.

References

[1] J. D. Anderson, P. A. Laing, E. L. Lau, A. S. Liu, M. M. Nieto and S. G. Turyshev, Study of the anomalous acceleration of Pioneer 10 and 11, Phys. Rev. D **65**, 082004 (2002).

[2] C. Lämmerzahl, O. Preuss and H. Dittus, Is the physics of the Solar system really understood ?, arXiv: gr-qc/0604052.

[3] S. G. Turyshev and V. T. Toth, The Pioneer Anomaly, Living Rev. Relativity **13** 4 (2010), http://www.livingreviews.org/lrr-2010-4.

[4] S. G. Turyshev, V. T. Toth, J. Ellis and C. B. Markwardt, Support for temporally varying behavior of the Pioneer anomaly from the extended Pioneer 10 and 11 Doppler data sets, Phys. Rev. Lett. **107** 081103 (2011).

[5] S. G. Turyshev, V. T. Toth, G. Kinsella, S.-C. Lee, S. M. Lok and J. Ellis, Support for the thermal origin of the Pioneer anomaly, Phys. Rev. Lett. **108** 241101 (2012).

[6] J. D. Anderson, J. K. Campbell, J. E. Ekelund, J. Ellis and J. F. Jordan, Anomalous Orbital-Energy Changes Observed during Spacecraft Flybys of Earth, Phys. Rev. Lett. **100**, 091102 (2008).

[7] P. H. Borcherds and G. P. McCauley, The gravitational three-body problem: optimising the slingshot, Eur. J. Phys. **15** (1994), pp. 162-129.

Modelling the academic performance in Spanish high school: an epidemiological approach with uncertainty

R. Cervelló-Royo [*], J.-C. Cortés[†], J.-A. Moraño[†], A. Sánchez-Sánchez[†] and J. Villanueva-Oller[‡]

(*) Department of Economics and Social Sciences

Universitat Politècnica de València, Spain

(†) Instituto Universitario de Matemática Multidisciplinar

Universitat Politècnica de València, Spain

(‡) Centro de Estudios Superiores Felipe II

Aranjuez, Madrid, Spain

November 30, 2012

1 Introduction

In this paper and focusing on Spain, we propose a gender–and–course–structured model where we consider the spread of good/bad academic habits (positive/negative transmission) between students belonging to the promotable and non–promotable group in the academic level of *Bachillerato* (16–18- year-old students). Notice that this approach is based on pedagogical strategies that consider mixing groups of students with bad and good academic results in order to induce improvement of them. Furthermore, we include the estimation of the abandon rates. Abandon is an important aspect still under debate in the pedagogical area which quantification is difficult. We have made a decision in order to include this issue in the model and this is to

[*]e-mail: rocerro@esp.upv.es

consider *abandon* when, during the academic year, the student leaves the academic system.

Once the model is stated, we will be able to monitor both, the promoted and non-promoted, and graduated students. Other new contribution is the introduction of uncertainty in the obtention of the value of the parameters of our model which will allow us to predict the evolution of the academic performance in specific confidence intervals.

The proposed approach will allow us to understand better the mechanism behind the academic performance as well as to predict how things will evolve in the Spanish *Bachillerato* over the next few years. This permits to provide relevant information to make appropriate decisions to policymakers.

2 Available data

The available data that we have considered in this paper correspond to the academic results belonging to the students of the First and Second Stage of *Bachillerato* during the academic years from $1999-2000$ to $2008-2009$, in both, state and private high schools all over Spain (see Table 1).

	First Stage (Girls | Boys)		Second Stage (Girls | Boys)	
Academic year	% Promote $(G_1 \mid B_1)$	% Non–Promote $(\overline{G}_1 \mid \overline{B}_1)$	% Promote $(G_2 \mid B_2)$	% Non–Promote $(\overline{G}_2 \mid \overline{B}_2)$
1999–2000	19.68 | 15.24	9.75 | 9.33	16.21 | 11.64	9.52 | 8.63
2000–2001	22.65 | 17.54	9.91 | 10.12	14.07 | 10.04	8.24 | 7.43
2001–2002	19.23 | 14.23	8.61 | 9.10	17.86 | 13.06	9.32 | 8.59
2002–2003	18.87 | 14.19	8.36 | 8.51	19.14 | 13.97	8.76 | 8.20
2003–2004	19.93 | 15.06	7.74 | 7.88	19.19 | 13.80	8.44 | 7.96
2004–2005	20.11 | 15.14	7.65 | 7.94	18.90 | 13.92	8.39 | 7.95
2005–2006	20.07 | 15.39	7.64 | 7.93	19.14 | 13.97	8.08 | 7.78
2006–2007	20.06 | 15.34	7.67 | 7.87	19.14 | 14.29	7.98 | 7.65
2007–2008	20.25 | 15.82	7.57 | 7.66	19.37 | 14.61	7.60 | 7.12
2008–2009	20.72 | 16.57	7.28 | 7.43	19.43 | 14.86	7.05 | 6.66

Table 1: The available data corresponding to the First and Second Stage of Bachillerato, in both, state and private high schools all over Spain from academic year $1999-2000$ to $2008-2009$. Each row shows the percentage of Girls and Boys who promote $(G_i \mid B_i)$ and do not promote $(\overline{G}_i \mid \overline{B}_i)$ for each level $i = 1, 2$ over the total Spanish *Bachillerato* students.

3 Model building and Predictions

We build our mathematical model based on a non-linear system of ordinary differential equations by following an epidemiological approach. We consider as main idea that the academic performance of a student, Girl (G) or Boy (B), is a mixture of her/his own study habits and the study habits, good or bad, of their classmates. In our model, we assume that the transmission of good and bad academic habits is caused by the social contact between students who belong to the same academic level [1, 2, 3, 4].

The subpopulations of the model will be (time t in years):

- $G_i = G_i(t)$ is the number of girls who are in condition to promote at time instant t, for $i = 1, 2$.
- $B_i = B_i(t)$ is the number of boys who are in condition to promote at time instant t, for $i = 1, 2$.
- $\overline{G}_i = \overline{G}_i(t)$ is the number of girls who are not in condition to promote at time instant t, for $i = 1, 2$.
- $\overline{B}_i = \overline{B}_i(t)$ is the number of boys who are not in condition to promote at time instant t, for $i = 1, 2$.

The flow diagram of the model is shown in Figure 1.

Then we proceed to estimate the parameters of model. This task has been performed by fitting the scaled model in the mean square sense to the available data collected in Table 1. Computations have been carried out with *Mathematica 8.0* [5]. The model has been built using *epiModel* software [5]. *epiModel* facilitates the implementation all the equations in *Mathematica* saving developing time. Then, with a simple *Mathematica* algorithm, the model built by *epiModel* is scaled.

The system of differential equations, in its scaled version, is numerically solved by taking as initial conditions the data of the academic year 1999-2000 (corresponding to $t = 0$).

Uncertainty is a key part of the real world and it should be considered in modelling, to be precise, in the data and the model parameters. Therefore, the assumption that parameters always are constant or the parameter estimation does not contain errors is not appropriate. Thus, it is natural to

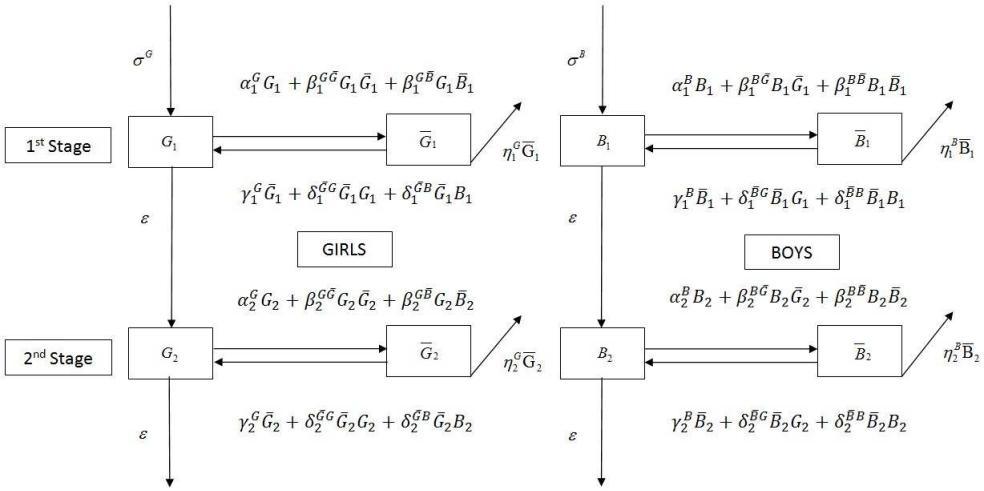


Figure 1: The flow diagraman of the model.

consider that the model parameters contain uncertainties. Hence, the deterministic prediction can give us an idea about the future trends but the obtained values may not be as accurate as expected.

Thus, we propose forecasting future evolutions using confidence intervals. In order to calculate these confidence intervals, let us use the technique called bootstrapping. Bootstrapping is a sophisticated and efficient method for determining a non–parametric probabilistic estimation of model parameters.

The obtained results are plotted in Figure 2.

4 Conclusion

In this paper we propose a dynamical model to study the students' academic performance in high school in Spain, taking into account sex, stages and academic results. The main idea behind our approach is to consider that academic performance depends on both student own study and their classmates' habits. The abandon is a crucial issue to analyse school failure that has also been contemplated in the model. To make more realistic our approach, we have considered uncertainty in the study. This fact allows us to predict the students' academic performance in the next few years through confidence intervals. The model predictions for $2009-2010$ have been compared with the recently available data with good predictions for promotable

student groups.

The results tell us that there is a slight decreasing of the number of students in the non–promotable groups and who leave the high school, and it seems to reach a stationary situation. The current and predicted scenarios are very worrying because around 30% of the students have bad academic results.

Finally, we would like to say that the modelling technique presented in this paper can be used in other educational levels in any region or country.

References

[1] L.S. Vygotsky, Mind in Society: The Development of Higher Mental Processes, Harvard University Press, Cambridge, 1978. ISBN-10: 0674576292.

[2] M. Lucci, La propuesta de Vygotsky: La psicología socio-histórica, Revista de Currículum y Formación del Profesorado. 10(2) (2006) 7-11. [The proposal of Vygotsky: socio-historical psychology]. (In Spanish). Available at `http://www.ugr.es/~recfpro/rev102COL2.pdf` (Accessed on April 24, 2012).

[3] N.A. Christakis, J.H. Fowler, Connected: The Surprising Power Of Our Social Networks And How They Shape Our Lives, Hachette Book Group, Brown and Company, 2009. ISBN-10:0316036145.

[4] R. Wentzel Kathryn, D. Watkins, Peer relationships and collaborative learning as contexts for academic enablers, School Psychology Review. 31(3) (2002) 366-377. Available at `http://www.bibsonomy.org/bibtex/2402c8d2eb020c0aa3f46f9f9ca7d69cb/clachapelle` (Accessed on April 24, 2012).

[5] `http://www.wolfram.com/products/mathematica.` (Accessed on April 24, 2012).

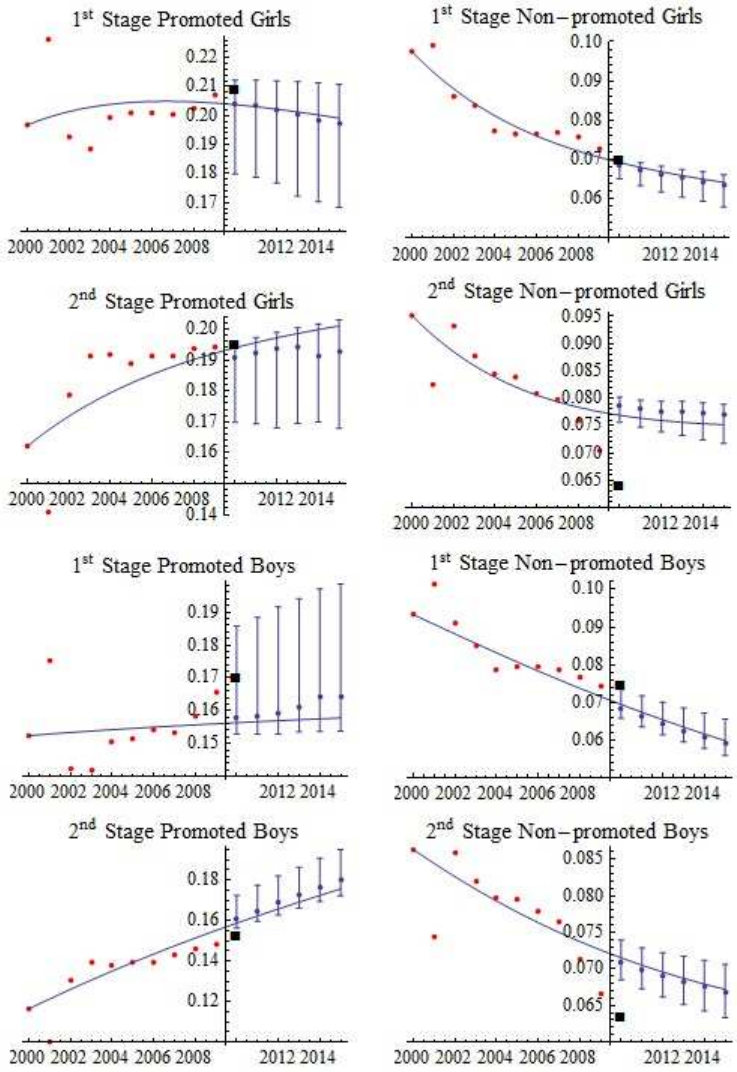


Figure 2: Real data (red points on the left side of vertical axis) and prediction (line) with confidence intervals (on the right side of vertical axis) of the academic performance of *Bachillerato* Spanish students over the academic years $1999-2000$ to $2014-2015$. The points in the middle of the confidence intervals are their medians. The square black point represents the last academic results published recently corresponding to the academic year $2009-2010$. A good prediction is done if the black point lies inside the confidence interval. Notice that each graph has its own scale.

Adaptive mapping routes of pipes in water supply systems using GPR and multi-agent approach *

D. Ayala–Cabrera[‡] †, M. Herrera[⋆], J. Izquierdo[‡] and R. Pérez–García[‡]

(‡) FluIng-IMM, Universitat Politècnica de València,

C. de Vera s/n, Edif 5C, 46022 Valencia, Spain,

(⋆) BATir - Université Libre de Bruxelles,

Av. F. Roosvelt, 50 CP 194/2 B-1050 Bruxelles, Belgique.

November 30, 2012

1 Introduction

Records of components, layouts and characteristics exhibit great inaccuracy or are even non-existent in many cases in water supply systems (WSS) companies. To cope with this situation, street surveys are usually undertaken through road excavation. This type of exploration involves high economic and social impact. As a consequence, growing interest for non-destructive methods in exploration of WSS components, instead of other destructive testing methods, is currently observed. However, even though information retrieval by non-destructive methods is worthwhile, the complexity of spatial

*This work has been supported by project IDAWAS, DPI2009-11591, of the Dirección General de Investigación of the Ministerio de Ciencia e Innovación of Spain, ACOMP/2011/188 of the Conselleria de Educación of the Generalitat Valenciana, and the FPI-UPV scholarship granted to the first author by the Programa de Ayudas de Investigación y Desarrollo (PAID) of the Universitat Politècnica de València.

†e-mail: daaycab@upv.es

layouts of WSS networks, along with the steady growth of cities, the huge volume of generated information, and the interpretation of data, usually require high levels of skill and experience. In this work, statistical methodologies are used to generate adaptive mapping routes in street surveys using ground penetrating radar (GPR). Mapping routes are based on the analysis of GPR images along with a multi-agent generation of pseudo-random walks, and a process to discard areas with less probability of pipe existence. This is an iterative procedure that we have integrated into a system that produces GPR sampling walks, and eventually set up a reliable location map of buried pipes. The purpose of this work is to improve the usual exhaustive scanning systems with GPR. As a result the survey time is optimized and the amount of data needed to conduct records of the components of WSS is minimized.

2 Proposed system

The proposed system for adaptive mapping routes is based on an iterative analysis process aimed at generate pipe layouts in WSS using GPR as a sensor. The proposed algorithm has been developed in MatLab and the activities involved are carried out within a multi-agent based system. These multi-agent activities are explained next.

1. *Image acquisition.* In this activity, agents work under conditions of completely unknown pipe layout. Firstly, the interest square area for the agents to walk and inspect is selected, and some features, namely origin, axes and coordinate system, established. Next, we use the Latin square sampling (LSS) technique to explore ground variation sources by means of randomly assigned treatments; this begins the process of finding the buried pipe. The LSS provides the proposed system with the initial prospection criterion for the agents to start their walks. The interest area is taken as an mxm matrix (M). M is filled by using m different symbols, each occurring exactly once in each row and exactly once in each column. Then, the symbol is randomly selected and thereby m cells for the test are chosen (in this paper m=4). Once the m cells have been chosen, one profile for every cell is randomly selected. The profile domain is {NS, SN, EW, WE}, and the starting location is the center of each cell. So, m profiles of tests are selected to be captured with the GPR in the interest area.

2. *Analysis and interpretation.* In the second agent activity, the pipe existence plausibility in the analyzed image is evaluated. This activity is composed of two stages. The first stage is used to improve the image visualization and to reduce the amount of data to handle. The underlying rationale behind this stage consists in cleaning zones where the presence of a pipe is less likely in the obtained image [1]. The cleaning of the zones in this stage considers that non-horizontal variations of the wave amplitude value correspond to 'no-pipe existence'. To this purpose, first, the obtained profiles are transformed to $T14$ and $T15$ images [2]. Then, the Hough transform, a segmentation technique, is used to detect and remove the horizontal lines from both images. These images are then merged; thereby a new image is generated. For this new image the detection and cleaning of horizontal lines is applied with the aim to minimize the noise in the final image. The second stage for this activity is used to interpret the final image in the last stage, and establish the plausibility of this image belonging to the pipe layout location sought. The interpretation was performed using images with and without pipes. Using 200 images in total, we established criteria about the pipe existence plausibility. Prior to any analysis, we flatten the matrix associated to the image onto a vector by row concatenation. Then, an autoregressive model of order 4 is applied to the vector, and it can be observed that the pipe existence in the image substantially increases the value of the model coefficients.

3. *Mapping of the route.* The n-th agent walk is performed in this activity. It takes into consideration: first, the directions in the walk (X_n) with domain {N, S, E, W}, and second, the 'plausibility' of pipe layout existence (C_n) with domain {'red', 'yellow', 'green'}. The general rules for the agent movements are: a) no repeat direction and orientation in the walks, b) exclude directions to zone with 'red' indicator, c) decrease walk probability to 'yellow' zones, and d) favor direction to 'green' zones. We have to mention that the initial agent walk is provided by the LSS profiles obtained before. In the adaptive mapping route, the profile with greater pipe layout plausibly existence will be selected as start point. Additionally, the agents criteria for indexing zones are: 1) 'red' (without pipe), when 3 coefficient values in the autoregressive model are not significantly different from zero, 2) 'yellow' (low plausible), if the considerations are not conclusive about the pipe existence, and 3)

'green', when the pipe existence is highly plausible.

3 Experimental study

The case-study corresponds to the urban area (see, Figure 1,a). After exploration with GPR through the LSS profiles and seven multi-agent additional walks, we eventually obtain the pipe layout in the interest area (Figure 1,b).

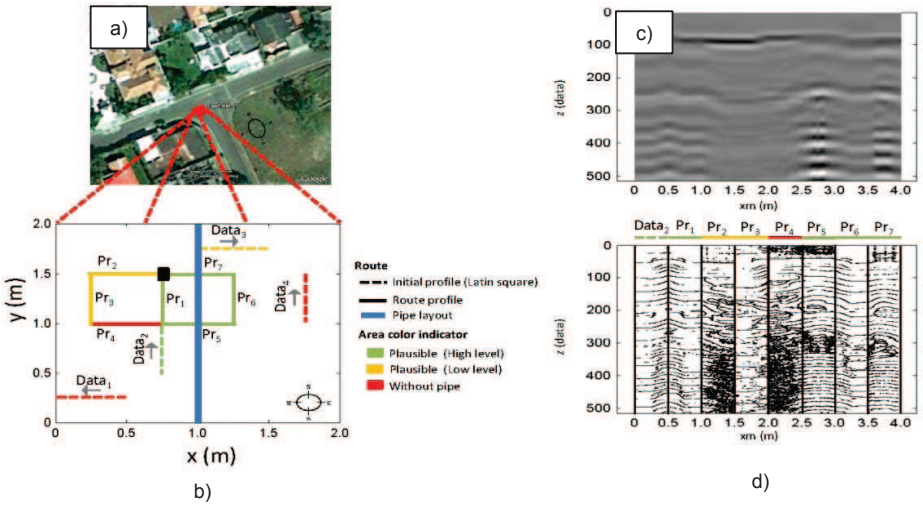


Figure 1: a) Interest area, b) result of application, c) raw profiles, d) pre-proc. profiles.

4 Conclusions

The application of the proposed system facilitates the analysis and interpretation of the images obtained with GPR in an interest area, favoring the mapping of unknown pipe layouts in WSS. This is because the system reduces the data capture time, the amount of dealt data is minimized, and the relevant data is suitably identified. These results allow more complete analysis since they also discard areas with low interest. This favors possible future automations and facilitates the study of pipes in WSS, with non-destructive methods, by non-highly qualified personnel in GPR analysis.

References

[1] Ayala-Cabrera, D., Pérez-García, R., Herrera, M., Izquierdo, J. Segmentación y limpiado de imágenes de GPR de tuberías enterradas. IX SEREA, Seminario Iberoamericano de planificación, proyecto y operación de abastecimiento de agua. Morelia (México), 2011.a

[2] Ayala-Cabrera, D., Herrera, M., Montalvo, I., Pérez-García, R. Towards the visualization of water supply system components with GPR images. Mathematical and Computer Modelling, vol. 54 (7-8), pp. 1818-1822, 2011.b

Improvements in the decay heat model in the thermalhydraulic code TRAC-BF1

A. Soler* ,* T. Barrachina*, R. Miró*, G. Verdú*,
A. Concejal† and J. Melara†

(*) Institute for Industrial, Radiophysical and Environmental Safety (ISIRYM),

Universitat Politecnica de Valencia - UPV,

Camí de Vera s/n, 46021 Valencia,

(†) Iberdrola Ingeniería y Consultoría,

Avenida Manoteras 20, 28050 Madrid

November 30, 2012

1 Introduction

Under normal operating conditions, the power obtained in a nuclear reactor comes mainly from the neutron-induced fission within the fuel of a reactor core and the subsequent conversion of mass to energy. When a reactor is shutdown, there is no energy produced by prompt-neutrons; however, the energy produced by delayed-neutrons remains and is directly related to the reactor power histories. This delayed energy is commonly labeled as decay heat. In a reactor working at nominal power conditions, the decay heat energy following a sudden shutdown contributes 7% of the total thermal energy which accounts for less than 1% after a 24 hour period. However, the decrease thereafter is quite small and the 1% power is sufficient to cause serious damage to the core if there is a loss of coolant. The amount of decay heat at various times after a sudden shutdown must be known and taken into account in the safety analysis because of its long-term effects.

*e-mail:asoler@iqn.upv.es

Being able to simulate accurately the different transients that can occur in a nuclear reactor is one of the main aims in nuclear safety analysis. The transient simulations involve both neutronic and thermalhydraulic calculations. The neutronic calculations refer to the quantification of the power generated by neutrons inside the reactor core. On the other hand, the thermalhydraulic calculations involve the resolution of the mass, momentum, and heat transfer hydraulic equations. Regardless of whether or not power calculations are performed, the decay heat energy must be obtained. All data concerning the decay heat of the main fissile nuclides of Light Water Reactors (LWRs) are published by the American Nuclear Society (ANS) as Standards which are used as reactor design and safety analysis guidelines.

The aforementioned calculations are solved by different computer codes, known as neutronic and thermalhydraulic codes respectively, which can be used stand-alone or coupled for more realistic and accurate results. One of the most used codes in nuclear industry is the thermalhydraulic code TRAC-BF1 which has already been proved against different transients in many nuclear power plants. The implemented decay heat models in TRAC-BF1 were based on the 1979 ANS standard, which was made completely obsolete due to the entry into force of the 1994 ANS standard, and the later review 2005 ANS. Besides, TRAC-BF1 solves the total decay heat energy system of ODE equations using the Runge-Kutta-Gill 4th order numerical method. However, the decay heat power system of equations can be analytically solved as is shown in this paper. Therefore a revision of the numerical solution methods and the implementation of ANSI/ANS-5.1-2005 in TRAC-BF1 are required.

2 Decay Heat Models Analysis

The first phase of the work presents a comparative study of decay heat models implemented in the thermo-hydraulic codes TRAC-BF1, TRACE, RETRAN and RELAP5. The study shows that all codes follow the directions on the American Standard ANS but differ in the initialization of the residual heat, in the evaluation at infinite time and in the numerical solution methods used. One of the key points in this first phase is the optimization of the numerical solution of the equation of the decay heat. Note that the code TRAC-BF1 solves this equation by the Runge-Kutta-Gill 4th order method. Finally, we note that the only code that has implemented the standard ANSI/ANS-5.1 1994 is TRACE.

The second phase involves the analytic resolution of the decay heat equation and its comparison with different numerical solution methods. Similarly, the influence of the short-term power histories in the final value of the decay heat is studied. This is particularly interesting in the simulation of severe transients like Anticipated Transients without Scram (ATWS).

The development of the analytical solution of the equations governing the calculation of the residual power allows the study of the impact of different numerical methods to analyze the possibility of replacing the Runge-Kutta-Gill method by the analytical method in TRAC-BF1. The selected methods are the analytical resolution, the quasi-analytic resolution, the 2^{nd} order ODE differential method and the 4^{th} order ODE differential method. The different operating reactor conditions for assessing the influence of the resolution methods on the decay heat power are listed as follows:

- Case 1 - Reference Case: 2 years of operation at 100% of nominal power + shutdown
- Case 2: 2 years of operation at 100% of nominal power + transient of 150 seconds at 150% of nominal power + shutdown
- Case 3: 2 years of operation at 100% of nominal power + transient of 2 seconds at 350% of nominal power + shutdown
- Case 4: 2 years of operation at 100% of nominal power + transient of 150 seconds at 50% of nominal power + shutdown

3 Discussion and Conclusion

The results showed that, for each case simulated, all of the numerical solutions provided virtually the same decay heat value. This fact confirmed that the implementation of the analytical solution implementation in TRAC-BF1 replacing the current method of Runge-Kutta-Gill 4^{th} order is feasible as well as recommended.

The results show that the residual heat obtained using the standard 2005 is greater than that obtained by the ANS 79 at the time of shutdown. However, during the decay, the values obtained using both standards have minimal deviations. Same trend was observed in all cases. In conclusion, the standard ANS 2005 provides more "best estimate" residual heat values than its predecessor and the update of the standard in TRAC-BF1 code is justified.

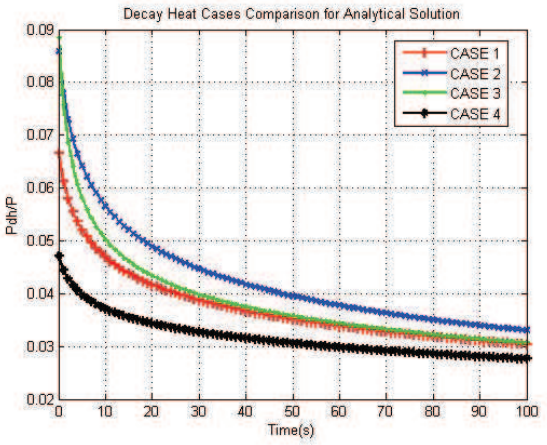


Figure 1: Influence of the operating reactor conditions in the decay heat.

From the study of the influence of the short-term power histories in the total decay heat power calculation show that the most extreme case is case 3 in which the second period only lasts two seconds while the first period lasts 2 years. A priori, one might assume that the contribution of the second period to the total residual heat is negligible compared to the contribution of the first period.

The following figure presents the simulation results of these different cases and it can be clearly observed how power changes (both increases and decreases) affect the residual heat, regardless of the duration of the variation.

In conclusion, the different operating reactor conditions, even if transients are of short duration, have high impact and, therefore, they cannot be neglected for the global calculation of residual heat in the safety analysis.

References

[1] ANSI/ANS- 5.1-1994: Decay Heat Power in Light Water Reactors, American Nuclear Society, 1994.

[2] ANSI/ANS- 5.1-2005: Decay Heat Power in Light Water Reactors, American Nuclear Society, 2005

[3] TRAC-BF1: An Advanced Best-Estimate Computer Program for BWR Accident Analysis, NUREG/CR-4356, volume 2, 1992.

[4] The RELAP5 Code Development Team (2001). RELAP5-3D Code Manual Volume 1: Code Structure, System Models and Solution Methods.

[5] US NRC (2007): TRACE V5.0 Theory Manual"

[6] RETRAN-3D: A Program for Transient Thermal-Hydraulic Analysis of Complex Fluid Flow Systems". Volume 1: Theory"

An Economic Dispatch Algorithm of Combined Cycle Units

L. Bayón*, P.J. García Nieto, J.M. Grau,
M.M. Ruiz and P.M. Suárez

E.P.I., Department of Mathematics, University of Oviedo

Campus of Viesques, Gijón, 33203, Spain

November 30, 2012

1 Introduction

The economic dispatch (ED) problem is defined as that of finding an optimal distribution of system load to the generators in order to minimize the total generation cost while satisfying the total demand and generating-capacity constraints. The different combinations of combustion turbines and steam turbines in a combined cycle (CC) unit produce multiple states and each state has its own unique cost curve. Therefore, in performing ED, we need to be able to shift between these cost curves. However, there is another, more serious problem: the cost curve is not convex for some of these states. As a result, ED of CC units must use special techniques like, for example: Complete Enumeration, Merit Order Loading, Genetic Algorithm (GA), Evolutionary Programming (EP), and Particle Swarm (PS) (see [1] and [2]). The three former techniques (GA, EP, and PS) involve a stochastic searching mechanism, and it should be noted that these heuristic methods do not always guarantee obtaining the globally optimal solution: they only provide an approximate solution for the non-convex optimization problem. In this paper we present a new technique for solving the ED problem of CC units. The technique, de-

*e-mail: bayon@uniovi.es

veloped to find the global solution, is based on the calculation of the Infimal Convolution (IC).

2 Mathematical Modelling of CC units

The literature [3] provides several alternatives to model CCs in electricity markets: Aggregated model, Pseudo unit model, Configuration-based model, and Physical unit model. The last two models are the most accurate ones. In general, the physical unit model is more suitable for power flow and network security analysis. However, the configuration-based model is more suitable for bid/offer processing and dispatch scheduling.

This paper focuses on the ED problem that a generation company with CC units faces when preparing its offers for the day-ahead market. We hence consider the configuration-based model. Each state has its own cost curve; for some states, this curve is not convex. The most widely-used model to represent the non-convexity of cost curves is a piecewise linear cost function ([1], [2]). This is the most flexible model and allows a greater approximation to reality. In the present paper, we shall use piecewise linear cost functions to represent the states of a CC unit.

3 Algorithm of Optimization

The classic ED problem can be described as an optimization problem:

$$\min \sum_{i=1}^{N} F_i(P_i) \text{ subject to: } \sum_{i=1}^{N} P_i = P_D;\ P_{i\min} \leq P_i \leq P_{i\max},\ \forall i = 1, ..., N$$

where $F_i(P_i)$ is the fuel cost function of the i-th unit, P_i is the power generated by the i-th unit, P_D is the system load demand, $P_{i\min}$ and $P_{i\max}$ are the minimum and maximum power outputs of the i-th unit, and N is the number of units. To solve this problem, we have designed an algorithm based on the mathematical concept of IC. The proposed recursive algorithm for calculating the analytic solution consists of 4 phases.

(Phase 1) Piecewise linear cost function of each CC unit.

The function $F_i(P_i)$ of each CC unit will be a piecewise linear function:

$$F_i(P_i) = \begin{cases} F_{i1}(P_i) & \text{if} \quad P_i \in [m_{i1}, M_{i1}] \\ ... & ... \quad ... \\ F_{ik(i)}(P_i) & \text{if} \quad P_i \in [m_{ik(i)}, M_{ik(i)}] \end{cases}, \quad i = 1, ..., N$$

(Phase 2) Infimal Convolution of 2 CC units.

In the following proposition, we shall express the IC for each pair of linear functions.

Proposition 1. *Let* $f_i(x_i) = a_i + b_i x_i$, $(i = 1, 2)$ *with domains* $[m_i, M_i]$. *Let us assume that* $b_1 \leq b_2$. *It is verified that:*

$$(f_1 \bigodot f_2)(\xi) := \begin{cases} f_1(\xi - m_2) + f_2(m_2) & \text{if} \quad \xi \in [m_1 + m_2, M_1 + m_2] \\ f_1(M_1) + f_2(\xi - M_1) & \text{if} \quad \xi \in [M_1 + m_2, M_1 + M_2] \end{cases}$$

We now need to perform all the possible combinations of pairs of linear functions between the 2 CC units, F_1, F_2, and then calculate the minimum of them all:

$$F_1 \bigodot F_2 = \min_{(i,j)}(F_{1i} \bigodot F_{2j}); \quad i = 1, ..., k(1); \quad j = 1, ..., k(2)$$

obtaining a piecewise linear function. This result will form the basis for the subsequent generalization to the case of N functions.

(Phase 3) Infimal Convolution of N CC units.

Bearing in mind the associative nature of the IC operation, the equivalent of N CC units may now be calculated by means of a recursive process, carrying out N operations of IC. We consider the next recurrence:

$$F_1 \bigodot F_2 \bigodot \cdots \bigodot F_N = (F_1 \bigodot F_2 \bigodot \cdots \bigodot F_{N-1}) \bigodot F_N$$

(Phase 4) Optimal solution of each CC unit.

Besides the minimum value of the total cost, the IC of the N CC units yields, for any P, the vector where said minimum value is reached. We shall now determine the distribution, for any P, of what each of the N CC units has to produce (the optimal solution of each CC unit). The ED problem of CC units is thus fully solved.

4 Example

Finally, the proposed method is applied to a test ED problem ([1] and [2]) and our solution is compared with three stochastic optimization techniques: GA, EP and PS. As already stated, these heuristic methods only provide an approximate solution for the non-convex optimization problem. The costs obtained using GA, EP and PS solutions in Table I are higher as they are

based on erroneous assumptions. It should be noted that our algorithm presents a much higher convergence speed than Complete Enumeration, as, in each phase, and after calculating the minimum, we shall only consider a very small fraction of the entire problem.

Table I: Comparison of the optimal solution.

	CC 1 (MW)	CC 2 (MW)	Demand (MW)	Cost ($/h)
GA	560	240	800	31888
EP	528.75	271.25	800	31544
PS	510	290	800	31460
AS	265	535	800	29871.2

5 Conclusions

In this paper we have presented a new technique, based on the calculation of the Infimal Convolution, for solving the ED problem of CC units. The technique consists in a recursive algorithm for calculating the global analytic solution. That is, we do not obtain the solution for only one value of demand, but solve a family of problems, varying P_D to obtain the solution for any value. This distinguishes our method from traditional heuristic methods. Furthermore, we have analytically obtained the solution for a test case that may serve as a comparison for subsequent studies using approximate methods.

References

[1] F. Gao, G. B. Sheble, Economic Dispatch Algorithms for Thermal Unit System Involving Combined Cycle Units, in *Proc. 15th Power Systems Computation Conference*, Liege, Belgium, (2005), pp. 1-6.

[2] F. Gao, G. B. Sheble, Stochastic Optimization Techniques for Economic Dispatch with Combined Cycle Units, in *Proc. 9th International Conference on Probabilistic Methods Applied to Power Systems KTH*, Stockholm, Sweden, (2006), pp. 1-8.

[3] J. Alvarez, R. Nieva, I. Guillen, Commitment of combined cycle plants using a dual optimization-dynamic programming approach, *IEEE Trans. Power Syst.* 26 (2) (2011) 728–737.

An epidemic model using parametric systems

B. Cantó* , C. Coll*, and E. Sánchez*

(*) Instituto de Matemática Multidisciplinar,

Universitat Politècnica de València, Camino de Vera, 14. 46022 Valencia. Spain.

November 30, 2012

1 Introduction

An epidemic is interpreted as an increase in the frequency of a disease in a population above the level expected in a given period. The need for more accurate predictions of how an epidemic affecting a population motivates the application of mathematical modelling techniques. In this way, one of the most important parameters in the study of mathematical epidemiology is the basic reproduction number. It is used to measure the transmission potential of a disease. That is, the basic reproduction number can be used to assess whether a newly infectious disease can invade a population [5]. Mathematical models of epidemic processes are given in [1], in the discrete-time case and [6], in continous-time case. The interest of this number lies in when the basic reproduction number is less than one the disease-free equilibrium is asymptotically stable and when the basic reproduction number is greater than one it is unstable [4].

Usually, in an epidemic model there are several unknown parameters which are often determined from observations of clinical disease. The determination of these parameters is critical for the design or control of epidemic models. For that, the study of the identifiability analysis is important. The problem of the structural identifiability of the model consists of the determination of all parameter sets which give the same input-output structure

*e-mail:bcanto@mat.upv.es

(more information in [3]). In [2] is shown a characterization of structural identifiability.

In this work we obtain a theoretical model to analyze an hypothetical epidemic. The proposed model has unknown parameters which depend on the considered epidemic.We analyze whether or not it is possible to determine the unknown parameters and we solve the identifiability problem. Moreover, we find the equilibrium of the system and we study the stability in the equilibrium point using the reproduction number.

The first step in building the model is to determine its structure, and the choice of variables and specific parameters of the process that we want to model and we will continue with the determination of the dynamical behavior of the system. We consider a discrete epidemic model that incorporates immature, mature and infectious individual in independent stages. Moreover each stage has male and female individuals. The population is divided into three classes: *immature* individuals, that is, individuals who cannot be infect, *susceptible mature* individuals, that is, individuals may become infected and finally, *infectious mature* individuals. Let $j_m(t)$, $j_f(t)$, $m_m(t)$, $m_f(t)$, $i_m(t)$ and $i_f(t)$ be the density of immature, mature infectious, males and females, respectively.

We assume that the susceptible individual become infectious after contact with infective individuals and recovery from disease does not give permanent immunity. Moreover, we suppose that only have ability of breeding the susceptible mature female.

The parameter k_1, k_2, k_3 are the survival rate of the immature individual, the mature individual and the infected mature individual, respectively, and c,g, σ are the probability of immature individual becoming mature individual, susceptible mature individual become infectious mature individual and infected individual recover, respectively. In our case parameters c,g, σ are unknown and we have to determine them from the input-output behavior of the epidemic process. Finally, $b(t)$ is the birth number of the male and female individuals, respectively. Using these assumptions, if $i = m, f$ we have the following discrete population model with stage structure,

$$\begin{aligned} j_i(t+1) &= k_1(1-c)j_i(t) + b(t) \\ m_i(t+1) &= k_1 c j_i(t) + k_2(1-g)m_i(t) + k_3\sigma i_i(t) \\ i_i(t+1) &= k_2 g m_i(t) + k_3(1-\sigma)i_i(t) \end{aligned}$$

If we obtain the system, if its matrices are nonnegative, the solutions

corresponding to nonnegative initial data remain nonnegative in the future. The nonnegativity condition means that the parameters satisfy $c \leq 1$, $g \leq 1$ and $\sigma \leq 1$ and this is true because these parameters are probabilities.

To analyze an epidemic model is determinant to know how the population increases. We consider two forms for $b(t)$. The first one is a linear birth function given by $b(t) = \alpha m_f(t)$. And second is a Beverton-Holt type. If $n(t)$ represents all the population, the function is given by $b(t) = \dfrac{\alpha m_f(t)}{1 + \beta n(t)}$ where $\alpha > 1$ and $\beta > 0$ are constant. This kind of function has the advantage that best describes the mechanism arising naturally from simpler models of epidemic processes but it has disadvantage that can be complicated to analyze.

2 Main results

Structural identifiability guarantees that the model parameters can be estimated uniquely, under ideal conditions. The parameter identification process for the structured system is followed from the structure of the vectors obtained by calculation of the Markov parameters of the system because they lead determine the input-output behaviour (i/o) of a model. We can assure that our model is structurally identifiable.

Now we obtain the disease free equilibrium points, x^*, considering that the probability that susceptible mature individual become infectious mature individual is proportional to infectious individuals. When the birth function is $b(t) = Fx(t)$, with an adequated feedback the non zero equilibrium point depends on the survival rate of non infectious individuals, the probability of moving from immature to mature and the number of births.

If the birth function is a Beverton-Holt type and the total population in the equilibrium, $n^* = \dfrac{(\alpha - 1)k_1 c - (1 - k_1 - k_2 k)}{\beta(1 - k_2)k}$, with $k = 1 - k_1(1 - c)$, we obtain the non zero disease-free equilibrium point and in this case it is related to the survival rate of non infectious individuals, the probability of moving from immature to mature and the population in the equilibrium.

In order to study the stability in the equilibrium $x^* = 0$ we consider the basic reproduction number, $\mathcal{R}_0$. From an asymptotically stable system, it is known (see [4]) that the disease-free equilibrium is asymptotically stable if $\mathcal{R}_0 < 1$ and it is unstable when $\mathcal{R}_0 > 1$.

When the birth function is a linear control $b(t) = Fx(t)$ the obtained reproduction number means that the disease is constantly present in the population.

And, when the birth function is a Beverton-Holt type control and we linearize the system around the equilibrium point, we obtain

$$\mathcal{R}_0 = \frac{(1 + (c-1)k_1)(1-k_2)}{\alpha c k_1},$$

which is related with the survival rate of non infectious individuals and the probability of immature individual becoming mature individual.

References

[1] L.S.J. Allen, P. van den Driessche, The basic reproduction number in some discrete-time epidemic models, *Journal of Difference Equations and Applications*, pp. 1–19, 2008.

[2] A. Ben-Zvi, P.J. McLellan, K.B. McAuley, Identifiability of Linear Time-Invariant Differential-Algebraic systems. 2. The Differential-Algebraic Approach, *Ind. Eng. Chem. Res.* 43:(5), 1251–1259, 2004.

[3] B. Cantó, C. Coll, E. Sánchez, Identifiability of a class of discretized linear partial differential algebraic equations, *Math. Problems Eng.*1–12, 2011.

[4] O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneus populations, *Journal Math. Biol.* 28: 365–382, 1990.

[5] J. Ma, J.D. Earn, Generality of the final size formula for an epidemic of a newly invading infectious disease, *Bull. Math. Biol.* 68: 679–702, 2006.

[6] W. Wang, X. Zhao, An epidemic model in a patchy environment, *Math. Biosci.* 190: 97–112, 2004.

Parameter Estimation Using Polynomial Chaos

B. Chen-Charpentier [*], and D. Stanescu[†]

(*) Department of Mathematics,

University of Texas at Arlington, Arlington, TX 76019-0408, USA,

(†) Department of Mathematics,

Department of Mathematics, University of Wyoming, Laramie, WY 82071-3036, USA.

November 30, 2012

1 Introduction

Parameter estimation is an inverse problem where model parameters are inferred from observations. It is a very important problem in engineering, physical, biological, economic and social models. Many methods have been developed to solve it, but is still a difficult and expensive problem.Among the most commonly used methods for parameter estimation are: least squares, Bayesian inference, minimum variance, methods using polynomial chaos which include our proposed method and methods using measure theory.

We will consider differential equations in which the unknown parameters are random variables and the solutions are stochastic processes. Polynomial chaos, as presented below, is a very useful and efficient way of dealing with random differential equations. It produces numerical solutions that are explicit in the random variables. Such solutions can be effectively used in combination with maximum likelihood methods to find the parameters that best fit a set of observed data. A simple model of biofilm growth provides a partial differential equation with unknown parameters provides an illustrative application.

[*]bmchen@uta.edu

2 Polynomial Chaos and Bayesian Inference

Here we consider a simple model of a microbial biofilm attached to the surface of an implant as given in [2]. We assume that there is no advection of microbes, since in general the microbes do not move with the fluid, and that there is a limited amount of nutrients that restricts the growth of the microbes due to same species competition for resources and modeled by a logistic term. The growth and diffusion of the biofilm microbes is governed by the partial differential equation

$$\frac{\partial}{\partial t}(c) - \frac{\partial}{\partial x}\left(D\frac{\partial c}{\partial x}\right) = r\,(1-c)\,c \tag{1}$$

Here c represents the mass concentration of microbes per unit volume, r is the growth rate and M is the saturation value. The diffusion coefficient $D(x,t)$ takes into account that the microbes displace if their concentration is high.All variables have been normalized to make them dimensionless. We will consider that r and D are random variables with a specified uniform or normal distribution, but independent of time and position.

There are many methods of dealing with differential equations with random coefficients. An efficient and also easy to implement is the polynomial chaos method.

To develop a general methodology for the numerical solution of the diffusion-reaction equation (1) as well as for estimating the various moments of the solution, we follow a Generalized Polynomial Chaos (GPC) approach [3]. The polynomial chaoses can be arranged in a sequence $\Phi_i(\boldsymbol{\xi}(\omega))$, such that the expansion of the random variables and the stochastic process appearing in equation (1) takes the form, i.e.:

$$r(\omega) = \sum_{j=0}^{\infty} r_j \Phi_j(\boldsymbol{\xi}(\omega)); D(\omega) = \sum_{j=0}^{\infty} d_j \Phi_j(\boldsymbol{\xi}(\omega)); c(t,x;\omega) = \sum_{i=0}^{\infty} c_i(t,x)\Phi_i(\boldsymbol{\xi}(\omega)),$$

where the Φ_i are properly chosen polynomial basis functions of the random variable vector $\boldsymbol{\xi}$. The number of variables in $\boldsymbol{\xi}$ represents the dimension of the chaos. A Galerkin projection using the orthogonality of the basis functions $\langle \Phi_i, \Phi_j \rangle = \delta_{ij}\langle \Phi_i, \Phi_i \rangle$, together with truncation of the polynomial chaos series to a finite number of terms $P+1$ will then lead to a system of partial differential equations governing the time evolution of the chaos

coefficients of the solution c_i of the reaction-diffusion differential equation (1).

$$\langle \Phi_L, \Phi_L \rangle \frac{\partial c_L}{\partial t} - \sum_{i=0}^{P}\sum_{j=0}^{P} D_j \frac{\partial^2 c_i}{\partial x^2} \langle \Phi_i \Phi_j, \Phi_L \rangle = \\ = \sum_{i=0}^{P}\sum_{j=0}^{P} c_i r_j \langle \Phi_i \Phi_j, \Phi_L \rangle - \sum_{i=0}^{P}\sum_{j=0}^{P}\sum_{l=0}^{P} c_i r_j c_p \langle \Phi_i \Phi_j \Phi_p, \Phi_L \rangle. \quad (2)$$

Note that both the space and time derivatives decouple, so we have a diffusion process of the "classical" type with a more complex right-hand side.

Equations (2) are a linear system of coupled partial differential equations in the $P + 1$ unknowns $c_0, \ldots, c_P$. The system can be solved using, for example, a finite difference method explicit in time and centered in x.

Here we develop a method for estimating the unknown parameters using Bayes inference and the polynomial chaos method. It is based on the ideas of [1]. First we give a prior distribution for the unknown parameters in terms of a random variable for each one. In our example we have two parameters, r and D, and we will assume that both are given by either uniform or normal distributions of the parameters $\boldsymbol{\xi}$. The differential equations of the model are then solved using the polynomial chaos method (2). We now have numerical solutions of the stochastic process c at different values of the time t. Note that since we solved for the coefficients of the polynomial chaos expansion, we have an explicit expression for c as a function of $\boldsymbol{\xi}$ at each time step. Next we maximize the likelihood function with respect to the random variables $\boldsymbol{\xi}$ and obtain the best parameter fit to the observed values.

Some advantages of this approach is that the maximization is done with respect to only one variable per unknown random parameter. Furthermore, it is not necessary to recalculate the numerical solutions since we have them explicitly in terms of the random variables. The usual calculus method is then used to solve the optimization problem.

We assume that the observations are the amount of bacteria at the mid-point in x at different times.Likelihood function for a mathematical model $y = \mathcal{M}(t_k, \boldsymbol{\xi})$ and observed data y_{obs} is given by

$$\mathcal{L}(y_{obs}, \boldsymbol{\xi}) = \frac{1}{n_{obs}} (\prod_{k=1}^{n_{obs}}) K(y_{obs}(k) - \mathcal{M}(t_k, \boldsymbol{\xi})).$$

A usual assumption is that the measurement errors are normally distributed and that the kernel function is the standard $K(x) = \exp(-x^2)$. Since the

ln is a monotonic function, instead of maximizing the likelihood function it is easier to minimize its logarithm. For this particular kernel, this method is similar to the method of least squares except that here the minimization variables are the random variables. The result of this differentiation are polynomial equations in ξ_1 and ξ_2 of the same order as of the polynomial chaos used. In our case they are cubic polynomials since we are using chaos of order 3. Substitution of the optimal values of $\boldsymbol{\xi}$ in the polynomial chaos expansion of the solution gives the best solution based on the observations.

We have explored the use of Hermite and Legendre expansions up to order three for system (2). Expansions of order two and three show very similar results. It is known from previous studies that long-time integration may lead to severe losses in the accuracy of the chaos expansion due to the accumulation of the non-linearity in stochastic space. From our experiments, it seems however that an expansion to order three is sufficient for the simple biofilm model studied here. The results for uniform and normal distributions of the parameters are almost identical. Since the distribution of the outputs is not known a-priori, except in simple cases, it is our opinion that one can use a polynomial chaos based on Hermite or Legendre polynomials for both the inputs and the outputs.

We have presented a method for estimating unknown parameters in differential equations by first assuming a distribution function in terms of a random variable for each unknown parameter, then using the method of polynomial chaos to solve the random differential equations. This gives numerical solutions of the unknown stochastic processes explicit in terms of the random variables. The maximum of the likelihood function is obtained in terms of the random variables and the optimal values of the parameters are then obtained solving a system of algebraic equations. This maximum likelihood method with random variables as the parameters is very efficient and easy to implement, The method of polynomial chaos requires the solution of a system of differential equations of the same form as the original deterministic one, and even though the number of equations can increase rapidly with the number of random variables and the order of the chaos, it is much faster than alternative methods such as Monte Carlo. A further advantage is that we have explicitly the solution in terms of the random variables which makes the maximization problem easy and fast. The method worked very well when both uniform and normal distributions were assumed and was independent of the distribution, or lack of it, of the errors.

References

[1] E, Blanchard, A. Sandu, and C. Sandu, *A polynomial-chaos-based Bayesian approach for estimating uncertain parameters of mechanical systems*, 19th International Conference on Design Theory and Methodology; 1st International Conference on Micro- and Nanosystems; and 9th International Conference on Advanced Vehicle Tire Technologies, Parts A and B; Vol. 3 (2008), pp. 1041–1048.

[2] B. M. Chen-Charpentier and D. Stanescu, *Biofilm growth on medical implants with randomness*, Mathematical and Computer Modelling 54 (2011) pp. 1682-1686.

[3] D. Xiu and G. E. Karniadakis, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput. 24 (2002), 619–664.

A new finite difference approach for partial integro-differential option pricing Merton model

M.-C. Casabán, R. Company, and L. Jódar*

Instituto de Matemática Multidisciplinar, Universitat Politècnica de València,

Camino de Vera s/n, 46022 Valencia, Spain.

macabar@imm.upv.es; rcompany@imm.upv.es; ljodar@imm.upv.es

November 30, 2012

1 Introduction

Since that empirical studies revealed that the normality of the log-returns, as assumed by Black and Scholes, could not capture features like heavy tails and asymmetries observed in market-data log-returns densities, a number of models try to explain these empirical observations. Jump-diffusion and infinite activity Lévy models allow to calibrate the model to market price of options and reproduce a wide variety of implied volatility skews/smiles. These models are characterized by partial integro-differential equations (PIDEs) that involve a second order differential operator, and a non-local integral term that requires specific treatment and presents additional difficulties. We recall that in a jump-diffusion model [1], the resulting PIDE for option price $V(S,t)$ in the case of a vanilla call option is given by

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2V}{\partial S^2}+(r-\lambda K)\,S\,\frac{\partial V}{\partial S}-(r+\lambda)V \\ +\lambda\int_0^\infty V(S\eta,t)\,g(\eta)\,d\eta=0,\qquad 0<S<\infty\,,\ 0\le t<T\,, \tag{1}$$

*This paper was supported by the Spanish M.E.Y.C. grant DPI2010-20891-C02-01.

$$V(S,T) = f(S) = \max(S-E,0)\,, \qquad 0 < S < \infty\,, \tag{2}$$

where S is the underlying stock price, σ is the volatility, r is the risk-free interest rate and $f(S)$ is the payoff function. The random variable representing the jump amplitude is denoted by η and the expected relative jump size is denoted by $K = \mathbf{E}[\eta - 1]$. The jump intensity of the Poisson process is denoted by λ and the probability density of the jump amplitude is given by $g(\eta)$. Merton's jump-diffusion model assumes that jump sizes are log-normally distributed with mean μ_J and standard deviation σ_J, i.e.

$$g(\eta) = \frac{1}{\sigma_J\,\eta\,\sqrt{2\pi}}\,\exp\left(-\frac{1}{2}\left(\frac{\ln(\eta)-\mu_J}{\sigma_J}\right)^2\right). \tag{3}$$

In order to solve the PIDE problem numerically, various authors have used finite difference schemes in [2, 3, 4]. The aim of the paper is the construction of a finite difference numerical scheme of the PIDE (1)–(2), with a different treatment of the integral part from previously quoted authors. Instead of assuming long term information about the solution we perform a full discretization of the integral part, involving the unknown function values in the numerical scheme.

2 Transformation of the integro-differential problem

We introduce a transformation of variables to remove both the advection and the reaction terms of the PIDE problem (1)–(2),

$$\left.\begin{array}{l} X = \exp\left((r-\lambda K)(T-t)\right)S, \quad \tau = T-t, \\ U(X,\tau) = \exp\left((r+\lambda)(T-t)\right)V(S,t), \end{array}\right\} \tag{4}$$

and note that problem (1)-(2) is transformed into the problem

$$\frac{\partial U}{\partial \tau} = \frac{1}{2}\,\sigma^2 X^2 \frac{\partial^2 U}{\partial X^2} + \lambda \int_0^\infty U(X\eta,\tau)g(\eta)\,d\eta, \tag{5}$$
$$0 < X < \infty\,,\ 0 < \tau \le T\,,$$

$$U(X,0) = f(X)\,, \qquad 0 < X < \infty\,. \tag{6}$$

In order to approximate the integral in (5) it is convenient the change of variable $\phi = X\,\eta$. Later, taking $A > 0$, we decompose the integral term into two integrals, one for the interval $]0, A]$ and the other for $[A, \infty[$. If we consider the substitution $z = \frac{A}{\phi}$ into the last one, an integral expression over the finite interval $]0, 1]$ is obtained and the problem (5)-(6) can be written as

$$\frac{\partial U}{\partial \tau} = \frac{\sigma^2 X^2}{2}\,\frac{\partial^2 U}{\partial X^2} + \frac{\lambda}{X}(J_1 + J_2), \quad 0 < X < \infty\,,\ 0 < \tau \le T\,, \tag{7}$$

$$U(X, 0) = f(X)\,, \qquad 0 < X < \infty\,, \tag{8}$$

where

$$\left.\begin{array}{c} J_1 = \int_0^A U(\phi, \tau)\, g\left(\frac{\phi}{X}\right) d\phi\,; \\ J_2 = \int_A^\infty U(\phi, \tau)\, g\left(\frac{\phi}{X}\right) d\phi \;= A \int_0^1 U\left(\frac{A}{z}, \tau\right)\, g\left(\frac{A}{Xz}\right) \frac{1}{z^2}\, dz\,. \end{array}\right\} \tag{9}$$

3 Numerical scheme construction

In this section a difference scheme for problem (7)-(8) is constructed. With respect to the time variable, let k be the time-step discretization $k = \frac{\tau}{L}$, and $\tau^n = nk$, $0 \le n \le L$. With respect to the spatial variable X, we construct an uniform grid in $[0, A]$, with the step $h = \frac{A}{N}$, with $X_j = jh$, $0 \le j \le N$. For the variable z, an uniform mesh of $]0, 1]$ with M points of the form $z_j = j\,\delta$, $1 \le j \le M$, $M\delta = 1$, with $M \ge 3$ implies a non uniform mesh for the original variable X in $[A, \infty[$,

$$X_j = \frac{A}{z_{N+M-j}} = \frac{A}{1 - (j - N)\delta}\,, \quad N \le j \le N + M - 1\,.$$

Let us denote the numerical solution by $u_j^n \approx U(X_j, \tau^n)$, and consider forward finite difference approximations in time and centered in space

$$\frac{\partial^2 U}{\partial X^2}(X_j, \tau^n) \approx \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{h^2} = \Delta_i^n\,, \tag{10}$$

for the internal points of $[0, A]$, and denoting $h_j = X_{j+1} - X_j > 0$,

$$\frac{\partial^2 U}{\partial X^2}(X_j, \tau^n) \approx 2\left(\frac{u_{j+1}^n}{h_j(h_j + h_{j-1})} + \frac{u_{j-1}^n}{h_{j-1}(h_j + h_{j-1})} - \frac{u_j^n}{h_j h_{j-1}}\right) = \Delta_i^n\,, \quad N \le j \le N + M - 2\,, \tag{11}$$

for the points X_j lying in $\left[A, \frac{A}{2\delta}\right]$. For the internal points we have the scheme

$$u_i^{n+1} = u_i^n + \frac{k}{2}\sigma^2 X_i^2 \Delta_i^n + \frac{k\,\lambda}{X_i}(J_{1,i}^n + J_{2,i}^n)\,, \quad 1 \le i \le N+M-2\,, \tag{12}$$

where $J_{1,i}^n$ and $J_{2,i}^n$ are approximations of the composite trapezoidal type of integrals appearing in (9). By denoting $g_{i,j} = g\left(\frac{X_j}{X_i}\right)$, one gets

$$J_{1,i}^n = h\left(\sum_{j=1}^{N-1} u_j^n\, g_{i,j} + \frac{1}{2}\, u_N^n\, g_{i,N}\right)\,, \quad 1 \le i \le N+M-2, \tag{13}$$

$$J_{2,i}^n = \frac{\delta}{A}\left(\frac{1}{2}\, u_N^n\, g_{i,N}\, X_N^2 + \sum_{j=N+1}^{N+M-1} u_j^n\, g_{i,j}\, X_j^2\right)\,, \quad 1 \le i \le N+M-2. \tag{14}$$

The numerical scheme (12)-(14) incorporates the initial and the boundary conditions

$$u_i^0 = f(X_i) = \max\left(X_i - E, 0\right)\,, \quad 1 \le i \le N+M-1\,, \tag{15}$$

$$u_0^n = 0\,, \qquad u_{N+M-1}^n = u_{N+M-1}^0\,, \qquad 0 \le n \le L\,. \tag{16}$$

Let us denote the vector $U^n = [u_1^n\ u_2^n\ \ldots\ u_{N+M-1}^n]^t$ and let P be the tridiagonal matrix related to the differential part and $B = (b_{ij})$ the full matrix related to the integral part. The scheme (12)-(16) can be written in the vector form

$$\left.\begin{aligned} U^{n+1} = (P+B)\,U^n = (P+B)^{n+1}\,U^0\,, \quad 0 \le n \le L-1; \\ U^0 = [f(X_1)\ f(X_2)\ \ldots f(X_{N+M-1})]^t\,. \end{aligned}\right\} \tag{17}$$

4 Positivity, Stability and Consistency of the numerical solution

Dealing with prices of contracts modeled by PIDE, the solution must be nonnegative. From the analysis of the entries of matrix P the next result is established.

Theorem 1 *With previous notation, assume that stepsizes* $k = \Delta\tau$, $h = \Delta X$ *in* $[0, A]$ *and* $0 < \delta \leq \frac{1}{3}$, $\delta = \Delta z$ *in* $]0, 1]$, *satisfy:*

$$(C_1): \quad \frac{k}{h^2} \leq \frac{1}{\sigma^2 A^2}, \qquad (C_2): \quad k \leq \min\left\{\frac{\delta^2}{\sigma^2(1-2\delta)}, \ \frac{\delta h}{\sigma^2}\right\},$$

then the solution $\{u_i^n\}$ *of the scheme* (12)-(16) *is nonnegative starting with initial values* $u_i^0 \geq 0$, $1 \leq i \leq N + M - 1$.

The next result establishes that the scheme (12)-(16) is conditionally strongly uniformly $\|\cdot\|_\infty$ stable in the sense that the numerical solution remains bounded in $\|\cdot\|_\infty$ with respect to the initial condition for all time levels independently of the step discretization sizes.

Theorem 2 *Under conditions* (C_1) *and* (C_2) *of theorem 1, the numerical solution* $\{u_i^n\}$ *of the scheme* (12)-(16) *satisfies*

$$\|U^n\|_\infty \leq \exp(\lambda C T) \left\|U^0\right\|_\infty, \qquad 0 \leq n \leq L, \tag{18}$$

where $C = 1 + \sqrt{\frac{2}{\pi}} \dfrac{\exp(-\mu_J + \sigma_J^2/2)}{\sigma_J} \left(1 + \exp(2\mu_J)\right)$.

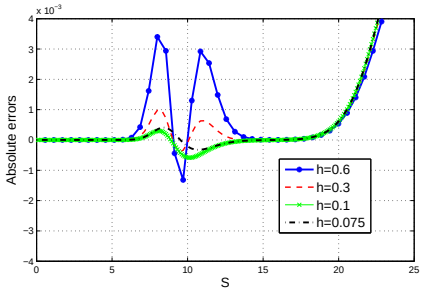
We say that a numerical scheme is consistent with a PIDE, if the exact theoretical solution of the PIDE approximates well the exact solution of the difference scheme as the stepsize discretization tends to zero. Using Taylor's expansions of the partial derivatives about (X_i, τ^n) one gets that (12)-(14) is consistent with the PIDE problem (7) of second order in both the spatial stepsizes h and δ and of first order in time stepsize k.

5 Numerical results

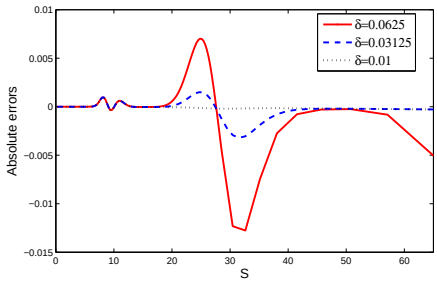
The next examples show the advantages of our double discretization technique in order to reduce the error of the numerical solution.

Example 1. Consider the vanilla call option problem (1)–(3) under Merton jump diffusion model with parameters $T = 1$, $r = 0.05$, $E = 10$, $\sigma = 0.1$, $\mu_J = 0$, $K = 0.00501$ and $\lambda = 0.1$. For $A = 30$, $k = 0.001$ and $\delta = 0.0625$, figure 1(a) shows the reduction of the absolute error about the strike when stepsize h decreases while the error close to the right boundary of the domain remains stationary. This fact agrees with facts illustrated in [4].

Example 2. Taking the problem of example 1 with fixed $h = 0.3$, figure 1(b) shows how the error close to the right boundary can be reduced with our double spatial discretization by decreasing the stepsize δ.



(a) Absolute errors with several values of h and a fixed δ in example 1.



(b) Absolute errors with several values of δ and a fixed h in example 2.

References

[1] Merton R.C., Option pricing when the underlying stocks are discontinuous, *Journal of Financial Economics*, 3(1-2):125–144, 1976.

[2] Cont R., and Voltchkova E., A finite difference scheme for option pricing in jump diffusion and exponential Lévy models, *SIAM Journal on Numerical Analysis*, 43(4):1596–1626, 2005.

[3] Toivanen J., Numerical valuation of european and american options under Kou's jump-diffusion model, *SIAM Journal on Scientific Computing*, 30(4):1949–1970, 2008.

[4] Almendral A., and Oosterlee C.W., Numerical valuation of options with jumps in the underlying, *Applied Numerical Mathematics*, 53(1):1–18, 2005.

A comparative analysis between some iterative methods from a dynamical point of view *

Francisco Chicharro § †, Alicia Cordero §, and Juan R. Torregrosa §

(§) Instituto de Matemática Multidisciplinar,

Universitat Politècnica de València,

Camino de Vera, s/n, 46022, Valencia, Spain

November 30, 2012

Abstract

In this paper, the dynamical behavior of different optimal iterative schemes for solving nonlinear equations with increasing order, is studied. The tendency of the complexity of the Julia set is analyzed compared with Newton's method.

1 Introduction

In this paper we are going to analyze, under the dynamical point of view, different derivative-free methods for solving a nonlinear equation $f(x) = 0$. The Steffensen's method is a well-known derivative-free scheme to solve nonlinear equations (see, for example, [5]),

$$z_{n+1} = z_n - \frac{f^2(z_n)}{f(v_n) - f(z_n)}. \tag{1}$$

This scheme replaces the derivative of Newton's scheme by the forward difference, holding the quadratic convergence. In Section 2 several optimal

*This research was supported by Ministerio de Ciencia y Tecnología MTM2011-28636-C02-02.

†e-mail: frachilo@upvnet.upv.es

derivative-free methods are introduced, obtained by composing Steffensen and Newton's schemes and estimating the involved derivative by Padé-Like approximants.

The stability of an iterative method can be analyzed by means of the dynamical study of the rational function associated to each fixed-point iteration. In Section 3, the dynamical planes of the iterative schemes applied to low-degree polynomials are developed.

2 Some optimal Steffensen-type methods

From now on, several iterative methods are introduced and compared by different features, as the convergence order, the Ostrowski efficiency index (see [1]) or the optimality in the sense of Kung-Traub's conjecture (see [4]).

The composition technique (described, for instance, in [5]) allows us to generate high-order methods. This technique, joint with Padé-like approximants, has allowed the authors to design in [6] optimal iterative schemes denoted by M4 (4th-order) and M8 (8th-order). Following this idea, an optimal sixteenth-order scheme (denoted by M16) can be developed, whose iterative expression is

$$z_{n+1} = w_n - \frac{f(w_n)}{f[z_n, w_n] + (z_n - w_n)\{c_5 f[z_n, w_n] - c_3 - c_4(z_n - w_n)\}}, \quad (2)$$

where $c_5 = -\frac{f[z_n, y_n, v_n, u_n, w_n]}{f[z_n, y_n, v_n, u_n]}$, $c_4 = f[z_n, y_n, v_n, w_n] + c_5 f[z_n, y_n, v_n]$, $c_3 = f[z_n, y_n, w_n] - c_4(z_n + y_n - 2w_n) + c_5 f[z_n, y_n]$, $v_n = z_n + f(z_n)$ and w_n, u_n and y_n are M8, M4 and Steffensen iterative steps, respectively. We denote by $f[\cdot,\cdot,\cdot,\cdot]$ and $f[\cdot,\cdot,\cdot,\cdot,\cdot]$ the divided differences of order three and four.

The efficiency index of the involved methods is 1.4142 for Steffensen, 1.5874 for M4, 1.6818 for M8 and 1.7411 for M16. All of them are optimal in the sense of Kung-Traub's conjecture.

3 Complex dynamics of iterative methods

The stability and convergence of a method can be analyzed from the dynamical plane associated to the rational function obtained by the evaluation of the fixed point iteration on quadratic and cubic polynomials. For a more

extensive and comprehensive review of dynamical concepts, see for example [2, 3].

Let $R : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ be a rational function, where $\hat{\mathbb{C}}$ is the Riemann sphere. The orbit of a point $z_0 \in \hat{\mathbb{C}}$ is defined as $\{z_0, R(z_0), \ldots, R^n(z_0), \ldots\}$.

A point $z_0 \in \mathbb{C}$ is a fixed point of R if $R(z_0) = z_0$. It is attracting, repelling or neutral if $|R'(z_0)|$ is less than, greater than or equal to 1, respectively. Moreover, if $|R'(z_0)| = 0$, the fixed point is superattracting.

Let z_f^* be an attracting fixed point of the rational function R. Its basin of attraction $\mathcal{A}(z_f^*)$ is defined as the set of pre-images of any order such that $\mathcal{A}(z_f^*) = \left\{ z_0 \in \hat{\mathbb{C}} : R^n(z_0) \to z_f^*, n \to \infty \right\}$.

The set of points whose orbits tend to an attracting fixed point z_f^* is defined as the Fatou set, $\mathcal{F}(R)$. Its complementary, the Julia set $\mathcal{J}(R)$, is the closure of the set consisting of its repelling fixed points, and establishes the boundaries between the basins of attraction.

If the rational function R is associated to the fixed point operator of the developed methods in Section 2 applied on a polynomial $f(z)$, denoted as $O_f(z)$, it is possible to find its fixed and critical points. The fixed points z_f verify $O_f(z) = z$, and the critical points z_c validate $O_f'(z) = 0$. The attracting fixed points z_f^* are those z_f such that $|O_f'(z_f)| < 1$. If $|O_f'(z_f)| = 0$, the fixed point is superattracting.

The expressions

$$S_f(z) = z - \frac{[f(z)]^2}{f(v) - f(z)}, \tag{3}$$

$$F_f(z) = y - \frac{f(y) f[z, v]}{f[z, y] f[y, v]}, \tag{4}$$

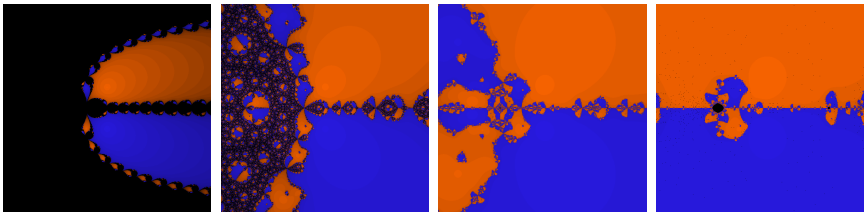
$$E_f(z) = u - \frac{f(u) f[z, y, v]}{f[y, v, u] f[u, z, y](u - z) + f[z, y, v] f[y, u]}, \tag{5}$$

$$X_f(z) = w - \frac{f(w)}{f[z, w] + (z - w)\{c_5 f[z, w] - c_3 - c_4(z - w)\}}, \tag{6}$$

are the fixed point operators of Steffensen's, M4, M8 and M16 methods, respectively, where $v = z + f(z)$, $y = S_f(z)$, $u = F_f(z)$ and $w = E_f(z)$.

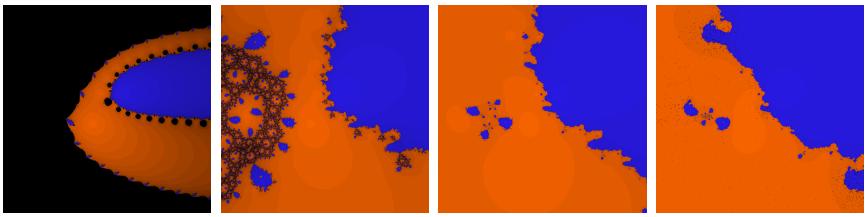
The behavior of each method has been analyzed on four different polynomials: $p_c(z) = z^2 + c$ and $q_c(z) = z^3 + c$, where $c \in \{1, i\}$.

The dynamical planes of $S_f(z)$, $F_f(z)$, $E_f(z)$ and $X_f(z)$, when they are applied to $p_{\{1,i\}}(z)$ and $q_{\{1,i\}}(z)$, are displayed in Figures 1 to 4. The represented region is $\Re\{z\} \in [-5, 5]$ in abscissas and $\Im\{z\} \in [-5, 5]$ in ordinates.



(a) $S_{p_1}(z)$ (b) $F_{p_1}(z)$ (c) $E_{p_1}(z)$ (d) $X_{p_1}(z)$

Figure 1: Dynamical planes of iterative methods on $p_1(z) = z^2 + 1$.



(a) $S_{p_i}(z)$ (b) $F_{p_i}(z)$ (c) $E_{p_1}(z)$ (d) $X_{p_i}(z)$

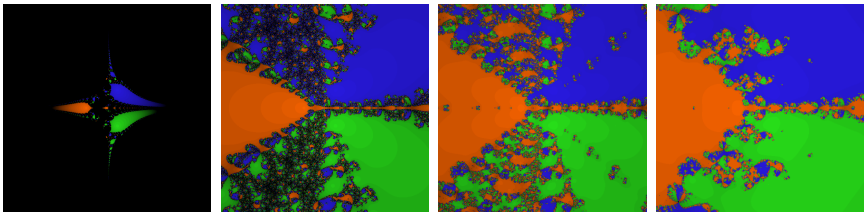
Figure 2: Dynamical planes of iterative methods on $p_i(z) = z^2 + i$.

The basins of attraction are colored in orange and blue – for quadratic polynomials – and also in green – for cubic polynomials –. The black points are those that do not converge to any of the attracting fixed points. Moreover, there are regions in each basin of attraction with lighter or darker color, depending on the amount of iterations needed to achieve the attracting fixed point. Each dynamical plane plots 600×600 points, and the maximum number of allowed iterations is 40.

As Figures 1 to 4 show, except for Steffensen's method, it seems that the complexity of the dynamical planes gets smoother as the order of convergence increases.

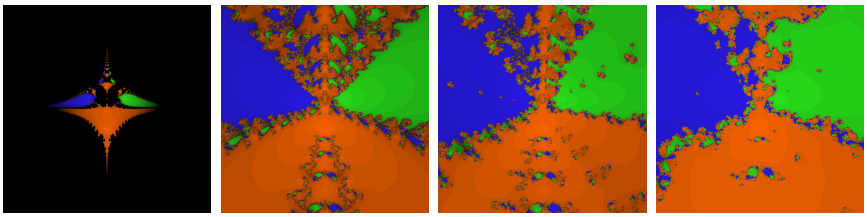
The fixed point operator associated to the Newton's method (**??**) is

$$N_f(z) = z - \frac{f(z)}{f'(z)}. \tag{7}$$



(a) $S_{q_1}(z)$ (b) $F_{q_1}(z)$ (c) $E_{q_1}(z)$ (d) $X_{q_1}(z)$

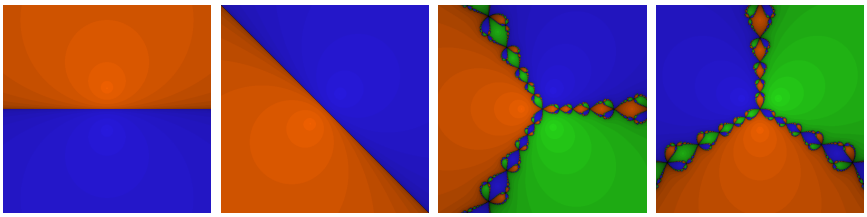
Figure 3: Dynamical planes of iterative methods on $q_1(z) = z^3 + 1$.



(a) $S_{q_i}(z)$ (b) $F_{q_i}(z)$ (c) $E_{q_i}(z)$ (d) $X_{q_i}(z)$

Figure 4: Dynamical planes of iterative methods on $q_i(z) = z^3 + i$.

The dynamical planes of (7) when they are applied to $p_c(z) = z^2 + c$ and $q_c(z) = z^3 + c$, with $c \in \{1, i\}$, are shown in Figure 5. If we focus on the evolution of M4 to M16, passing through M8, it looks like they "tend" to is the Newton's dynamical planes, for each polynomial.



(a) $N_{p_1}(z)$ (b) $N_{p_i}(z)$ (c) $N_{q_1}(z)$ (d) $N_{q_i}(z)$

Figure 5: Dynamical planes of Newton's scheme on $p_c(z)$ and $q_c(z)$.

References

[1] A. M. Ostrowski. Solutions of Equations and Systems of Equations. Academic Press, New York-London (1966).

[2] R. L. Devaney. An Introduction to Chaotic Dynamical Systems. Addison-Wesley Publishing Company (1989).

[3] P. Blanchard. Complex Analytic Dynamics on the Riemann Sphere. Bulletin of the AMS, 11 (1): 85–141 (1984).

[4] H. T. Kung, J. F. Traub. Optimal order of one-point and multi-point iteration Applied Mathematics and Computation, 21: 643–651 (1974).

[5] J. M. Ortega, W. G. Rheinboldt. Iterative solutions of nonlinear equations in several variables. New York, Academic Press (1970).

[6] A. Cordero, J. L. Hueso, E. Martínez, J. R. Torregrosa. A new technique to obtain derivative-free optimal iterative methods for solving nonlinear equations. Journal of Computational and Applied Mathematics, doi: 10.1016/j.cam.2012.03.030 (2012).

New advances on matrix exponential computation for engineering problems*

J. Sastre* ¦ J. Ibáñez†, P. Ruiz†, and E. Defez‡

(*) ITEAM, (†) I3M, (‡) IMM, Universitat Politècnica de València.

November 30, 2012

1 Introduction

Matrix exponential plays a fundamental role in many areas of science [1]. This work improves the scaling and squaring Taylor algorithm **exptayns** from [2] with an improved version of [2, Th. 1] to bound the norm of matrix power series, a new formula for the Taylor approximation forward relative error, and the increase of the allowed bounds for the forward relative error in exact arithmetic with matrix size and approximation order, as Taylor approximation computation roundoff error also tends to increase with both of them. Matlab version http://personales.upv.es/∼jorsasma/software/exptaynsv2.m provided higher accuracy in tests than Padé algorithms from [1, 5] at similar or lower cost. For a complete description and proofs see [3]. Throughout this paper $\mathbb{C}^{n\times n}$ denotes the set of $n \times n$ complex matrices, $\rho(A)$ is the spectral radius of matrix A, $\mathbb{N}$ is the set of positive integers, $\|\cdot\|$ denotes any subordinate matrix norm, and $\|\cdot\|_1$ is the 1-norm. Theorem 1.1 unifies the two cases in which [2, Th. 1] is divided into and avoids needing a bound for $||A^{p_0}||$:

Theorem 1.1 *Let $h_l(x) = \sum_{k\geq l} b_k x^k$ be a power series with radius of convergence w, and let $\tilde{h}_l(x) = \sum_{k\geq l} |b_k| x^k$. For any matrix $A \in \mathbb{C}^{n\times n}$ with $\rho(A) < w$, if a_k is an upper bound for $||A^k||$, $||A^k|| \leq a_k$, $p \in \mathbb{N}$, $1 \leq p \leq l$,*

*Work supported by the *Universitat Politècnica de València* grant PAID-06-11-2020.

†e-mail: jorsasma@iteam.upv.es

$p_0 \in \mathbb{N}$ *is the multiple of* p *with* $l \leq p_0 \leq l+p-1$, *and* $\alpha_p = \max\{a_k^{\frac{1}{k}} : k = p, l, l+1, l+2, \ldots, p_0-1, p_0+1, p_0+2, \ldots, l+p-1\}$, *then* $||h_l(A)|| \leq \tilde{h}_l(\alpha_p)$.

Now we show that Theorem 1.1 can provide sharper bounds than [5, Th. 42]. Using (a) [5, Th. 42] if $\tilde{\alpha}_q = \min\{\max\{\|A^x\|^{\frac{1}{x}}, \|A^{x+1}\|^{\frac{1}{x+1}}\}, x = 2, 3 \ldots, r\} = \max\{\|A^q\|^{1/q}, \|A^p\|^{1/p}\} = \|A^q\|^{1/q}$, where $r = \max\{r_0 : l \geq r_0(r_0-1)\}$, $2 \leq q \leq r+1$, $2 \leq p \leq r+1$, and p can be $p = q-1$ or $p = q+1$ depending on the matrix A, then $\|h_l(A)\| \leq \tilde{h}_l(\tilde{\alpha}_q)$.

By the proof of (a) [5, Th. 4.2] we can write $\|A^k\| \leq \|A^q\|^i \|A^p\|^j, k \geq l$ where $k = qi + pj$ with i and j in $\mathbb{N} \cup \{0\}$, and note that for k non multiple of q one gets $j > 0$. Let t be the lowest multiple of q in $[l, l+p-1]$. Then, if $\|A^p\|^{1/p} < \|A^q\|^{1/q}$ and $t \geq qp$ (note that for the cases where $q \leq r$ and $p = q-1$ one gets $t \geq l \geq r(r-1) \geq qp$), taking $a_p = \|A^p\|$, $a_k = \|A^q\|^u \|A^p\|^q$, $u \in \mathbb{N} \cup \{0\}$ for each value of $k = (u+p)q$ multiple of q in $[l, l+p-1]$, and $a_k = \|A^q\|^i \|A^p\|^j$ for the remaining values of $k \in [l, l+p-1]$ non multiple of q, where $k = qi + pj$ with $j > 0$, one gets $a_k^{1/k} < \|A^q\|^{1/q}$ for $k = p$ and $k \in [l, l+p-1]$. Hence, α_p from Theorem 1.1 satisfies $\alpha_p < \tilde{\alpha}_q$ and then $\|h_l(A)\| \leq \tilde{h}_l(\alpha_p) < \tilde{h}_l(\tilde{\alpha}_q)$.

The same can be proven similarly if $qp > t$ and there exist certain constants $i_1, i_2, \ldots, i_{r+1}$ in $\mathbb{N} \cup \{0\}$ such that $\|A\|^{i_1} \|A^2\|^{i_2} \cdots \|A^{r+1}\|^{i_{r+1}} < \|A^q\|^{t/q}$ where $i_1 + 2i_2 + \ldots + (r+1)i_{r+1} = t$, taking $a_k = \|A^q\|^u \|A\|^{i_1} \|A^2\|^{i_2} \cdots \|A^{r+1}\|^{i_{r+1}}$, $u \in \mathbb{N} \cup \{0\}$ for each value of k multiple of q in $[l, l+p-1]$, $k = qu + t$.

Analogously, it is easy to obtain similar conditions for obtaining sharper bounds than (b) [5, Th. 4.2] with Theorem 1.1 for even h_l.

2 Error analysis

The roundoff error in evaluation of matrix exponential Taylor approximation of the scaled matrix $T_m(2^{-s}A)$ (10) from [2] can be studied by using a componentwise analysis [4, pp. 18-19] giving asymptotically for large n

$$\frac{\|fl\left(T_m(2^{-s}A)\right) - T_m(2^{-s}A)\|_1}{T_m(2^{-s}\|A\|_1)} \leqslant mnu, \quad (u \text{ unit roundoff}). \tag{1}$$

If we denote the actual roundoff error from (1) as $\phi(m,n)u$, and noting that minimum roundoff errors of u are expected, for large n one gets $1 \leq \phi(m,n) \leq mn$. Applying the well-known rule of thumb from [6, p. 52], supported by assuming that the rounding errors are independent random variables, a

realistic error estimate in (1) is $\phi(m,n)=\sqrt{mn}$. However, rounding errors do not necessarily behave like independent random variables [6, p. 52] and sometimes this rule can be pessimistic or optimistic. However, it allowed reducing **exptayns** cost maintaining accuracy in most of test matrices.

If we denote the remainder of the truncated exponential Taylor series of $A\in\mathbb{C}^{n\times n}$ as $R_m(A)=\sum_{k\geq m+1}A^k/k!$, for a scaled matrix $2^{-s}A$ we can write

$$\left(T_m(2^{-s}A)\right)^{2^s}=e^A\left(I+g_{m+1}(2^{-s}A)\right)^{2^s}=e^{A+2^s h_{m+1}\left(2^{-s}A\right)},\quad s\in\mathbb{N}\cup\{0\},\ (2)$$

$$g_{m+1}(2^{-s}A)=-e^{-2^{-s}A}R_m(2^{-s}A),\ h_{m+1}\left(2^{-s}A\right)=\log\left(I+g_{m+1}(2^{-s}A)\right),(3)$$

see [5, sec. 3], where log denotes the principal logarithm and $h_{m+1}(X)$ is defined in the set $\Omega_m=\left\{X\in\mathbb{C}^{n\times n}:\rho\left(e^{-X}T_m(X)-I\right)<1\right\}$, and both $g_{m+1}(2^{-s}A)$ and $h_{m+1}\left(2^{-s}A\right)$ are holomorphic functions of A in Ω_m. The main improvement in the new scaling algorithm with respect to [2] is the selection of the minimum value of s so that, see [3]

$$\left\|h_{m+1}\left(2^{-s}A\right)\right\|\leq\left\|2^{-s}A\right\|u,\ \text{or}\ \ \left\|g_{m+1}\left(2^{-s}A\right)\right\|\leq\phi(m,n)u,\ \phi(m,n)=\sqrt{mn}. \quad (4)$$

The algorithm in [2] is equivalent to taking $\phi(m,n)=1$ in (4) and then it selects higher values of parameters s and/or m for some matrices giving higher costs. Using exponential Taylor series in (3) one can obtain [3]

$$g_{m+1}(2^{-s}A)=-\sum_{k\geqslant 0}\frac{(-1)^k(2^{-s}A)^{m+1+k}}{k!m!\,(m+1+k)}, \quad (5)$$

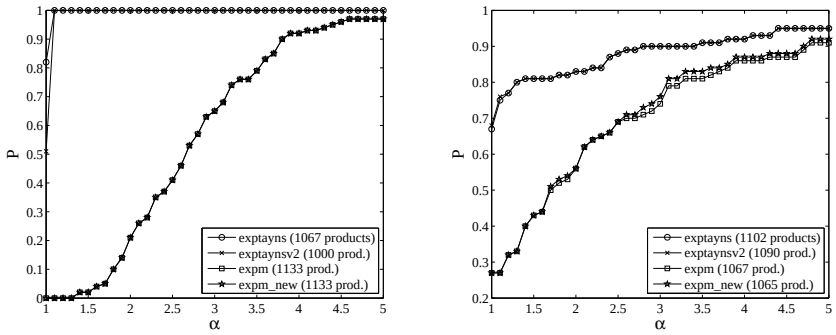
and using α_{min} from [2, Sec. 2.2] for $q\in\mathbb{N}\cup\{0\}$ it is possible to obtain [3]

$$\begin{aligned}||g_{m+1}(2^{-s}A)||\leq&\left\|\sum_{k=0}^{k=q}\frac{(-1)^k(2^{-s}A)^{m+1+k}}{m!k!\,(m+1+k)}\right\|+\frac{\|(2^{-s}A)^{m+q+2}\|}{m!(q+1)!(m+q+2)}\\&+\frac{(2^{-s}\alpha_{min})^{m+1}}{m!(m+q+3)}\left[e^{2^{-s}\alpha_{min}}-\sum_{k=0}^{q+1}\frac{(2^{-s}\alpha_{min})^k}{k!}\right], \quad (6)\end{aligned}$$

and accurate enough approximations to this new bound can be computed using Matlab function **exp** for the values of q and $2^{-s}\alpha_{min}$ used in [3].

3 Numerical experiments

Figure 1 shows the cost in terms of matrix products, taking 4/3 matrix products for solving the multiple linear system in Padé algorithms [2], and the performance profiles [1] of **exptaynsv2**, **exptayns**, **expm** [1] and **expm_new**



(a) Diagonalizable 128. (b) Multiple eigenvalues 256.

Figure 1: Performance profile and cost using $m_M = 30$ in Taylor functions.

[5] for two sets of 100 random 128×128 diagonalizable and 256×256 multiple eigenvalued matrices, where α coordinate varies between 1 and 5 in steps equal to 0.1, and p coordinate is the probability that the considered method has a relative error E lower than or equal to α-times the smallest error over all the methods, where probabilities are defined over all matrices and $\mathrm{E} = \|e^A - \tilde{Y}\|_1 / \left\|e^A\right\|_1$, where $\tilde{Y}$ is the approximation and e^A was computed using high precision arithmetic. Taylor functions had the highest accuracy, with **exptaynsv2** cost similar to (2.34% greater) or up to 11.74% lower than **expm_new**. See [3] for tests with other matrix sets.

References

[1] N.J. Higham, Functions of Matrices: Theory and Computation, SIAM, USA, 2008.

[2] J. Sastre, J. Ibáñez, E. Defez and P. Ruiz, Accurate matrix exponential computation to solve coupled differential models in engineering, *Math. Comput. Model.*, 54:1835-1840, 2011.

[3] J. Sastre, J. Ibáñez, P. Ruiz and E. Defez, Accurate and efficient matrix exponential computation, Submitted to *Int. J. Comput. Math.*

[4] C. Fassino, Computation of Matrix Functions, PhD Thesis TD-7/93, Università di Pisa, Genova, 1993.

[5] A.H. Al-Mohy, N.J. Higham, A new scaling and squaring algorithm for the matrix exponential, *SIAM J. Matrix Anal. Appl.* 31(3):970–989, 2009.

[6] N.J. Higham. Accuracy and Stability of Numerical Algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002.

Mathematical modeling of the propagation of fitness practice in Spain

M. Alkasadi[†] , E. De la Poza[‡] [*], L. Jódar[†] and A. Pricop[†]

([†])Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València. Camino de Vera s/n. 46022, Valencia, España,

([‡])Facultad de Administración y Dirección de Empresas, Universitat Politècnica de València,

Universitat Politècnica de València, Building 7J. Camino de Vera s/n. 46022, Valencia, España.

November 30, 2012

1 Introduction

Our society is concerned with people's physical appearance and an ideal body image. As a result of this trend, the physical fitness prevalence especially among men has extended which leads to health benefits. However, its excess can be associated to the development of psycho-medical diseases such as bigorexia or a disorder in which the person becomes obsessed with the idea that he or she is not muscular enough. This disorder affects mainly to men, [1]. The individuals who suffer from it think their body shape and size is skinny or small,[2]. They develop obsessive and negative behaviors related with their appearance [3], affecting not only the adult men, but also young boys [4].

The aim of this study is to develop a discrete population mathematical model to forecast the future bigorexia addicts in Spain in the next five years. For this purpose, economic and psychological motivations are taken into account in order to quantify the dynamic behavior of men gym users. Personal and

[*]elpopla@esp.upv.es:

social consequences of this addiction are studied and public health recommendations are suggested.

2 Mathematical model and results

Our population objective consists of Spanish men who attend the gym, aged in the interval [18, 40].

We classified the population into three categories depending on their level of activity and emotional dependence to fitness practice both measured through their score obtained from the adapted questionnaire of muscle dysmorphia[5].

We defined three subpopulations for building the mathematical model:
N_n: gym users at year n whose emotional score was equal or lower than 2 points.
S_n: gym users at year n whose emotional score was equal to 3 or 4 points.
A_n: gym users at year n whose emotional score was higher than 5 points.

The individuals can only transit from one category or subpopulation to another. Thus, the N can transit to S and S to A. Also, it is assumed a possible recovery transit from S to N. The drivers considered in the development of bigorexia are the influence of personal experiences during childhood (bullying)[6], sexual orientation [7] and also private life situations (such as divorce) [8]. Secondly, the contagion effect produced by the influence of personal relationships between gym users, especially with addicts, it is a determinant of people's behaviors' [9]. Also, there is the effect of opposite drivers such as the emotional factors and the economic worsening. Finally, the influence of an economic improvement (decrease of unemployment rate) combined with the rebuilt of the personal life can produce a recovery transit from an over-gym user (S) to a normal gym one (N),[10].

The individuals come into the model at year "n" as N becoming gym users older than 17 years old or by an economic improvement. On contrast, the gym users leave the model at year "n" when they get older than 40 years old, pass away and also if they emigrate abroad giving up their fitness practices. Also, N can leave the model due to the economic crisis, decreasing their gym attendance, while S and A will not decrease their gym practice.

The dynamic of population can be described by the following equations:

$$P_n = N_n + S_n + A_n$$

$$N_{n+1} - N_n = b_1 N_n - d_1 N_n + \alpha_2(n) S_n - \alpha_e N_n - \gamma_1 N_n - \alpha_f N_n + \alpha_u N_n$$

$$S_{n+1} - S_n = b_2 S_n - d_2 S_n + \alpha_e N_n - \alpha_2(n) S_n + \gamma_1 N_n - \alpha_t(n) S_n - \gamma_2 S_n$$

$$A_{n+1} - A_n = -d_3 A_n + \alpha_t(n) S_n + \gamma_2 S_n \quad (1)$$

The values of all parameters were estimated from different sources of information and hypothesis with the exception of the transit rate γ_1 that was adjusted by the model, considering the assumptions of previous studies of social contagious[8, 11] also we applied the hypothesis $\gamma_2 = 2 * \gamma_1$.
We simulated four possible scenarios according to different annual levels of unemployment based on the economic forecast from OECD and FUNCAS for 2012 and 2013. For 2014 and 2015 we built a double economic scenario, one more optimistic than the other, named OECD+, FUNCAS+ and OECD-, FUNCAS-.

Then, the results were obtained by the computation of the system. The percentage of addicts increases over time independently of the economic scenario, with scarce differences between them. In particular the rate of addicts in 2015 was in the interval [11.53%, 11.60%]what confirms the robustness of our model.

3 Conclusions

Results show that the rate of addicts evolves robustly to economic changes from 1% in 2011 to over 11% in 2015. The intensive gym practice can produce muscle dysmorphia as the abuse of dangerous substances as anabolic steroids. We can conclude some recommendations to public health authorities and families, as it is to pay attention to young male sportive practices, encouraging the practice of team sports.Finally, we should teach teenagers how to spend their spare time healthily.

References

[1] K.A.Phillips, Understanding body dysmorphic disorder an essential guide. New York,Oxford University press ,: 49, 2009.

[2] J.L.Rubio, and J.I.Baile."Tendencia vigorexia en un grupo de usuarios de gimnasio". Congreso de la Sociedad Iberoamericana de Psicología delDeporte, México, 2006.

[3] C. D.Lantz, D. J.Rhea, and J. L.Mayhew. The Drive for size: Apsycho - behavioral model of muscle dysmorphia *International Journal of Sports sciences*, 5(1):71-86, 2001.

[4] G. H.Cohane, and H. G., Jr.Pope. Body image in boys:A review of the literature *International Journal of Eating Disorders*, 26(4): 373-379, 2001.

[5] J.L.Baile,Vigorexia Cómo Reconocerla y Evitarla. Madrid,Editorial Sintesis, 2005.

[6] D.Wolke, and M.Sapouna. Big men feeling small: Childhood bullying experience, muscle dysmorphia and other mental health problems in bodybuilders *Journal of Psychology of Sport and Exercise*, 9:595-604, 2008.

[7] J.Brown, and D.Graham. Body satisfaction in Gym-active males: An exploration of sexuality, gender, and narcissism *Journal of Sex Roles*, 59 :94-106, 2008.

[8] R.Duato, and L.Jódar. Mathematical modeling of the spread of divorce in Spain *Mathematical and Computer Modelling*, http:// dx.doi.org/10.1016/j.mcm.2011.11.020, 2011.

[9] N.A.Christakis, and J.H.Fowler, Connected: The Surprising Power of Our Social Networks and How they Shape Our Lives, Hachette Book Group, 2009.

[10] B.M.Popkin. The nutrition transition in the developing world. *Journal of Development Policy Review*,21(5-6):581-597,2003.

[11] I.García, L. Jódar, P.Merello, and F.J.Santonja. A discrete mathematical model for addictive buying:predicting the affected population evolution *Mathatematical Computer Modelling*, 54 (7-8): 1634-1637, 2011.

Wavelet based approach for singular perturbation problems

Dana Černá and Václav Finěk*

Department of Mathematics and Didactics of Mathematics, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic.

November 30, 2012

1 Introduction

We consider here the following singularly perturbed two-point boundary value problem

$$-\epsilon u''(x) + p(x)u'(x) + q(x)u(x) = f(x) \quad \forall x \in (a,b) \text{ and } u(a) = \alpha,\ u(b) = \beta,$$

where a, b, α, β are constants and $0 < \epsilon \ll 1$. This equation represents simple mathematical model of a convection-diffusion problem and it can be used to model many practical problems. For example linearized Navier-Stokes equations at high Reynolds number provides an accurate model of dynamics of transition in the problem of turbulence suppression in channel flow. Further applications can be found in survey paper [6]. Problems of this types have solutions which are discontinuous as ϵ is approaching to zero and typically possess boundary or interior layers, i.e. regions of rapid change in the solution near the endpoints or some interior points. Many numerical methods have been suggested to solve such types of problems and a lot of them requires informations about locations and widths of different layers. One of possibilities, how to solve them without these informations, is an application of adaptive methods. We employ here an asymptotically optimal

*e-mail: Vaclav.Finek@tul.cz

adaptive wavelet scheme. It generates the approximate solution comparable with the best N-term approximation and the number of arithmetic operations needed to compute this solution is proportional to N. Next advantage of this scheme consists in the efficient diagonal preconditioning of stiffness matrices. To further improve the quantitative properties of the used scheme for small values ϵ, we propose an improved diagonal preconditioning.

2 Wavelet based methods

To use wavelets for the solution of differential equations has several advantages, namely:

- Due to the vanishing wavelet moments (the cancellation property) the adaptivity in the context of a wavelet discretization is simple. It consists in keeping the large wavelet coefficients and discarding the smaller ones. Moreover vanishing moments lead to sparse representations of functions.

- There are asymptotically optimal algorithms. See for example [2, 3, 4, 5]. It means that the number of floating point operations depends linearly on the number of nonzero wavelet coefficients.

- They allow to characterize various function spaces such as Sobolev or Besov spaces by weighted sequence norms of the corresponding wavelet coefficient. A consequence of these equivalences between function norms and weighted sequence norms is efficient diagonal preconditioning for stiffness matrices.

These properties of wavelets can be exploited considerably in numerical solution of differential equations. To be able to solve realistic problems, it is necessary to use adaptive methods with highly nonuniform meshes to keep the number of unknowns at a reasonable level. Key ingredients of adaptive methods are a posteriori error estimators and adaptive refinement strategies. In wavelet methods, wavelet expansions of the current residual can serve as reliable a posteriori estimators. One of the fundamental adaptive wavelet refinement strategies is based on iterations in the infinite-dimensional spaces which are given by choosing accuracies in an approximation of the right-hand side and in an approximation of exact matrix-vector multiplications. New

elements of the unknown solution are then generated by increasing accuracy in both approximations.

3 Results and conclusion

Using wavelet discretization of singularly perturbed two-point boundary value problems, one can observe that condition numbers of arising stiffness matrices are growing with decreasing parameter ϵ when an unsymmetric part starts to dominate. We propose the diagonal preconditioning which significantly improves condition numbers of stiffness matrices for small value of parameter ϵ. For $\epsilon < 10^{-8}$, proposed preconditioning leads to significantly smaller condition numbers of stiffness matrices in wavelet coordinates and also to smaller numbers of used iterations. For more details, we refer to [1].

Acknowledgements

V. Finěk has been supported by the SGS project "Construction of multiwavelets for numerical solution of differential equations" and special thanks belongs to Z. Ondračková for her help with numerical experiments, D. Černá and V. Finěk have been partially supported by the project ESF "Constitution and improvement of a team for demanding technical computations on parallel computers at TU Liberec" No. CZ.1.07/2.3.00/09.0155.

References

[1] D. Černá, V. Finěk, *Wavelet based approach for singular perturbation problems*, submitted.

[2] D. Černá, V. Finěk, *Approximate Multiplication in Adaptive Wavelet Methods*, to appear in CEJM.

[3] A. Cohen, W. Dahmen and R. DeVore, *Adaptive Wavelet Schemes for Elliptic Operator Equations - Convergence Rates*, Math. Comput. 70 (2001), pp. 27–75.

[4] A. Cohen, W. Dahmen and R. DeVore, *Adaptive Wavelet Methods II - Beyond the elliptic case*, Found. Math. 2 (2002), pp. 203–245.

[5] T. J. Dijkema, Ch. Schwab, and R. Stevenson, *An adaptive wavelet method for solving high-dimensional elliptic PDEs*, Constr. appr. 30 (2009), pp. 423–455.

[6] M.K. Kadalbajoo, and V. Gupta, *A brief survey on numerical methods for solving singularly perturbed problems*, Appl. Math. and Comp. 217 (2010), pp. 3641–3716.

A new method for the simulation of non-linear parabolic equations in cylindrical coordinates.

V. Macián, A. Gil, J.P.G. Galache, I. Blanquer

CMT-Motores Térmicos, Universitat Politècnica de València

Valencia 46022, Spain,

November 30, 2012

Elliptic and parabolic equations are widely used in science and engineering. The problems to be solved for these disciplines becomes bigger and complexer as time goes on. To be able to achieve them, there are two ways. In one hand, a higher performance machine can be used; in the other hand there is the algorithm improvement. Of course, a mix of both of them gives the best results. In this paper it is presented a new method to solve elliptic and parabolic equations in circular domains. The algorithm uses a mixed Fourier-Compact Finite Difference method, whose main advantage is achieved by a new concept of mesh. The topology of this grid keeps constant the aspect ratio of the cells, avoiding the typical clustering for radial structured meshes at the center. The reduction of the number of nodes has as a consequence the reduction in memory consumption. Particularizing for fluid mechanics problems, this technique also increases the time step for a constant Courant number.

The resolution of non-linear parabolic and elliptic equations in circular domains are of great interest for several branches of knowledge. Particularising in a specific case, the understanding of the kinematics and dynamic of turbulent pipes is one of the challenges of the next decade. As an example, 50% of the energy losses in big pipes are originated in the first millimetres from the wall. About this issue, see for instance [1, 2, 3]. Spectral or spectral-

like methods are frequently chosen to solve these sort of problems, due to both their great precision and their high mesh size - computational cost ratio. These methods, when applied on circular domains, are usually formulated in polar coordinates. One of the main advantages of this is that the boundary conditions can be straightforwardly imposed. Nevertheless, there are two particularities to be deal with. In one hand, the origin is a pole and it needs a special treatment, usually working with artificial boundary conditions. In the other hand, for structured radial grid, there is a mesh size reduction at the centre. If dealing with a turbulent flow, for instance, the shorter structures, those that define the mesh structure, are close to the wall, whereas the greater ones are at the centre of the pipe. Apart from the memory requirements that this may cause, Courant-Friedrichs-Lewy condition (CFL from now on) would impose a very small time step.

The problem at the origin has been solved in different ways. Chen et. al. [4] used spectral collocation methods in order to increase the order of the equation. Li et. al. [5] simulated Navier-Stokes equations with three Chebycheff-Fourier spectral collocation methods. Pure spectral methods are used too, as Z. Qiu et. al. [6], who used Fourier-Legendre discretization. Matsushima and Marcus [7] used Fourier-generic orthogonal polynomials. A coordinate system transformation is another possible technique. Heinrichs [8] used conformal mapping to convert a Cartesian coordinate system in a polar system. Something similar was done by Hansen et. al. [9]. The transformation was made at the nodes closest to the centre; the components, already in Cartesian form, are averaged. The result is the value of the pole. Regarding this work, the same method followed by Lai [10] is employed. In Lai's paper, a Fourier-Compact Finite Difference (CFD) discretization [11] was used, with symmetry and antisymmetry conditions at the middle (asymptotic behavior) applied on phantom points for odd and even values of each wave number.

The accumulation of grid points at the centre of the domain can also be compensated. Kwan [12] proposed a spectral-Galerkin method with a quadratic transform in the radial direction to improve the clustering at the center. Akselvoll and Moin [13] solved Navier-Stokes dividing the circular domain in two separated regions, the "core region" and the "outer region". At the "core region" some of the terms are treated explicitly, whereas at the "outer region" all the terms are solved implicitly. They use different time

schemes for each region to improve the performance without modifying the accuracy. In this paper, a new algorithm is presented. It uses a sixth-order CFD method in the radial direction, and spectral Fourier in the azimuthal one. When applied to an elliptic equation, the algorithm produces a set of 1D radial equations. Each equation system is represented by a compact sixth-order finite difference discretization. As a consequence, the error is bounded by the maximum wave number and the radial spacing. The aim of this work is to optimize the CFD-Fourier refinement to solve the system as accurately as possible, at the same time than the computational resources requirements are minimized. This objective is achieved by the creation of a structured radial mesh with thresholded aspect-ratio and compatible with the Fourier decomposition in azimuthal direction. The main characteristic of the grid is the reduction on the number of computed wave lengths for the internal radii. As a consequence, the memory storage is halved and the computational time is reduced by some orders of magnitude.

References

[1] O. Shishkina, C. Wagner, "A fourth order finite volume scheme for turbulent flow simulations in cylindrical domains", Computers & Fluids **36** (2) (2007) 484–497

[2] X. Wu, P. Moin, "A direct numerical simulation study on the mean velocity characteristics in turbulent pipe flow", Journal of Fluid Mechanics **608** (2008) 81–112

[3] C. Chin, A. S. H. Ooi, I. Marusic, H. M. Blackburn, "The influence of pipe length on turbulence statistics computed from direct numerical simulation data", Physics of Fluids **22** (11)

[4] H. Chen, Y. Su, B. D. Shizgal, "A direct spectral collocation Poisson solver in polar and cylindrical coordinates", Journal of Computational Physics **160** (2) (2000) 453–469.

[5] B. W. Li, Y. R. Zhao, Y. Yu, Z.-D. Qian, "Three-dimensional transient Navier-Stokes solvers in cylindrical coordinate system based on a spectral collocation method using explicit treatment of the pressure",

International Journal for Numerical Methods in Fluids **66** (3) (2011) 284–298.

[6] Z. Qiu, Z. Zeng, H. Mei, L. Li, L. Yao, L. Zhang, "A Fourier–Legendre spectral element method in polar coordinates", Journal of Computational Physics **231** (2) (2012) 666–675.

[7] T. Matsushima, P. S. Marcus, "A spectral method for polar coordinates", Journal of Computational Physics **120** (2) (1995) 365–374.

[8] W. Heinrichs, "Spectral collocation schemes on the unit disc", Journal of Computational Physics **199** (1) (2004) 66–86.

[9] M. O. L. Hansen, J. N. Sørensen, W. Z. Shen, "Vorticity-velocity formulation of the 3D Navier-Stokes equations in cylindrical coordinates", International Journal for Numerical Methods in Fluids **41** (1) (2003) 29–45.

[10] M. C. Lai, "A simple compact fourth-order poisson solver on polar geometry", Journal of Computational Physics **182** (1) (2002) 337–345, cited By (since 1996) 15.

[11] S. Lele, "Compact finite-difference schemes with spectral-like resolution", Journal of Computational Physics **103** (1) (1992) 16–42

[12] Y. Y. Kwan, "Efficient spectral-galerkin methods for polar and cylindrical geometries", Applied Numerical Mathematics **59** (1) (2009) 170–186.

[13] K. Akselvoll, P. Moin, "An efficient method for temporal integration of the Navier-Stokes equations in confined axisymmetric geometries", Journal of Computational Physics **125** (2) (1996) 454–463.

[14] P. R. Spalart, R. Moser, M. Rogers, "Spectral methods for the Navier-Stokes equations with one infinite and two periodic directions", Journal of Computational Physics **96** (2) (1991) 297–324.

[15] S. Hoyas, J. Jiménez, "Scaling of the velocity fluctuations in turbulent channels up to Re_τ=2003", Physics of Fluids, **18**, 011702, (2006).

Electricity demand forecasting with multiple seasonal patterns: an application to Spanish data

J.Carlos García-Díaz* and Oscar Trull†

Applied Statistics, Operations Research and Quality Department.

Universitat Politècnica de València

November 30, 2012

1 Introduction

Electricity power supply systems and their distribution network operators must provide every day short-term electricity demand forecasts for an efficient electricity production management. Electricity cannot be saved in big quantities, thus they must match their production to the real demand. But the production must be established at least the day before, so they need to predict the demand based on the experience. Time Series forecasting has become a very powerful tool for such purpose, and in special the exponential smoothing methods[1, 2, 3]. Hence Holt-Winters exponential smoothing[4] have been widely used, basically due to its simplicity and robustness.

Newer works[5, 6] have opened a new stage in exponential smoothing using multiple seasonal patterns for making forecasts. In this paper we analyse the performance of double seasonal Holt-Winters models while forecasting Spanish electricity demand, based on multi-step ahead forecast mean squared errors.

Table 1: Methods classification. This frameworks gathers all implemented models. The first letter determines the trend, N:none, A:additive,d:damped additive, M:multiplicative and D:damped multiplicative. The second letter is used for Seasonality, with only N,A and M option. The third letter is used for the AR(1) adjustment, L:no adjustment and C:adjusted.

Seasonality / Trend	Normal			Corrected		
	None	Additive	Multiplicative	None	Additive	Multiplicative
None	NNL	NAL	NML	NNC	NAC	NMC
Additive	ANL	AAL	AML	ANC	AAC	AMC
Damped Additive	dNL	dAL	dML	dNC	dAC	dMC
Multiplicative	MNL	MAL	MML	MNC	MAC	MMC
Damped Multiplicative	DNL	DAL	DML	DNC	DAC	DMC

2 Implemented models

The Double Seasonal Holt Winters method (HWT) was first presented by Taylor[5], which handles series with more than one seasonal pattern. An evolution to Triple Seasonal Holt-Winters methods were presented in [6].

A generalisation of the model is shown in equations (1) to (4). X_t are the observed values, S_t is the level, T_t is the trend, $I_t^{(i)}$ are the seasonal indexes for seasonal pattern i.$\hat{X}_t(k)$ are the k-step ahead forecasted values. The involved parameters α,γ are the smoothing parameters for the level and trend. $\delta^{(i)}$ are the smoothing parameters each seasonal index. s_i represent the period length each seasonal pattern.

$$S_t = \alpha \left(\frac{X_t}{\prod_i I_{t-s_i}^{(i)}} \right) + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{1}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{2}$$

$$I_t^{(i)} = \delta^{(i)} \left(\frac{X_t}{S_t \prod_{j \neq i} I_{t-s_j}^{(j)}} \right) + (1-\delta^{(i)}) I_{t-s_i}^{(i)} \tag{3}$$

$$\hat{X}_t(k) = (S_t + kT_t) \prod_i I_{t-s_i+k}^{(i)} \tag{4}$$

This model apply for an additive trend and multiplicative seasonality method. Table 1 summarises all methods from Pegel's classification. Additionally, Chatfield[7] showed in some cases the first-order autocorrelation error adjustment(AR(1) adjustment) improve the results. The regular expression is shown in (5), where φ_{AR}

*e-mail: juagardi@eio.upv.es
†e-mail: otrull@adp.com.es

is the parameter for that adjustment, and e_t is the one-step ahead error.ε_t stands for white noise.

$$\hat{X}_t(k, \varphi_{AR}) = \hat{X}_t(k) - \varphi_{AR} e_t + \varepsilon_t \tag{5}$$

These methods are included at the right side of Table 1. The notation used is very simple: the first letter indicates the trend model, the second letter the seasonality and the third whether it has AR(1) adjustment(C) or not(L).

3 Spanish Case Study

Empirical analysis is performed to compare the forecasting accuracy of the models. We report the results of the implementation of model (1 - 4) for the hourly Spanish electricity demand. We set $s_1 = 24$ to model the within-day seasonal pattern of 24 hours (intraday cycle), and $s_2 = 168$ to model the within-week seasonal pattern of 168 hours (intraweek cycle). The data set used values from June 23, 2009 to four weeks later compiled by the operator Red Eléctrica Española.

Within the double seasonal additive methods, the non-trend(NAC) or damped additive trend(dAC) was found to be the best for multi-step ahead forecasting. Forecasting accuracy was improved by the inclusion of first-order autocorrelation error adjustment, confirming Taylor's known results.

We show the comparison between its forecasts versus observed values in Figure 1. as well as the whole model along the time in Figure 2 to check the model performance.

4 Conclusion

In this paper, we developed an automatic procedure implemented in the MATLAB®software package based on univariate multiple seasonal Holt-Winters models, by generalising the Taylor's proposed methods to n-seasonal patterns.

To illustrate the availability of the developed computer software for the short-term demand forecasting, an application for hourly Spanish electricity demand is presented. A comparison of the AIC and sMAPE of several models was also carried out. We found that, in general, the models with the best in-sample as well as out-of-sample forecasting performance of the hourly Spanish electricity demand are those that do not include a trend. Forecasting accuracy was improved by the inclusion of the first-order autocorrelation error adjustment. One of the conclusions of the study was that the accuracies of various methods depend upon

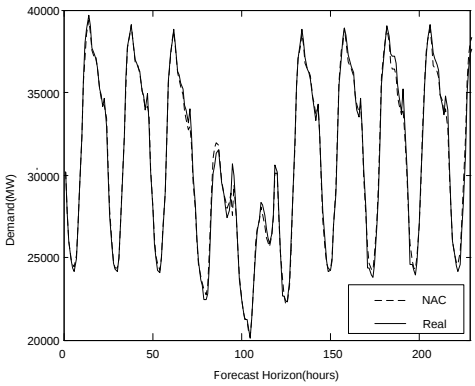


Figure 1: NAC model forecasted values versus observed values comparison graph.

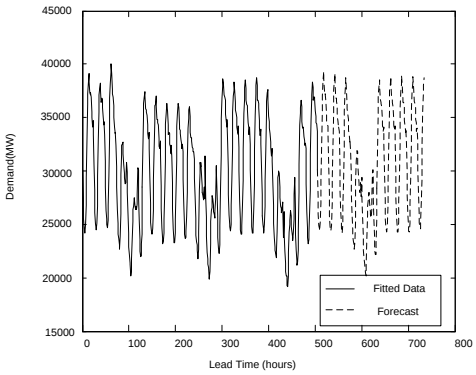


Figure 2: Hourly Spanish electricity demand. Actual data (solid), and two weeks ahead forecast produced by NAC model (dashed).

the length of the forecasting horizon involved. The improvement is stabilised with the forecasting horizon.

The double seasonal additive methods outperformed the double seasonal multiplicative methods. Within the double seasonal additive methods, the non-trend (NAC) or damped additive trend (dAC) methods was found to be the best for multi-step ahead forecasting. The empirical results suggest that the seasonality additive model without trend with first-order autocorrelation error adjustment (NAC) is the model with the best predictive ability to Spanish electricity short-term demand forecasting. AIC criteria guarantee that those models are the best without overfitting. Forecasting accuracy was improved by the inclusion of first-order autocorrelation error adjustment, confirming Taylor's known results.

References

[1] C. Chatfield and M. Yar. Holt-Winters forecasting: some practical issues *The Statistician*, Volume 37:129–140, 1998.

[2] E.S. Gardner Exponential smoothing: The state of the art *Journal of Forecasting*, Volume 4:1–28, 1985.

[3] E.S.Gardner. Jr. Exponential smoothing: The state of the art,part ii, *International Journal of Forecasting, Volume22:637–666, 2008.*

[4] P.R. Winters Forecasting sales by exponentially weighted moving averages Management Science, Volume6:324–342, 1960.

[5] J.W. Taylor, Short-term electricity demand forecasting using double seasonal exponential smoothing *Journal of Operational Research Society, Volume54:799–805, 2003.*

[6] J.W. Taylor, Triple seasonal methods for short-term electricity demand forecasting European Journal of Operational Research, Volume204:139–152, 2010.

[7] C. Chatfield, Comments on exponential smoothing: The state of the art by e. s. gardner jr. *Journal of Forecasting, Volume4:30–30, 1985.*

Computing Survival Functions of the Sum of Two Independent Markov Processes. An Application to Bladder Carcinoma Treatment

B. García–Mora[†*], C. Santamaría[†], G. Rubio[†] and J. L. Pontones[‡]

(†) Instituto Universitario de Matemática Multidisciplinar

Building 8G, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia (Spain).

(‡) Hospital Universitari i Politècnic La Fe, Bulevar Sur s/n, 46026 Valencia (Spain).

November 30 2012

.

1 Introduction

Many situations in the fields of biomedical research and engineering can be modeled by stochastic processes that evolve through different states over time. The nature of the problem may suggest the decomposition of the overall process into two (or more) subprocesses. We have focused on situations that can be described with two (or more) consecutive homogeneous Markov processes, each one with an absorbing state and several transient states. The aim is to study new aspects in the modeling of bladder cancer, but the approach can be applied to other areas.

Bladder carcinoma is a highly aggressive neoplasm and the second most common malignancy. Approximately 80% of patients present non muscle invasive bladder cancer (NMIBC), which can be managed with transurethral resection (TUR), a surgical endoscopic technique. However, more than 50% of the patients will have *recurrences* (reappearance of a new superficial tu-

*e-mail: magarmo5@imm.upv.es

mor) and 10-30% of patients will have *progression* (MIBC: muscle invasive bladder cancer) with the possibility of the bladder extirpation.

We are interested in to establish predictions of recurrences and progression after the *TUR* of the *primary tumor*. In this regard, *Markov models* have proven to be useful in the study of the course of chronic diseases. For our application two independent database of bladder carcinoma were avalaible. This suggested to consider two independent Markov processes, each one with an absorbing state: the progression to a MIBC in the first process and the extirpation of the bladder in the second one.

2 The distribution function of the sum of two Markov processes

Let X_1 and X_2 be nonnegative random independent variables representing the absorption times in two homogeneous Markov processes. We assume that both variables are *Phase-Type (PH)*-distributed with representation (α, T) and (β, S) respectively and distribution function $F_1(\cdot)$ and $F_2(\cdot)$. A *Phase-Type (PH) distribution* is the absorption distribution time in a homogeneous Markov process in a finite state space with one absorbing state [1, Chapter 2]. We develop the distribution function of the sum variable $X = X_1 + X_2$, also PH–distributed, making use of the Fréchet derivative and the Kronecker matrix form. We arrive to the following expression

$$\begin{aligned} F(x) &= 1 - \alpha\, \exp(Tx)\mathbf{e_m} - \alpha_{m+1}\beta \exp(Sx)\mathbf{e_n} - \\ &\quad (\mathbf{e_n}' \otimes \alpha)[S'x \oplus (-Tx)]^{-1}(\exp(S'x) \otimes \mathrm{I_m} - \mathrm{I_n} \otimes \exp(Tx))\mathrm{vec}(T^0\beta x) \end{aligned}$$

3 An application to Bladder carcinoma

We apply the model to the sum of two different periods of bladder carcinoma according to the aggressiveness of the tumors (see Diagram). In the *first period* there are three transient states and one absorbing state: the *primary non-muscle invasive tumor* (NMIBC), a *first recurrence* and a *second recurrence* (reappearance of a new NMIBC similar to the primary tumor). The absorbing state is the appearance of a *muscle invasive tumor* (MIBC) much

more aggressive. The *second period* has two transient states and one absorbing state: two *muscle invasive tumors* (MIBC) and the *bladder extirpation*.

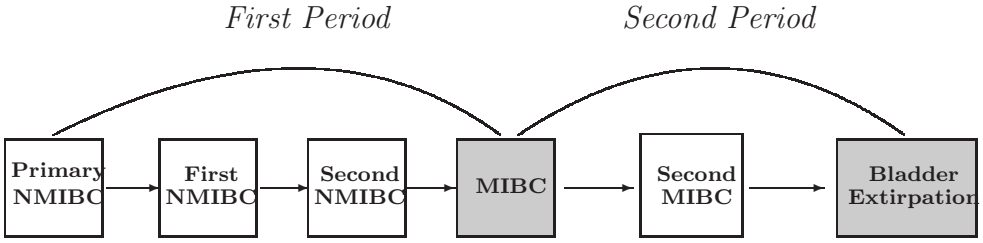


Diagram: Sum of two Markov processes in bladder carcinoma.

Two independent databases, of 526 and 106 patients respectively, have been considered from the Department of Urology at *La Fe University Hospital* in Valencia (Spain), covering each of these two periods of the disease with two well differentiated follow-up protocols. Both bases are ranged from 1995 to 2009. The patient transitions between states are shown in the Table 1.

Table 1: Transitions between states for the two periods.

	First Period					**Second Period**		
	Prim NMIBC	1 NMIBC	2 NMIBC	MIBC		1 MIBC	2 MIBC	Bladder Extirp.
Prim NMIBC	306	216	0	4	1 MIBC	59	38	9
1 NMIBC	0	77	102	5	2 MIBC	0	16	5
2 NMIBC	0	0	86	2				

Let X_1 be the time from the primary NMIBC until the first MIBC and X_2 the time from the first MIBC until the *bladder extirpation*. The initial probabilities are $\alpha = (1, 0, 0)$ and $\beta = (1, 0)$. For both periods, patients are initially in the primary NMIBC and the first MIBC respectively, and the initial probability of being in an absorbing state is zero, so $\alpha_4 = \beta_3 = 0$.

The random sum variable $X = X_1 + X_2$ is the total time duration from the primary NMIBC until the *bladder extirpation* passing through the first MIBC (first absorbing state). The new Markov process is structured as follows: a primary NMIBC, a first and a second recurrences of a NMIBC,

the first MIBC, a second MIBC and the *bladder extirpation* where this last one is the only absorbing state of the whole of the process (see Diagram).

The variable X is PH-distributed and we compute the survival function $S(x) = 1 - F(x)$ for X. The probabilities that a patient does not undergo a bladder extirpation have been calculated for 3 years, $S(36) = 0.99669$, and 5 years approximately, $S(60) = 0.98778$. Although the survival function has been calculated for high risk patients, (that is to say, with worse prognosis), these survival probabilities are high. This result is coherent with only between 10-30% of patients after *TUR* will evolve to a MIBC with a possibly later of *bladder extirpation* [2], so the probability to reach the absorbing state ofthe whole process is small. In fact, in our case, only 11 patients of 526 in the first cohort reached a MIBC.

4 Concluding remarks

Many authors have considered the mixture of distribution functions as a procedure to obtain new distributions but, in general, the analytic expression of the mixture is not mathematically manageable. In this context, the approach of modeling the process by considering several sections strongly benefits from the use of phase-type distributions. The expression of the distribution function $F(x)$ is easily tractable, and the fact that convolution of PH-distributions leads again to a PH-distribution is algorithmically useful.

Acknowledgement

This study has been funded by the *Vicerrectorado de Investigación de la Universitat Politècnica de València. Code 2406.*

References

[1] M. F. Neuts, Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach, John Hopkins University Press, Baltimore, 1981.

[2] S. Hölmang, H. Hedelin, C. Anderström and S. L. Johansson. The relationship among multiple recurrences, progression and prognosis of patients with stage ta and t1 transitional cell cancer of the bladder followed for at least 20 years, *J. Urol.*, 153: 1823–1827, 1995.

Support vector machines and multilayer perceptron networks used to evaluate the cyanotoxins presence from experimental cyanobacteria concentrations in the Trasona reservoir (Northern Spain)

P.J. García Nieto[*] [,*], J.A. Vilán Vilán[†], J.R. Alonso Fernández[‡], F. Sánchez Lasheras[+], F.J. de Cos Juez[**], and C. Díaz Muñiz[‡]

(*) Department of Mathematics, Faculty of Sciences,

University of Oviedo, 33007 Oviedo, Spain,

(†) Department of Mechanical Engineering,

University of Vigo, 36200 Vigo, Spain,

(‡) Cantabrian Basin Authority,

Spanish Ministry of Agriculture, Food and Environment, 33071 Oviedo, Spain,

(+) Department of Construction and Manufacturing Engineering,

University of Oviedo, 33204 Gijón, Spain,

(**) Mining Exploitation and Prospecting Department,

University of Oviedo, 33004 Oviedo, Spain.

November 30, 2012

Abstract

The aim of this study is to build a cyanotoxin diagnostic model by using support vector machines and multilayer perceptron networks from experimental cyanobacteria concentrations in Trasona reservoir

[*]e-mail: lato@orion.ciencias.uniovi.es

(recreational reservoir used as a high performance training centre of canoeing in the Northern Spain). For this purpose, some biological parameters (phytoplankton species expressed in biovolume and the chlorophyll concentration) in combination with the most important physical-chemical parameters are considered. The results of the present study are two-fold. In the first place, the significance of each biological and physical-chemical variables on the cyanotoxins presence in the reservoir is presented through the model. Secondly, a predictive model able to forecast the possible presence of cyanotoxins is obtained. The agreement of the model with experimental data confirmed its good performance. Finally, conclusions of this innovative research work are exposed.

Keywords: Statistical learning techniques; Cyanobacteria; Cyanotoxins; Support vector machines (SVM); Multilayer perceptron networks (MLP); Regression analysis.

1 Introduction

Cyanobacteria, a kind of bacteria, can be found in almost every possible environment. Sometimes become injurious due to their uncontrolled growth giving place to the formation of extensive harmful algal blooms (HABs) [1-3]. Some cyanobacteria produce toxins called cyanotoxins. The aim of this research is to construct a support vector regression (SVR) model with different kernels and a multilayer perceptron (MLP) model to identify cyanotoxins in the Trasona reservoir (Principality of Asturias, Northern Spain) [4,5]. A multilayer perceptron (MLP) is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate output [4].

2 Materials and methods

2.1 Experimental data set

The data used for the SVM and MLP analyses were collected over five years (2006 to 2010) from lots of samples in Trasona reservoir. The total number of data processed was about 151 values (see `http://dl.dropbox.com/u/36679320/Trasona_reservoir_data_sc.xls`). In this research work, we

have taken into account the two dominant species of the cyanobacteria community: *Microcystis aeruginosa* and *Woronichinia naegeliana*. The main goal of this research work was to obtain the dependence relationship of cyanotoxins of the Trasona reservoir (output variable), expressed in micrograms per liter, as a function of the following eight biological and fifteen physical-chemical input variables [1]: (1) biological parameters: *Microcystis aeruginosa* (mm^3/L); *Woronichinia naegeliana* (mm^3/L); other cyanobacteria species (mm^3/L); diatoms (mm^3/L); chrysophytes (mm^3/L); chlorophytes (mm^3/L); and other phytoplankton species (mm^3/L); (2) physical-chemical parameters: water temperature (^oC), ambient temperature (^oC), secchi disk depth (m), turbidity (NTU), total phosphorus (mg P/L), phosphates concentration (mg PO_4^{3-}/L), total nitrogen concentration (mg N/L), nitrate concentration (mg NO^{3-}/L), nitrite concentration (mg NO^{2-}/L), ammonium ion concentration (mg/L), dissolved oxygen concentration (mg O_2/L), conductivity (Ω/cm), alkalinity (mg $CaCO_3/L$), calcium concentration (mg/L) and pH.

2.2 Support vector machines for regression

The SVM problem can be formulated as follows [4,5]:

$$\begin{array}{l} \min\limits_{\mathbf{w},b,\xi} \ \frac{1}{2}\left\{\|\mathbf{w}\|^2 + C\sum\limits_{i=1}^{L}\left(\xi_i^+ + \xi_i^-\right)\right\} \\ \left(\begin{array}{c} \langle \mathbf{w}, \psi(\mathbf{x}_i)\rangle + b - y_i \leq \varepsilon + \xi_i^+ \\ y_i - (\langle \mathbf{w}, \psi(\mathbf{x}_i)\rangle + b) \leq \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geq 0 \end{array}\right) \quad i = 1, ..., L \end{array} \tag{1}$$

where $\psi : X \to Z$ is a transformation of the input space into a new space (feature space), usually a larger dimension space, where we define an inner product by means of a positive definite function Z (kernel trick):

$$\langle \psi(\mathbf{x}), \psi(\mathbf{x}')\rangle = \sum_i \psi_i(\mathbf{x}) \cdot \psi_i(\mathbf{x}') = k(\mathbf{x}, \mathbf{x}') \tag{2}$$

3 Analysis of results and conclusions

Table 1 shows the coefficients of determination for all the models studied here: the fitted MLP and the fitted SVM for different kernels. According

Table 1: Coefficients of determination (R^2) for all the fitted models.

Model name	R^2
MLP	0.64
SVM (linear)	0.57
SVM (PUK)	0.91
SVM (RBF)	**0.92**

to this statistic, the SVM with the radial basis kernel function (with RBF-Kernel) is the best model for estimating the cyanotoxins concentration in the Trasona reservoir, since the fitted SVM with RBF-Kernel has a coefficient of determination equal to 0.92 and a correlation coefficient equal to 0.96. These results indicate an important goodness of fit, that is to say, a good agreement is obtained between our model and the observed data.

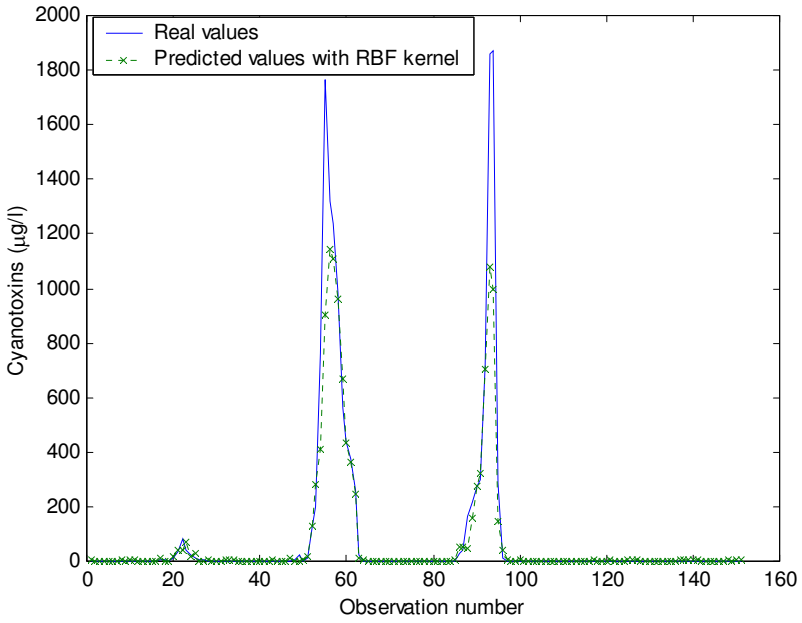


Figure 1: Comparison on the Trasona reservoir from 2006 to 2010 between the three blooms of cyanobacteria observed and predicted by the SVM model with RBF kernel function.

Figure 1 shows the comparison on the Trasona reservoir from 2006 to

2010 between the three blooms of cyanobacteria observed and predicted by the SVM model with RBF kernel function. Therefore, it is necessary the use of a SVM model with RBF-Kernel in order to achieve an effective approach to nonlinearities present in the regression problem. Obviously, these results coincide again with the outcome criterion of "goodness of fit" (R^2) so that the SVM model with a radial basis kernel function (RBF-Kernel) has been the best fitting. This research work was able to estimate the presence of cyanobacteria blooms from 2006 to 2010 in agreement to the actual cyanobacteria blooms observed. Moreover, this methodology can be applied to other reservoirs with similar or different sources of pollutants, but it is always necessary to take into account the specificities of each location.

Acknowledgements

Authors wish to acknowledge the computational support provided by the Department of Mathematics at University of Oviedo.

References

[1] Allman, E.S., Rhodes, J.A. Mathematical Models in Biology: An Introduction. New York, Cambridge University Press, 2003.

[2] Brönmark, C., Hansson, L.-A., The Biology of Lakes and Ponds. New York, Oxford University Press, 2005.

[3] Reynolds, C.S. Ecology of Phytoplankton. New York, Cambridge University Press, 2006.

[4] García Nieto, P.J., Martínez Torres, J., Araújo Fernández, M., Ordóñez Galán, C. Support vector machines and neural networks used to evaluate paper manufactured using Eucalyptus globulus. Appl. Math. Model. 36 (12), 6137–6145, 2012.

[5] Steinwart, I., Christmann, A. Support Vector Machines. New York, Springer, 2008.

Influence of candidate qualities and previous president performace in voting intentions

Gabriela Ribes-Giner

Faculty of Business Administration and Management, Universidad Politécnica de Valencia. Camino de Vera s/n, 46022 (Valencia), Spain. gabrigi@omp.upv.es

Maria Fuentes-Blasco

Business Management and Marketing Department, Pablo de Olavide University. Ctra. de Utrera, km. 1, 41013 (Sevilla), Spain. mfuebla@upo.es

Keywords: political marketing, voting intention, Forecasting, Structural Equation Methodology

AMS 2010 Mathematics Subject Classification codes: 91, 62

Introduction

This article aims to measure empirically the influence of the main variables affecting the voting intention of the electorate, taking as reference the polls result obtained the previous months of the 2008 American General Elections, which are provided by the American National Election Studies (ANES). Our research is an aproach to political marketing with causal methodology, for that purpose, Structural Equation methodology is used, confirming some concepts such as the personalisation of politics, main pillar of the current political marketing strategies, and the retrospective voting.

Political marketing is fairly recent as a discipline and its emergence is directly related to the history of political communication in the United States. Several authors (Maarek, 1995; Beresford, 1998) date its appearance to the 1952 presidential election and Eisenhower's campaign. In this election, both Democrats and Republicans earmarked part of their budgets to political communication and produced the first televised political advertising, direct mail marketing and opinion surveys (Martin, 2002). Subsequent growth in this area has been an inevitable consequence of the development of the media and politics.

Political marketing can be seen as a psychological purchase by the voter through which they expect to obtain future benefits based on the election promises or message conveyed by the candidate, who in this case functions as a product.

We chose four variables in our model -political candidate qualities, the previous president's performance, the concerns of the electorate and voting intention- are made up of a number of issues specifically selected from the questionnaire to form our measurement scale. However, while the candidate qualities and voting intention variables varied depending on the candidate, the other two were the same in both the Obama and McCain models. The choice of variables and their measurement scales was supported by the theory outlined above, thus providing the model with conceptual consistency.

Firstly, the presence of the political candidate qualities variable is warranted given the recent trend in the political arena towards the personalisation of politics (Martín, 2002). Politics increasingly focuses on candidates and not on political parties, which means a candidate who has qualities that are highly valued by voters is more likely to be elected than another who does

not. Thus, this model rates qualities such as morality, leadership, intelligence, honesty and optimism (Martín, 2002; Nimo and Savage, 1976).

Secondly, the inclusion of the previous president's performance as a variable is justified by the theory of retrospective voting (Fiorina, 1978; Lichtman, 1996) which argues that the electorate votes according to how well they think the previous incumbent did. If they think that his presidency has been successful, they reward him with their vote. If the reverse is the case, they punish him by voting for the opposition. In this respect it is interesting to see how the electorate's assessment of Bush affected the voting intention for each candidate, since it involved rating his administration in terms of the economy, foreign affairs, environment and/or health and also touched on controversial issues in his last term such as the Iraq War.

Thirdly, the concerns of the electorate are a very important part of the political product because they define the candidate's political manifesto which has to fit in with what the voters want. It should be recalled that the candidates do not have complete freedom in choosing their message but instead are constrained by the voters' preferences, interests and expectations. Thus, in this case the political manifesto was limited to focus on economic issues (Martín, 2002; Kiewiet, 1983) – the economy in general, unemployment and inflation. The study analysed the impact of the economic crisis on the issues that concerned the electorate and how they affected the latter's tendency to vote for one candidate or another.

In terms of voting intention, the proposed measurement scale sought to evaluate the way people felt about the candidates (Nimo and Savage, 1976). This meant returning to the personalisation of politics (Harris, 2001; Martín, 2002) and focusing on the candidates, their qualities and how the electorate feels about them. This is what Nimo and Savage (1976) defined as the affective component of the candidate's image and refers to the feelings that candidates generate in their voters.

Empirical research methodology and results assessment

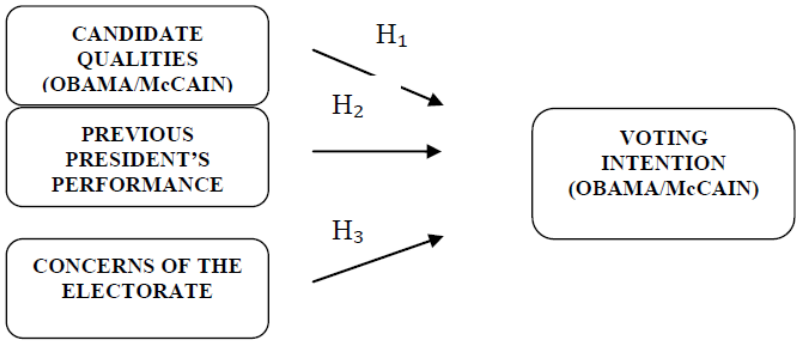
The 2008 American elections turned out to be a historic event since it was the first time in U.S. history that a black candidate had run for President. Our study attempted to empirically predict the voting intention for both candidates in the months leading up to the election using questionnaires provided by ANES (American National Election Studies). The questionnaires were face-to-face surveys with a sample of 2,323 voters. Two models, one for each candidate (Obama & McCain), were constructed based on these questionnaires. The models analysed some of the variables that influenced the voting decision-making process using SEM methodology.

SEM (Structural Equation Modelling), also called "covariance structure analysis", is a confirmatory technique that is mainly a way of checking whether a particular theoretical model is valid. Examples of this technique are factor analysis, regression and path analysis (Hair et al., 1999). The main difference between SEM and other multivariate relationship techniques is the use of different relationships for each set of dependent variables. SEM estimates a series of multiple regression equations which are interrelated by specifying a structural model. The study

of these causal relationships seeks to determine the effect of an explanatory variable on the explained one and to what extent the observed variation in the latter is due to changes produced by the former. Hence in SEM there are both manifest variables which are measurable and observable and also latent or unobservable variables.

At any event, the main feature of SEM is its confirmatory approach. While in an exploratory approach the structure of the factors underlying the data matrix is not known or specified a priori, in a confirmatory approach there is a theory and a series of hypotheses that suggest a relationship pattern between variables. The question asked when applying SEM is how well do the data fit the proposed model? (Hair et al., 1999).

Figure 1. Proposed empirical model



Source: authors' own compilation.

Results

Both structural models illustrate that only two out of the three variables that were originally supposed to influence voting intention were shown to be significant in the case of both candidates. These were the candidate's qualities and the performance of the previous President George W. Bush.

The first factor reflects the qualities of the political candidate. This aspect encompasses measurements about morality, leadership skills, intelligence, honesty and optimism among others. According to the results obtained from estimating the causal model, this variable behaved differently for the two candidates. While for McCain it was directly proportional to voting intention (the more highly rated McCain was, the greater the voting intention), for Obama it was quite the reverse. The variable has a negative coefficient, which means a higher rating of the Democratic candidate leads to lower voting intention.

As noted above, politics is increasingly focused on candidates. In this respect, political marketing is based on the concept of the ideal candidate: voters have an ideal and the candidate who comes closest to this prototype is the one who is elected. In the case of our model, a hypothesis for the divergent behaviour of the variable could be established using this theory. It might be that in the case of our sample, McCain is better rated than Obama, which would explain to some extent the different behaviour of the variable which is positive for the

Republican candidate and negative for the Democratic representative. However, it would also be useful to examine the socio-demographic features of the sample due to the highly polarised nature of American politics.

An examination of the impact of the second factor on voting intention for both candidates confirms the retrospective voting theory (Fiorina, 1978) set out in the theoretical framework, which argues that the electorate votes based on how they view the previous incumbent's performance. If he has done well, he is rewarded with their vote. If the reverse is the case, they punish him by voting for the opposition. Both models prove that this theory holds true for two reasons. Firstly, the electorate's assessment of the Bush administration had a negative impact on the intention to vote for Obama since the coefficient associated with the structural equation was negative. Thus, the better the assessment of the previous president of the United States, the lower the tendency to vote for the Democratic candidate. People who thought that the Bush administration had done a good job did not see any reasons to change party and so in order to keep up the "good work" they voted for continuity and the Republican candidate, in this case McCain. Secondly, the worse the assessment of Bush's policy the greater the tendency to vote for the Democratic candidate, as people's disgruntlement led them to change candidates in an attempt to improve political management. In the case of McCain, the reverse was true. As the structural equation shows, the coefficient associated with the variable was positive, so the assessment of the Bush administration had a directly proportional effect on the voting decision. If things go well, the model demonstrates that the impact of the previous Republican candidate's performance has a direct and positive effect on voting.

The third factor, concerns of the electorate, has no impact on voting intention (Martín, 2002). One possible hypothesis to explain this is the model's conceptual nature. As noted above, all the variables except for the concerns of the electorate focus on the political candidate, be it the previous president or the candidates aspiring to the White House, and each issue is related to the candidate. However, the third variable is not related to the candidates but directly affects the concerns of voters.

Voting intention is also measured with the electorate's perceptions of the candidates, as is the case with the first two variables. Perhaps this lack of relationship with the candidates is the reason why the variable is not significant in the model.

As an initial assessment of this research it can be concluded that SEM methodology is effective in measuring political marketing processes. However, in terms of future research a questionnaire designed specifically to study voting intentions would enable the creation of much more consistent latent variables that can measure other effects which may affect a person's vote.

Two preconditioning techniques for the time dependent neutron diffusion equation

S. González-Pintor[†], D. Ginestar[*], and G. Verdú[†]

(*) Departamento de Ingeniería Química y Nuclear, Universidad Politécnica de Valencia,

(†) Instituto de Matemática Multidisciplinar, Universidad Politécnica de Valencia,

Camino de Vera, 14. 46022 Valencia, Spain.

November 30, 2012

1 Introduction

When solving the time dependent neutron diffusion equation a spatial discretization of the equations is carried out [4], obtaining a semidiscrete system of equations which is, in general, stiff. Then, a one step implicit method is used to integrate the system [2]. Thus, it is necessary to solve a sequence of linear systems

$$A^n x^n = b^n \ , \quad n = 1, 2, \ldots \tag{1}$$

In many transients can be assumed that the spectral properties of the system matrix vary slowly in time, and they are similar from one linear system to the next one. In this way, the invariant subspace associated with the preconditioned matrix at linear system n will be used to precondition the linear system $n + 1$.

2 Spectral preconditioners

Here, we study methods for preconditioning a sequence of linear systems based on modifying the eigenvalues distribution of the coefficient matrices,

*e-mail:dginesta@mat.upv.es

by using the information provided by the Krylov subspaces previously computed. Two strategies have been implemented and its performance has been compared to solve the systems associated with the numerical solution of the time dependent neutron diffusion equation.

2.1 Low-rank transformations preconditioning

Let us consider a system, $Ax = b$ and M an initial preconditioner. Considering V the matrix associated with a right invariant subspace of the matrix AM of dimension k, $AMV = VJ_k$, where the eigenvalues of J_k are $\{\lambda_1, \ldots, \lambda_k\}$. Then, a result from [3] is used to update the preconditioner shifting the smallers (in magnitude) k eigenvalues of the system matrix, and can be applied in a recursive way as follows,

$$M^{(l+1)} = M + \sum_{j=1}^{L_{max}} M^{(j)} V^{(j)} \left(W^{(j)} A M^{(j)} V^{(j)} \right)^{-1} W^{(j)}$$

To maintain the required memory by the preconditioner under a user-defined bound, the L_{max} terms associated with the L_{max} smallest eigenvalues of the matrix to optimize the performance of the preconditioner.

2.2 FGCRO-DR method

Flexible GCRO-DR method [1] is an inner-outer method that combines GCRO as the outer method and Flexible GMRES-DR as the inner method, and it allows deflated restarting and subspace recycling.

GCRO method is used to compute optimal approximation over a given set of search vectors in the sense that the residual is minimized, and the inner method FGMRES-DR computes a new search vector by approximately solving the residual equation. GCRO uses two matrices, U_k and $C_k = AU_k$, with the property that $C_k^T C_k = I_k$, and solves a minimization problem.

To preserve optimality with respect to the search directions of the outer method (GCRO), the inner method (FGMRES-DR) uses the operator $(I - C_k C_k^T)A$ instead of A, preserving the orthogonality relations from GCRO also in the inner method.

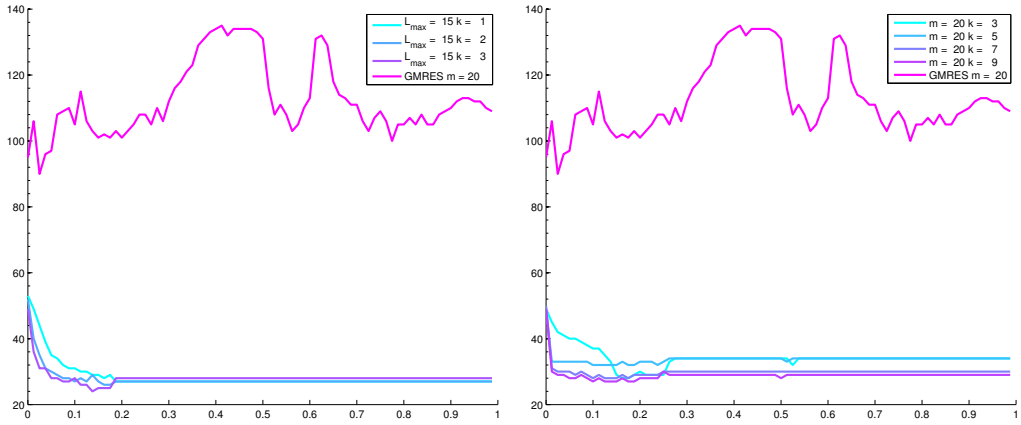
During the process the Ritz pairs of the system matrix can be computed and this spectral information can be easily adapted to a new linear system as the initial search direction of the outer methods thus, the spectral information can be recycled from one system to the next one in a natural way.

3 Numerical Results

To test the preconditioning strategies exposed above for the neutron diffusion equation, we have considered a transient in a small reactor of type VVER. A detailed discussion of the problem is reported at [5].

From this transient we have considered 80 systems of linear equations corresponding to the time interval $[0, 1]$s with a time step $\Delta t = 0.0125$s.

The the low rank transformations preconditioning is used to update an initial preconditioner M, which is taken initially as the diagonal of the matrix. This preconditioner is used in combination with FGMRES-DR(m,k) method. The number of iterations needed to solve each of the systems values of $L_{\max} = 15$ using different values of k are shown in Figure 1, for linear and quadratic spatial discretizations. These results have been obtained with a fixed dimension $m = 20$ of the restarting Krylov subspace. For comparison, the results of solving the same transient with the classical GMRES(m) method, that is the same that taking $L_{\max} = 0$ and $k = 0$ with the FGMRES-DR(m,k), is included. There can be observed the improvement of using FGMRES-DR(m,k) with the low rak transformations preconditioning, even when recycling little information. It can be also observed than the improvement scales well with the size of the matrix systems.



(a) preconditioned FGMRES-DR(m,k) (b) FGCRO-DR(m,k)

Figure 1: Number of iterations used by the preconditioned FGMRES-DR(m,k) and the FGCRO-DR(m,k) methods.

The same set of systems has been solved using FGCRO-DR(m,k) method

and an approximate invariant subspace associated with the smallest eigenvalues of one system matrix is recycled to precondition the solution of the next system. The number of iterations needed to solve each system setting the dimension of the restarted Krylov subspace to $m = 20$ for different values of the dimension of the recycled space, k, are shown in Figure 1.

The use of the spectral preconditioners is efficient to reduce the total number of iterations needed to solve the total set of linear systems associated with the transient in the reactor.

Acknowledgments

This work has been partially supported by the Spanish Ministerio de Educación y Ciencia under projects MTM2010-18674 and ENE2011-22823.

References

[1] L. M. Carvalho, S. Gratton, R. Lago, X. Vasseur, A flexible generalized conjugate residual method with inner orthogonalization and deflated restarting, *SIAM J. Matrix Anal. Appl.* 32(4) 1212–1235, 2011.

[2] D. Ginestar, G. Verdú, V. Vidal, R. Bru, J. Marín, J.L. Muñoz-Cobo, High Order Backward Discretization of the Neutron Diffusion Equation, *Annals of Nuclear Energy* 25(1-3) 47–64, 1998.

[3] L. Giraud, S. Gratton, E. Martin, Incremental spectral preconditioners for sequences of linear systems, *Applied Numerical Mathematics* 57, 1164–1180, 2007.

[4] S. González-Pintor, D. Ginestar, G. Verdú, High Order Finite Element Method for the Lambda Modes problem on hexagonal geometry, *Annals of Nuclear Energy* 36, 1450–1462, 2009.

[5] Sebastián González-Pintor, Damian Ginestar, Gumersindo Verdú, Preconditioning Multiple Right Sides for the Time Dependent Neutron Diffusion Equation, *Proceeding of the ANS Annual Conference: Nuclear Science and Technology: Managing the Global Impact of Economic and Natural Events*, June 2012, Chicago, Illinois.

Behavioural models for temporary disability in primary health care centres

N. Guadalajara, I. Barrachina, C. Sancho
nguadala@omp.upv.es;
Centro de Ingeniería Económica, Universitat Politècnica de València

1. Introduction

In most western European countries the increase in Temporary Disability (TD) for mental disorders was higher than in the case of other diseases, as were the costs relating to TD for mental disorders. This might be due to the high prevalence of mental illnesses in patients who visit primary health care centres, which ranges from 20% to 34.4 % depending on the country. Therefore, mental illness is not only one of the main causes of TD but also one of the illnesses with the greatest degree of recurrence, disability, and duration.

Anxiety, dissociative and somatoform disorders with the diagnosis code 300 (cod300) are the second highest ranking disorders for which TD leave days are taken in Spain. However, little is known about the factors which allow TD for mental disorders to be assessed, although a higher incidence has been observed in women than in men, as in the case of the consumption of drugs to treat this disorder, and in the case of the elderly. On the contrary, other studies have noted a higher incidence and recurrence in men than in women, but also a shorter duration of the TD.

The aim of this study is to analyse the factors relating to the Primary Health Care Centre (PHCs) and Health Districts (HDs), which have an impact on the prescription of TD for illnesses included under cod300 of the ICD-9-CM, and their modelling.

2. Material and Methods

2.1 Participants

The entire population of the Valencian Community (VC) in 2009 was analysed, and consisted of a 5,117,190 inhabitants, with a working population of 2,395,598. The population was attended to in the 739 total PHCs, grouped into 23 HDs.

In these PHCs, 480,755 TD were prescribed for all diagnoses in 2009, 25,859 of which corresponded to the cod300, and the related total days of absence amounted to 2,347,394. The number of leaves finalized in 2009 for cod 300 illnesses, regardless of the year in which they were started, amounted to 20,194, which represents an increase of 5,665 leaves compared to the year 2008. In other words there were close to 8 more leaves prescribed per PHC, which evidences the magnitude of this problem.

One of the novelties of the present study is the unit of analysis used, formed by the 739 PHCs, whereas in all the previous studies analysed the unit of analysis was either the individuals themselves or the patients.

2.2 Information sources

The data relating to TD were obtained from the Ambulatory Information System (AIS) integrated in the ABUCASIS information system of the Valencian Regional Government's Department of Health (Conselleria de Sanitat de la Generalitat Valenciana).

Population-based data (age and sex) was taken from the Population Information System (PIS) of the Valencian Health Care Agency (Agencia Valenciana de Salud).

The information relating to the PHCs (accreditation for health-care training) was taken from the HC Record of Accreditation for specialized health care training of the District of Health. The remainder of the HC data (age of health care professionals, diagnoses of hospitalized patients, use of the electronic medical record, etc.) was taken

from the Cyrus Information System for human resource management of the Valencian Health Department.

The information relating to hospital stays was taken from the Valencian Region's Minimum Basic Data Set (MBDS) of hospital admission services.

2.3. Mathematical model

In all the previous studies analysed in the literature review on DT modelling, uniequational regression models were used (models with a single dependent variable), in which the cause-effect relationship was unidirectional. However, in reality this is often not the case, and therefore it is preferable to use simultaneous equation models, which identify a set of variables simultaneously through a remaining set of variables. In this way the model parameters are estimated taking into consideration the other equations. Simultaneous equation models have been used in health areas, such as in the patient's relationship of trust in PCPs, and its influence on the responsibility they take for follow-up of treatment.

2.4. Explanatory variables and variables to be explained

Two dependent or endogenous variables for the year 2009 were used: The PHCs' Incidence Rate (IR) for cod300 and the PHC's Absence Rate (AR) for cod300.

The PHC variables considered were: Total population assisted by the PHC, as well as their gender and age, divided into intervals; province and the presence of a coastline; average age of PHC' physicians; accreditation for health training and the use of electronic medical records; average duration of DT sick leaves for cod300 and for all diagnoses, the absence rate for all diagnoses, psychiatric services, psychiatric hospital stays and the days of delay in the HD in the case of both primary health care and specialized care.

3. Results

In order to obtain a simultaneous equation model, crossing the various sources of information allowed a total of 485 PHC of the initial 739 PHC to be assessed, i.e. 66 %, due to a lack of information relating to some of the variables. The PHCs analysed served a total working population of 2,163,456 (90%), which ensures the representativeness of the data analysed.

Firstly, based on the ratios obtained between the dependent variables and the rest of the explanatory variables, a following system of 2 simultaneous equations was posed. The two equations are overidentified, and therefore the 2SLS method was applied.

The AR for cod300 in the PHC, the coastline location of the PHC and the number of psychiatric stays in hospitals of the PHC population, allow almost 60% of the TD initiated by the PHC for cod300 to be explained. In PHCs close to coastlines, the number of TDs is higher since there is a higher population density, while a higher number of hospital stays for psychiatric disorders reduces the TDs, possibly due to the improved treatment of the disease, although its influence is very small. Clearly, an increased number of TDs leads to a higher AR.

This AR for cod300, is also due to the fact that the population is located closer to the coastline, the increase being even greater if the PHCs are within the province of Valencia, the most populated area in the VC. In PHCs with a higher AR for all diagnoses, there was also a greater AR for cod300, which demonstrates that the higher or lower number of TD prescribed is linked to the operation of the PHC. However, there is also an inverse ratio between the duration of the TD for all diagnoses in a PHC and its AR for cod300, which proves that the behaviour is different in terms of the duration of the TD for cod300 compared to the rest of diseases observed in the descriptive analysis.

Higher quality management by the PHC physicians, which is indicated by the increased use of electronic medical records, leads to a higher AR for cod300, which is most likely due to the fact that the use of electronic medical records leads to a better diagnosis of the disease.

Finally, although it is less explanatory than the rest of the variables, the number of days a first specialised care visit is delayed in all the PHCs of a HD, raises the AR for cod300, which can only be expected, since a delay in the provision of specialized care to the patient increases the number of TD leave days.

4. Discussion

Like in Western European countries, an increase in TD for mental disorders was observed, evidencing that an occupational health problem exists. For this reason alone, an average of 1.0794% of working force population initiated a TD, which supposed a total of 2,347,365 days of absence.

The average duration of the TD was also higher than for the whole of the diagnoses, as was concluded in other studies. The introduction of changes in sick leave pay, as well as changes in position in the economic market and changes in company policies could possibly lead to a reduction in this rate, as in the case of Holland in the years 2004 to 2007.

The PHC population variables had very little effect on the TD variability, and did not appear in the simultaneous equation model, in contrast to in previous studies, where they were offset by the variables relating to PHCs and their HDs. This demonstrates the need to carry out a more extensive longitudinal cross study of the characteristics of the PHCs, in order to take measures to reduce the number of TDs, such as increased assistance to physicians in PHCs on the coastline and in the province

of Valencia, and speedier specialised care services. The reasons why an increased use of electronic medical record leads to a greater AR should be extensively explored, since although it is true that the disease may be better diagnosed, it is also true that this might lead to the propensity to lengthen leaves in the event of substitutions in these PHCs.

Modelling Driving Behaviour and its Impact on the Energy Management Problem in Hybrid Electric Vehicles

C. Guardiola, B. Pla, D. Blanco-Rodriguez, A. Reig *†

CMT Motores Térmicos, Universidad Politécnica de Valencia, Valencia, Spain

November 30, 2012

1 Introduction

The Energy Management Problem (EMP) in HEV consists in finding the control policy which minimises the fuel consumption of the vehicle under a set of restrictions over a defined driving cycle. The complexity of the system to optimise, the charge-sustainability condition and the unknown driving cycle involve the main difficulties to optimally solve the EMP.

Since the driving style have a strong impact on fuel efficiency and Optimal Control [1] and no *a priori* information is available the present work proposes and describes different methods to estimate future driving conditions in order to address the EMP from an Optimal Control approach. Stochastic techniques and sophisticated algorithms are evaluated with a HEV backwards model over a set of urban and highway real driving cycles.

*This research has been supported by Ministerio de Ciencia e Innovación through Project TRA2010-16205 uDiesel

†e-mail: {carguaga, benplamo, dablarod, alreiber}@mot.upv.es

2 Methods to estimate driving behaviour

The formulation of the methods proposed in this paper is oriented so that they may be combined with the ECMS [2] to solve the EMP. The ECMS method permits to exchange the integral optimisation problem for a local minimisation of a cost function including a fuel cost associated to battery electrical energy consumption. Accordingly, the cost may be re-defined as:

$$f\left(u\left(t\right),E_b\left(t\right),t\right)=P_f\left(u\left(t\right),t\right)+s\cdot P_b\left(u\left(t\right),E_b\left(t\right),t\right) \tag{1}$$

where the parameter s is an equivalence factor to transform electrical into an equivalent fuel power. This parameter can only be calculated when the entire driving cycle is known in advance. Thus, the present paper proposes different methods to obtain an optimal estimated $\overline{s}$ value verifying the charge sustainability condition within a limited time horizon.

Markov chains. Markov chains stand for a particular discrete stochastic event in which the current state only depends on the previous one. This property has been used to solve automotive problems in the past [3] since it allows a system to be characterised with a probability matrix. In this paper current vehicle speed and power requirements from past driving cycles are evaluated and used as Markov chain states to train the probability matrix, like approached in [4].

Montecarlo method in combination with Markov probability matrix offers a set of *a priori* known random driving cycles containing the same driving behaviour than those from the past. Then, an average $\overline{s}$ value may be calculated and used with ECMS to optimally solve the EMP when no future information is available.

Histogram-based models (S-ECMS). Assuming a quasi-static powertrain behaviour (i.e. zero-order system), the cost at any time depends only on the state at that point so a driving pattern may be characterised by a probability distribution of its power requirements.

For a particular probability distribution, the required battery power can be calculated as:

$$\overline{P_b}(s)=\sum_n \mathbf{Pr}(P_{req,n})\cdot P_b(P_{req,n},s) \tag{2}$$

where P_b may be mapped offline for computational economy.

The charge-sustainability condition may require a particular battery power, $\widehat{P}_b(t)$ to reach the desired SoC (state of charge) along an horizon, depending on the current SoC. Therefore, the optimal $\overline{s}$ choice is that verifying that $\widehat{P}_b(t) = \overline{P}_b(\overline{s})$.

Geotagging (Geo-S-ECMS). A power demand histogram contains information regarding driving style and road type, but it can only store one pattern (or several mixed together). However, very different styles may be found in just one trip. Therefore, a geotagging technique to store different driving patterns is proposed as an upgrade of S-ECMS method.

Via in-car sensors and a GPS receiver a probability distribution grid may be easily build. If on-board navigation system is present, driver could preprogram the route so a set of expected probability distributions is known in advance. Then, the estimated battery power requirement along the remaining trip may be calculated as:

$$\overline{P_b}(s) = \sum_m \mathbf{Pr}(\ell_m) \left(\sum_n \mathbf{Pr}_m(P_{req,n}) \cdot P_b(P_{req,n}, s) \right) \tag{3}$$

where m is the histogram index, n refers to the power requirements and ℓ_m is the geographic position.

Finally, the S-ECMS method is applied as previously described.

3 Results

Strategies efficiency was evaluated over two driving cycles (urban and highway) and with two different non-professional drivers. First, Dynamic Programming (DP) was calculated as a benchmark solution. Following Markov chain and probability distributions were trained with a set of four driving cycles belonging to the same route than the evaluated trip. Finally, all three proposed methods EMP solutions were calculated online. Results are shown in figure 1.

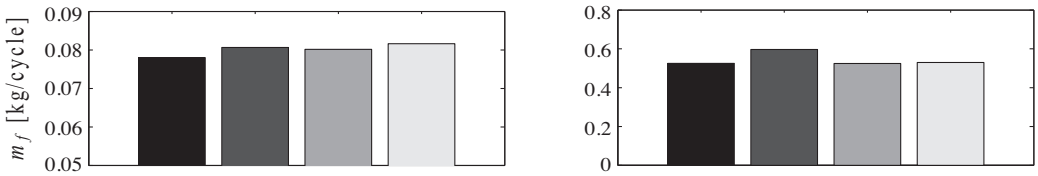


Figure 1: Fuel consumption for urban (left) and highway (right) cycles. Strategies are (from left to right) DP, Markov, S-ECMS and Geo-S-ECMS

4 Conclusions

Markov chain-based estimation shown to be not robust enough for HEV application. Even not being possible to assure that a driving pattern verifies Markov property in advance, this issue is mostly due to the lack of feedback of the battery state during operation. Histogram-based methods are close to DP optimality in both urban and highway cycles with a reasonable computational burden. Geo-S-ECMS do not show any improvement over S-ECMS since the evaluated cycles took place in homogeneous roads with no advantage in storing different driving styles. However, this is not the usual case so a better fuel efficiency is expected for Geo-S-ECMS method. Therefore, ECMS method in combination with histogram-based techniques shown a significant level of robustness reaching an online quasi-optimal solution for the EMP.

References

[1] R Wang and S M Lukic. Review of Driving Conditions Prediction and Driving Style Recognition Based Control Algorithms for Hybrid Electric Vehicles, *Vehicle Power and Propulsion Conference (VPPC)*, 2011.

[2] G Paganelli *et al.* General Supervisory Control Policy for the Energy Optimization of Charge-Sustaining Hybrid Electric Vehicles, *JSAE Review*, 2001.

[3] I Kolmanovsky *et al.* Optimization of Powertrain Operating Policy for Feasibility Assessment and Calibration: Stochastic Dynamic Programming Approach, *American Control Conference, Anchorage, USA*, 2002.

[4] J Liu and H Peng. Modeling and Control of a Power-Split Hybrid Vehicle, *IEEE Transactions on Control Systems Technology*, 2008.

Application of the finite-element method within a two-parameter regularised inversion algorithm for electrical capacitance tomography

DORIS HINESTROZA G.
Departamento de Matemáticas
Universidad del Valle, Cali, Colombia
Doris.hinestroza@correounivalle.edu.co

CARLOS GAMIO
School of Eng. & Computing,
Glasgow Caledonian University,
70 Cowcaddens Road,
Glasgow G4 0BA,
United Kingdom
Carlos.Gamio@gcu.ac.uk

In ECT, we have a sensor (fig. 1) consisting of an electrically non-conducting pipe with a circular array of contiguous rectangular sensing electrodes, separated by small gaps, attached to its outer surface. The ring of electrodes is located between two grounded cylindrical end-guard electrodes that eliminate any end-effects, allowing the use of 2-dimensional (2-D) modelling [2]. The whole assembly is in turn surrounded by a grounded screen to avoid external interference. The aim of ECT is to reconstruct, using a suitable algorithm, a cross-sectional image of the unknown electrical permittivity distribution inside the non-conducting pipe at the zone of the electrode ring, from the knowledge of the mutual capacitance that exists between all possible sensing-electrode pairs, which must be previously measured using a suitable instrument. The electrical permittivity distribution obtained will reflect the phase distribution of a mixture contained in the sensor.

To measure the mutual capacitance values $c_{j,i}$, for a 12-electrode sensor like that of figure 1, first a known excitation voltage is applied to electrode 1 while keeping all the others at zero potential and the charge on electrodes 2 to 12 is measured. These measurements divided by the excitation voltage directly represent $c_{2,1}$ to $c_{12,1}$. Next, the excitation voltage is applied to electrode 2 while keeping all the others at zero and the charge on electrodes 3 to 12 is measured, representing $c_{3,2}$ to $c_{12,2}$. This procedure is repeated, applying voltage to electrode n and measuring the charge on electrodes (n + 1) to 12, until, as a final step, voltage is applied to electrode 11 and the charge of electrode 12 is measured. In this way, 66 independent mutual capacitance values are obtained

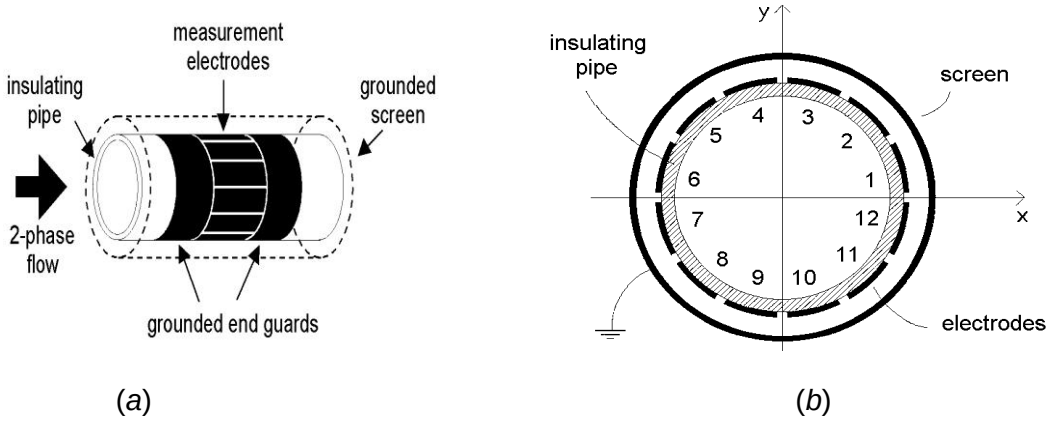


Figure 1 ECT sensor: (*a*) complete assembly (*b*) cross-section view

The rest of this paper is organised as follows: Section 2 presents the mathematical model of the problem, Section 3 describes the two-parameter regularised inversion method, and Section 4 deals with the application of the finite element method in this context.

2. Mathematical Model

The tomography sensor of figure 1(b) can be modelled (figure 2) as the circular region Ω_1 (with radius R_1) corresponding to the imaging area, surrounded by two annular regions: Ω_2 (with inner and outer radiuses R_1 and R_2, respectively) corresponding to the insulating pipe, and Ω_3 (with inner and outer radiuses R_2 and R_3, respectively) corresponding to the area between the pipe and the external screen.

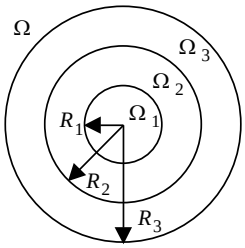


Figure 2 Sensor model

Let us then consider the bounded domain $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$ (figure 2) with $\Omega_1 = \{ z = (x, y) : \| z \| < R_1 \}$, $\Omega_2 = \{ z = (x, y) : R_1 < \| z \| < R_2 \}$ and $\Omega_3 = \{ z = (x, y) : R_2 < \| z \| < R_3 \}$. We shall neglect the width of the inter-electrode gaps. Then, we can consider that electrode i is described by the arc $S_i = \{ z = (x, y) : \| z \| = R_2,\ 2\pi (i\text{-}1)/N \le arg(z) \le 2\pi i/N \}$.

The following partial differential equation describes the sensor

$$\nabla \cdot (\varepsilon(x, y)\ \nabla u^{(i)}(x, y)) = 0 \quad \text{for} \quad i = 1, ..., N \tag{1}$$

under the boundary conditions

$$u^{(i)}(x, y) = \begin{cases} 0 \ \text{ in } \ |(x, y)| = R_3 \\ 1 \ \text{ in } \ (x, y) \in S_i \\ 0 \ \text{ in } \ |(x, y)| = R_2, \ (x, y) \notin S_i \end{cases} \tag{2}$$

where $\varepsilon(x, y)$ is the relative electric permittivity (or simply permitivity, for short).
The mutual capacitances are given by the formula

$$c_{j,i} = K \int_{S_j} \varepsilon \frac{\partial u^{(i)}}{\partial n} ds \tag{3}$$

where K is a constant, and $u^{(i)}$ is the solution of the equation (1)-(2). The arc S_j corresponds to the curve surrounding electrode j.

The inverse problem is: *given* ½$N(N$-1) *values* $c_{j,i}$, $(i, j = 1,2, \ldots, N)$, $(i < j)$, *of the mutual capacitances between the electrodes* S_i *and* S_j, *determine approximately the value of* $\varepsilon(x, y)$. N is the number of electrodes (12 in our particular case).

In order to find $\varepsilon(x, y)$ we will consider the spaces $X = H^1(\Omega)$, $E = \{\varepsilon \in H^1(\Omega) : 1 \le \varepsilon \le \varepsilon_{max}\}$, $Y = \bigotimes_{i=1}^{N} H^1(\Omega)$, $Z = R^{\frac{1}{2}N(N-1)}$, and the function

$$\boldsymbol{\lambda}(\varepsilon) \ = \ \mathbf{f} \circ \mathbf{u}(\varepsilon) \ = \ \mathbf{f}(\mathbf{u}(\varepsilon)) \tag{4}$$

where

$$\mathbf{u}(\varepsilon) = \left[u^{(1)}(\varepsilon),\ \cdots, u^{(N)}(\varepsilon) \right] \tag{5}$$

and

$$\boldsymbol{\lambda}(\varepsilon) \ = \ \frac{1}{K}\left[c_{2,1} \cdots c_{N,1}, c_{3,2} \cdots c_{N,2}, \cdots, c_{(N-2),(N-3)} \cdots c_{N,(N-3)}, c_{(N-1),(N-2)}, c_{N,(N-2)}, c_{N,(N-1)} \right] \tag{6}$$

3. Inversion Method

The proposed inversion method is based on the minimization of the following functional that the depends on two parameters α and β.

$$M(\varepsilon) \ = \ M_{\alpha}^{\beta}(\varepsilon) \ = \ \frac{1}{2} \ \left(\ \| \mathbf{f} \circ \mathbf{u}(\varepsilon) - \boldsymbol{\lambda}_{obs} \|_Z^2 \ + \ \alpha \| \varepsilon \|_X^2 \ + \ \beta \| \mathbf{u}(\varepsilon) \|_Y^2 \ \right) \tag{7}$$

Here, $\boldsymbol{\lambda}_{obs} = \dfrac{\mathbf{c}_{obs}}{K}$, where $\mathbf{c}_{obs}$ is the vector of ½N(N-1) observed mutual capacitances for a given permittivity distribution in the imaging area.

A main problem related with this functional has to do with finding its derivative.

Applying the Gateaux derivative, given by

$$\delta M\left(\varepsilon\right) = \lim_{t \to 0} \frac{M(\varepsilon + t\delta\varepsilon) - M(\varepsilon)}{t} = M'(\varepsilon) \cdot \delta\varepsilon \tag{8}$$

Then, after doing all the technical calculations involved we get

$$\delta M(\varepsilon) \ = \ \left\langle \delta\boldsymbol{\lambda},\ \boldsymbol{\lambda}(\varepsilon) - \boldsymbol{\lambda}_{obs} \right\rangle \ + \ \beta \left\langle \delta\mathbf{u},\ \mathbf{u}(\varepsilon) \right\rangle \ + \ \alpha \left\langle \delta\varepsilon,\ \varepsilon \right\rangle \tag{9}$$

where $\delta\boldsymbol{\lambda} = \boldsymbol{\lambda}'(\varepsilon) \cdot \delta\varepsilon$ and $\left\langle \delta\mathbf{u},\ \mathbf{u}(\varepsilon) \right\rangle = \sum_{i=1}^{N} \iint_{\Omega} u^{(i)} \delta u^{(i)} \, dx \, dy$.

Using the sensitivity equations, we have to solve the differential equation

$$\nabla \cdot (\varepsilon \nabla \delta u^{(i)}) \ + \ \nabla \cdot (\delta\varepsilon \nabla u^{(i)}) \ = \ 0 \tag{10}$$

with

$$\delta u^{(i)} \ = \ 0 \quad \text{ on } \partial\Omega \tag{11}$$

Solving the previous problem for $\delta u^{(i)}$, requires the knowledge of $\nabla u^{(i)}$ and $\delta\varepsilon$. By solving the equation $\nabla \cdot (\varepsilon(x,y) \nabla u^{(i)}) = 0$, we obtain $\nabla u^{(i)}$. To find $\delta\varepsilon$ we shall introduce the concept of the ‘‘adjoint equations’’, choosing a function $\Psi^{(i)}(x,y)$ that satisfies the equation

$$\frac{\partial}{\partial x}\left(\varepsilon\frac{\partial\Psi^{(i)}}{\partial x}\right)+\frac{\partial}{\partial y}\left(\varepsilon\frac{\partial\Psi^{(i)}}{\partial y}\right) = \nabla\cdot(\varepsilon\nabla\Psi^{(i)}) = -\beta u^{(i)}(\varepsilon) \tag{12}$$

with the boundary conditions

$$\Psi^{(i)} = 0 \quad \text{on } \partial\Omega. \tag{13}$$

From equations (9) to (13), it follows that

$$dM = \langle \lambda'(e)\times de, \lambda(e) - \lambda_{obs}\rangle - \sum_{i=1}^{N}\int\!\!\int_{W} \tilde{N}Y^{(i)}\times\tilde{N}u^{(i)} de\, dx\, dy + a\langle de, e\rangle \tag{14}$$

Having obtained the differential $\delta M = M'(\varepsilon)\cdot\delta\varepsilon$, we now need to find the minimum of M, which must satisfy $M'(\varepsilon) = 0$. In practice, the process of minimisation can be carried out using the Gauss-Newton Method. In order to do this, the problem must first be discretised. For this, we shall use the finite-element method.

4. Application of the Finite-Element Method

In order to apply the finite element method to approximate the solution of the equation

$$\nabla\cdot(\varepsilon(x,y)\nabla u) = 0 \tag{15}$$

Using Green's formula, we have that

$$\int\!\!\int_{W} \tilde{N}\times(v(e\tilde{N}u))\,dx\,dy = \int\!\!\int_{W} v\tilde{N}\times(e\tilde{N}u)\,dx\,dy + \int\!\!\int_{W} e\tilde{N}u\times\tilde{N}v\,dx\,dy = \int_{\P W} e v\frac{\P u}{\P n}\,ds \tag{16}$$

Taking u = $u^{(i)}$ and since $\nabla\cdot(\varepsilon(x,y)\nabla u^{(i)}) = 0$, we have that

$$\int_{\Omega}\varepsilon\nabla u^{(i)}.\nabla v\,dx\,dy = \int_{\partial\Omega}\varepsilon v\frac{\partial u^{(i)}}{\partial n}\,ds \quad \text{for } i = 1,...,N. \tag{18}$$

If we define a L-dimensional space $V_L \subseteq H^1(\Omega)$, the weak form of the differential equation requires that $u^{(i)}$ and v be in the space V_L instead of considering it in $H^1(\Omega)$. Since V_L is finite-dimensional, there exists a finite basis $\{\phi_j\}_{j=1}^{L}$ with $\phi_j \in V_L$. Making $v = \phi_j$, $j = 1,2,...,L$, in equation (18) we have

$$\int_{\Omega}\varepsilon\nabla u^{(i)}.\nabla\phi_j\,dx\,dy = \int_{\partial\Omega}\varepsilon\phi_j\frac{\partial u^{(i)}}{\partial n} \tag{19}$$

By considering the expansion of the unknown solution in the form

$$u^{(i)}(x,y) = \sum_{l=1}^{L} u_l^{(i)}\phi_l(x,y) \tag{20}$$

We denote the vector of the components of the function $u^{(i)}$ in the basis $\{\phi_j\}_{j=1}^{L}$ as $\mathbf{u^{(i)}} = \begin{bmatrix} u_1^{(i)} \\ \vdots \\ u_L^{(i)} \end{bmatrix}$

Substituting (20) in (19) we obtain the linear system of equations

$$\sum_{l=1}^{L} \left(\int_{\Omega} \varepsilon \nabla \phi_l \cdot \nabla \phi_j \right) u_l^{(i)} = \int_{\partial\Omega} \varepsilon \phi_j \frac{\partial u^{(i)}}{\partial n}, \qquad j = 1,...L. \tag{21}$$

We can write these system equations in the form

$$\mathbf{A u^{(i)}} = \mathbf{b^{(i)}}, \quad \text{for} \quad i = 1, \ldots, N \tag{22}$$

with $\mathbf{A} = (a_{lj})$ where $a_{lj} = \int_{\Omega} \varepsilon \nabla \phi_l \cdot \nabla \phi_j$, with $l, j = 1, \ldots, L$, and $\mathbf{b^{(i)}}$ is the column vector with components $\int_{\partial\Omega} \varepsilon \phi_j \frac{\partial u^{(i)}}{\partial n}$, $j = 1, \ldots, L$.

Let's consider a triangular mesh on the region Ω, made by sub-dividing it using a set of T non-overlapping triangles $\{P_k\}_{k=1}^{T}$, such that the vertex of any triangle does not lie on the edge of another triangle.

We shall consider as the basis for V_L the set of 'hat' functions ϕ_j which are linear on each triangle and take the value 0 at all nodes, except for x_j where $\phi_j(x_j) = 1$. This construction implies that

$$u^{(i)}(x_j) = \sum_{l=1}^{L} u_l^{(i)} \phi_l(x_j) = u_j^{(i)} \tag{26}$$

This means that if we apply the finite element method, we obtain the value of the function $u^{(i)}$ at the mesh nodes. Note that the base functions ϕ_j vanish in all triangles that do not have the node x_j. Then, the entries a_{lj} of matrix **A** can be calculated only in the triangles that contain the node x_l. This means a_{lj} is zero except when x_l and x_j are nodes of the same triangle. Therefore, **A** is very sparse.

The permittivity takes the form

$$\varepsilon(x, y) = \sum_{k=1}^{NL} \varepsilon_k \, \chi_{P_k}(x, y) \tag{27}$$

and therefore,

$$a_{lj} = \int_{\Omega} \varepsilon \nabla \phi_l \cdot \nabla \phi_j \, dx \, dy = \sum_{k=1}^{T} \varepsilon_k \int_{P_k} \nabla \phi_l \cdot \nabla \phi_j \, dx \, dy. \tag{29}$$

By consider to estimate the coefficient $\varepsilon(x, y)$ as

$e(x,y) = \sum_{k=1}^{M} e_k c_h(x,y)$ where χ_h is the characteristic function, defined by $\chi_h(x,y) = \begin{cases} 1 & (x,y) \in P_k \\ 0 & \text{Otherwise} \end{cases}$.

Then, using (15) we find that

$$\frac{\partial M}{\partial \varepsilon_k} = 2\frac{\partial \boldsymbol{\lambda}}{\partial \varepsilon_k} \cdot (\boldsymbol{\lambda}(\varepsilon) - \boldsymbol{\lambda}_{obs}) - \sum_{i=1}^{N} \iint_{P_k} \nabla \Psi^{(i)} \cdot \nabla u^{(i)} + 2\alpha\varepsilon_k \tag{60}$$

It is not difficult to prove that

$$\frac{\partial \boldsymbol{\lambda}_{ij}}{\partial \varepsilon_k} = \int_{P_k} \nabla u^{(i)} \cdot \nabla u^{(j)} dxdy \tag{61}$$

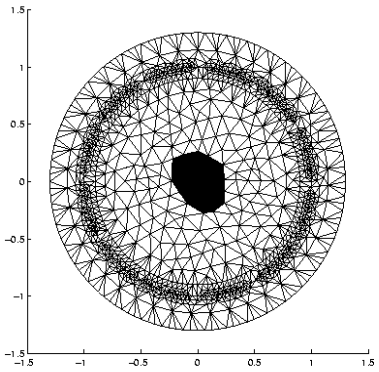
5. Numerical Examples

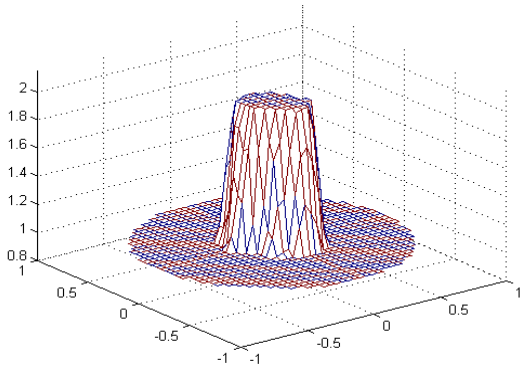
In this work we use some of the subroutines of the software package *EIT-2D* for two-dimensional electrical capacitance tomography (EIT) image reconstruction, developed as part of the *EIDORS* project by researchers at the University of Kuopio, Finland, and the University of Manchester Institute of Science and Technology (UMIST), UK [10]. The package consists of a series of routines written in MATLAB. *EIT-2D*. We built new programs to adapt the package to our two-parameter minimization problem, and we use the G-Newton method for the minimization procedure.

All the examples are simulated, and thus for a known permittivity distribution, we solve the forward problem (3) - (4), and calculate the capacitances. Then, given these capacitances with an added 2% random normal-distribution error (to simulate the measurement errors), we find the permittivity. We describe some examples and report the image-reconstruction error in L_2 for a particular selection of values for the parameters α and β. We select, for each example, the best values for these parameters, the ones that result in the smallest reconstruction error. In the figures, we have the exact solution, and we show the optimal α and β (selected from the table of errors) as well as the reconstructed images obtained using these optimal parameters.

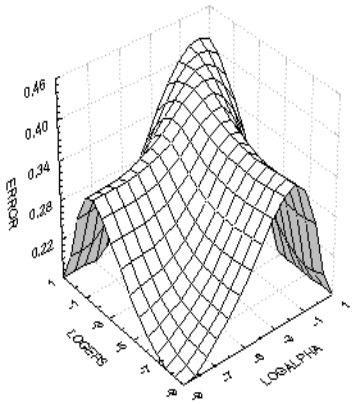
Example 1 consists of an object with permittivity $\varepsilon = 2$ placed at the center using an area of 29 triangles. In the rest of the grid the permittivity is $\varepsilon = 1$. We take for this example, $\alpha = 0$ and $\beta = 10^{-8}$. It is very interesting to see that the only non-zero parameter actually has a regularizing effect with respect to the potential (and not the permittivity). See Figure 3 and table of errors 1.

Example 2 consists in stratified flow (approximately 30%) with permittivity $\varepsilon = 2$. In the rest of the grid the permittivity is $\varepsilon = 1$. For this case $\alpha = \beta = 10^{-8}$. See Figure 4 and table of errors 2.

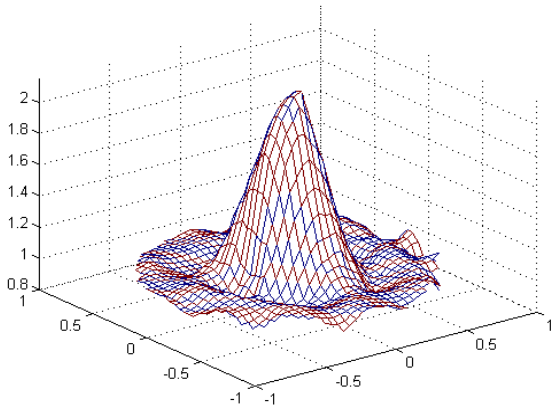




True image



L-surface

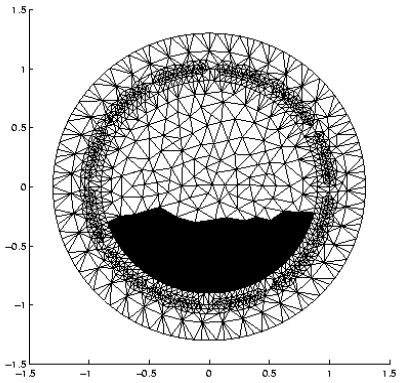


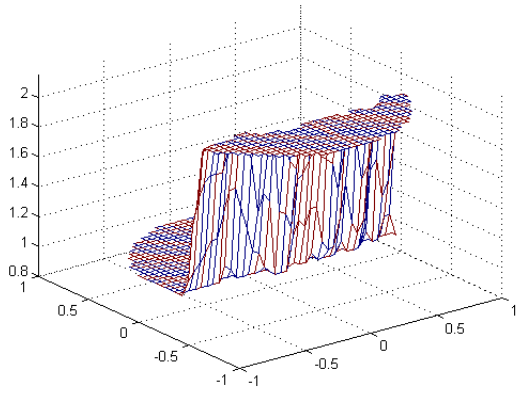
Reconstructed image $\alpha = 0, \beta = 10^{-8}$

Figure 3 - One object at the center (example 1)

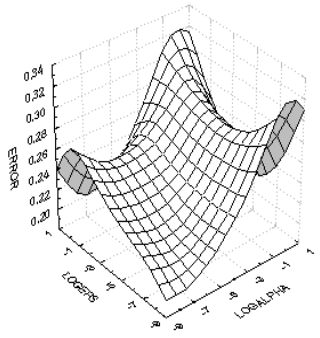
β \ α	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	0
10^{-1}	0.4244	0.4058	0.4014	0.4009	0.4008	0.4008	0.4008	0.4008	0.4008
10^{-2}	0.4201	0.3879	0.3518	0.3421	0.3410	0.3409	0.3409	0.3409	0.3409
10^{-3}	0.4197	0.3869	0.3351	0.3045	0.2984	0.2977	0.2977	0.2977	0.2977
10^{-4}	0.4196	0.3868	0.3350	0.2950	0.2580	0.2486	0.2475	0.2474	0.2474
10^{-5}	0.4196	0.3868	0.3350	0.2948	0.2450	0.2218	0.2179	0.2175	0.2174
10^{-6}	0.4196	0.3868	0.3350	0.2948	0.2452	0.2203	0.2043	0.1999	0.1994
10^{-7}	0.4196	0.3868	0.3350	0.2948	0.2452	0.2204	0.1970	0.1822	0.1798
10^{-8}	0.4196	0.3868	0.3350	0.2948	0.2452	0.2204	0.1970	0.1816	0.1775
0	0.4198	0.3869	0.3340	0.2945	0.2465	0.2202	0.1914	0.1777	

Table of Errors 1

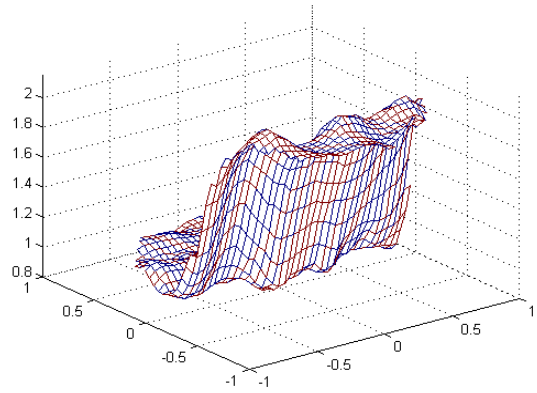




True image



L-Surface



Reconstructed image $\alpha = \beta = 10^{-8}$

Figure 4 - Stratified flow (example 2)

β \ α	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	0
10^{-1}	0.3220	0.2845	0.2792	0.2787	0.2786	0.2786	0.2786	0.2786	0.2786
10^{-2}	0.3243	0.2776	0.2550	0.2512	0.2509	0.2508	0.2508	0.2508	0.2508
10^{-3}	0.3245	0.2781	0.2509	0.2384	0.2377	0.2377	0.2377	0.2377	0.2377
10^{-4}	0.3246	0.2781	0.2511	0.2383	0.2374	0.2427	0.2435	0.2435	0.2435
10^{-5}	0.3246	0.2781	0.2511	0.2382	0.2254	0.2171	0.2187	0.2191	0.2191
10^{-6}	0.3246	0.2781	0.2511	0.2382	0.2258	0.2107	0.2157	0.2248	0.2263
10^{-7}	0.3246	0.2781	0.2511	0.2382	0.2258	0.2108	0.2046	0.2232	0.2580
10^{-8}	0.3246	0.2781	0.2511	0.2382	0.2258	0.2108	0.2052	0.2016	
0	0.3246	0.2781	0.2511	0.2382	0.2258	0.2108	0.2052		

Table of Errors 2

References

Gamio J.C., A High-Sensitivity Flexible-Excitation Electrical Capacitance Tomography System, PhD Thesis, UMIST, 1997.

Gamio J.C., C. Ortiz-Alemán and R. Martin. Electrical capacitance tomography two-phase oil-gas pipe flow imaging by the linear back-projection algorithm. Geofísica Internacional (2005), Vol. 44, No. 3, pp. 265-273.

Hinestroza D., Murio D. A., Zhan, S., Regularization Techniques for Nonlinear Problems. Journal of Comput. Math. Applic., 1999.

Lifeng Zhang, Huaxiang Wang. Identification of oilgas two-phase flow pattern based on SVM and electrical capacitance tomography technique. Elsevier, Journal Flow Measurement and Instrumentation, 21(1):20-24. DOI:10.1016/ j.flowmeasinst.2009.08.006.

Lifeng Zhang, Pei Tian, Xiuzhang Jin, Weiguo Tong. Numerical simulation of forward problem for electrical capacitance tomography using element-free Galerkin method. Elsevier, Journal Engineering Analysis with Boundary Elements, 34 (2010) 477–482.

Spink, D.M. and Noras, J.M., Recent Developments in the Solution of the Forward Problem in Capacitance Tomography and Implications for Iterative Reconstruction, Nondestr. Test. Eval., 1998, Vol. 14, pp. 115-142.

Yang W.Q., Beck M.S., and Byars M., Electrical capacitance tomography: From design to applications, Measurement + Control, 28, 1995, pp. 261-266

Vauhkonen M., Lionheart W.R.B., Heikkinen L.M., Vauhkonen P.J. and Kaipio J.P., 2001, A MATLAB package for the EIDORS project to reconstruct two-dimensional EIT images, Physiological Measurement, 22, pp. 107-111.

The LES modeling of diesel injectors: the spray first instants

J.M. Desantes, S. Hoyas*, X. Margot and J.M. Mompó-Laborda

CMT-Motores Térmicos, Universitat Politècnica de València

Valencia 46022, Spain,

November 30, 2012

There is a common agreement on the increasing awareness about polutant emissions [1]. In the case of Diesel engines, there are two problems: soot and NoX. It is well known that there exist a direct relation between emissions and instantaneous values of both temperature and fuel concentration fields inside engines. As fuel injection and spray characterisation have been investigated thoroughly during the last decades, there exists many techniques to model diesel spray. The traditional Eulerian method performs well in the liquid phase while the Lagrangian drop method describes accurately the dispersed region. Transition between both zones is not particularly well resolved, mainly due to time and computational power restrains.

Until very recently Reynolds-Average Navier Stokes methods (RANS), where the only tool able to modelate the behaviour of the spray in Diesel engine, and thus modelling the combustion process. However the increasing in computational power and accuraccy of the models, make Large Eddy Simulations (LES) affordable.

The proved capability of LES to predict unsteady characteristics of turbulent flow fields can be, for instance, coupled with DDM, in order to increase the modelling of the atomization. Nevertheless most of the work in this field has been restricted to scientific use of LES (i.e. dense grids, simple flow configurations) due to the requirements of most of the turbulent models.

The present work aims to contribute to the effort of bringing the potential

*e-mail: serhocal@mot.upv.es

of LES close to the industry and more applied research. In this regard, this study focus on the engineering LES of non-evaporating spray.

Lagrangian spray parcels models were originally developed for RANS [4] included momentum, heat and mass transfer and specific models for atomization, breakup and collision. Its use with LES models require full interaction between spray an the gas phase including effects of gas flow on spray as well as spray effects on the gas in both grid and subgrid leves. Since subgrid interaction is specific of LES, its influence on classic RANS models behaviour must be clarified. From a general point of view, [2] provided a review on the physical models used in atomization and spray for RANS, LES and DNS. Closer to internal engine simulation, a group at Doshish university has explored spray models using KIVA with OEE turbulent model [3] demostrating that RANS correlations for spray modelling can provide reasonable results with very high grid resolutions. However this turbulent model presents several drawbacks for spray modelling:

- OEE turbulent model assumes an isotropic turbulence spectrum at subgrid level which can be hard to fulfil close to the nozzle for VOF type of simulations under realistic injection conditions even for unreasonable high grid resolutions (DNS requirements).

- For Lagrangian-Eulerian spray simulations the interaction between the parcel and the gas phase only occurs at grid level. Therefore the parcel has no effect on the sub-grid scale turbulent kinecit energy.

Dynamic one-equation turbulent model developed by Rutland, [5, 6] solved the first objection since abandoned the eddy viscosity concept. Instead, the sub-grid shear stress tensor is estimated directly by means of sub-grid kinetic energy, and the tensor coefficient based on Leonard stress tensor. This model was improved later [7], [8].

The CFD code for E-E simulations used in this study are based on the rhoReactingFoam solver (OpenFOAM 2.0.0). Since there is no parcel-gas liquid interaction, only the part of the dS turbulet model published in 2002 was implemented [5]. For L-E spray simulations dieselFoam solver was modified by including the fully developed dS turbulent model. The detailed derivation of the model as well as the parcel gas phase sub-grid interaction can be found on citeBharadwaj10. Regarding drag calculation, the effect of dropplet distortion on the droplet drag is taken into account by OpenFOAM's default

formulation. Also small modifications on KHRT breakup model (including primary breakup) and the gas turbulence velocity for LES were implemented as explained below.

The computational domain is a cylindrical constant volume vessel (D = 40mm, L = 100mm) that represents the shape of an injection test rig chamber. The meshing methodology is fairly the same for the gas jet and diesel spray calculatons, with the same grid densities for both RANS and LES calculations. In the spray calculatiions there is no inlet boundary condition and any external surface is defined as wall.

Tests were done in fully and partially refined meshes and no evidence on differente jet behavior were found. Therefore only partially refined mesh calculations are shown in this study. Dynamically refined meshes are out of the scope of the present study but will be included in future works.

Although symmetric shape is expected in RANS calculations, turtbulent inlet boudary condition generates asymmetry at the outer region of the jet (1% of injected N2) and fluctuating distribution of gas velocity over that surface. Regarding gas jets simulated by LES, the so called viscous models require a physical length to fully develope turbulence, given that the boundary contion only applies a random perturbation on the reference field value $U_{\text{bar}} = (0, 0, 373.27)$m/s (fluctuationscale $= (0.05, 0.05, 0.1)$, $\alpha = 0.0001$ this should go to inside numerical setup B.C.).

Unlike viscous models, when dS was used to simulate turbulence, this lack of turbulent consistency at the inlet boundary condition did not require an extra length to develope turbulence, showing in this way a greater independence on the inlet boundary condition. The dS turbulent model shows a turbulent behavior closer to the nozzle due to the randomness of the boundary condition. Consistent with the other turbulent models, when simulated with a non-perturbed constant value the gas jet did not generate any turbulent motion in the calculated domain. As expected, a minimum level of perturbation is required for LES simulations. In adition, cell size and time step have a lower impact on penetration, regardless of the increase of turbulent scales seen with smaller cell size mesh. Compare with viscous models, this independence includes the increase of the temporal schemes

Part of this work was developed by JMML at the ERC of the University of Wisconsin-Madison. The authors are in debt with Professors Rutland and Trujillo for their help in this work. This work was supported by the Spanish Government in the frame of the Project "Métodos LES para la simulación de chorros multifásicos", Ref. ENE2010-18542.

References

[1] European Commission Air Environment standards: Transport & Environment of Road Vehicles, Available at: http://ec.europa.eu/environment/air/transport/

[2] Jiang, X., Siamas, G. A., Jagus, K., and Karayiannis, T. G. Physical modelling and advanced simulations of gasliquid two-phase jet flows in atomization and sprays. Prog. Energy Combust. Sci., 2010, 36(2), 131167

[3] Hori, T., Kuge, T., Senda, J., and Fujimoto, H. Effect of convective schemes on LES of fuel spray by use of KIVALES. SAE paper 2008-01-0930, 2008.

[4] Amsden AA, ORourke PJ, Butler TD. KIVA-II: A computer program for chemically reactive flows with sprays. DE89-012805. Los Alamos, NM: Los Alamos National Laboratory; 1989.

[5] Pomraning, E. and Rutland, C. J. Dynamic one- equation nonviscosity large-eddy simulation model. AIAA J., 2002, 40(4), 689701.

[6] Lu, H., Rutland, C. J., and Smith, L. M. A priori test of one-equation LES modeling of rotating turbulence. J. Turb., 2007, 8(37), pp. 127, DOI: 10.1080/14685240701493947.]

[7] Bharadwaj, N. and Rutland, C. J. A large eddy simulation study of sub-grid two-phase interac- tion in particle laden flows and diesel engine sprays. Atomization Sprays, 2010, 20(8), 673695.

[8] Vuorinen, V., Larmi, M., and Fuchs, L. Large- eddy simulation on the effect of droplet size dis- tribution on mixing of passive scalar in a spray. SAE paper 2008-01-0933, 2008.

Discontinuous Galerkin method for the numerical solution of the exotic option pricing model

J. Hozman*

(*) Faculty of Science, Humanities and Education, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic.

November 30, 2012

1 Introduction

During the last decade, financial models have acquired increasing popularity in option pricing. The valuation of different types of option contracts is very important in modern financial theory and practice, especially exotic options have become very popular speculation instruments in recent years. The problem of determining the fair price of such an option is standardly formulated in the well-known Black–Scholes equation, firstly presented in [2].

A huge amount of literature has been devoted to the solving of this equation or its modification. The performance demands on the valuation process are very high in this case. Moreover, most of the analytical formulas for these options is limited by strong assumptions, which led to the application of numerical methods instead. Therefore, the main goal of this paper is to develop an efficient, robust and accurate method for the exotic option pricing problem, which arises from the concept of the discontinuous Galerkin (DG) approach (cf. [4]) and enables better resolving of occurred special properties of certain types of exotic options, in comparison with the standard finite element approach, see e.g. [1, 5] and the references cited therein.

*e-mail: jiri.hozman@tul.cz

2 Discrete barrier option pricing problem

In this paper, we focus only on one family of exotic options such as discrete barrier options. Furthermore, we shall concentrate only on a discrete double time–independent barrier knock-out option, i.e. an option that expires worthless if one of the two barriers has been hit at a monitoring date, for more details see [1, 5]. Let $M := \{0 = t_0^M < t_1^M < \ldots < t_{l-1}^M < t_l^M = T\}$ be the set of monitoring dates and B_- be the lower barrier and B_+ the upper barrier active only at discrete instances $t_l^M \in M$.

Let $\Omega := (S_{min}, S_{max})$, $0 < S_{min} < B_- < B_+ < S_{max}$, be a bounded open interval and T stands for the maturity. We denote by x the price of an underlying asset (e.g. stock) and by t the time to expiry of the option. The price $u : Q_T := \Omega \times (0,T) \to \mathbb{R}$ of the discrete barrier option satisfies the Black–Scholes partial differential equation with initial and boundary conditions, i.e.

$$\frac{\partial}{\partial t}u(x,t) - \frac{1}{2}\sigma^2 x^2 \frac{\partial^2}{\partial x^2}u(x,t) - rx\frac{\partial}{\partial x}u(x,t) + ru(x,t) = 0 \quad \text{in } Q_T, \tag{1}$$

$$u(S_{min},t) = 0 \quad \text{and} \quad u(S_{max},t) = 0, \tag{2}$$

$$u(x,0) = \begin{cases} \max(x-K,0)\cdot\chi_{[B_-,B_+]}, & \text{(call)} \\ \max(K-x,0)\cdot\chi_{[B_-,B_+]}, & \text{(put)} \end{cases}, \quad x \in \Omega, \tag{3}$$

where $\sigma > 0$ and $r > 0$ are constant model parameters denoting the volatility of stock price and the risk–free interest rate, respectively.

From the mathematical point of view the problem (1)–(3) represents a convection–diffusion-reaction equation equipped with a set of two homogeneous Dirichlet boundary conditions (2) prescribed at the endpoints of Ω and with the initial condition (3), where the symbol K stands for the strike price and $\chi_{[B_-,B_+]}$ denotes the characteristic function of the barrier interval.

Moreover the discrete monitoring of the contract introduces an updating of the solution $u(x,t)$ at the monitoring dates $t_l^M \in M$, i.e.

$$u(x,t_l^M) = \lim_{\varepsilon\to 0+} u(x,t_l^M - \varepsilon)\cdot\chi_{[B_-,B_+]}. \tag{4}$$

3 DG discretization

The discontinuous Galerkin approach is suitable for problems with irregular solutions, because its framework originally arises from a generally discontinuous piecewise polynomial approximation $u_h(t)$ describing the global solution

$u(x,t)$ on the whole domain Ω, i.e.

$$u_h(t) \in S_h = \left\{ v_h \in L^2(\Omega);\ v_h|_I \in P^p(I)\ \forall\, I \in \mathcal{T}_h \right\} \subset H^1(\Omega) \tag{5}$$

where $\mathcal{T}_h$ is a family of partitions of the closure $\overline{\Omega} = [S_{min}, S_{max}]$ into closed mutually disjoint subintervals I, and $P^p(I)$ denotes the space of all polynomials of degree $\leq p$ on element I.

In order to obtain a space semi–discrete DG scheme from [4], we multiply (1) by a test function $v_h \in S_h$, integrate over an element $I \in \mathcal{T}_h$ and use integration by parts in the diffusion and convection terms of (1) subsequently. Further, we sum over all $I \in \mathcal{T}_h$ and add some artificial terms vanishing for the exact solution such as penalty and stabilization terms, which replace the inter–element discontinuities and guarantee the stability of the resulting numerical scheme, respectively. Consequently, we employ a concept of an upwind numerical flux (see [3]) for the discretization of the convection term and end up with the following DG formulation for the semi–discrete solution $u_h(t)$ represented by a system of ordinary differential equations, i.e.

$$\frac{d}{dt}\left(u_h(t), v_h\right) + \mathcal{A}_h(u_h(t), v_h) = 0 \quad \forall\, v_h \in S_h,\ \forall\, t \in (0,T) \tag{6}$$

where a form $\mathcal{A}_h(\cdot,\cdot)$ stands for the semi–discrete variant of the linear differential operator in (1), see [4].

In order to obtain the discrete solution, it is necessary to equip the scheme (6) with suitable solvers for the time integration. The suggested implicit time discretization is suitable for avoiding the strong time step restriction of explicit time schemes. Moreover, a bilinearity of the form $\mathcal{A}_h(\cdot,\cdot)$ directly implies that the used implicit treatment in (6) corresponds to a system of linear algebraic equations without employing any additional linearisation, cf. [1, 5].

For the sake of clarity, we use the simplest implicit method — backward Euler method — for the time discretization and introduce the fully discrete scheme. We now partition $[0,T]$ as $0 = t_0 < t_1 < t_2 < \ldots < t_N = T$, denoting each time step by $\tau_l = t_l - t_{l-1}$. We compute the approximate values u_h^l of the exact solution $u(t_l)$ only at given time levels t_l according the following formula, i.e.

$$\left(u_h^l, v_h\right) + \tau_l \mathcal{A}_h\left(u_h^l, v_h\right) = \left(u_h^{l-1}, v_h\right) \quad \forall\, v_h \in S_h,\ \ l = 1, 2, \ldots, N \tag{7}$$

with initial state u_h^0 as S_h-approximation of (3) and monitoring constraints $u_h^l := u_h^l \cdot \chi_{[B_-,B_+]}$ valid only at monitoring dates $t_l^M \in M$. Finally, the system (7) is then solved by a suitable linear algebraic solver.

4 Conclusion

We have dealt with the numerical solution of the discrete barrier option pricing model, represented by the linear convection–diffusion–reaction equation. We have derived the above mentioned numerical scheme: from the continuous problem, over the semi–discrete one to the fully discrete one. The whole method is based on the space semi–discretization by the discontinuous Galerkin method in space and on the implicit Euler method used for discretization in time. For the future work, we intend to extend this concept to a simple theoretical analysis and also to the multivariate Black–Scholes equation describing basket options.

Acknowledgement. The paper was supported by the ESF Project No. CZ.1.07/2.3.00/09.0155 "Constitution and improvement of a team for demanding technical computations on parallel computers at TU Liberec".

References

[1] Y. Achdou, and O. Pironneau, Computational Methods for Option Pricing. Philadelphia, SIAM, 2005.

[2] F. Black, and M. Scholes, The pricing of options and corporate liabilities, *J. Political Economy*, vol. 81: 637-?659, 1973.

[3] M. Feistauer, J. Felcman, and I. Straškraba, Mathematical and Computational Methods for Compressible Flow. Oxford, Oxford University Press, 2003.

[4] B. Riviére, Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation. Philadelphia, Frontiers in Applied Mathematics, SIAM, 2008.

[5] R. Seydel, Tools for Computational Finance, 4^{th} edition. Berlin, Springer, 2008.

DAGSVM Multiclass algorithm based on SVM binary classifiers with 1vsAll approach to the slate tile classification problem

J. Martínez*, C. Iglesias†, J.M. Matías‡, and J. Taboada†

(*) Defense University Center,

General Military Academy, Zaragoza, Spain,

(†) Department of Environmental Engineering,

University of Vigo, Spain,

(‡) Department of Statistics,

University of Vigo, Spain.

November 30, 2012

1 Introduction

Support Vector Machines technique is a reference in the field of pattern recognition [2] due to its flexibility, predictive capacity, parsimony and the global optimum nature of the optimization problem.

SVM can be applied to solve binary classification problems, as well as multiclassification problems. One-versus-one and one-versus-all approaches combined with the construction of SVM binary classifiers allow their solving. Furthermore, there is another classification strategy: Directed Acyclic Graphs (DAGs) [9] with one perceptron at each node and its adaptative version [5].

A new strategy for multiclassification problem solving is presented. It is based on the construction of a Directed Acyclic Graph with SVM binary

*e-mail: jmtorres@unizar.es

classifiers with one-versus-all approach.

The UCI Machine Learning Repository [4] is frequently used in the machine learning field for the evaluation of new algorithms [10], [11], [6]. In order to validate this new strategy, four databases from the UCI Machine Learning Repository were used. Likewise, available data from the slate tile classification problem were used for the validation. The named problem is to classify slate slabs in three classes of quality according to the information gathered by an artificial vision system. The first stage of the slate project is developed in [8].

2 Materials and Methods

2.1 The classification problem

Given a group of individuals X to be classified into different classes $Y \in \mathcal{Y} = \{1, 2, ..., c\}$, the classification problem is finding the classification rule g which matches each individual with its corresponding class, minimizing the probability of error.Thus, the objective of the training phase is to find the classifier g which minimizes the probability of error in the classification.

One-versus-one approach [7] builds $c(c-1)/2$ SVM binary classifiers and evaluates every pair of distinct classes. In the testing phase, classification is done by a Max Wins voting strategy [3].

One-versus-all approach [1] builds c SVM binary classifiers, evaluating each class against all the others in the training phase. c Decision functions are obtained, and each element is assigned the class which maximizes the decision function.

Directed Acyclic Graphs (DAG) are tree-structured graphs with no directed cycles, whose nodes have two acyclic edges [9]. DAGSVM methodology can be used to solve a multiclassification problem building the directed acyclic graphs with SVM binary classifiers.

2.2 Validation

Five different databases were used: Iris, Wine, Glass and Abalone from the UCI Machine Learning Repository [4] and Slates data [8]. Error rates were calculated through a cross validation process for the following SVM classification approaches: one-versus-one, one-versus-all, DAG-one-versus-one and

Table 1: Error Rates (%) obtained for the different methodologies and Standard Deviation of the initial node analysis.

Data set	$1vs1$	$1vsAll$	$DAG-1vs1$	$DAG-1vsAll$	STD
Iris	0.0001	0.0001	0.00008	0.06	0
Wine	0.005	0.005	0.08	0.015	0
Glass	0.0025	0.0029	0.01	0.0022	0.0006
Abalone	0.0012	0.0015	0.0083	0.0027	0.0005
Slates	0.0112	0.0125	0.0138	0.0098	-

DAG-one-versus-all (proposed in this work). Furthermore, the standard deviation of the results was studied in order to analyse the influence of the first node of the classification.

3 Results and conclusions

Table 1 shows the error rates of the databases presented previously. It includes as well the standard deviation, used to study the different results obtained with each initial node of the model.

With regard to Wine and Abalone data sets, one-versus-one and one-versus-all approaches have the lowest error rates. Nevertheless, the proposed approach shows satisfactory results, better than DAG-one-versus-one.

Except for Iris database, the error rate of the new approach is lower than that for DAG-one-versus-one approach. In addition, DAG-one-versus-all approach is the best classifying Glass and Slates data sets.

Hence, the approach proposed in this work obtains good results classifying databases with a wide number of output classes (Glass and Abalone), as good as or even better than that for the approaches without DAG.

When the number of output classes and attributes is short, DAG-one-versus-all results are not optimal. Nevertheless, when the number of attributes is higher, the ER for the DAG-one-versus-all approach is similar as the ER of the approaches without DAG (Wine) or even lower (Slates). Moreover, the standard deviation obtained shows the first node has no influence in the classification process.

Appart from the good results of the classification process, the DAG-one-versus-all approach has several advantages compared to the DAG-one-versus-one approach. Since it requires the construction of a lower number of binary

classifiers, the graph has less inner nodes. In addition, the process of classification of each individual might require a different number of steps, in contrast to the DAG-one-versus-one approach. Thus, the result is a more efficient calculation. Furthermore, the class of every element participates in the binary classifier, either as one or as all, and the errors in each binary classifier do not propagate.

Given the above, the approach proposed here is presented as an improvement of the existent methods for the multiclassification with Support Vector Machines. Furthermore, this new approach has demonstrated its suitability for the resolution of the slate tile classification problem.

References

[1] Bottou L., Cortes C., Denker J., Drucker H., Guyon I., Jackel L., LeCun Y., Muller U., Sackinger E., Simard P. y Vapnik V. Comparison of classifier methods: a case study in handwriting digit recognition.*Proceedings of Int. Conf. Pattern Recognition*: 77-87, 1994.

[2] Burges C. A tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2: 121-167, 1998.

[3] Friedman J. Another approach to polychotomous classification. Dept. Statist., Stanford University, 1996.

[4] Frank, A. and Asuncion, A. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2010.

[5] Kijsirikul, B. and Ussivakul, N. Multiclass Support Vector Machines using Adaptative Directed Acyclic Graph. *Proceedings of the 2002 International Joint Conference on Neural Network. Honolulu*, 1990.

[6] Kraipeerapun, P., Nakkrasae, S., Fung, C.C. and Amornsamankul, S. Solving regression problem with complementary neural networks and an adjusted averaging technique *Memetic Computing*,2(4): 249-257, 2010.

[7] Knerr S., Personnaz L. y Dreyfus G. Single-layer learning revisited: A stepwise procedure for building and training a neural network. *Neurocomputing: Algorithms, Architectures and Applications Springer*, 1990.

[8] López, M., Martínez, J., Matías, J.M., Vilán, J.A. and Taboada, J. Application of a hybrid 3D-2D laser scanning to the characterization of slate slabs. *Sensors*, 10: 5949-5961, 2010.

[9] Platt J.C., Cristianini N. and Shawe-Taylor J. Large margin DAG´s for multiclass classification. *Adavances in Neural Information Processing Systems*. Cambridge, MA:MIT Press, 2000.

[10] Suresh, S., Savitha, R. and Sundararajan, N. A sequential learning algorithm for complex-valued self-regulating resource allocation network-CSRAN. *IEEE Transactions on Neural Networks*, 22(7): 1061-1072, 2011.

[11] Wang, D., Zheng, J., Zhou, Y. and Li, J. A scalable support vector machine for distributed classification in ad hoc sensor networks. *Neurocomputing*, 74 (1-3): 394-400, 2010.

$\{K, -1\}$-potent matrices and applications in image encryption

Leila Lebtahi*, Oscar Romero†, and Néstor Thome*

(*) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València. E–46022 Valencia, Spain,

(†) Departamento de Comunicaciones,

Universitat Politècnica de València. E–46022 Valencia, Spain,

November 30, 2012

1 Introduction

Involutory matrices have been widely studied. In cryptography, for example, they were first used in 1920's by Hill [1]. The Hill's idea was to use the same matrix for encrypting and decrypting avoiding the computation of an inverse matrix. The Hill cipher's keyspace consists of all matrices of a given size that are invertible over the ring $\mathbb{Z}_m$ of the integers modulo m. The number of such matrices was computed in [2].

Let J be the square matrix with ones on the cross diagonal and zeros elsewhere; note that J is often called the centrosymmetric permutation matrix. This matrix J allows to introduce the centro-invertible matrices as those matrices X such that its inverse coincides with the rotation of all the elements of the matrix through 180 degrees about the mid-point of the matrix, that is JXJ [3]. The author studied these matrices computing the total number of them by means of a bijection with the involutory matrices of the same size.

Notation: $\mathbb{Z}^{n\times n}$ denotes the set of $n \times n$ integer matrices and $\mathbb{Z}_m^{n\times n}$ the set of $n \times n$ matrices with coefficients in $\mathbb{Z}_m$. Throughout, for two given matrices $A = [a_{ij}] \in \mathbb{Z}^{n\times n}$, $B = [b_{ij}] \in \mathbb{Z}^{n\times n}$, and a positive integer m, the notation $A \equiv B(\mathrm{mod}\,(m))$ will denote the equivalence relation given by $a_{ij} \equiv b_{ij}(\mathrm{mod}\,(m))$, for all i and j. Moreover, if $M \equiv N(\mathrm{mod}\,(m))$ for $M, N \in \mathbb{Z}^{n\times n}$ then $AM \equiv BN(\mathrm{mod}\,(m))$.

In what follows we introduce a new type of matrix, namely the $\{K, -1\}$-potent matrices. They are an extension of the centro-invertible matrices where an involutory matrix K is used instead of the centrosymmetric permutation matrix J.

*e-mail: {leilebep,njthome}@mat.upv.es

†e-mail: oromero@dcom.upv.es

Definition 1 *Let $K \in \mathbb{Z}^{n\times n}$ be an involutory matrix. A matrix $A \in \mathbb{Z}^{n\times n}$ is called $\{K,-1\}$-potent if it satisfies $KAK = A^{-1}$.*

Our main goal is the construction of members of this class in an effective form. In order to compute them we will design an algorithm. Further, an application in image encryption will be developed and its advantages with respect to the centro-invertible matrices will be indicated. In addition, a numerical example is presented to show the performance of our algorithms and to demonstrate their applicability.

2 Constructing $\{K,-1\}$-potent matrices

Firstly, the next algorithm allows us to construct involutory matrices K. Moreover, for each computed matrix K, the corresponding $\{K,-1\}$-potent matrices are also performed.

Algorithm 1

Inputs: The size n of A and K.

Outputs: An involutory matrix K and a $\{K,-1\}$-potent matrix.

Step 1 Generate $n \times n$ random integer matrices R_K, R_A and Q_K, Q_A such that R_K, R_A are lower triangular and Q_K, Q_A are upper triangular, all of them with 1's in the main diagonal.

Step 2 Compute $P_K = R_K Q_K$ and $P_A = R_A Q_A$.

Step 3 Choose an arbitrary integer $1 \leq r(K) \leq n$ and set

$$K = P_K(I_{r(K)} \oplus -I_{n-r(K)})P_K^{-1}.$$

Step 4 Choose an arbitrary integer $1 \leq r(A) \leq n$ and set

$$A = KP_A(I_{r(A)} \oplus -I_{n-r(A)})P_A^{-1}.$$

End

3 Application in image encryption

In order to encrypt an image, we apply a set of encryption key matrices by means of matrix multiplications. This will alter the gray level or RGB levels, for B&W or color image, respectively. In case of color images, the following development will be applied to each one of the three RGB components.

In this section, we assume that a $\{K,-1\}$-potent matrix A has been constructed by means of Algorithm 1. This matrix will be used to encrypt and then its inverse will be needed to decrypt.

Algorithm for encrypting

Let us introduce the encryption function $e_A : \mathbb{Z}_m^{n\times n} \to \mathbb{Z}_m^{n\times n}$ defined by

$$e_A(X) = AX(\mathrm{mod}\,(m)) \text{ for } X \in \mathbb{Z}_m^{n\times n}.$$

Our procedure consists of restricting the function e_A to the set $\mathcal{X} = \{X_1, \ldots, X_t\}$ where X_i denotes an arbitrary $n \times n$ sub-image of the original $p \times q$ image for an arbitrary positive integer t. Once we have computed $Y_i = e_A(X_i) = AX_i$ for $i = 1, \ldots, t$, we perform the matrices $Y_{i,m} \in \mathbb{Z}_m^{n\times n}$ such that $Y_{i,m} \equiv Y_i(\mathrm{mod}\,(m))$. These last matrices $Y_{i,m}$ contain the encrypted sub-images corresponding to the starting sub-images X_i for $i = 1, \ldots, t$.

Algorithm for decrypting

In order to decryption we proceed as follows. Let us now consider the decryption function $d_A : \mathbb{Z}_m^{n\times n} \to \mathbb{Z}_m^{n\times n}$ defined by

$$d_A(Y) = (KAK)Y(\mathrm{mod}\,(m)) \text{ for } Y \in \mathbb{Z}_m^{n\times n}.$$

We first restrict the function d_A to the set $\mathcal{Y}$. Then, for every $i = 1, \ldots, t$, we get $d_A(Y_{i,m}) = X_i(\mathrm{mod}\,(m))$ Thus, the decrypted sub-images coincide with the matrices X_i.

4 Full/partial image encryption

For a full image encryption, the original image X is divided in square sub-images in the following form: Let p_x and p_y be the horizontal and vertical pixels of the original image. Let r denote the (positive integer) number of subdivisions in both horizontal and vertical of X. Then all sub-images $X_{k,j}$ are the squares $L_k \times H_j$ of length $\frac{p_x}{r}$ given by

$$L_k = \left[(k-1)\frac{p_x}{r} + 1, k\frac{p_x}{r}\right] \cap \mathbb{N} \qquad H_j = \left[(j-1)\frac{p_x}{r} + 1, j\frac{p_x}{r}\right] \cap \mathbb{N}$$

where $k \in \{1, \ldots, r\}$, $j \in \{1, \ldots, r\frac{p_y}{p_x}\}$, $\frac{p_x}{r} \in \mathbb{N}$, $r\frac{p_y}{p_x} \in \mathbb{N}$. Now, it is clear that the original image can be recovered as

$$X = \bigcup_{k=1}^{r} \bigcup_{j=1}^{r\frac{p_y}{p_x}} X_{k,j} = \bigcup_{k=1}^{r} \bigcup_{j=1}^{r\frac{p_y}{p_x}} (L_k \times H_j).$$

However, in some situations, it is necessary to blur an image partially, which is called partial image encryption. For this purpose, in the original image X we select rectangular sub-images $Z_1, \ldots, Z_s$ of adequate sizes. In each sub-image Z_i we apply the full image encryption method, that is, Z_i is divided into $(Z_i)_{k,j}$. If all the matrices $(Z_i)_{k,j}$ have the same size we can use the same $\{K, -1\}$-potent matrix A for all of them. Otherwise, different $\{K, -1\}$-potent matrices will be computed via Algorithm 1. Even more, in order to improve the security, we can use different key matrices A_i for each subimage X_i. The described method can be also applied to data encryption. In this case, it will be used the algorithm for encrypting/decrypting black and white images.

5 An application

Our algorithms can be easily implemented on a computer. We have used the MATLAB R2010b package. In this example we obtain an involutory matrix $K \in \mathbb{Z}^{4\times 4}$ and a random integer $\{K, -1\}$-potent matrix $A \in \mathbb{Z}^{4\times 4}$ applying Algorithm 1. Figure 1 (a) shows the original image partially encrypted via the algorithm for encrypting. The algorithm is applied to three subimages X_1, X_2 and X_3 of different sizes. Figure 1 (b) shows the decrypted image, that obviously coincides with the original one.

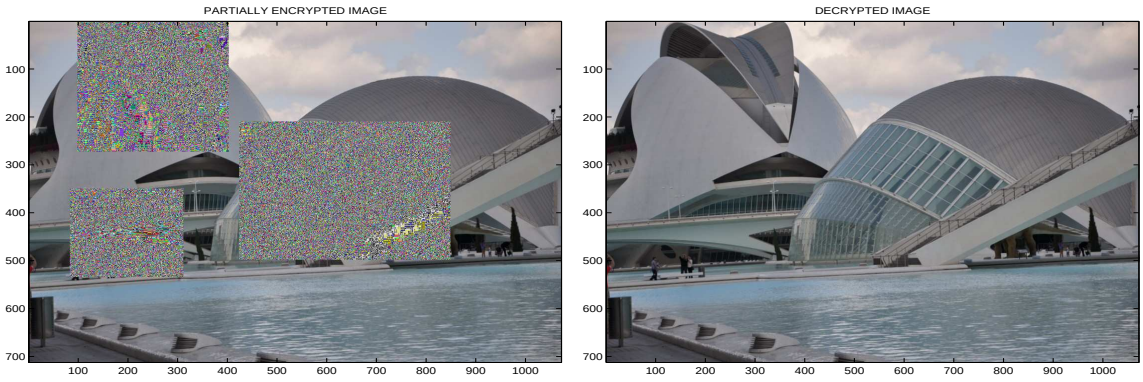


Figure 1: Partially encrypted and decrypted image

We can conclude that our encryption method provides a large number of keys because a large number of matrices A and K can be chosen for encrypting and decrypting. This large quantity of matrices is due to the randomness of the selection of these matrices as Algorithm 1 shows. Moreover, our procedure provides a more robust method than a single matrix method because several matrices are involved in the computation.

It is important to remark that our decrypting algorithm does not compute inverse matrices. An additional advantage is that multiple key matrices of different sizes can be used to encrypt/decrypt the sub-images $X_1, \ldots X_t$ from the same image.

Acknowledgements

This work has been partially supported by the Ministry of Education of Spain (Grant DGI MTM2010-18228).

References

[1] L.S. Hill, Cryptography in an algebraic alphabet, *American Mathematical Monthly*, 36(6): 306–312, 1929.

[2] J. Overbey, W. Traves, and J. Wojdylo. On the keyspace of the Hill Cipher. *Cryptologia*, 29(1): 59–72, 2005.

[3] R.S. Wikramaratna, The centro-invertible matrix: A new type of matrix arising in pseudo-randon number generation, *Linear Algebra and its Applications*, 434(1): 144–151, 2011.

Statistics and analytic compatibility to joint catalogs with a set of common ICRF defining sources

F.J. Marco*, J.A. López*, and M.J. Martínez†

(*) Universitat Jaume I,

Dep. Matemáticas, Castellón

(†) Universidad Politécnica de Valencia,

Dep. Matemática Aplicada, Valencia

November 30, 2012

1 Introduction

The construction of quasar catalogs from other catalogs with few points but having all of them a good number of ICRF defining sources represents an interesting way to obtain an increasingly extended and accurate catalog (see [1], [2] and [3]). Anyway, there are several questions that should be taken into account in order to successfully reach this aim. For example, the residuals of each catalog should be normally distributed and this property should be also translated to the final catalog. This final property must be compatible with other related with the "defining sources" which should be well corrected in such a way that, geometrically, conform a rigid structure. Also, the total residuals in the final catalog should not decrease while the variance increases too much, in order to avoid the introduction of excessive deformations. There are several ways to correct catalogs. Roughly speaking we can classify the different methods in parametric and non parametric. The

*e-mail:marco@mat.uji.es

parametric methods subdivide in geometrical (including corrections for rotation or rotation+deformation) and analytical (which consider developments in different sets of functions, such as spherical harmonics or Legendre-Fourier functions). All the parametrical methods make aprioristic suppositions about the function of residuals, in order to assure that this function belong to a certain functional space. On the other hand, non parametrical methods do not need to make any supposition about the function of residual, because they start off from the statistical properties of the data. Our method uses both techniques related with each procedure in order to get the best possible properties. We will set our attention in the problems that arise when we apply improperly the parametrical methods. We could remark that the application of the discrete least squares method can provide erroneous results due to different causes, such as: 1. Analytical causes: a lack of homogeneity in the data, because the functional orthogonality of the set of functions employed in the adjustment is not preserved 2. Statistical causes: The hypothesis of the Gauss-Markov theorem should be fulfilled in order to assure that the least squares method provides the least variance estimator in the class of unbiased estimators. All the former problems could be avoided if we use a non parametrical method as an intermediary. If this is the case, a good method is to compute an estimation of the function over fiducial points homogeneously distributed over the celestial sphere. This procedure has additional advantages: in the first place, each coefficient of a harmonic development can be computed with independence of the other. Secondly, each coefficient may be obtained from the numerical approximation of different integrals allowing that the widthband for each coefficient used in the non parametrical (local) adjustment to be the most suitable.

References

[1] Marco F. et al. A critical discussion on parametric and nonparametric regression methods applied to Hipparcos-FK5 residuals *Astronomy and Astrophysics*, Volume(418), 2004.

[2] Marco F. et al. Accurate Analytical and Statistical Approaches to Reduce O-C discrepancies in the Precessional Parameters *Pub. of the Astron. Soc. Pacific*, Volume(121), 2009.

[3] Marco F. et al. Determination of the Parameters of a Gaussian Mixture Distribution by Means of a Derivative-based Method *Under Review*

Improving CFD compressible segregated solvers by optimizing updates-equations sequence

Raúl Payri[†] [*], Santiago Ruiz[†],
Jaime Gimeno[†], and Pedro Martí-Aldaraví[†]

(†) CMT - Motores Térmicos, Universitat Politècnica de València,

Edificio 6D, Camino de Vera s/n, 46022, Valencia, Spain

November 30, 2012

1 Introduction

Focusing on automotive problems, concretely fuel injection problems without combustion, it is common to divide the study in two parts depending on the area of interest: internal flow and external flow. This division is made because of the different flow nature in the two areas, in the internal part the flow is *continuous*, mono-phase (or multi-phase if cavitation is considered), and in the external area far from the nozzle exit the flow is *dispersed* multi-phase.

If the whole injection process, internal and external flow, is going to be simulated at the same time, a mixture model Eulerian approach seems to be the best option. A mixture model with a liquid mass fraction that defines the percentage of liquid in the cell rather than a two-fluid model is desired due to the fuel droplets are very small downstream far from the nozzle exit and then tracking the interphase becomes very expensive in computational cost. The main awkwardness of this approach is the break-up model, in other words, how to get a dispersed phase from a continuous phase. This issue was solved in the Eulerian model presented in [6], where the dispersion is taken

[*]e-mail: rpayri@mot.upv.es

into account with a diffussion term in the mass fraction equation and the interphase surface is calculated by a new balance equation with convection, diffusion, production and destruction terms.

However, Vallet model [6] and its further improvements calculate the pressure with the equation of state (EOS) or an isentropic relationship between density and pressure, adding then certain hypothesis to the simulation. A new Eulerian two-phase model has been developed with the aim of simulating internal and external flows at once. This new solver uses a pressure equation derived from compressible continuity and momentum conservation equations as described in [7].

2 Methodology

The model proposed here is based in the same four principles than the one proposed by Vallet et al. [6]: (1) high Reynolds and Weber numbers, (2) the difference between the mean velocity of the liquid fluid and gaseous fluid particles can be calculated, (3) the dispersion of the liquid phase into the gas phase can be computed by a balance equation, and (4) the mean size of the liquid fragments can be calculated through the mean surface area of the liquid-gas interface per unit of volume.

Instead of the classical PISO algorithm, a PIMPLE approach is used. This algorithm combines the loop structures of PISO and SIMPLE, including $\partial/\partial t$ terms in equations but not limited by Courant number. In every outer loop, the sequence of transport equations is solved as follows: mass fraction equation with a Fick's law as diffusion closure term, continuity equation, momentum conservation equation, full energy equation by means of total enthalpy, the pressure equation derived from continuity and momentum conservation equations, and if desired turbulence equation (either RANS or LES). Density, viscosity, heat capacity and other fluid properties of the mixture are calculated as a linear interpolation of gas and liquid properties depending on the liquid mass/volume fraction [6].

The mass flux included in transport equations is obtained as the inner product of the interpolated density times velocity with the cell face vector, but usually only updated after the pressure equation inside the PISO loop. Nevertheless, fluxes can be updated in three different positions: (1) after mass fraction equation where the density changes because the amount of liquid inside cells changes or, if not, because the density has been updated in

the previous time step, (2) after continuity equation, and (3) after velocity equation. Notwithstanding, updating fluxes using the velocity field does not enforce the mass conservation principle because conservation is not enforced on U exactly, but on Φ (the flux is the conservative variable, not the velocity). Thus, conservation errors could be introduced by this way.

These three positions give up to 8 different sequences of updates-equations sequence.

3 Results

Accuracy and computational cost of the 8 sequences have been measured in three different steady state problems: (1) a one-dimensional problem where air initialized with zero velocity reaches the value set at the inlet which correspond to Mach number equal to 0.5; (2) an axi-symmetric convergent-divergent nozzle filled with incompressible and inviscid water; and (3) the same nozzle filled with compressible air.

For the one dimensional problem, though errors are always below 1% updating fluxes after the momentum equation reduces the accuracy and increases the computational cost. Update after the continuity equation does not affect the errors but requires 1 iteration per time step more to reach convergence. And the update after the mass fraction equation does not modify the performance of the solver.

For the incompressible nozzle problem the maximum error is always around 20% (in the pressure field) no matters the updates-equations sequence. However the maximum error in the temperature field grows from 0.1% to more than 2% if the update after the momentum equation is used.

For the compressible case, the most remarcable result is that sequences with the update after momentum equation do not converge. Despite, the same trends than for the one dimensional case are observed.

4 Conclusions

A new transient multi-phase compressible solver has been developed. Obtained solutions with this solver match the analytical ones for incompressible and compressible problems.

Up to 8 updates-equations sequences have been calculated for incompressible and compressible problems. It has been seen that updating fluxes in a non-conservative way leads to divergence in compressible problems, so this option can only be used inside the PISO loop where the internal corrector loop ensures convergence and the flux is calculated at the end in a conservative way from the pressure corrector.

For incompressible problems, none of the sequences changes the accuracy of the solution neither the computational cost, though the temperature drop is bigger with non-conservative updates. However, for compressible problems the accuracy is the same in all cases that converge, but sequences with no updates or only update 1 are slightly faster.

So, the final proposal for this model is updating the fluxes after the mass fraction equation. Thus, for multi-phase simulations mass fraction and density fields will be consistent in every time-step. No multi-phase results have been presented because any non-evaporative mixing reference problem (with analytical or high accuracy solution) has been found in the literature to compare with, though preliminary results carried out until now agree with this conclusion.

References

[1] J. M. Desantes, R. Payri, J. M. Pastor and J. Gimeno, Experimental characterization of internal nozzle flow and diesel spray behavior. Part 1: Non-evaporative conditions *Atomization and Sprays*, Volume (15):489–516, 2005.

[2] F. Payri, R. Payri, F. J. Salvador and J. Martínez-López, A contribution to the understanding of cavitation effects in Diesel injector nozzles through a combined experimental and computational investigation *Computer and Fluids*, Volume (58):88–101, 2012.

[3] R. Payri, F. J. Salvador, J. Gimeno and R. Novella, Flow regime effects on non-cavitating injection nozzle over spray behavior *International journal of Heat and Fluid Flow*, Volume (32):273–284, 2011.

[4] J. M. Desantes, R. Payri, F. J. Salvador and J. De la Morena, Influence of cavitation phenomenon on primary break-up and spray behavior at stationary conditions *Fuel*, Volume (89):3033–3041, 2010.

[5] R. Payri, J. Gimeno, P. Martí-Aldaraví and J. Manin, Fuel concentration in isothermal Diese sprays through structured planar laser imaging measurements *International Journal of Heat and Fluid Flow*, Volume (34):98–106, 2012.

[6] A. Vallet and A. A. Burluka and R. Borghi, Development of a eulerian model for the "atomization" of a liquid jet *Atomization and Sprays*, Volume(11):619–642, 2001.

[7] Hrvoje Jasak, Error Analysis and Estimation for the Finite Volume Method with Applications to Fluid Flows. Ph.D. thesis, Department of Mechanical Engineering, Imperial College of Science, Technology and Medicine (June 1996).

An explicit difference scheme for time dependent heat conduction models with delay

M.A. Castro *, F. Rodríguez, J. Cabrera, and J.A. Martín

Dep. Matemática Aplicada, Universidad de Alicante,

Apdo. 99, E03080, Alicante, Spain.

October 30, 2012

1 Introduction

Non-Fourier models of heat conduction have a long history, originally proposed to avoid the paradox of the infinite speed of propagation implied by the classical diffusion equation [1, 2], and later to account for phenomena as heat waves [3], or the propagation of second sound in helium [4]. In the last years, the huge technical progress in ultrafast lasers and in micro and nanoscale engineering, has lead to a parallel increase in the use of non-classical heat conduction equations, to properly model and analyze microscale heat transfer in a variety of engineering problems, as in ultrashort time laser processing of thin-film structures, heat transfer in nanofluids, or in laser irradiation of biological tissues (see, e.g., [5, 6, 7, 8]).

A family of non-Fourier heat conduction equations arise from the dual-phase-lag model [9, 10], where time lags are incorporated into Fourier law, which relates heat flux $\mathbf{q}(\mathbf{r}, t)$ and temperature gradient $\nabla T(\mathbf{r}, t)$, for a point $\mathbf{r}$ at time t,

$$\mathbf{q}(\mathbf{r}, t + \tau_q) = -k\nabla T(\mathbf{r}, t + \tau_T), \tag{1}$$

where τ_q and τ_T are the corresponding phase lags. An equation for heat conduction is obtained by combining (1) with the principle of energy conservation, and it reduces to the classical diffusion equation when $\tau_q = \tau_T = 0$.

*e-mail: ma.castro@ua.es

In other cases, it results into a partial differential equation with delay, either retarded, the so called delayed model, when $\tau_T = 0$, i.e., in the single-phase-lag model [11], and also if $\tau = \tau_q - \tau_T > 0$, or an advanced partial differential equation with delay when $\tau_q - \tau_T < 0$ [12, 13]. Some other models derive from the use of approximations with respect to the lags in (1). When first order approximations are used, the resulting heat conduction equation is usually referred to as the DPL model [9], and includes as a particular case, when $\tau_T = 0$, the Cattaneo-Vernotte model [1, 2], but higher order approximations have also been considered [14, 15].

There is a certain discussion about the validity of these non-Fourier models, as some of them may present stability problems [15, 16] or violate physical principles [17]. It should also be taken into account that non-classical effects only appear in transient behaviours, when the response time of interest is of the same order of magnitude as the relaxation time [10], whereas, for much longer times, the classical diffusion model is able to provide accurate descriptions of the evolution of the system. Therefore, it seems reasonable to consider a model of heat conduction combining both classical and delayed terms, with time varying strength, so that the term with delay eventually becomes negligible. Thus, in the long term, the model would approach classical diffusion, and its stability would be guaranteed.

In this work, we consider the following model for heat conduction,

$$\frac{\partial T}{\partial t}(x,t) = a(t)\frac{\partial^2 T}{\partial x^2}(x,t) + b(t)\frac{\partial^2 T}{\partial x^2}(x,t-\tau),\ 0 \le x \le l,\ t > \tau, \qquad (2)$$

a partial differential equation with delay and non-delay terms, and with time dependent coefficients, where $\tau > 0$, and $a(t)$, $b(t)$ are positive real functions. We will consider an initial condition

$$T(x,t) = \phi(x,t),\ 0 \le x \le l,\ 0 \le t \le \tau, \qquad (3)$$

and Dirichlet boundary conditions

$$T(0,t) = T(l,t) = 0,\ t \ge 0. \qquad (4)$$

Difference schemes for problem (2) in the constant coefficients case have already been proposed [18, 19]. The aim of this work is the construction of an explicit difference scheme for the time dependent problem (2)–(4), characterizing its convergence properties.

2 Explicit difference scheme

We consider a bounded domain $[0,l]\times[0,M\tau]$, and an uniform mesh $\{(x_p,t_n), p = 0\dots P, n = 0\dots MN\}$, defined by the increments $h = \Delta x$ and $k = \Delta t$, satisfying $l = Ph$ and $\tau = Nk$. The values of a function w at the points of the mesh, $w(x_p,t_n)$, is denoted $w_{p,n}$.

Let $u(x,t)$ denote the numerical approximation to $T(x,t)$ which is to be obtained with the difference scheme. By using finite differences approximations to partial derivatives in (2), as in the classical explicit scheme for the diffusion equation [20, pp. 6-8], and writing $\alpha = k/h^2$, $a_n = a\,(nk)$, and $b_n = b\,(nk)$, the system of difference equations

$$u_{p,n+1} = u_{p,n}+\alpha a_n\delta_x^2 u_{p,n}+\alpha b_n\delta_x^2 u_{p,n-N},\ p = 1,2,\dots,P-1,\ n = N,N+1,\dots,MN-1, \tag{5}$$

is obtained, together with the initial conditions

$$u_{p,n} = \phi\,(ph,kn,)\,,\ p = 0,1,\dots,P,\ n = 0,1,\dots,N, \tag{6}$$

and boundary conditions

$$u_{0,n} = u_{P,n} = 0,\ n = 0,1,\dots,MN. \tag{7}$$

These equations defining the scheme can be expressed in matrix form by introducing the vectors $\mathbf{u}_n = [u_{1,n},u_{2,n},\cdots,u_{P-1,n}]^T$, for $n = 0,1,\dots,MN$, where the superindex T denotes the transpose, and the $(P-1)\times(P-1)$ tridiagonal matrix

$$\mathbf{M} = \begin{bmatrix} -2 & 1 & 0 & \cdots & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 \\ 0 & 0 & 0 & \cdots & 1 & -2 \end{bmatrix},$$

as follows,

$$\mathbf{u}_{n+1} = (\mathbf{I}+\alpha a_n\mathbf{M})\,\mathbf{u}_n + \alpha b_n\mathbf{M}\mathbf{u}_{n-N},\ n = N,N+1,\dots,MN-1, \tag{8}$$

where $\mathbf{I}$ is the $(P-1)\times(P-1)$ identity matrix. Initial and boundary conditions (6) and (7) result in the matrix initial condition

$$\mathbf{u}_n = \phi_n,\ n = 0,1,\dots,N, \tag{9}$$

where $\phi_n = [\phi(h, nk), \phi(2h, nk), \ldots, \phi((P-1)h, nk)]^T$, for $n = 0, 1, \ldots, N$.

Introducing appropriate stack vectors, the scheme can also be expressed as a two-level scheme, which facilitates the analysis of its properties. In this way, it can be proved that if $b(t) < a(t)$, and $2\alpha(a(t) + b(t)) \leq 1$, for $\tau < t \leq M\tau$, then the scheme is convergent, and the local truncation error is of the order $O(k) + O(h^2)$.

References

[1] Cattaneo C. Sur une forme de l'équation de la chaleur éliminant le paradoxe d'une propagation instantanée. *C. R. Acad. Sci.*, 247:431–433, 1958.

[2] Vernotte, P. Les paradoxes de la théorie continue de l'équation de la chaleur. *C. R. Acad. Sci.* 246:3154–3155, 1958.

[3] Joseph D.D., and Preziosi L. Heat waves. *Rev. Mod. Phys.*, 61:41–73, 1989.

[4] Bertman B., and Sandiford D.J. Second sound in solid helium, *Sci. Am.*, 222: 92, 1970.

[5] Qiu, T. Q., and Tien, C. L. Short-pulse laser heating on metals. *Int. J. Heat Mass Transfer*, 35:719–726, 1992.

[6] Qiu, T. Q., and Tien, C. L. Heat transfer mechanisms during short-pulse laser heating of metals. *ASME J. Heat Transfer*, 115:835–841, 1993.

[7] Wang, L., and Wei, X. Heat conduction in nanofluids. *Chaos Solitons Fractals*, 39:2211–2215, 2009.

[8] Zhou, J., Chen, J.K., and Zhang, Y. Dual-phase-lag effects on thermal damage to biological tissues caused by laser irradiations. *Comput. Biol. Med.*, 39:286–293, 2009.

[9] Tzou, D.Y. The generalized lagging response in small-scale and high-rate heating. *Int. J. Heat Mass Transfer*, 38:3231-3240, 1995.

[10] D.Y. Tzou, Macro- to Microscale Heat Transfer: The Lagging Behavior. Washington, Taylor & Francis, 1996.

[11] Tzou, D.Y. On the thermal shock wave induced by a moving heat source. *J. Heat Transfer*, 111:232–238, 1989.

[12] Kulish, V.V., and Novozhilov, V.B. An integral equation for the dual-lag model of heat transfer. *ASME J. Heat Transfer*, 126:805–808, 2004.

[13] Xu M., and Wang, L. Dual-phase-lagging heat conduction based on Boltzmann transport equation. *Int. J. Heat Mass Transfer*, 48:5616-5624, 2005.

[14] Tzou, D.Y. A unified approach for heat conduction from macro to micro-scales. *ASME J. Heat Transfer*, 117:8–16, 1995.

[15] Quintanilla, R., and Racke, R. A note on stability in dual-phase-lag heat conduction. *Int. J. Heat Mass Transfer*, 49:1209-1213, 2006.

[16] Jordan, P.M., Dai, W., and Mickens, R.E. A note on the delayed heat equation: Instability with respect to initial data. *Mech. Res. Comm.*, 35:414-420, 2008.

[17] Christov, C.I., and Jordan, P.M. Heat conduction paradox involving second-sound propagation in moving media. *Phys. Rev. Lett.*, 94:154301, 2005.

[18] García, P., Castro, M.A., Martín, J.A., and Sirvent, A. Numerical solutions of diffusion mathematical models with delay. *Math. Comput. Modelling*, 50:860–868, 2009.

[19] García, P., Castro, M.A., Martín, J.A., and Sirvent, A. Convergence of two implicit numerical schemes for diffusion mathematical models with delay. *Math. Comput. Modelling*, 52:1279–1287, 2010.

[20] J.W. Thomas, *Numerical Partial Differential Equations*, Springer-Verlag, New York, 1995.

An integral optimization using a Gravitational Search Algorithm (GSA). An application to on shore wind farm

I. Marton[*], A. Sanchez[†], S. Carlos[*], and S. Martorell[*]

(*) Departament d' Enginyeria Química i Nuclear,

(†) Departament de Estadística i Investigació Operativa Aplicades i Qualitat,

Universitat Politècncia de València. Camino de Vera, 14. 46022 València. Spain.

November 30, 2012

1 Introduction

In Spain, the actual wind power capacity is around the 16 percent of the total capacity available to supply the electricity demand. Due to the random nature of the wind, it is of great importance to assure wind farms will be in optimal conditions to produce the maximum amount of electricity possible [1]. The energy produced by a wind turbine is directly proportional to its availability and wind farm profit can increase with an appropriate preventive and corrective maintenance management. Furthermore, suitable maintenance intervals and spare parts are planned to increase turbine availability [2]. The objective of this paper is focused on the optimization the maintenance strategies under unavailability and cost criteria considering as decision variables maintenance intervals and material resources needed in the maintenance tasks.

[*]e-mail: ismarllu@upv.es

2 Cost and unavailabilty models

The total cost is modeled using the contribution of performing preventive and corrective maintenance and the necessary spare parts to perform such maintenance. The preventive and corrective maintenance cost are evaluated using Eqn.(1) and Eqn.(2), respectively:

$$c_{pm}(x) = \sigma/M \cdot c_{hpm}, \tag{1}$$

$$c_{cm}(x) = \lambda \cdot \mu \cdot c_{hpm}, \tag{2}$$

where,x is the decision variable, σ is the preventive maintenance duration, M is the preventive time interval, λ is the failure rate, μ is the corrective maintenance duration and c_{hpm} and c_{hcm} are the man-hour cost. The cost of spare parts $c_r(x)$ includes: the cost of storage, $c_a(x)$, the purchase cost $c_c(x)$, the cost due to the lack of spare parts, $c_{nr}(x)$, and the cost of excess $c_{ex}(x)$ and is calculated as:

$$c_r(x) = c_a(x) + c_c(x) + c_{nr} + c_{ex}(x), \tag{3}$$

The total cost, c_t, is evaluated summing Eqn.(1), Eqn.(2) and Eqn.(3).

The unavailability of a system is evaluated by solving the fault tree determining the minimal cut sets, MCS, which represent a unique combination of component failures that cause the system failure. The total system unavailability is given by:

$$u_t(x) = \sum_j \prod_k u_{jk}(x), \tag{4}$$

where j is the number of minimal cut sets and k the number of basic events in a given MCS. In this work,the component is normally in operation. The main unavailability contributions are due to preventive and corrective maintenance evaluated using Eqn.(5) and Eqn.(6) :

$$u_{pm}(x) = \sigma/M, \tag{5}$$

$$u_{cm}(x) = \lambda \cdot (\mu + \rho + \delta), \tag{6}$$

where,ρ is the delay time due to lack of spare parts in stock and δ is the logistic delay.In the case of preventive maintenance δ and ρ are zero. To model corrective maintenance two types of components are considered:

Table 1: Components reliability, maintainability and costs.

Reliability					
Component	λ (hrs^{-1})	μ (hrs^{-1})	ρ (weeks)	δ (weeks)	Component type
$C1$	$0.55E-6$	72	–	16	No repairable
$C2$	$5.48E-6$	144	16	–	Repairable
$C3$	$0.55E-6$	72	–	1	No repairable
$C4$	$5.48E-6$	10	16	–	Repairable
Costs					
Component	c_a *euros*	c_c *euros*	c_{nr} *euros*	c_{ex} *euros*	c_h *euros*
$C1$	75000	400	–	7500	50
$C2$	400000	–	120000	40000	50
$C3$	1500	400	–	150	50
$C4$	120000	–	36000	12000	50

repairable and non-repairable. In first type of components δ in Eqn.(6) is always zero and for the second type ρ is always zero.

3 Gravitational Search Algorithm (GSA)

Gravitational Search Algorithm, GSA,[3] is a heuristic algorithm based on the gravity laws and mass interactions. This algorithm works with a multi-dimensional coordinate space defined by the problem, each point in space is a possible solution of it. In the GSA the agents are considered objects and their performance is measured by its mass. The objects attract each other by the force of gravity that causes an overall movement of all the objects to those heavier.

4 Application case

The application case is focused on the optimization of the number of spare parts and the preventive maintenance time interval of a wind turbine. The wind farm is composed by 20 wind turbines consisting by four components in a serial configuration. Table 1 shows the characteristics of reliability and costs of the wind turbine components. The optimization problem is raised as a single-objective problem being the vector of decision variables: $x = (M, n_1, n_2, n_3, n_4)$. Where M is the preventive maintenance time interval, and n_1,n_2,n_3 y n_4 the number of spare for components 1 to 4. Two optimization problems were considered, in the first one the cost is minimized

Table 2: Results.

Objective Function	*Constrain*	*Initial values*	*Min* $c_t(x)$	*Min* $u_t(x)$
$u_t(x)$	$c_t(x) < 16244$	16244		0.052
$c_t(x)$	$u_t(x) < 0.01$	0.01	3191	
$M(hours)$		180	11616	11972
n_1		1	1	3
n_2		1	1	1
n_3		1	3	3
n_4		1	1	2

using unavailability as a constrain, and the second one unavailability is minimized, using cost as a constrain. The optimization is performed using a MatLab implementation of the GSA, considering 50 agents, 1000 iterations, and a ramdom initial population. Table 2 shows the results obtained.

5 Concluding remarks

The optimization of maintenance planning is a key aspect in industrial plants. Traditionally, maintenance planning was formulated as an optimization problem with constrains unavailability and/or cost under criteria. In this work the frequency of maintenance and material resources available to perform these maintenance activities are included. The problem has been solved by a gravitational search algorithm demonstrating the feasibility of its application to maintenance optimization problems.

References

[1] J., Wen, Y., Zheng,F., Dongham. A review on reliability assessment for wind power. *Renewable and Sustainable Energy Reviews.*, Volume(13):2485–2494, 2009.

[2] C.Walfrod. Wind turbine reliability: understanding and minimizing wind turbine operation and maintenance costs. Sandia National Laboratories. Albuquerque, NM, 2007.

[3] E., Rashedi, H., Nezamabadi-Pour, S., Saryazdi. SA: A Gravitational Search Algorithm. *Information Sciences.*, Volume(179 (13)):2232–2248, 2009.

Mathematical modeling of workaholism in Spain: analyzing its economic and social impact

E. de la Poza[†] [*], M. del Líbano[‡], I. García[○],
L. Jódar[⋆] and P. Merello[⋆]

(†) Departamento de Economía y Ciencias Sociales,

(⋆) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València.

(‡) Facultad de Ciencias Humanas y Sociales, WONT Team, Universitat Jaume I de Castelló.

(○) Departamento de Comunicación Audiovisual y Publicidad.

Universidad del País Vasco.

November 30, 2012

1 Introduction

One of the addictions not produced by the consumption of psychoactive substances is workaholism, a syndrome characterized by a tendency to work excessively in a compulsive way [1].

In the context of the current economic and financial crisis, Spanish companies recruit and promote those employees capable to work intensely and commit to help their firms; they are known as 'work-engaged' employees [2]. Workaholism is associated with negative consequences such as stress [3], psychosomatic symptoms [4], physical exhaustion [5], burnout [6], poor social relationships [7], family problems [8] and poor job performance [9].

[*] elpopla@esp.upv.es

The aim of this study is to develop a mathematical model to predict workaholism for the next four years (2011-2015) in Spain considering as main variables that can influence the addiction: the social contact [10], [11], the economic situation [12], and the marital termination [13]. This will lead us to estimate the social effects and possible public health recommendations.

2 Mathematical model and results

In order to analyze the level of addiction to work, we passed a validated questionnaire [14] at two different dates, obtaining two samples first one (S1) in May 2011 and second one (S2) in April 2012.

Thus, three subpopulations are defined for the construction of the proposed model:

- Rational workers (N): those individuals who work 40 or less hours per week and obtain a score lower than 3.25 in the short Spanish version of the DUWAS [14].
- Overworkers (S): those individuals who work more than 40 hours per week and obtain a score lower than 3.25 in the short Spanish version of the DUWAS.
- Workaholics (A): those individuals characterized by obtaining a score higher than 3.25 in the short Spanish version of the DUWAS.

The total population (P) at any time n (months) is expressed as follows:

$$P_n = N_n + S_n + A_n \,. \tag{1}$$

The factors considered in the development of workaholism are the economic scenario (with different unemployment rate values), transition due to emotional impacts (embracing marital dissolution through divorce, separation and annulment) and the social contagion. The model was built considering the assumptions of previous studies of social contagious [11]. The values of all parameters were estimated from different sources of information and some assumed hypothesis, with the exception of β_a (considered in β_1, which represents the proportion of concerned rational workers (N) who decide to increase their working hours) and the social contagion rate γ_2 both adjusted by the model with the Nelder-Mead algorithm.

The dynamic of the population can be described by the following system of difference equations (n, time in years):

$$\begin{aligned} N_{n+1} - N_n &= b(N_n + S_n + S_n) - dN_n - pN_n - \beta_1 N_n - \alpha_1 N_n \,, \\ S_{n+1} - S_n &= -dS_n - pS_n + \beta_1 N_n - g_3 S_n - \gamma_2 S_n A_n - \alpha_2 S_n + \epsilon_2 A_n + \alpha_1 N_n \,, \\ A_{n+1} - A_n &= -pA_n - dA_n + \alpha_2 S_n + \gamma_2 S_n A_n - \epsilon_2 A_n + g_3 S_n \,, \end{aligned} \tag{2}$$

Different simulations are developed considering that the value of unemployment rate (modifying parameters p and β_1) evolves during the next four years assuming different economic scenarios OECD, FUNCAS, and a pessimistic and optimistic ones.

In the course of four years the prevalence of workaholism will be almost tripled, from 4% in July 2011 to approximately 11.5% in December 2015. Particularly, the percentage of workaholics in the Spanish population is higher for the optimistic economic scenario.

3 Conclusions

The difference equations model presented in this work has allowed us to forecast the evolution of workaholics in Spain from 2011 to 2015. The results show an increasing trend of the addiction for the four economic scenarios considered. In order to prevent this evolution, public health and educational recommendations jointly with healthy organizational cultures have been proposed.

References

[1] Schaufeli W. B., Taris T. W. and Van Rhenen W. Workaholism, burnout and engagement: Three of a kind or three different kinds of employee well-being, *Appl Psychol-Int Rev*, 57, pp. 173–203, 2008.

[2] Schaufeli W. B. and Bakker A. B. Job demands, job resources and their relationship with burnout and engagement: A multi-sample study, *J Organ Behav*, 25, pp. 293–315, 2004.

[3] Andreassen C. S., Ursin H. and Eriksen H. R., The relationship between strong motivation to work, "workaholism", and health, *Psychol Health*, 22, pp. 615–629, 2007.

[4] Burke R. J., Oberklaid F. and Burgess Z., Workaholism among Australian women psychologists: antecedents and consequences, *Women in Management Review*, 5, pp. 252–259, 2004.

[5] Sonnentag S. Recovery, work engagement, and proactive behavior: A new look at the interface between non-work and work, *J Appl Psychol*, 88 , pp. 518–528, 2003.

[6] Schaufeli W. B., Bakker A. B., Van Der Heijden F. and Prins J. T. Workaholism, Burnout and well-being among junior doctors: the mediating role of role conflict, *Work Stress*, 23, pp. 155–172, 2009.

[7] Burke R. J. and Koksal H. Workaholism among a sample of Turkish managers and professionals: An exploratory study, *Psychol Rep*, 91, pp. 60–68, 2002.

[8] Robinson B. E. and Post P. Risk of addition to work and family functioning, *Psychol Rep*, 81, pp. 91–95, 1997.

[9] Shimazu A., Schaufeli W. B. and Taris T. W. How does workaholism affect worker health and performance?, *Int J Behav Med*, 17, pp. 154–160, 2010.

[10] N.A. Christakis and J.H. Fowler, Connected: The Surprising Power of Our Social Networks and How they Shape Our Lives, Hachette Book Group, New York, 2009.

[11] De la Poza E., Guadalajara N., Jódar L. and Merello P. Modeling Spanish anxiolytic consumption: Economic, demographic and behavioral influences, *Math Comput Model*, Article in Press, 2011.

[12] Fry L.W. and Cohen M-P. Spiritual leadership as a paradigm for organizational transformation and recovery from extended work hours cultures, *J Bus Ethics* 84, pp. 265–278, 2009.

[13] Yaniv G. Workaholism and marital estrangement: a rational-choice perspective, *Math Soc Sci* 61 (2), pp. 104–108, 2011.

[14] Del Líbano M., Llorens S., Salanova M. and Schaufeli W.B. Validity of a brief workaholism scale, *Psychothema*, 22, pp. 143–150, 2010.

Application of the Level Set Method for the Visual Representation of Continuous Cellular Automata Oriented to Anisotropic Wet Etching

C. Montoliu[†] [*], N. Ferrando[‡], J. Cerdá[†],
R. J. Colom[†] and M. A. Gosálvez[§]

([†]) Instituto de Instrumentación para Imagen Molecular (I3M).
Centro mixto CSIC Universitat Politècnica de València CIEMAT,
Camino de Vera s/n, 46022 Valencia, Spain,

([‡]) Centro de Física de Materiales,
Centro mixto CSIC-UPV/EHU and Donostia International Physics Center (DIPC),
20018 Donostia-San Sebastian, Spain.

([§]) Department of Materials Physics, University of the Basque Country (UPV-EHU),
Donostia International Physics Center (DIPC), and Centro de Física de Materiales,
Materials Physics Center (CFM-MPE), 20018 Donostia-San Sebastian, Spain

November 30, 2012

1 Introduction

Anisotropic wet chemical etching is one of the most popular bulk micromachining methods for the fabrication of Micro-Electro-Mechanical Systems (MEMS). By using this process it is possible to get suspended microstructures with both flats and smooth surfaces. Another interesting point of this

[*]e-mail:carmonal@upv.es

process is its low cost. However, the final result of a particular etching process depends on many factors so it is hard to predict the resulting structure.

In order to ease the design of MEMS, an important effort has been made through last years in order to accurately model and simulate the process for microengineering applications. First simulators of the wet etching process were based on geometrical models [1] which understand the etching process as a set of moving flats.

On the other hand, Cellular Automata (CA) models the moving surface as a collection of points that represent the substrate. The etching process is simulated by making surface atoms to disappear according to some microscopic rules, letting the neighbouring sites to emerge into the surface. The Continuous CA (CCA) model is currently accepted as the most adequate model for simulating this process in terms of performance and accuracy.

Despite the accurate results obtained with CCA, the final result is a cloud of unconnected points, so it is hard to visualize correctly some details of the surface, specially at complicated topologies. The method currently used in [3] to improve the visual representation, is to shade the points depending on its normal vector, but in complex morphologies of some structures the shading accuracy of the method is not good enough and adds too much noise, so the visualization quality decreases greatly. Thus, in order to improve the final visualization it is necessary to obtain a continuous surface from the information of unconnected points. Also, it is important to consider all the different kinds of topologies that can be obtained due to anisotropic wet etching. In this study an implementation of the Level Set (LS) method is presented due to the robustness that it has proven in the recent years.

2 Level Set method applied to surface reconstruction

The LS method was introduced for capturing moving fronts [5]. The main idea of this method consists in embed the front $x(t)$ in a signed distance function ϕ such as $\phi\left(x(t), 0\right) = 0$. Therefore, the movement is applied to the function ϕ instead of the front itself.

In the LS method, the zero-level of the ϕ function is the propagating front between both phases, and the main idea when LS is applied to surface reconstruction is to build an initial surface exterior to the cloud of points,

embed it in the ϕ function and move it to the cloud until it is close enough. In this study the minimal energy model presented in [9] has been chosen due to this model does not need any additional information, only the data points (from now on referred as set S), and also because CCA resulting data set follows a crystallographic structure, so the density of the points is uniform.

The LS method has a high computational cost ($O(N^3)$), where N is the number of grid points in each dimension, since every point of the three-dimensional grid needs to be updated. In order to reduce computational cost to $O(N^2)$, the technique presented in this study implements an optimization based on Sparse Field Method (SFM) [8], which consists in update only those points that are strictly necessaries to evolve the front.

3 Results

In this section some reconstructed surfaces with our implementation are shown, as well as performance results. In addition, results obtained with CCA simulator [3] are also shown in order to compare the improvement obtained with the proposed technique. Three examples of experiments are collected in figure 1.

	Grid size	Points of S	Iterations	Execution time (s)
a)	274x800x26	1109390	73	40.6
b)	212x212x117	287034	70	19.4
c)	352x352x49	785471	60	35.1

Table 1: Summary of the parameters used to obtain simulations results shown in figure 1.

Performance data of these simulations is collected in table 1. The first column shows the size of the built grid, the second one shows the number of the points produced by CCA simulator, the third column is the number of iterations needed to converge and finally time required by LS algorithm to converge is shown.

Due to SFM and the low order derivatives that have been used, time required by LS algorithm to converge is only about a few tens of seconds even at example c) which has a grid of more than 6 millions points. Thus, taken into account the improvement of details visualization obtained, our proposed technique can be considered a fast and accurate tool.

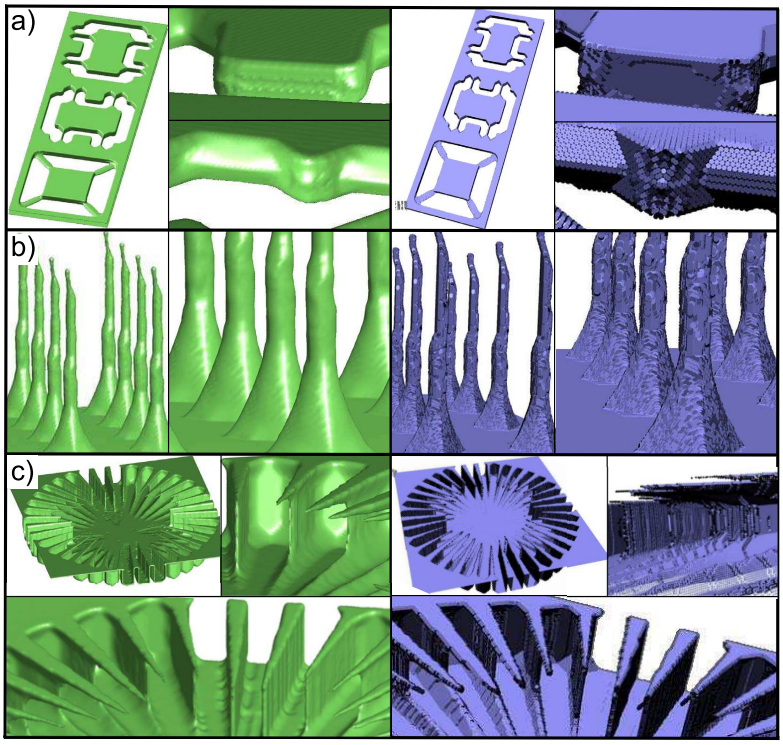


Figure 1: Green figures: surfaces reconstructed with our implementation. Blue figures: shaded point cloud CCA simulator solutions [3]. a)Three-axis accelerometer [6], b)microneedles [7] and c)wagon wheel [2].

4 Conclusion

In this study it has been demonstrated that by using our implementation of the LS method, the visual representation of CA based simulators is improved, so it can be very useful to ease the design of MEMS. Also, it has been probed that first order derivatives are sufficient to get a good level of details. Another interesting fact is that our implementation has been able to reconstruct all the tested topologies types with only tens of seconds, so it can be considered a fast and robust technique. In addition, execution time could be reduced even more if the algorithm is implemented in a parallel way to be executed on a many-core architecture like GPUs, as it has been demonstrated previously on similar LS methods [4].

References

[1] K. Asaumi, Y. Iriye, and K. Sato, *Anisotropic-etching process simulation system microcad analyzing complete 3d etching profiles of single crystal silicon*, IEEE Micro Electro Mechanical Systems (1997), pp. 412–417.

[2] M.A. Goslvez, P. Pal, N. Ferrando, H. Hida, and K. Sato, *Experimental procurement of the complete 3d etch rate distribution of si in anisotropic etchants based on vertically micromachined wagon wheel samples*, J. Micromech. Microeng. 21 (2011), p. 125007.

[3] IntelliSense-Corp., *Intellietch web page:* `http://www.intellisensesoftware.com/modules/IntelliEtch.html` (2010).

[4] A.E. Lefohn, J.E. Cates, and R.T. Whitaker, *Interactive gpu-based level sets for 3d segmentation*, Proc. of Medical Image Computing and Computer Assisted Intervention (2003), pp. 564–572.

[5] S. Osher and J.A. Sethian, *Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations*, Journal of Computational Physics 79 (1988), pp. 12–49.

[6] G. Schrpfer, M. de Labachelerie, S. Ballandras, and P. Blind, *Collective wet etching of a 3d monolithic silicon seismic mass system*, J. Micromech. Microeng 8 (1998), pp. 77–79.

[7] M. Shikida, M. Ando, Y. Ishihara, T. Ando, K. Sato, and K. Asaumi, *Non-photolithographic pattern transfer for fabricating pen-shaped microneedle structures*, Journal of Micromechanics and Microengineering 14 (2004), pp. 1462–1467.

[8] R.T. Whitaker, *A level-set approach to 3d reconstruction from range data*, International Journal of Computer Vision 29 (1998), pp. 203–231.

[9] H.K. Zhao, S. Osher, B. Merriman, and M. Kang, *Implicit and non-parametric shape reconstruction from unorganized data using a variational level set method*, Computer Vision and Image Understanding 80 (2000), pp. 295–319.

Forecasting the protection provided by the current vaccination schedule against meningitis C

L. Acedo [*], J. Díez-Domingo[†], J.-A. Moraño[*],
L. Pérez-Breva[†], R.-J. Villanueva[*] and J. Villanueva-Oller[‡]

(*) Instituto Universitario de Matemática Multidisciplinar

Universitat Politècnica de València, Spain

(†) Centro Superior de Investigación en Salud Pública, Valencia, Spain

(‡) Centro de Estudios Superiores Felipe II,

Aranjuez, Madrid, Spain

November 30, 2012

1 Introduccion

Meningococcal disease is caused by the bacterium Neisseria meningitidis, also called meningococcus. About 10% of people have this type of bacteria in the back of their nose and throat with no signs or symptoms of disease, being called 'a carrier'. But sometimes Neisseria meningitidis bacteria can invade the body causing certain illnesses, which are known as meningococcal diseases [1].

Neisseria meningitidis bacteria (fig. 1) are spread through the exchange of respiratory and throat secretions like spit (e.g., living in close quarters, kissing, sharing drinks). Fortunately, these bacteria are not as contagious as what causes the common cold or the flu. Besides, the bacteria are not spread by casual contact or by simply breathing the air where a person with meningococcal disease has been. Sometimes Neisseria meningitidis bacteria spread to people who have had close or lengthy contact with a patient with meningococcal disease. People in the same household, roommates, or anyone with direct contact with a patient's oral secretions, meaning saliva or spit, (such as a boyfriend or girlfriend) would be considered at increased risk of getting the infection [1] (fig. 2).

Meningitis is an infection of the brain and spinal cord and can even infect the blood. Before 1990 the main cause was the bacterium Haemophilus influenzae: (almost completely eradicated by the Hib vaccine). Nowadays the main cause of Meningitis is the bacterium

[*]e-mail: luiacrod@imm.upv.es

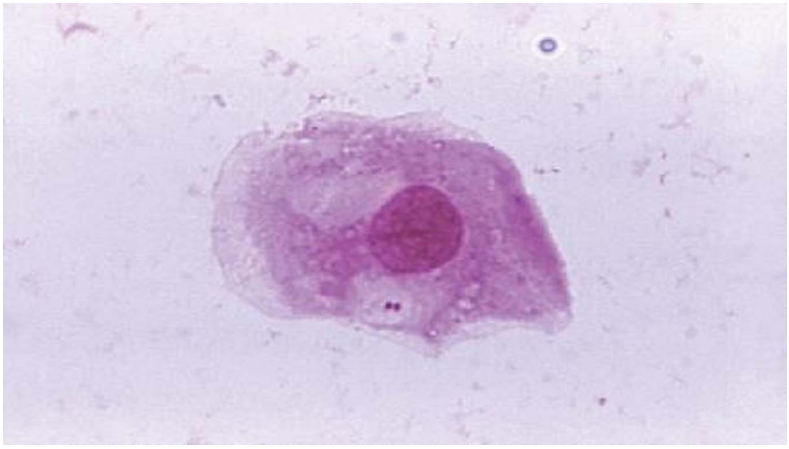


Figure 1: A photomicrograph of Neisseria meningitidis recovered from the urethra of an asymptomatic male; Magnified 1125X.

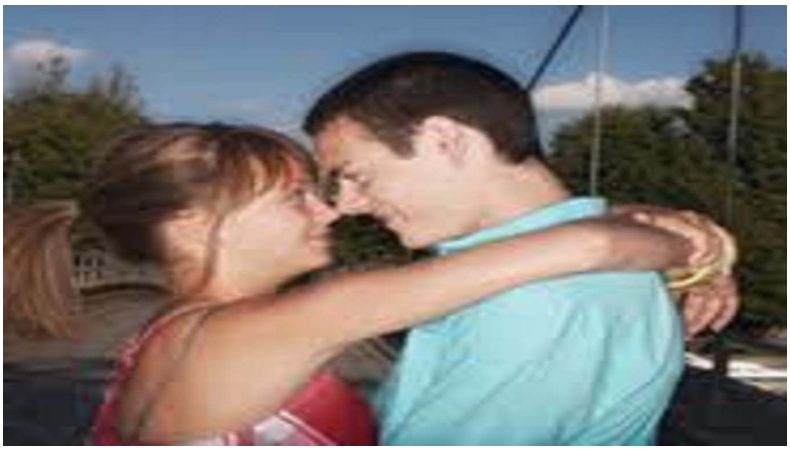


Figure 2: Relations among adolescents is the main cause of transmission of Neisseria meningitidis.

Figure 3: Hands with gangrene due to meningococcemia of a 4 month old female.

Neisseria meningitidis which is transmitted exclusively among humans, but mainly during adolescence [2].

This transmission is made by healthy carriers. The treatment is made with specific antibiotics but even properly treated, there is up to 10% of mortality and 10% of survivors have sequelae [3] (fig. 3).

1.1 Meningococcal C: Incidence, serogroups and vaccines

Low protection levels in adolescence increases the transmission to children under one year old, who may get infected more easily. There are several serogroups being the main ones A, B, W135, C and Y. We are interested in serogroup C, the responsible of meningococcal C. Nowadays, there are several types of vaccines: Simple polysaccharides against serogroups A and C, Simple polysaccharides against serogroups A, C, Y and W135, Meningococcal serogroup C conjugate (MCC) vaccine. There is still no vaccine against serogroup B.

In the Community of Valencia (CV) there have been different vaccination campaigns. In 1997, 85% of population between 18 months and 19 years of age was immunized with the bivalent polysaccharide vaccine A+C. From 2000 the Conjugate Vaccine C has been used in campaigns with different strategies: In 2001, it was incorporated at vaccination schedule of children under 6 months of age and 1 dose for children between 1 and 6 years old. In 2002, this dose was extended to 19 years old (fig. 4). In 2006 is fixed the current vaccination schedule with three doses at 2, 6 and 18 months of age [4].

Recent studies on the MCC-vaccination have determined that levels of protection provided by this vaccine are lower than expected, in particular, in toddles. Doctors conjecture that, in 5 or 10 years, there will be an increase of cases in children younger than a year

Figure 4: A young girl is being vaccinated against Neisseria meningitidis.

because the herd immunity provided among the adolescents by the current vaccination schedule will decrease. Joint Committee on Vaccination and Immunization of DH in UK has recommended in January 2012 a change in the vaccination schedule of this country. They conclude that an adolescent dose of MCC-vaccine should be introduced and a dose in infants should be removed [5]. This change needs to ensure that coverage is high enough to maintain the herd immunity. This recommendation is not based on a mathematical modelling study. Our objective is to support the schedule change with a mathematical model. Here, as a first step, we try to support the doctors conjecture.

2 How to state the model: Data hunt

There are no data of carriers in literature, except in [6], although there are data of cases (currently very few). Moreover, the period of carriage is very short and it is difficult to count carriers. Besides, we cannot assume a stationary situation because few years ago, in Spain, serogroup B was substituted partially by serogroup C. These reasons and the lack of data lead us to no propose a typical SIS model.

Most of data in the recent literature are based on analysis of the Serum Bactericidal Activity (SBA) in blood. SBA is related to the immunity against meningococcal disease ($SBA \geq 1/8$), not with the carriage state. The studies analysing SBA in blood samples give a general trend about the population protection against meningococcal C, but they are not comparable and do not allow a quantification of the lose of the protection over the time. In 2011, under a research project doctors in the Centro Superior de Investigación en Salud Pública and Health Institute Carlos III have measured the SBA in 1800 individuals

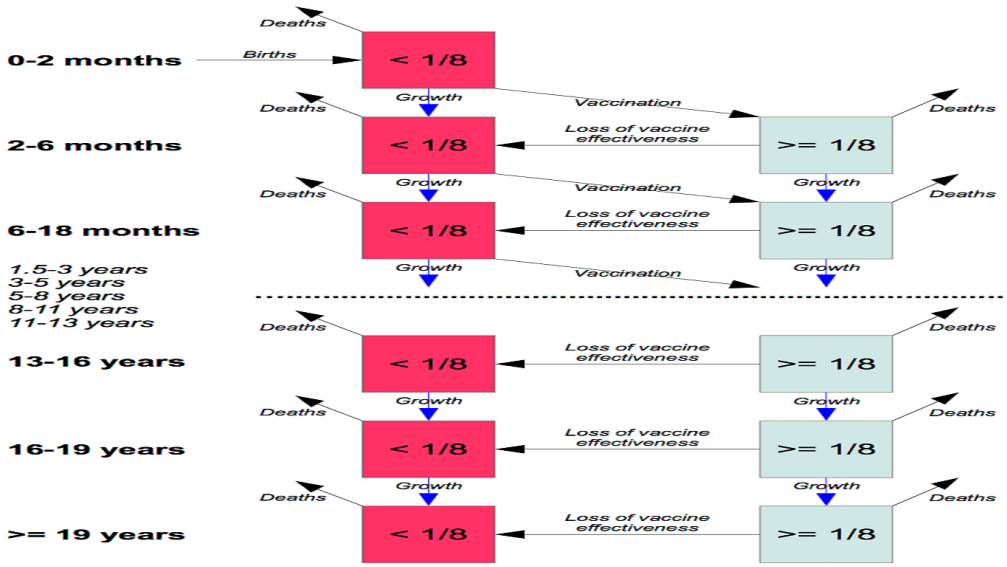


Figure 5: Flow diagram of the model.

of different ages in CV (SBA-CV data). Taking the above information, we are going to present a model in order to predict the people with protection against meningococcal disease over the next few years.

3 Model building

We consider a continuous model with 11 age-groups determined by SBA-CV data. This data are initial conditions of the model and we take into account the current vaccination schedule (2, 6, 18 months of age). Following the trends pointed out by the literature we define different loss of the vaccine effectiveness with decreasing values according to the age. In this model there are not unknown parameters. We can see a representation of this model in the following figure 5:

A person of a determined age group may die or may grow and move to the next age group. Besides, an individual can also change his/her protection or no protection respect to the disease. If the individual belongs to the protected group $\left(SBA \geq \frac{1}{8}\right)$ he/she can pass to the unprotected group due to the loss of vaccine effectiveness. If, however, the individual already belongs to the unprotected group $\left(SBA < \frac{1}{8}\right)$ he/she can move to the protected group by means of vaccination with a rate determined by the vaccine coverage.

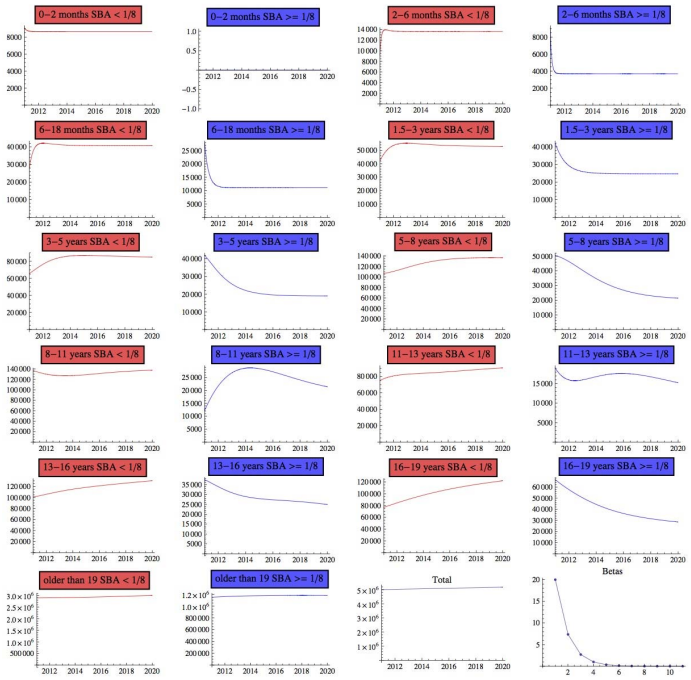


Figure 6: It can be seen how the adolescent age groups have a decreasing trend in the protected $\left(SBA \geq \frac{1}{8}\right)$ subpopulations.

4 Simulation

In our simulation graphs (fig. 6), we can see that the groups with $SBA \geq \frac{1}{8}$ decrease, in particular, the corresponding to adolescents. Doctors suppose that the reduction in the protection may lead to an increase in the meningitis cases in children younger than a year old (herd immunity decreases).

Computations were carried out with Mathematica [7].

5 Conclusion

We present a first approach to forecast the protection of the population against the meningococcal C over the next few years. We can see a decreasing in the protected population for all ages.

Our future research is based on the thorough study of the database SBA-CV and obtain parameters of losing vaccination protection over the time adapted to CV. Also we want to propose an agent-based model with similar characteristics to the one presented here and, finally, study how to change the vaccination schedule in order to maintain as long as possible high levels of protection against meningococcal C.

References

[1] http://www.cdc.gov/meningococcal/index.html

[2] Cartwright K.,Meningococcal carriage and disease. In: Cartwright K., editor. Meningococcal disease. Chichester, UK: John Wiley & Sons; 1995. p.71-114.

[3] De Walls P. Immunization strategies for the control of serogroup C meningococcal disease in developed countries. Expert Rev Vaccines 2006; 5: 269-75.

[4] http://www.sp.san.gva.es/rvn/calendario.jsp?perfil=ciudadano

[5] http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@ab/documents/digitalasset/dh_132443.pdf

[6] Trotter C.L., Gay N.J. and Edmunds W.J. The natural history of meningococcal carriage and disease. Epidemiol. Infect. 134. 556-566. (2006)

[7] http://www.wolfram.com/products/mathematica.

NewFriends: An Algorithm for Computing the Minimum Number of Friends required by a User to Get the Highest PageRank in a Social Network

Francisco Moreno [*], Andrés González[*], and Andrés Valencia[*]

(*) Universidad Nacional de Colombia, Sede Medelln,

Carrera 80 No 65-223.

November 30, 2012

1 Introduction

The PageRank[1] is one of the most well-known methods for classifying web pages. The PageRank of a web page p represents the probability that a web surfer is visiting p after a considerable time of navigation. Web pages with high PageRank are probably very popular and influencial in the web. This method has also been applied for classifying social network users[2]. In the field of computational biology, the PageRank also has been used to classify species[3]. Recently, Herrera[4] applied this method to rank nodes in water supply networks and Radicchi[5] to rank tennis players.

The PageRank method considers the number of links (incoming and outgoing) of the web pages and the structure of their navigation network (graph). In this method, the web page with the largest number of links (incoming) is not necessarily the web page with the highest PageRank, since the network structure plays a decisive role.

[*]fjmoreno@unal.edu.co

In this paper, we propose a novel algorithm based on PageRank method, to determine the minimum number of new friends required by a user u of a social network to become the user with the highest PageRank. The idea is to add new users to the social network one at at time. Each new user is connected to u and then we analyzed if u has become the user with the highest PageRank. This analysis can be applied in several domains. For example, in politics it could help to determine the minimum number of electors that a candidate require to become the most popular candidate in a social network. In marketing, it could help to determine the minimum number of customers that a company require to become popular, and in this way attract and reach more customers. The paper is organized as follows. In Section 2, we introduce the main concepts of the PageRank method. In Section 3, we present our proposal for computing the minimum number of friends that a user requires to get the highest PageRank.

2 Basic definitions

Consider a social network with $n = 5$ users represented with a directed graph, see Figure 1.

The idea is to classify the nodes (users) of the social network according to their links and structure. To achieve this, we apply the PageRank method[2]. To apply the PageRank method, first, we build a connectivity matrix $H = (h_{ij}) \in R^{nxn}, 1 \leq i, j \leq n$, that represents the links of each node. If there exists a link from node i to node j, $i \neq j$, then $h_{ij} = 1$, otherwise $h_{ij} = 0$; if $i = j$ then $h_{ii} = 0$.

Now, from H matrix we build a row stochastic matrix $P = (p_{ij}) \in R^{nxn}, 1 \leq i, j \leq n$. A matrix is row stochastic if the sum of the elements of each of its rows is 1. P is calculated by dividing each element h_{ij} by the sum of the elements of row i of H.

Note that this sum may not be zero because each user has at least one friend. Note that in a social network it is reasonable to assume that each user has at least one friend, i.e., there do not exist dangling nodes [6].

The PageRank method requires that the P matrix in addition to be row stochastic must be primitive. In order to ensure this property (and at the same time preserving the row stochastic property), we can apply the following transformation [6].

$$G = \alpha P + (1 - \alpha)ev^T \tag{1}$$

Where G is known as Google matrix[6], α is a damping factor, $0 < \alpha < 1$. This factor represents the probability with which the surfer of the network moves among the links of the matrix H, and the term $(1 - \alpha)$ represents the probability of the surfer to randomly navigate to a link (page) that is not among the links of H. Usually, α is set to 0.85, a value that was established by Brin and Page, the creators of the PageRank method[1, 6].

On the other hand $e \in R^{nx1}$ is the vector of all ones and $ev^T = 1$. v is called *personalization* or *teletransportation* vector and can be used to affect (to benefit or to harm) the ranking of the nodes of the network[6]: $v = (v_i) \in R^{nx1} : v_i > 0, 1 \leq i \leq n$.

3 Computing the minimum number of friends that a user requires to get the highest PageRank in a social network

Our goal is to determine the minimum number of new friends that a user requires to become the user with the highest Pagerank in the network. Formally, let $G = (N, E)$ be the initial graph that represents the network, where N is the set of nodes and E is the set of links of the network. Each link is represented as (n_1, n_2) where $n_1, n_2 \in N$, $n_1 \neq n_2$. Let $\pi_i(G)$ denote the i component of the PPR for some personalization vector v, where $i \in N$. Let $Newfriends(i, m) = (k_1, i), (k_2, i), ..., (k_m, i)$ with $k_y \notin$ N, $1 \leq y \leq m$. Newfriends(i,m) represents the set of new nodes that will be connected to node i: $N' = N \cup k_1, k_2, ..., k_m$, $E' = E \cup Newfriends(i, m)$, and $G'(Newfriends(i, m)) = (N', E')$; where m is the smallest positive integer such that $\pi_i(G'(Newfriends(i, m))) = max(\pi_j(G'(Newfriends(i, m))))$, $j \in N'$. That is, m is the minimum number of new friends that i requires to get the highest PageRank in the network.

4 Conclusions

In this paper, we analyzed how the PageRank of a node of a network is affected when it is connected to new nodes of the network. We proposed a formal definition and an algorithm to determine the number of new friends that a node requires in order to become the node with the highest PageRank

in the network.

Our method has applications in fields such as marketing, politics, sales, and entertainment, among others; where their users want to gain visibility and be leaders of the network. As future work we plan to develop our proposal with other centrality measures. In particular, we expect to modify our method considering weighted networks, i.e., networks where every link is associated with a weight. For example, the weight of a link may indicate the influence of a node on another. Finally, we would like to determine the best potential friends for a node w. That is, which are the nodes of the network that if connected to w, they will increase the most the PageRank of w.

References

[1] Page L, Brin S, Motwani R, Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. Stanford Digital Library Technologies Project. 1999.

[2] Pedroche F. Ranking nodes in Social Network Sites using biased PageRank. Instituto de Matemática Multidisciplinaria, Universidad Politécnica de Valencia 2010; E-46022 Valencia.

[3] Allesina S, Pascual M. Googling food webs: can an eigenvector measure species importance for coextinctions? PLoS Computational Biology 2009.

[4] Herrera M, Gutiérrez-Pérez J, Izquierdo J, Pérez-García R. Ajustes en el modelo PageRank de Google para el estudio de la importancia relativa de los nodos de la red de abastecimiento. In: Proceedings of X Ibero-American Seminar SERE 2011.

[5] Radicchi F. Who is the best player ever? A complex network analysis of the history of professional tennis. PLoS ONE 2011.

[6] Pedroche F. Métodos de cálculo del vector PageRank. Instituto de Matemática Multidisciplinaria, Universidad Politécnica de Valencia 2007; 7-30.

Compressible Flow Turbomachinery Simulations with OpenFOAM

J. Benajes, J. Galindo, P. Fajardo and R. Navarro*†

CMT-Motores Térmicos,

Universitat Politècnica de València

Camino de Vera S/N, 46022 Valencia, Spain.

Tel. +34-963977650 Fax.+34-963877659

November 30, 2012

1 Introduction

OpenFOAM® is a CFD open source toolbox which is becoming a real alternative to well-know commercial codes. However, the code is still under development and some existing capabilities have not been completely validated.

The objective of this work is to develop a Multiple Reference Frame and Sliding Mesh compressible pressure-based segregated solvers in OpenFOAM®, aiming to simulate the flow in an automotive turbocharger. The pressure corrector equation approach is avoided to take advantage of implicit underrelaxation. The solvers have been implemented in OpenFOAM-1.6-ext version [1].

*Corresponding author

†e-mail: jbenajes, galindo, pabfape, ronagar1@mot.upv.es

2 Multiple Reference Frame solver

The solver structure is sketched in Fig. 1. Following the SIMPLE [2] approach, each iteration begins with momentum predictor, in which velocity equation is defined, implicitly under-relaxed and solved. Then, a pseudo-flux is computed, in which the velocity employed does not carry the pressure gradient contribution [3]. However, since implicit under relaxation has changed the matrix coefficients of momentum equation, an update of the coefficients is required. Afterwards, pressure equation is assembled, implicitly relaxed and solved. Since the equations are implemented in a conservative form, in

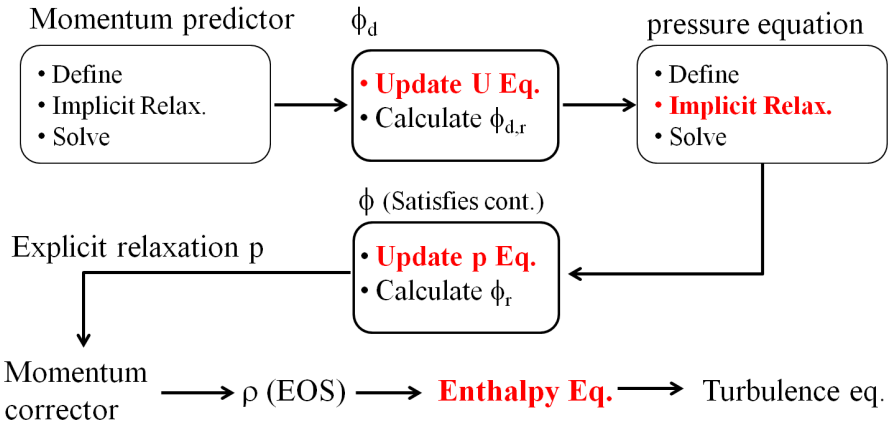


Figure 1: Flowchart of the proposed MRF solver. Major changes regarding currently available OpenFOAM® solvers are highlighted in red

which continuity is implicitly considered, it is important that the calculated fluxes satisfy continuity. In order to obtain a set of face fluxes that fulfill this requirement, two things must be taken into account: pressure equation should be updated before obtaining the flux, for the same reason as in momentum equation, and the implicit under-relaxation factor should be close to 1. Then, pressure field is explicitly relaxed and momentum corrector is applied taking into account the new pressure contribution.

Finally, density is obtained using the equation of state (EOS) and enthalpy and turbulent equations are solved. After that, a new iteration is performed. Regarding enthalpy equation, it has been placed after the pressure-velocity coupling to have a consistent set of pressure and velocity fields and fluxes when solving it. Moreover, energy equation has been considered using

total enthalpy because it provides a conservative formulation. If the fluid cell considered belongs to the rotor, the corresponding non-inertial terms must be added prior to solving momentum and energy equations and the fluxes should be made relative to the rotating reference frame after its computation.

The developed solver was used to simulate the variable geometry turbine analyzed by Galindo et al. [4] under different operating conditions, providing good convergence behavior. The solution obtained by this solver is compared to the one computed using ANSYS-FLUENT, showing good agreement in terms of flow patterns but underpredicting pressure drop and efficiency.

3 Sliding Mesh solver

The solver structure is sketched in Fig. 2. Since is quite similar to the one described in Fig. 1, only major differences will be pointed out.

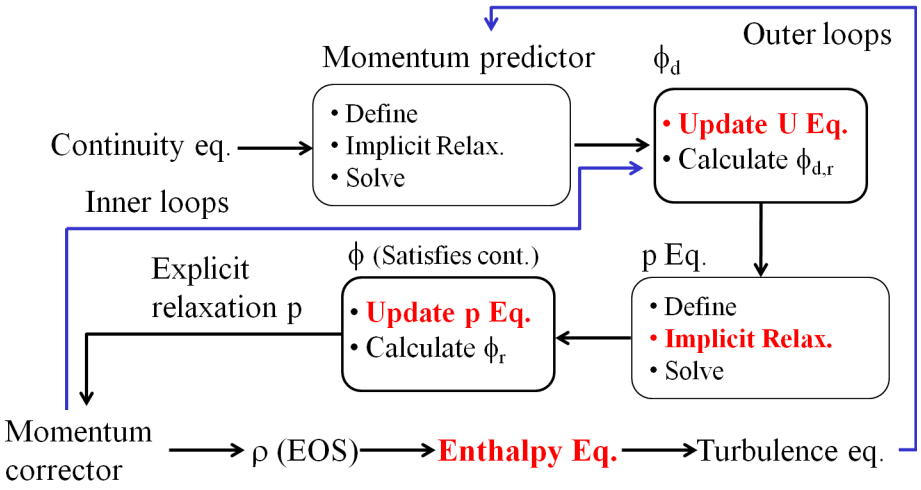


Figure 2: Flowchart of the proposed SM solver. Major changes regarding already available OpenFOAM® solvers are highlighted in red.

When starting a new time-step, the mesh is correspondingly rotated and continuity equation is firstly solved. Then, the so-called PIMPLE loop is applied. Momentum predictor is followed by a number of pressure equation and momentum corrector loops, known as inner loops. Afterwards, density is updated and enthalpy and turbulent equations are solved. The whole set

of equations, excluding continuity, form an outer loop which is repeated a predefined number of times to overcome the segregated approach. After that, a new time-step can be performed.

This solver has been used in one simulation, checking that the residuals drop by several orders of magnitude at every time-step. The solution obtained by this solver is compared to the one computed using ANSYS-FLUENT. Pressure and velocity fields predicted by both codes are almost identical.

4 Conclusions

In this work, a MRF compressible solver and a SM one have been developed in OpenFOAM® to simulate turbomachinery. Pressure equation has been derived following the work by Jasak [3] instead of using a pressure corrector equation. In this way, implicit under-relaxation can be performed, with the corresponding increase of linear solver stability. Energy equation has been implemented using total enthalpy because it provides a conservative formulation.

Both solvers have been used to simulate a turbocharger turbine, showing good convergence behavior. Their solutions have been compared to corresponding ones computed using ANSYS-FLUENT, as a means of validation. The MRF solver predicts proper flow features. However, it still needs some work since the provided values can differ up to 10 % compared to ANSYS-FLUENT ones. The SM solver obtains a more accurate solution, even though some differences exist compared to ANSYS-FLUENT. Finally, OpenFOAM® is found to be a very promising code, although it seems that it still needs some effort by the developers and the community of users to become a reliable tool.

References

[1] OpenFOAM Extend Project. URL `http://sourceforge.net/projects/openfoam-extend/`.

[2] S. V. Patankar and D. B. Spalding. A calculation procedure for heat, mass and momentum transfer in three-dimensional parabolic flows. *International Journal of Heat and Mass Transfer*, 15(10):1787–1806, 1972.

[3] Hrvoje Jasak. *Numerical Solution Algorithms for Compressible Flows (Lecture notes)*. Faculty of Mechanical Engineering and Naval Architecture, 2006-2007.

[4] J. Galindo, P. Fajardo, R. Navarro, and L. M. García-Cuevas. Characterization of a radial turbocharger turbine in pulsating flow by means of CFD and its application to engine modeling. *Applied Energy*, 2012. doi: 10.1016/j.apenergy.2012.09.013.

CFD modeling of reacting Diesel sprays with tabulated detailed chemistry

J.M. García–Oliver, R. Novella *, J.M. Pastor, J.F. Winklinger

CMT–Motores Térmicos,

Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain

November 30, 2012

1 Introduction

The prediction of the auto–ignition process and the structure of the fully developed reacting fuel spray is still one of the main challenges in modelling partially premixed combustion. Detailed chemical mechanisms have to be employed to correctly reproduce the mentioned characteristics of Diesel spray combustion [1]. In order to overcome problems with excessive calculation times which arise with the use of such mechanisms, tabulation techniques such as FPI or FGM [2, 3] have been developed.

In this research, a similar approach to the FPI (Flame Prolongation of ILDM—intrinsic low dimensional manifold) method, where chemical reaction paths obtained from Perfectly Stirred Reactor (PSR) calculations are stored in tables as proposed in [4, 5], has been implemented. These tables are used to obtain the progress variable reaction rate and the species mass fractions in the Homogeneous Auto–Ignition (HAI) model. Furthermore, in order to account for subgrid turbulence–chemistry interaction, the modified FPI method is coupled in a subsequent step with the Presumed Conditional Moment (PCM) approach [4]. This approach intends to model the mixture fraction distribution by the adoption of Probability Density Functions (pdf)

*e-mail: rinoro@mot.upv.es

and results in the Auto–Ignition–Presumed Conditional Moment (AI–PCM) model.

Then, both models have been evaluated for Diesel spray combustion simulation by means of a parametric study of the "Spray H" configuration from the Engine Combustion Network (ECN) with varying initial temperatures and oxygen levels.

2 Methodology

The two combustion models, based on those described in [4, 5], have been implemented in a compressible RANS solver of the open source CFD platform OpenFOAM® in its version 1.6. The PSR calculations are conducted adopting a chemical mechanism for n-heptane with 110 species and 1170 reactions [6].

The tabulated chemistry approach requires an efficient data handling procedure due to the big amount of data stored in the multidimensional tables and the permanent data retrieval. For this reason, the storage of the tables is realised by the adoption of hash tables. The hash tables are entered with given lookup parameters, which provide direct and consequently fast access to a data entry in the tables.

The setup of "Spray H" consists of a single spray injected into a constant volume vessel and is described in [7]. A discrete droplet method (DDM) approach is applied to model the liquid fuel spray. This approach as well as the settings of the spray submodels and the k-ε turbulence model are described in [8].

3 Results

Experimental data of the Ignition Delay (ID) taken from the ECN data base [7] are compared with CFD results obtained with the two combustion models in Figure 1. The ID obtained with the AI–PCM model generally agrees better with the experimental data, both for varying temperatures and for varying oxygen levels.

A comparison of experimental data of Lift-Off Length (LOL), taken from the same source, with the CFD results is shown in Figure 2. As already seen for the ID, the results from the AI–PCM model expectedly agree better with

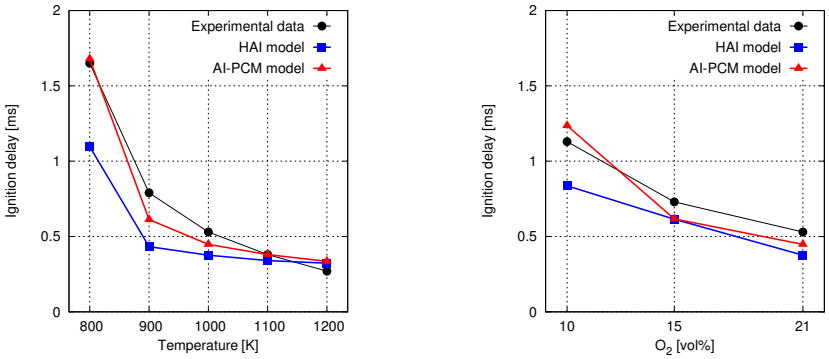


Figure 1: Comparison of the ignition delay for different initial temperatures at $X_{O_2} = 21\%$ (left) and different oxygen levels at $T_{init} = 1000$ K (right)

the measured LOLs than those from the HAI model. This demonstrates the importance of considering turbulence–chemistry interaction when modelling partially premixed combustion and its effect on the structure of the reacting spray.

The temperature fields obtained with the two models shown in Figure 3 (left) represent the structure of the reacting spray at different initial temperatures. Especially in the case of 800 K initial temperature, the results of the two models differ clearly from each other. The reason for this can be seen in the ϕ–T–map in Figure 3 (right, top), which illustrates the reduced reactivity of high equivalence ratios. This effect is less distinctive at higher ambient temperatures, since the reactivity is increased in general with increasing temperature.

In the HAI model the reaction path of a given mean mixture fraction is directly represented by the reactivity of its corresponding homogeneous reactor. The fluctuations of the mixture fraction considered in the AI–PCM model, in contrast, lead to variance–weighted contributions from all PSRs to

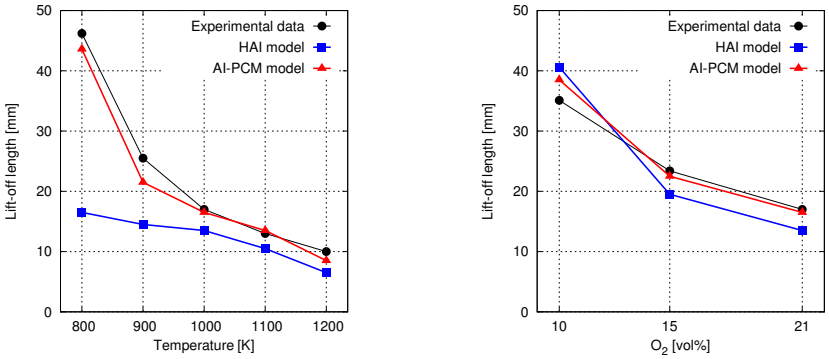


Figure 2: Comparison of lift-off length for different initial temperatures at $X_{O_2} = 21\%$ (left) and different oxygen levels at $T_{init} = 1000$ K (right)

the reaction path of the same mean mixture fraction.

4 Conclusions

Two different combustion models, both use a tabulated detailed chemistry approach based on the FPI method, were implemented in a RANS CFD environment. A comparison of the CFD model results with experimental data from the Engine Combustion Network shows, that ignition delay and lift-off length obtained with the AI–PCM model generally agree better with the experimental data then those predicted by the HAI model. This demonstrates the influence of turbulence on the chemistry of the combustion process and hence the importance of including its effects in advanced combustion modelling.

The HAI model shows a lack of physics, since it goes back to pure chemistry data obtained from homogeneous reactor calculations, which do not contain any information about subgrid turbulence effects on the chemistry of

the mixture. However, the AI–PCM model allows to reproduce part of the phenomena caused by turbulence. The included fluctuations of the mixture fraction lead to a reduction of the reactivity, being more distinct at lower ambient temperatures and rich mixtures.

References

[1] O. Colin , J.-B. Michel, P.E. Vervisch, *8th International ERCOFTAC Symposium on Engineering Turbulence Modelling and Measurements*, Marseille, France, June 9th to 11th, 2010

[2] O. Gicquel, N. Darabiha, D. Thevenin, Proceedings of the Combustion Institute 28 (2000), pp. 1901–1908

[3] J.A. Van Oijen, F.A. Lammers, L.P.H. De Goey, Combustion and Flame 127 (2001), pp. 2124–2134

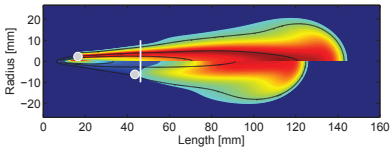
[4] J.-B. Michel, O. Colin, D. Veynante, Combustion and Flame 152 (2008), pp. 80–99

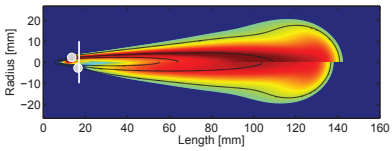
[5] C. Pera, O. Colin, S. Jay, Oil & Gas Science and technology – Rev. IFP 64 (2009), No. 3, pp. 243–258

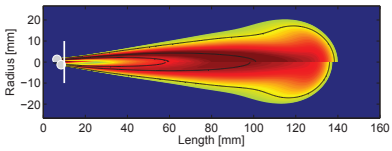
[6] Th. Zeuch, G. Moréac, S. S. Ahmed, F. Mauss, Combustion and Flame 155 (2008), pp. 651–674

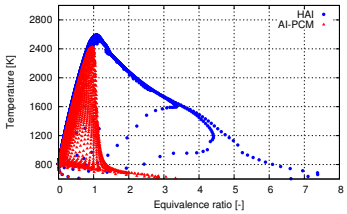
[7] Data Search Utility of the ECN, http://www.sandia.gov/ecn/cvdata/dsearch/frameset.php

[8] R. Novella, A. García, J.M. Pastor, V. Domenech, Mathematical and Computer Modelling 54 (2011), pp. 1706–1719









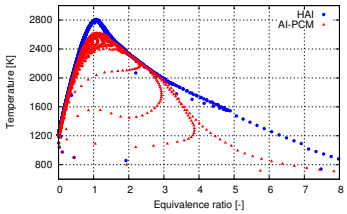


Figure 3: Left: temperature fields of reacting sprays at quasi–stationary state with $X_{O_2} = 21\%$ and $T_{init} = 800$ K (top), $T_{init} = 1000$ K (middle), $T_{init} = 1200$ K (bottom). Black lines mark $\phi = 0.2, 1, 2$, vertical white lines state experimental LOL, grey dots mark calculated LOL. Top half: HAI, bottom half: AI–PCM. Right: ϕ–T–maps for $X_{O_2} = 21\%$ and $T_{init} = 800$ K (top) and $T_{init} = 1200$ K (bottom)

Uncertainties in power computations in a turbocharger test bench

P. Olmeda [*], A. Tiseira, V. Dolz, and L.M. García-Cuevas

(*) CMT Motores Térmicos. Universidad Politècnica de València.,

Camino de Vera s/n, 46022 Valencia (Spain)

November 30, 2012

1 Introduction

Improving the experimental characterisation of centrifugal compressors and centripetal turbines is a topic of high interest for both researchers and engine manufacturers. To ensure minimum errors in the experimental studies done in turbocharger test rigs, experimental measurement standards [5, 6] are developed and a high technical knowledge and experience in this particular area is needed, as well as high quality experimental facilities. Nevertheless, as both researchers and engine manufacturers are interested into getting experimental results at operating conditions typical of urban driving cycles, the uncertainty of the measurements done within these experimental facilities becomes excessively high for practical purposes and the test rig designer needs to invest in newer equipment.

The aim of this work is the development of a flexible methodology to estimate the measurement uncertainties in the particular case of the experimental characterisation of turbochargers and to give hints to the test rig designer to select new transducers optimising the results with a minimum cost using non-linear mixed integer programming.

[*]e-mail: pabolgon@mot.upv.es

2 Propagation of uncertainty in turbine power

The turbine power can be computed by means of total enthalpy difference between its outlet and its inlet, which is a function of the composition of the air, the mass flow rate and the total temperature. The total temperature is a function of the measured temperature, the composition of the air, the pressure, the mass flow rate and the kinetic energy recovery factor of the sensor. The full expression of turbine power uncertainty expressed as standard deviation can be computed using:

$$\begin{aligned} u_{\dot{W}_t} = & \left[c_p \left(T_{04} - T_{03}\right) u_{\dot{m}_t}\right]^2 \\ & + \left[\dot{m}_t \left(T_{04} - T_{03}\right)\right]^2 \cdot \left[\left(\frac{\partial c_p}{\partial Y} u_Y\right)^2 + \left(\frac{\partial c_p}{\partial T} u_T\right)^2\right] \\ & + \left(\dot{m}_t c_p\right)^2 \left\{ \left(\frac{\partial T_{03}}{\partial T_3} u_{T_3}\right)^2 + \left(\frac{\partial T_{04}}{\partial T_4} u_{T_4}\right)^2 \right. \\ & + \left[\left(\frac{\partial T_{03}}{\partial \dot{m}_t}\right)^2 + \left(\frac{\partial T_{04}}{\partial \dot{m}_t}\right)^2\right] u_{\dot{m}_t}^2 + \left[\left(\frac{\partial T_{03}}{\partial k}\right)^2 + \left(\frac{\partial T_{04}}{\partial k}\right)^2\right] u_k^2 \\ & \left. + \left[\left(\frac{\partial T_{03}}{\partial Y}\right)^2 + \left(\frac{\partial T_{04}}{\partial Y}\right)^2\right] u_Y^2 + \left(\frac{\partial T_{03}}{\partial p_3} u_{p_3}\right)^2 + \left(\frac{\partial T_{04}}{\partial p_4} u_{p_4}\right)^2 \right\} \end{aligned} \tag{1}$$

where $\dot{W}$ is power, u is uncertainty, c_p is mean specific heat capacity, $\dot{m}$ is mass flow rate, T is temperature, p is pressure, Y represents the composition of the working fluid, k is the kinetic energy recovery factor, t refers to the turbine, 0 means total or stagnation quantity, 3 refers to the turbine inlet and 4 to the turbine outlet.

When computing the propagation of uncertainty in turbine output power, the more important terms are that of mass flow rate and temperature measurement. Nevertheless, high fluctuations in the working fluid composition are expected in gas stands and should be taken into account and at medium and high expansion ratios a bad kinetic energy recovery factor estimation propagates to high errors in the turbine power computation. Also, when low uncertainties are achieved by means of high accuracy and high precision temperature and flow rate transducers, all the terms become important.

3 Optimisation methodology

In order to optimise the investment in new experimental equipment, a database of prices and technical specifications of sensors from different suppliers has to be compiled. Also, the expression of turbine output power uncertainty and a representative experimental dataset are needed. After that, the test rig designer has one of the next targets:

- Minimise the uncertainty for a given maximum acceptable economic cost.
- Minimise the cost for a given target uncertainty.

In the former case, the cost is a constraint of the optimisation problem and the uncertainty is weighted so it is more important to reduce it in some zone of interest, whereas in the latter the cost is the objective function to optimise.

As all general non-linear integral programming problems, both described situations are NP-hard, but some cases can be simplified. In the first case, when the uncertainty is to be minimised, if the test right designer has no sensors in stock, the constraints becomes linear and some efforts can be done into obtaining a better solving algorithm, as seen in [3]. Nevertheless, any NMIP algorithm may be used, such as a genetic algorithms [7] or branch and bound methods as described in [2] or, more recently, in [1] or [8].

The proposed algorithm to solve the problem is as follows:

- First, the relaxed non-integer problem is solved by means of sequential quadratic programming, giving results with low computational cost.
- Last, the ceilings and floors of the parameters are probed to get an approximation of the real optimum. The cost of this last stage is only $\mathrm{O}\left(2^n\right)$ and gives better results than just rounding the result to the nearest integer.

Solving the relaxed non-integer problem with sequential quadratic programming usually gives results with less iterations than using other methods such as genetic algorithms. The authors have succesfully used the SLSQP algorithm by D. Kraft [4], but other SQP algorithms may apply.

4 Application to a real case

The former methodology has been applied to a real testing campaign with a typical sensors arrangement, computing uncertainties for the power output of an automotive turbine, giving the results shown in figure 1, using a coverage factor of 3.

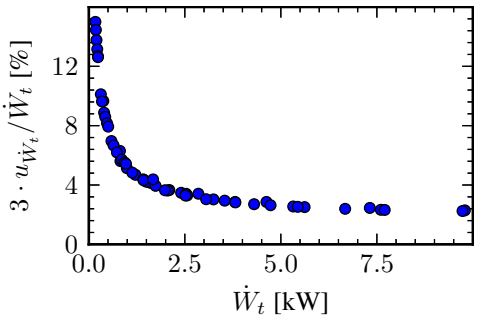


Figure 1: Typical power output uncertainty

The figure shows big uncertainties at low turbine output powers. These results are due to measuring both small temperature differences between the inlet and the outlet of the turbine and small mass flow rates.

At high powers, the expected uncertainty shows an asymptotic behaviour, ruled by the flow rate meter characteristics.

This dataset is used during an optimisation process for a new test rig with minimum uncertainty and the following results arise:

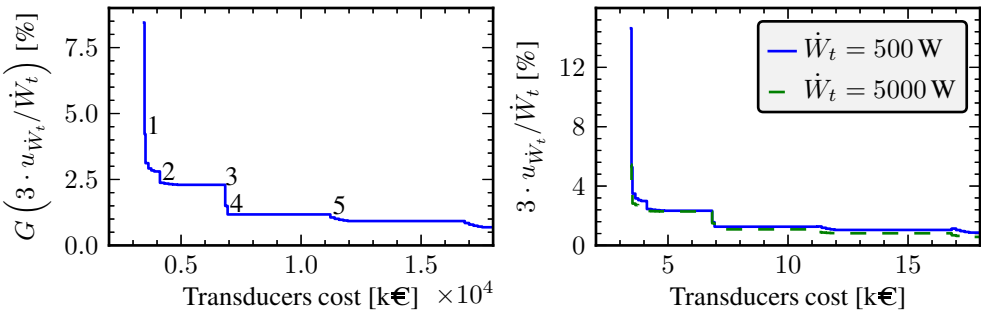


Figure 2: Results of an optimisation process

In figure 2, the optimisation has been done supposing no sensors in stock and minimising the uncertainty with increasing maximum costs. The left

graph shows the geometric mean of the results and the right chart the uncertainty obtained at two different output power levels. In both figures only the transducer cost is shown, not taking into account other costs such as ducts or joints. The major improvements are found at low powers with relatively small costs, investing in better temperature sensors. In figure 2, the first point is obtained with low cost mass flow rate sensors, one unshielded type K thermocouple per measurement section, low cost pressure transducers and no humidity measurement; the step 1 is obtained with one class B PT-100 transducer per measurement section and mantaining the same sensors as before; step 2 is obtained with 4 class A, 4 wire shielded PT-100 transducers per measurement section; in step 3, the mass flow rate sensor used has a lower uncertainty of 1 % of the measured value; step 4 introduces humidity sensors; step 5 is obtained with two low uncertainty mass flow rate sensors but reducing the number of temperature transducers; the rest of the optimisation is done with better mass flow rate sensors (the large steps in cost) and recovering temperature transducers up to the saturation point and investing in better pressure transducers (the smaller steps).

5 Conclusions

Turbine power uncertainties grow at a very fast rate when low powers are measured. The most important contributions are that derived of temperature measurement at the inlet and the outlet. When high accuracy in turbine power measurement is wanted, special care is needed in the thermocouples arrangement. Better results are expected by using lower uncertainty sensors, such as RTDs. When using the best measurement techniques available for temperature and mass flow rate, the effects of humidity and pressure become to be of importance.

The authors conclude that the first way to improve current power measurement techniques in turbocharger test rigs is to focus on mass flow rate sensors at high mass flow rates and on temperature sensors arrangement and selection at low mass flow rates, but if a global optimisation of the test rig is required, the methodology explained in section 3 may give better results with lower costs by using a good transducers database and the expression derived in section 2.

References

[1] Piere Bonami, Jon Lee, Sven Leyffer, and Andreas Wächter. More branch-and-bound experiments in convex nonlinear integer programming. 2011.

[2] Omprakash K. Gupta and A. Ravindran. Branch and bound experiments in convex nonlinear integer programming. *Management Science*, 31(12):1533–1546, 1985.

[3] M. Jnger, T. Liebling, D. Naddef, G. Nemhauser, W. Pulleyblank, G. Reinelt, G. Rinaldi, and L. Wolsey. *50 Years of Integer Programming 1958–2008: The Early Years and State-of-the-Art Surveys*. 2009.

[4] D. Kraft and Deutsche Forschungs und Versuchsanstalt für Luft-und Raumfahrt Köln. *A software package for sequential quadratic programming*. Forschungsbericht / Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt. Dt. Forschungs- u. Versuchsanst. für Luft- u. Raumfahrt (DFVLR), 1988.

[5] J Luján, V Bermúdez, J Serrano, and C Cervelló. Test bench for turbocharger groups characterization. *SAE International*, (2002-01-0163), 2002-03-04 2002.

[6] SAE. Turbocharger gas stand test code. *S. of Automotive Engineers Inc*, (SAE J1826), Mar. 95. 1995 1995.

[7] Takao Yokota, Mitsuo Gen, and Yin-Xiu Li. Genetic algorithm for nonlinear mixed integer programming problems and its applications. *Computers and Industrial Engineering*, 30(4):905 – 917, 1996.

[8] Daisuke Yokoya and Takeo Yamada. A mathematical programming approach to the construction of bibds. *Int. J. Comput. Math.*, 88(5):1067–1082, 2011.

Adaptation of finite difference numerical methods to the solution of governing equations in wall-flow diesel particulate filters

J. R. Serrano, F. J. Arnau, P. Piqueras*, O. García-Afonso

Universitat Politècnica de València, CMT-Motores Térmicos,

Camino de Vera s/n, 46022 Valencia, Spain.

November 30, 2012

1 Introduction

The increasingly restrictive soot emission regulations require the automotive industry to adopt the extensive use of wall-flow diesel particulate filter (DPF) in internal combustion engines [1]. Besides the influence of the DPF on the gas flow path, the competitive nature of the market demands to provide solutions with high timing restrictions. As a consequence, the inclusion of wall-flow DPF models into gas dynamic codes is becoming widespread.

A wall-flow DPF consists of a ceramic monolith with small axial parallel channels separated by a porous wall. At the inlet cross-section, the channels are alternatively plugged defining the inlet and outlet channels. The flow enters to the monolith through the inlet channels, which are plugged at its outlet cross-section. As a consequence, the flow inside the inlet channels is forced to flow across the porous substrate walls, where the soot particulates are filtrated and accumulated until the regeneration process takes place. Finally, the clean gas flow leaves the monolith through the outlet channels, which have the outlet cross-section open to the exhaust tailpipe. According to this architecture, the flow inside wall-flow monoliths is modelled as

*e-mail: pedpicab@mot.upv.es

one-dimensional. However, finite difference numerical methods, which are traditionally applied in gas dynamic codes [2] to obtain the numerical solution, need to be adapted. It is due to the flow exiting or entering to the inlet and outlet channels respectively, which means the coupling between the system of governing equations of both channels.

This work deals with the adaptation and evaluation of the two-step Lax & Wendroff method [3] and the CE-SE method [4, 5] to be applied in square channels of wall-flow monoliths. A shock tube test affecting the one-dimensional domain of a pair of inlet and outlet channels is proposed to compare the performance of both numerical methods.

2 Adaptation of the numerical methods

The solution of the governing equations is performed by means of shock-capturing finite difference schemes, with the only exception of the boundary conditions, which are solved applying the MoC [6]. The governing equations of a pair of inlet and outlet channels are coupled by the flow across the porous wall. Therefore, the numerical solvers need to be properly adapted from its traditional formulation for 1D elements with non-porous wall.

2.1 The two-step Lax & Wendroff method

The formulation of the two-step Lax & Wendroff method in porous wall channels is given by equations (2) and (3):

- First step

$$\begin{aligned}\mathbf{W}_{k,j+\frac{1}{2}}^{n+\frac{1}{2}} &= \frac{\mathbf{W}_{k,j}^{n}+\mathbf{W}_{k,j+1}^{n}}{2}-\frac{\Delta t}{2\Delta x}\left(\mathbf{F}_{k,j+1}^{n}-\mathbf{F}_{k,j}^{n}\right) \\ &\quad -\frac{\Delta t}{4}\left(\mathbf{C}_{k,j}^{n}+\mathbf{C}_{k,j+1}^{n}\right)-\frac{\Delta t}{4}\left(\mathbf{C}_{w_k,j}^{n}+\mathbf{C}_{w_k,j+1}^{n}\right)\end{aligned} \tag{1}$$

- Second step

$$\begin{aligned}\mathbf{W}_{k,j}^{n+1} &= \mathbf{W}_{k,j}^{n}-\frac{\Delta t}{\Delta x}\left(\mathbf{F}_{k,j+\frac{1}{2}}^{n+\frac{1}{2}}-\mathbf{F}_{k,j-\frac{1}{2}}^{n+\frac{1}{2}}\right) \\ &\quad -\frac{\Delta t}{2}\left(\mathbf{C}_{k,j-\frac{1}{2}}^{n+\frac{1}{2}}+\mathbf{C}_{k,j+\frac{1}{2}}^{n+\frac{1}{2}}\right)-\frac{\Delta t}{2}\left(\mathbf{C}_{w_k,j-\frac{1}{2}}^{n+\frac{1}{2}}+\mathbf{C}_{w_k,j+\frac{1}{2}}^{n+\frac{1}{2}}\right)\end{aligned} \tag{2}$$

In equations (2) and (3) $\mathbf{W}_k$ and $\mathbf{F}_k$ represent the solution and the flux vectors in channel k respectively. The source terms are divided into two contributions. On the one hand, the source terms related to the cross-section area change, friction and heat transfer are included in vector $\mathbf{C}_k$; on the other hand, the source terms characteristic of the porous medium are arranged in vector $\mathbf{C}_{w_k}$. Subscripts j and n define the space-time mesh identifying the node and the time level respectively.

2.2 The CE-SE method

The CE-SE method [4] solves the governing equations subdividing the space-time mesh into rhombic regions (Solution Element (SE)), in which the solution is provided by Taylor's approximation, and into rectangular regions (Conservation Element (CE)), in which the conservation equations are fulfilled.

The existence of flow through the porous walls of the channels only affects the formulation of the source term vector accounting for the mass and enthalpy flow through the porous walls. These terms are governed by the filtration velocity:

$$c_{w_k,1} = -\left(-1\right)^k \frac{4\sigma_{k,1} u_{w_k}}{\alpha - 2w_p k} \tag{3}$$

$$c_{w_k,2} = 0 \tag{4}$$

$$c_{w_k,3} = -\left(-1\right)^k \frac{4h_{0w}\sigma_{k,1} u_{w_k}}{\alpha - 2w_p k} \tag{5}$$

3 Discussion of the results

This section is devoted to assess the performance of the two-step Lax & Wendroff method and the CE-SE method when solving the governing equations in channels of wall-flow DPF monoliths. The evaluation of the numerical methods is based on the performance of the solution in a shock-tube tests adapted to the specific case of wall-flow monoliths. The approached shock-tube test consists of a pair of inlet and outlet channels. A diaphragm in the inlet channel separates two regions where the flow has different conditions in pressure and temperature. An additional diaphragm is placed on the porous

wall so that source terms due to porous medium flow are inhibited. The flow conditions in the outlet channel are equal to those in right side of the inlet channel.

The separation between both flow regions in the inlet channel is imposed in the centre of the domain defined as $D = \{x : x \in [-1, 1]\}$ m, which is expressed in meters. This diaphragm and the one which is placed on the porous wall are removed at time $t = 0$ s. The initial conditions in every region are detailed in Table 1.

Table 1: Initial conditions at inlet and outlet channel.

Property	Inlet channel		Outlet channel	
	Left	Right	Left	Right
p [bar]	1.15	1	1	1
T [K]	655	290	290	290
u [m/s]	0	0	0	0

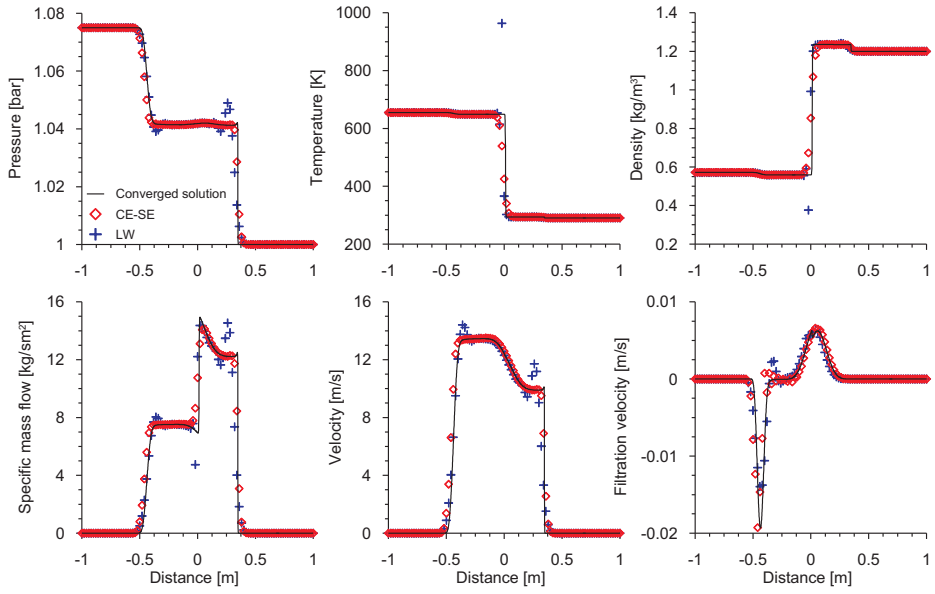


Figure 1: Comparison between the two-step Lax & Wendroff method and the CE-SE method. Inlet channel properties at $t = 0.001$ s with $\Delta x = 20$ mm.

Figures 1 and 2 show the comparison of the solution provided at time 0.001 s by the CE-SE and the two-step Lax & Wendroff methods in the inlet and outlet channels respectively. Both of the methods are able to reproduce with good accuracy the flow properties in time and space. Nevertheless, the two-step Lax & Wendroff methods is characterized by a dispersive solution with spurious oscillations around the discontinuities.

In the case of the CE-SE method, the solution does not show any kind of oscillations which may lead to divergences in the solution as the spatial mesh size increases. However, the solution is diffusive, what avoids the method to reproduce small discontinuities. This effect appears around the contact discontinuity in the solution of the specific mass flow in the inlet channel, which is shown in Figure 1. Both the maximum and minimum peak in specific mass flow due to the change in density and the mass flow across the porous medium are not predicted properly. Similar considerations can be done with respect to the value of the temperature around the contact discontinuity in the outlet channels, which is shown in Figure 2.

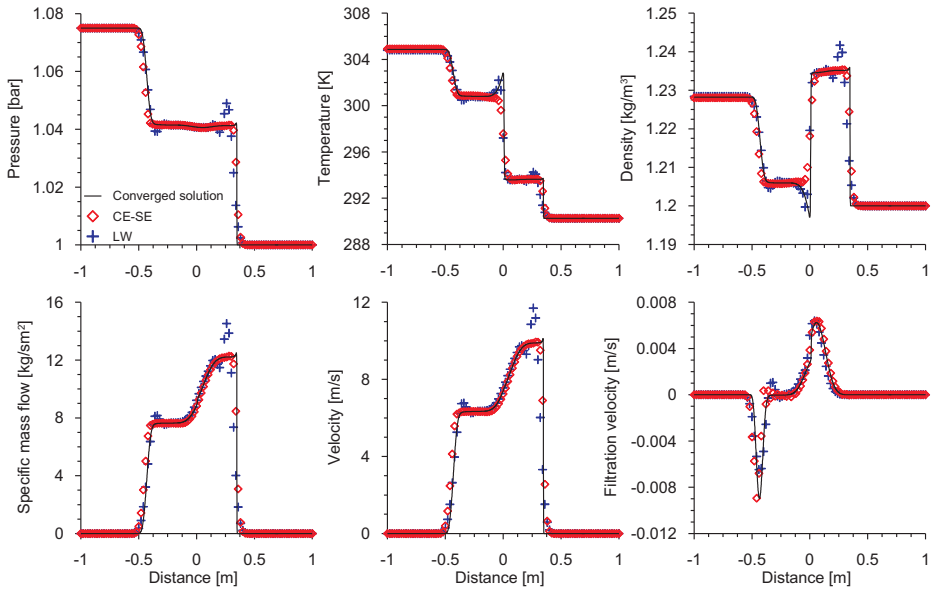


Figure 2: Comparison between the two-step Lax & Wendroff method and the CE-SE method. Outlet channel properties at $t = 0.001$ s with $\Delta x = 20$ mm.

4 Conclusions

Compressible unsteady flow transport along channels of wall-flow monolith can be solved applying shock capturing methods. However, the solution in the inlet and outlet channels is coupled because of the porous nature of the walls. It leads to the inclusion of source terms, which are dependent on the porous wall permeability and the flow properties, in the governing equations.

This work has dealt with the adaptation of the two-step Lax & Wendroff method and the CE-SE method in order to be applied to the solution of the governing equations in this kind of 1D structures. The two-step Lax & Wendroff method shows the typical spurious oscillations around discontinuities generated by second order symmetric schemes. The use of the CE-SE method leads to remove non-physical overshoots, although at the expense of an increase of the computational effort.

References

[1] V. Bermúdez, J.R. Serrano, P. Piqueras and O. García-Afonso. Assessment by means of gas dynamic modelling of a pre-turbo diesel particulate filter configuration in a turbocharged HSDI Diesel engine under full-load transient operation, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 225(9) 1134–1155, 2011.

[2] D. Winterbone and R. Pearson. Theory of engine manifold design: wave action methods for IC engines. Professional Engineering Publishing, 2000.

[3] P. Lax and B. Wendroff. Systems of conservation laws. *Communications on Pure and Applied Mathematics* 17, 381–398, 1964.

[4] S. Chang and W. To. A new numerical framework for solving conservation laws – The method of space-time Conservation Element and Solution Element. NASA Technical Memorandum 104495, 1991.

[5] G. Briz and P. Giannattasio. Applicazione dello schema numerico Conservation Element - Solution Element al calcolo del flusso intazionario nei condotti dei motori a C.I. In *Proc. 48th ATI National Congress*, 1993, pp. 233–247.

[6] J.M. Desantes, J.R. Serrano, F.J. Arnau and P. Piqueras. Derivation of the method of characteristics for the fluid dynamic solution of flow advection along porous wall channels. *Applied Mathematical Modelling* 36, 3134–3152, 2012.

Linear quadratic methods for the optimal regulator of an unmanned air vehicle*

P. Bader†, S. Blanes and E. Ponsoda

Instituto de Matemática Multidisciplinar, Universitat Politècnica de València.

1 Introduction

The optimal control of Unmanned Air Vehicles (UAV) has attracted a great attention in recent years [7]. Helicopters are classified as Vertical Take Off Landing (VTOL) aircraft and are among the most complex flying objects because their flight dynamics is nonlinear and their variables are strongly coupled. Specifically, a quadrotor is a small air vehicle with four propellers, whose rotational speeds are independent, placed around a main body [5, 7, 9]. Linear techniques to control the system have been frequently used. However, to improve the performance, the nonlinear nature of the quadrotor has to be taken into account.

The controllers are designed based on a simplified description of the system behavior (linearized models). While this is satisfactory at hover and low velocities, it does not predict correctly the system behavior during fast maneuvers (most controllers are specifically designed for low velocities). In order to reach the desired final position as fast as possible, real time calculations are necessary and hence more efficient and elaborated algorithms have to be designed.

Linear quadratic (LQ) optimal controllers are widely used, in particular for the control of small aircrafts [13], where they have shown to produce better

*This work has been partially supported by Ministerio de Ciencia e Innovación (Spain) under projects MTM2007-61572 (co-financed by the ERDF of the European Union) and MTM2009-08587, and the Universitat Politècnica de València throughout the project 2087. The authors also acknowledges the support trough the FPU fellowship AP2009-1892.

†e-mail: phiba@imm.upv.es

results than other standard methods, like proportional integral derivative methods (PID) [5]. The techniques presented here, however, are valid for the general optimal LQ control problem.

For the illustration of our methods, we consider a VTOL quadrotor, based on the model presented in [7, 13] (and references therein). In general, one assumes some standard general conditions: symmetric and rigid structure of the flying robot, the center of mass is in the center of the planar quadrotor and the propellers are rigid. However, more realistic problems have time-varying parameters [15], require a time dependent state reference [9] or involve non-linear equations [8, 13].

Let us consider the non-linear problem

$$\min_{u \in L^2} \int_0^{t_f} \left(X^T(t) Q(t, X(t)) X(t) + u^T(t) R(t, X(t)) u(t) \right) dt \,, \tag{1}$$

subject to

$$\dot{X}(t) = f_A\left(t, X\left(t\right)\right) + f_B\left(t, X\left(t\right), u\left(t\right)\right), \quad X(0) = X_0. \tag{2}$$

The dynamics of a quadrotor, as well as most UAV, show nonlinear dynamics [7] and thus lead to an optimal control problem of the form (1)-(2). We remark that inhomogeneities $f_A(t, 0) = b(t)$, e.g. from gravitational forces, can be treated as disturbances, by adding new state variables or by taking advantage of non-vanishing states, e.g. the altitude of the UAV when hover is searched [8].

In this paper we solve optimal control problems of type (1)-(2) that appears in the flight control of a quadrotor. After linearization of the problem in section 2, we obtain a linear problem where, by application of linear quadratic methods, is necessary to solve a symmetric matrix Riccati differential equation (RDE) whose solution is a symmetric positive definite matrix. In section 3, numerical methods based on the Magnus expansion are applied in order to solve the RDE. Magnus integrators are Lie group methods which provide accurate results while preserving some important qualitative properties of the original problem. Magnus integrators can be used to solve the equation for the state vector and the matrix RDE. The performance of the methods are illustrated by numerical examples in the last section.

2 Linear quadratic (LQ) methods in optimal control problems

Let us consider the linear control problem

$$\min_{u\in L^2} \int_0^{t_f} \left(X^T(t)Q(t)X(t) + u^T(t)R(t)u(t) \right) \, dt \,, \tag{3}$$

subject to

$$\dot{X}(t) = A(t)X(t) + B(t)u(t), \quad X(0) = X_0, \tag{4}$$

where $\dot{X}(t)$ is the time-derivative of the state vector $X(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$ is the control, $R(t) \in \mathbb{R}^{m\times m}$ is symmetric non-negative, $Q(t) \in \mathbb{R}^{n\times n}$ symmetric positive definite, $A(t) \in \mathbb{R}^{n\times n}$, $B(t) \in \mathbb{R}^{n\times m}$, and M^T denotes the transpose of a matrix M.

The optimal control problem (3)-(4) is solved [11], assuming some controllability conditions, by the linear feedback controller

$$u(t) = -R^{-1}(t)B^T(t)P(t)X(t), \tag{5}$$

with $P(t)$ verifying the matrix RDE

$$\dot{P}(t) = -P(t)A(t) - A^T(t)P(t) + P(t)B(t)R^{-1}(t)B^T(t)P(t) - Q(t), \tag{6}$$

with the final condition $P(t_f) = 0$. It can be shown that the solution $P(t)$ is a symmetric and nonnegative matrix [1]. To compute the optimal control $u(t)$, we solve for $P(t)$ and plugging the control law into (4) yields a linear equation for the state vector with which the control is readily computed

$$\dot{X}(t) = \left(A(t) - B(t)R^{-1}(t)B^T(t)P(t) \right) X(t), \quad X(0) = X_0 \,.$$

Then, our first goal consist in the linearization of the non-linear problem in order to apply LQ methods for the optimal control of the quadrotor's dynamics. There are various strategies to control the nonlinear problem and we present three of them in the following.

Quasilinearization. For $f_A(t, 0) = 0$ and $f_B(t, X, u) \neq 0$ for all t, X in the appropriate domains, the state equation (2) can be written in a non unique way as

$$\dot{X}(t) = A(t, X)X(t) + B(t, X, u)u(t), \quad X(0) = X_0. \tag{7}$$

The formulation (7) is the basic ingredient for the State Dependent Riccati Equation (SDRE) control technique [8]. Then, formal similarity to the linear problem (3)-(4) motivates the imitation of the optimal LQ controller by defining

$$u(t) = -R^{-1}(t)B^T(t, X(t))P(t, X)X(t) \tag{8}$$

where $P(t, X)$ solves the now state-dependent algebraic Riccati equation

$$0 = -PA(t, X) - A(t, X)^T P + PB(t, X)R(t, X)^{-1}B(t, X)^T P - Q(t, X). \tag{9}$$

Combining (8) with (2), the closed-loop nonlinear dynamics are given by

$$\dot{X} = \big(A(t, X) - B(t, X)R(t, X)^{-1}B(t, X)^T P(t, X)\big)\, X, \quad X(0) = X_0. \tag{10}$$

The usual approach is to start from $X(0) = X_0$, and then to advance step by step in time by first computing P from (9) at each step and then applying the Forward Euler method on (10). The application of higher order methods, such as Runge-Kutta schemes, requires to solve implicit systems with (9) and can thus be costly.

Waveform relaxation. Alternatively, we can linearize (10), by iterating

$$\frac{d}{dt}X^{n+1} = \big(A(t, X^n) - B(t, X^n)R(t, X^n)^{-1}B(t, X^n)^T P(t, X^n)\big)\, X^{n+1}, \tag{11}$$

where starting from a guess solution, $X^0(t)$, one obtains iteratively, by solving linear systems of non autonomous ordinary differential equations (ODEs), a sequence of solutions, $X^1(t)$, $X^2(t), \ldots, X^n(t)$ to be stopped once consecutive solutions differ by less than a given tolerance. Here, $P(t, X^n(t))$ at each iteration is obtained from

$$\dot{P} = -PA^n(t) - A^n(t)^T P + PB^n(t)R^n(t)^{-1}B^n(t)^T P - Q^n(t), \quad P(t_f) = 0,$$

with $A^n(t) \equiv A(t, X^n(t))$, $B^n(t) \equiv B(t, X^n(t))$, etc.

This procedure is similar to what is known as waveform relaxation [14], however, the backward integration for P limits the parallelizability in this application. This approach corresponds to freezing the nonlinear parts in (7) at the previous state and then applying the optimal control law (5). It is worth noting, that this technique can handle inhomogeneities by slightly adapting the control law, at the cost of solving an inhomogeneous linear system.

Taylor-type linearization. Similarly to [12], we can Taylor-expand the vector field in (2) around an approximate solution $X^n(t)$ and use optimal LQ controls for the approximated equation. The iteration step reads then

$$\dot{X}^{n+1}(t) = \bar{A}^n(t)X^{n+1}(t) + \bar{B}^n(t)u^{n+1}(t) + \bar{C}^n(t), \tag{12}$$

where

$$\begin{aligned}
\bar{A}^n(t) &= D_X f_A\left(t, X^n(t)\right) + D_X f_B\left(t, X^n(t), u^n(t)\right) \\
\bar{B}^n(t) &= D_U f_B(t, X^n, U^n) \\
\bar{C}^n(t) &= f_A(t, X^n) + f_B(t, X^n, u^n) - \left(\bar{A}^n(t)\cdot X^n + \bar{B}^n(t)\cdot u^n\right),
\end{aligned}$$

and D_X denotes the derivative whit respect to X, etc. The inhomogeneity C^n can be treated as a disturbance input and compensated by the controller [6]. The optimal control then becomes

$$u^{n+1}(t) = -R^n(t)^{-1}\bar{B}^n(t)^T\left(P^n(t)X^{n+1}(t) + V^n(t)\right),$$

where $P^n(t)$ satisfies (6) with replacements $A \to \bar{A}^n$ and $B \to \bar{B}^n$, etc. and $V^n(t)$ is given by

$$\dot{V} = \left(P\bar{B}R^{-1}\bar{B}^T - \bar{A}^T\right)V - P\bar{C}, \quad V(t_f) = 0\,, \tag{13}$$

at each iteration.

3 Magnus integrators to solve linear ODEs

For all presented methods, after substitution of a control law, we need to solve an initial value problem (IVP)

$$\dot{X}(t) = D(t, P(t))X(t) + C(t), \quad X(0) = X_0.$$

Here, $P(t)$ is a matrix function satisfying (6) with final condition $P(t_f) = 0$. If $C \neq 0$, the linear non homogeneous equation can formulated as a homogeneous one in the following way,

$$\frac{d}{dt}\begin{bmatrix} X(t) \\ 1 \end{bmatrix} = \begin{bmatrix} D(t, P(t)) & C(t) \\ 0_n^T & 0 \end{bmatrix}\begin{bmatrix} X(t) \\ 1 \end{bmatrix}, \quad [X(0), 1]^T = [X_0, 1]^T,$$

where $0_n = [0, \ldots, 0]^T \in \mathbb{R}^n$. The RDE can also be written as a linear problem

$$\frac{d}{dt}\begin{bmatrix} V(t) \\ W(t) \end{bmatrix} = \begin{bmatrix} -A(t)^T & -Q(t) \\ -B(t)R^{-1}(t)B^T(t) & A(t) \end{bmatrix} \begin{bmatrix} V(t) \\ W(t) \end{bmatrix}, \quad \begin{bmatrix} V(0) \\ W(0) \end{bmatrix} = \begin{bmatrix} 0 \\ I \end{bmatrix}, \tag{14}$$

where the solution $P(t)$ of problem (6) is given by

$$P(t) = V(t)W(t)^{-1}; \quad P(t)\,,\ V(t)\,,\ W(t) \in \mathbb{R}^{n\times n}\,, \tag{15}$$

in the region where $W(t)$ is invertible (see, for instance, [10] and references therein). Hence, both the IVP for $X(t)$ and the RDE for $P(t)$ reduce to a matrix linear homogeneous equation to which we focus our attention.

The matrix RDE appears frequently in linear optimal control problems and has been extensively studied in the literature (see [1] for instance). In this paper, Lie group methods (see [4] and references therein) are proposed in order to solve the RDE (6). They are geometric integrators which have shown in many cases a high performance for the numerical integration of linear differential equations because they preserve some of the qualitative properties of the exact solution and also frequently provide accurate results. Regulator problems and their solutions have an important qualitative structure, see [2], which is not preserved by most classical numerical methods, and this can seriously affect the numerical solution.

Magnus integrators, a special class of exponential integrators as well as Lie group integrators, when used to numerically solve, e.g. the eq. (14), can be interpreted as exactly solving a slightly perturbed matrix RDE with a similar structure (i.e. replacing the matrices Q, R by perturbed ones $\tilde{Q}, \tilde{R}$ close to Q, R which are also symmetric, non negative, etc.) and then, the numerical solution for $P(t)$ will also be a symmetric and nonnegative matrix, etc.

Let us first present an explicit symmetric second order Magnus integrator to solve a general linear equation

$$y = S(t)\, y\,, \qquad y(t_0) = y_0\,; \tag{16}$$

with $y \in \mathbb{R}^p$. Let us denote by $\Phi(t, t_0) \in \mathbb{R}^{p\times p}$ the fundamental solution, such that $y(t) = \Phi(t, t_0)y(t_0)$. Then

$$\exp\left(\int_t^{t+h} S(t)\, dt\right) = \Phi(t + h, t) + \mathcal{O}(h^3),$$

corresponds to the first order approximation (second order in the time step, h) for most exponential methods like e.g. the Magnus, Fer or Wilcox expansions, see [3] for details. Here, it suffices to approximate the integral by a second order symmetric rule. From the computational point of view, we found it useful for the numerical algorithm to consider the trapezoidal rule, so

$$\exp\left(\frac{h}{2}\left[S(t+h)+S(t)\right]\right) = \Phi(t+h,t) + \mathcal{O}(h^3). \tag{17}$$

Firstly, let us consider the RDE (6) that corresponds to (16) with the data (14). Let us consider an equidistant time grid $t_n = t_0 + nh\,,\ \ 0 \le n \le N\,,$ with constant time step $h = (t_f - t_0)/N\,.$ By (15), applying (17) and taking into account that this equation has to be solved backward in time, we obtain

$$\left[\begin{array}{c} V_n \\ W_n \end{array}\right] = \exp\left(-\frac{h}{2}\left[S(t_n)+S(t_{n+1})\right]\right)\left[\begin{array}{c} V_{n+1} \\ W_{n+1} \end{array}\right] \ \Rightarrow\ P_n = V_n\, W_n^{-1}\,, \tag{18}$$

where P_n is a time symmetric second order approximation to $P(t_n)$ (for the non linear problems $P^k(t)$ corresponds to the time dependent solution at the k-th iteration). In this way, the matrix functions $A(t_n)$, $B(t_n)$, $Q(t_n)$, $R(t_n)$ are computed at the same mesh points as the approximations P_n of $P(t)$.

The final step is the integration of the dynamics, i.e, depending on the linearization (10), (11) or (12) and (13), the last two of which are inhomogeneous. The equation of motion is integrated forward in time. To illustrate the method, we write explicitly how to solve the waveform relaxation (11) with the exponential integrator (17) in one iteration

$$X_{n+1}^{k+1} = \exp\left(\frac{h}{2}\left[D_{n+1}^k + D_n^k\right]\right) X_n\,, \quad D_m^k = A_m^k - B_m^k (R_m^k)^{-1}(B_m^k)^T P_m^k \tag{19}$$

$m = n, n+1$, with $A_n^k = A(t_n, X_n^k)$, $B_n^k = B(t_n, X_n^k)$, etc., and where $X_n^{k+1} = X^{k+1}(t_n) + \mathcal{O}(h^3)$.

For the linear case this is only computed for $k = 0$ while in the non linear case this has to be repeated for $k = 0, 1, 2, \ldots$ until convergence. As a result, the controls which allow us to reach the final state in a nearly optimal way are

$$u_n \ = \ -R^{-1}(t_n) B^T(t_n) P_n\, X_n\,.$$

In (18) and (19) one has to compute the action of a matrix exponential on a vector. To evaluate the exponential of a large matrix can be computationally expensive but, for sparse matrices, its action on a vector can be

approximated efficiently. In general, the matrices A and B are sparse and Q, R are diagonal. We can take diagonal Padé matrix approximations preserving the Lie group structure. These rational approximations can be easily computed using a fixed point iteration algorithm.

If accurate results are required for linear problems, one can use a high order Magnus integrator. They usually require to compute matrix commutators. However, as mentioned, since in general the matrices A and B are sparse and Q, R are diagonal, it is faster to use a commutator-free Magnus integrator (see [3]). For coupled systems of equations it is convenient to consider an equispaced mesh and, for one time step one can replace the exponential in (18) by

$$\exp\left(\frac{h}{12}(-S_1 + 4S_2 + 3S_3)\right) \exp\left(\frac{h}{12}(3S_1 + 4S_2 - S_3)\right),$$

where $S_1 \equiv S(t_n), S_2 \equiv S(t_n + h/2), S_3 \equiv S(t_n + h)$. Note that a linear combination of sparse matrices is a sparse matrix with the same non-zero elements. We can use the same method to solve the equation for the state vector, $X(t)$, but using a time step twice larger.

4 Numerical experiments

An analysis of the dynamics of the quadrotor shows that the control of the attitude can be separated from the translation of the UAV [13] and we focus our attention on the stabilization of the attitude, neglecting the gyroscopic effect. The state vector is given by

$$X(t) = \left(\phi(t), \dot{\phi}(t), \theta(t), \dot{\theta}(t), \psi(t), \dot{\psi}(t)\right)^T \in \mathbb{R}^6,$$

and the input vector $u(t) \in \mathbb{R}^3$ is formed by linear combinations of the thrust of each propeller.

There are several ways to solve (6), among the simplest is the so called Pearson method that proposes $\dot{P}(t) = 0$ and thus simplifies (6) to an algebraic Riccati equation whose symmetric and nonnegative solution is chosen. For time dependent problems, however, better results are obtained with the Sage-Eisenberg method [5], i.e. to compute $P(t)$ by integrating (6) backwards in time. Note that this requires to store the values $P(t)$ on an appropriate time mesh.

The system designer can choose the weight matrices to tune the behavior of the control according to the requirements, $R(t)$ is used to suppress certain movements and $Q(t)$ limits the use of the control inputs. Usually, these matrices are chosen constant, nonnegative definite, and often even diagonal, see [7, 9]. For the numerical experiments we consider the following values taken from [5, 13]

$$\begin{array}{lll} a_{1,2} = a_{3,4} = a_{5,6} = 1, & a_{2,4} = \lambda\alpha_1 I_1 \dot{\psi}, & a_{2,6} = \lambda(1-\alpha_1) I_1 \dot{\theta} \\ a_{4,2} = \lambda\alpha_2 I_2 \dot{\psi}, & a_{4,6} = \lambda(1-\alpha_2) I_2 \dot{\phi}, & a_{6,2} = \lambda\alpha_3 I_3 \dot{\theta}, \\ a_{6,4} = \lambda(1-\alpha_3) I_3 \dot{\phi}, & b_{2,1} = L/I_x, \quad b_{4,2} = L/I_y, & b_{6,3} = 1/I_z \end{array} \tag{20}$$

where α_i reflects the non-uniqueness in the SDRE formulation, λ denotes the inflow ratio, L is the length of the arms connecting the propellers with the center, $I_1 = (I_y - I_z)/I_x$, $I_2 = (I_z - I_x)/I_y$, $I_3 = (I_x - I_y)/I_z$. Here, $m_{i,j}$ denotes the element located at i-th row and j-th column of the matrix M. Other entries of $A(t) \in \mathbb{R}^{6\times 6}$ and $B(t) \in \mathbb{R}^{6\times 3}$ not indicated in (20) are null elements.

The numerical values are extracted from [5] and are given in the SI units

$$I_x = 0.0075, \quad I_y = 0.0075, \quad I_z = 0.0130, \quad L = 0.23, \quad \lambda = 1\,.$$

The weight matrices are fixed at

$$Q = 0.01 \cdot \mathrm{diag}\{\,1, 0.1, 1, 0.1, 1, 0.1\} \in \mathbb{R}^{6\times 6}, \quad R = \mathrm{diag}\{1, 0.1, 1\} \in \mathbb{R}^{3\times 3}.$$

We set the time frame to $t_f = 10$ seconds, with a stepsize of $h = 0.125s$ and initial state

$$X_0 = (70,\, 10,\, 70,\, 20,\, -130\,,\, -1)^T\,,$$

where the angles are given in degrees. That corresponds to a disadvantageous orientation and high rotational velocities that is sought to be stabilized at $0 \in \mathbb{R}^6$.

We have implemented a variety of methods to test against the Magnus integrators presented in section 3. For the waveform relaxation technique, the parameters α_i are set to 1. As initial condition, we have taken $X^0(t) = (1 - t/t_f)X_0$ and the iteration is stopped when $||X^n - X^{n-1}||_2 < 10^{-3}$.

Figure 1 shows the dynamics of the quadrotor subject to the obtained controls for the schemes `S2` (quasilinearization), `W3` (waveform), `T3` (Taylor). We can appreciate how the Magnus methods maximize the use of the controls to reach an overall minimum of the cost functional.

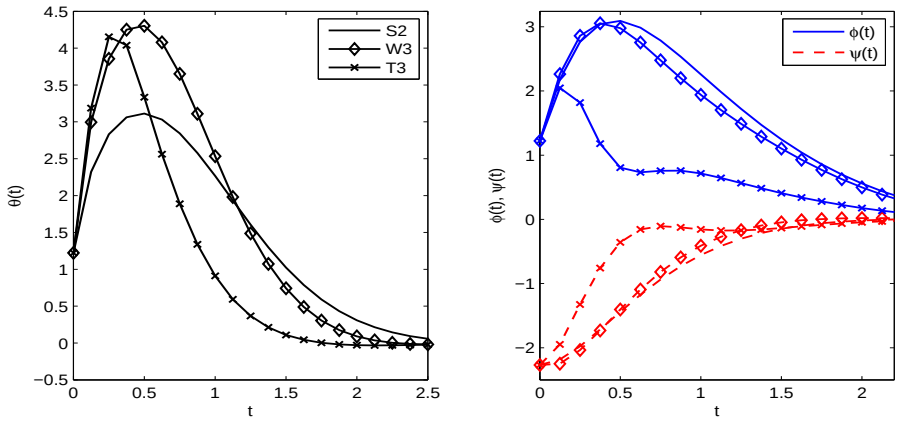


Figure 1: Evolution of the orientation of the quadrotor. The left column shows the coordinate $\theta(t)$ that has been less penalized. All curves are given for all methods `S2` (line), `W3` (diamond) and `T3` (cross).

From the numerical experiments we conclude that Lie group methods like Magnus integrators are very useful tools for solving optimal control problems of UAV. The results shown for a quadrotor easily extend to other helicopters. In addition, for more involved trajectories the structure of the equations will play a more important role and Lie group methods can provide efficient numerical algorithms.

References

[1] H. Abou-Kandil, G. Freiling, V. Ionescu & G. Jank. Matrix Riccati equations in control and systems theory. Virkäuser Verlag, Basel, 2003.

[2] B. D. O. Anderson & J. B. Moore. Optimal control. Linear quadratic methods. Dover Publications. New York, 2007.

[3] S. Blanes, F. Casas, J. A. Oteo & J. Ros. The Magnus expansion and some of its applications. *Physics Reports*, 470: 151–238, 2009.

[4] S. Blanes & E. Ponsoda Time-averaging and exponential integrators for non-homogeneous linear IVPs and BVPs. *Appl. Num. Math*, 62: 875–894, 2012

[5] S. Bouabdallah. Design and control of quadrotors with application to autonomous flying. Ph.D. dissertation, EPFL, 2006.

[6] A. Bryson Jr. & Y. C. Ho. Applied Optimal Control, Halsted, 1975.

[7] P. Castillo, R. Lozano & A. E. Dzul. Modelling and control of mini-flying machines. Advances in Industrial Control Series. Springer. London. England. 2005.

[8] T. Çimen. State-dependent Riccati equation (SDRE) control: A survey. *Proc. of the 17th IFAC World Congress(IFAC'08) Seoul, Korea*, pp. 3761–3775, 2008.

[9] I. D. Cowling, J. F. Whidborne & A. K. Cooke. Optimal trajectory planning and LQR control for a quadrotor UAV. *Proc. UKACC Int. Conf. Control (ICC 2006)*, Glasgow, UK, 2006.

[10] L. Jódar & E. Ponsoda. Non-autonomous Riccati-type matrix differential equations: existence interval, construction of continuous numerical solutions and error bounds. *IMA J. Num. Anal.*, 15: 61-74, 1995.

[11] D. Kirk. Optimal control theory, an Introduction. Dover Publ., Mineola, New York, 2004.

[12] E. Ponsoda, S. Blanes & P. Bader. New efficient numerical methods to describe the heat transfer in a solid medium. *Math. Comput. Mod.*. 54: 1858-1862, 2011.

[13] H. Voos. Nonlinear state-dependent Riccati equation control of a quadrotor UAV. *Proc. Int. Conf. Control Appl.,* Munich, Germany, pp. 2547–2552, 2006.

[14] J. White, F. Odeh, A.S. Vincentelli & A.Ruehli. Waveform relaxation: theory and practice. *Trans. Soc. Comput. Simulation*, 2: 95–133, 1985.

[15] R. Zhang, Q. Quan & K.-Y. Cai. Attitude control of a quadrotor aircraft subject to a class of time-varying disturbance. *IET Control Theory Appl.*, 5: 1140–1146, 2011.

Ensemble of naïve Bayesian approaches for the study of biofilm development in drinking water distribution systems

E. Ramos-Martínez [†] [*], M. Herrera [‡], J. Izquierdo [†] and R. Pérez-García [†]

(†) FluIng-IMM, Universitat Politècnica de València,

C. de Vera s/n, Edif 5C, 46022 Valencia, Spain

(‡) BATir - Université libre de Bruxelles,

Av F. Roosevelt, 50 CP 194/2 B-1050 Brussels, Belgium

November 30, 2012

1 Introduction

Biofilms develop in drinking water distribution systems (DWDSs) as layers of microorganisms bound by a matrix of organic polymers and attached to pipe walls. Biofilm growth within a DWDS could lead to operational problems, generation of bad tastes and odors, proliferation of macroinvertebrates, biocorrosion, and residual chlorine consume. In recent years it has also become evident that biofilms in DWDSs can become transient or long-term habitats for hygienically relevant microorganisms. Biofilms in DWDSs can serve as an environmental reservoir for pathogenic microorganisms and represent a potential source of water contamination, resulting in a potential health risk for humans if left unnoticed. Besides, a number of additional problems associated with biofilm development in DWDSs can be identified. Among others, aesthetic deterioration of water, proliferation of higher organisms,

[*]e-mail:evarama@upv.es

biocorrosion and disinfectant decay are universally recognized. Although in most countries regulated quantities of residual disinfectant are present in the DWDSs, these are not enough to avoid the biofilms formation. So nowadays, biofilms represent a paradigm in the management of water quality in all DWDSs.

Survival and regrowth of microorganisms in DWDSs is affected not only by biological aspects but also by the interaction of various other factors. Numerous studies have been approached in relation to the influence that a number of characteristics of the DWDSs have in biofilm development. Nevertheless, their joint influence, apart from few exceptions, has been scarcely studied, due to the complexity of the community and the environment under study [4]. This work aims to approach this problem studying the effect that the interaction of the relevant hydraulic and physical characteristics of the DWDSs has in biofilm development. As a consequence, it achieves deeper understanding of the cause-effect relations involved in biofilm assessment. To address the difficulties usually found in data relative to this work environment, we propose focusing our analysis on a naïve Bayes approach. It supposes a simplification of unrestricted Bayesian networks. However, it often achieves good accuracy even when compared with decision trees or neural networks classifiers [2].

Four different alternatives are proposed: a tree augmented naïve Bayes classifier (TAN) [1], a bagging combination of naïve Bayes approaches (BNB), a naïve Bayesian tree [3] (NBT) and an ensemble of these approaches by a modified version of a naïve Bayesian tree, where a bagging process is applied in their leaf nodes, generating, what we call, a Bagging naïve Bayesian tree (B-NBT).

2 Discussion

We have approached the problem of studying the influence that the whole set of characteristics of the DWDSs has on biofilm development through the naïve Bayes algorithm, showing that the intricacy of the problem under study is a big handicap to get this aim.

It has been demonstrated that ensemble techniques are more useful in this complex case, obtaining better results than the simpler methods because the iterations increased the robustness of the process. However, this has not been enough to get a good model. Hybrid ensemble techniques have been

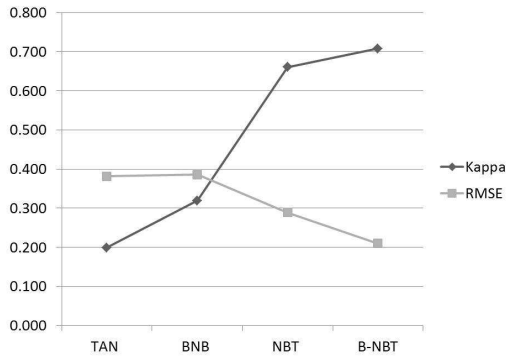


Figure 1: Kappa statistic value and RMSE for TAN, BNB, NBT and B-NBT

necessary to achieve good results (Figure 1). The cumulative experience on the performance of multiple applications of different learning systems is the adequate way to achieve our aim, thus, reducing the uncertainty and improving the overall prediction accuracy of the model. Furthermore, the approach proposed in this paper, has demonstrated to be a suitable way to achieve a good model in this case. It has shown to be able to exploit the advantages of the different techniques used. Avoiding bias and decreasing the uncertainty with the classification trees, improving the efficiency through the naïve Bayes classifier and, finally, gaining accuracy by applying bagging.

The improvement of the output is not shown only in the goodness indexes, but also in the results. Although, in the cases with normal biofilm development, the error percentage of the B-NBT method is a little bit bigger (7.19%) than the obtained with the NBT (6.02%), the error rate of the cases with high biofilm development, in which we are interested in due to their implication in numerous DWDSs problems, is greatly reduced (NBT 29.55%; B-NBT 22.81%). As a consequence, we claim that the methodology that we have developed is able to deal suitably with the problem tackled in this paper, and outperforms previous approaches found in the literature.

3 Conclusions

This work is characterized by offering an innovative perspective in the study of biofilms development in DWDSs with the introduction of intelligent data analysis techniques in this field.

Until now, the effect that the different physical and hydraulic characteristics of the DWDSs have on biofilms development were studied individually in a majority of cases, due to the complexity of the community and the environment under study, together with the scarcity of data. These are the main reasons to propose simple algorithms to approach biofilm assessment in DWDSs. To gain robustness and accuracy different combinations of simple processes, which produce good performance, have been introduced. Thus, by an ensemble algorithm we have achieved deeper understanding of the consequences that the interaction of the relevant hydraulic and physical factors of the DWDSs has in biofilm development. Also we have gone further, increasing even more the accuracy of the obtained model by B-NBT, reaching better results while the process still remains simple and computationally efficient.

This paper represents the base of a more complex tool of decision support system in DWDSs. The problems related to biofilm development in these systems could be solved or mitigated thanks to it. The present work is an advance in the study of biofilms development in DWDSs as it allows deeper understanding of the ecology of these communities and facilitates better understanding of the processes and interactions that occur in DWDSs related to the development of these communities.

References

[1] Friedman N. , Geiger D., and Goldszmidt M. Bayesian network classifiers,*Machine Learning*, Volume(29):131–169, 2012.

[2] Ham J. and Kamber M., Data Mining: Concepts and Techniques. Elsevier, 2006.

[3] Kohavi R. Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid,*2nd International Conference on Knowkedge Discovery and Data Mining*, 202–207, 1996.

[4] Ramos-Marínez Eva. Evaluación del desarrollo de biofilms en los sistemas de distribución de agua potable mediante la extracción de conocimiento a través de los datos (Knowledge Discovery in Databases - KDD),*Master s thesis*, 2012.

Consensus Networks with Signed Graphs to Solve Coherence Problems

M. Rebollo[*,**], A. Palomares[*,†], C. Carrascosa[*,‡], F. Pedroche[†,§]

(*) Dept. Sistemes Informàtics i Computació. Universitat Politècnica de València

(†) Inst. de Matemàtica Multidisciplinaria. Universitat Politècnica de València

November 30, 2012

1 Introduction

In many social networks, determining the relationship between users is not a simple problem to solve because existing methods do not scale well when the networks get bigger. Furthermore, there are cases that allow negative relations explicitly, such as having a negative feedback on eBay. This problem can be considered as the so-called coherence problems. Given a set of elements, we want to create a partition that divides it into two subsets: one with the accepted (coherent) elements and another with the rejected ones.

The problem is represented in a weighted, undirected graph [3]. Let N be a finite set of elements and set of constraints defined as weighted edges $w_{ij} \in [-1, 1]$. The coherence problem tries to find a partition of N into two sets A and R, such that maximizes the compliance of this two coherence conditions:

1. if $w_{ij} > 0$ then $e_i \in A$ iff $e_j \in A$, and
2. if $w_{ij} < 0$ then $e_i \in A$ iff $e_j \in R$.

[*]mrebollo@dsic.upv.es
[†]apalomares@dsic.upv.es
[‡]carrasco@dsic.upv.es
[§]pedroche@imm.upv.es

Let C_A the subset of constraints than fulfill conditions 1 and 2. The coherence value of the partition is $C = \frac{1}{N} \sum_{w_{ij} \in C_A} |w_{ij}|$.

Coherence is maximized if no other partition has greater total weight, so $C^{opt} = \max_{A \subseteq N} C$.

In this work we propose a consensus process to determine the partition that maximizes the coherence value. In a consensus, a network of entities tries to reach a common value for a variable x exchanging information with the direct neighbors. Olfati–Saber et al.[2] has proved that the consensus networks converges to the average value of the network without any additional control nor any kind of knowledge about the structure of the network. And it can be used to converge to any other (common) function. The dynamic of the system is ruled by the equation $x_i(k+1) = x_i(k) + \varepsilon \sum_{j \in N_i} w_{ij}(x_j(k) - x_i(k))$, where N_i denotes the neighbors of e_i. This equation means that each node e_i updates its value x_i as the weighted average of the values received from its neighbors. The complete dynamics of the system can be described by the Laplacian matrix $x(k+1) = (I - \varepsilon L)x(k)$.

However, this model is not useful to solve coherence problems yet. On the one hand, there is a strong constraint about the weights in the matrix, which must be positive values so as to the Laplacian to be positive semidefinite (a requirement to guarantee the convergence). On the other hand, to converge to one unique value does not help identify the partition (A, R).

2 Consensus Networks for Coherence Problems

The obstacle to apply the consensus mechanism to the coherence problem is that the Laplacian matrix must be positive semidefinite and that's not possible if negative weights are allowed. But, if we use the signed laplacian $\bar{L}$ this problem is solved [1]. Let's define $\bar{L} = \bar{D} - W$, where $\bar{D} = \sum_{j \neq i} |w_{ij}|$ and W is the adjacency matrix. $\bar{L}$ is positive semidefinite, so the convergence is guaranteed [2]. The new dynamics is defined by $x(k+1) = (I - \varepsilon \bar{L})x(k)$. If each node is initialized with a random value, when the model converges, a common $|x_i|$ has been reached, but some $x_i > 0$ whereas others $x_i < 0$. The difference in the sign of x_i indicates whether e_i belongs to A or to R. Figure 1 shows the evolution of the values until the networrk converges and the partition is identified.

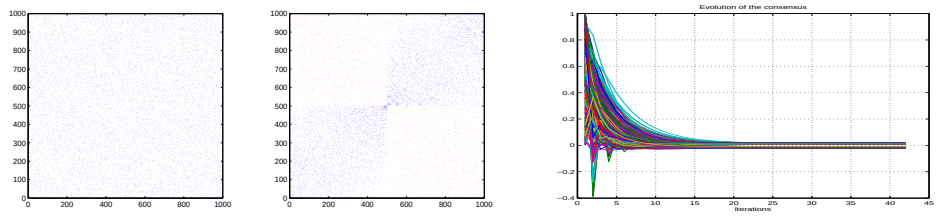


Figure 1: (color online) Convergence of a small-world network with 1000 agents. Initial network (left), final group partition (middle) and convergence process (right)

To analyze the performance of the consensus process, this method has been compared with other techniques used traditionally to solve the coherence problem: **incremental**, **hill climbing**, **Hopfield** networks and **maxcut** (solved by quadratic programming). Two other methods have been added to the tests: the **spectral** analysis, using the eigenvector associated to the smallest eigenvalue, and the **consensus** networks. All these methods have been checked in different network topologies: **complete** networks, 1–dimensional **lattice**, **random** networks and **small-world** networks.

The graphics showed in this paper resume the most relevant obtained results. The parameters used to run the experiments have been the following ones.

1. The number of agents vary from 10 to 100. Only results with the biggest networks have been included here. All of them present the same tendency. Eventually, some isolated experiments with bigger networks have been tested, until 100,000 agents for coherence problems, showing the same tendencies.

2. The size of A and R tend to be the same. It has low impact in the networks with guaranteed solution.

3. Series of 100 runs over different networks

4. The maxcut algorithm used is the Octave[1] implementation

5. In the consensus network, a precision of 10^{-5} is used. The consensus ends when the difference in all the nodes between the current solution

[1]http://www.gnu.org/software/octave/

and the previous one is under this limit.

6. x is initialized with a random solution. No significative change has been observed in the final solution with other initial values.

If the optimal solution is known, a good measure about the quality of the solution is to calculate the proportion W/W^{opt}, where W is the obtained coherence value and W^{opt} is the coherence value of the best solution. But this best solution can not be known unless it is a very small network. In those cases, an proximation is used, being the ratio of the solution calculated as W/W^*, where W^* is the sum of all the edges in the network. This measure is related with the optimal in the sense that those solutions that are nearer to W^{opt} are nearer to W^* too. If the network has a solution, then $W^* = W^{opt}$ and the ratio $W/W^* = W/W^{opt}$ is calculated with respect to the actual optimal solution [3].

Figure 2 shows the ratio W/W^*, that compares the closeness to the optimal solution. The value is not important as a performance measure, but it is useful as comparison among the different algorithms. The experiments show that the quality of the solutions are at the same level that the solutions generated with other methods, computationally more expensive that the consensus process.

An interesting conclusion observed in the experiments is that, under the same coherence value W, consensus networks tend to isolate the communities, whereas other methods (as Hopfield networks) are exclusively based on the utility and they do not take into account additional considerations.

3 Conclusions

This work presents a modification of consensus network to deal with a matrix with negative weights. In this way, consensus networks can be applied to solve coherence problems. When the solution is not guaranteed, all tested methods obtain similar solutions in complete networks, lattices and random networks. The bigger difference is obtained in small-world networks, where consensus processes obtain significative better results than quadratic optimization techniques and spectral analysis.

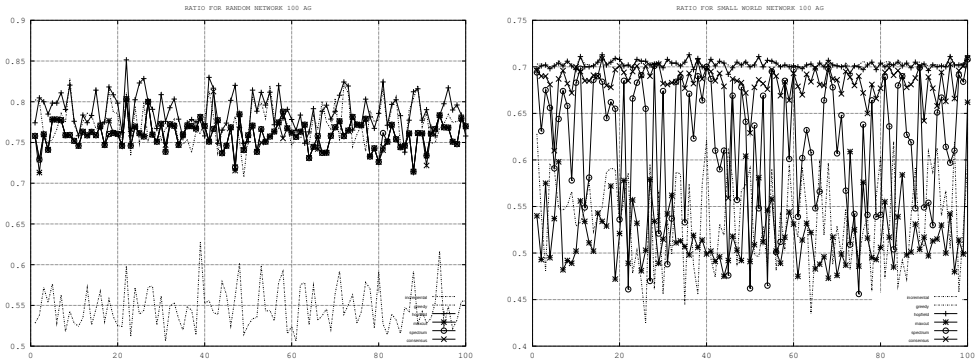


Figure 2: Ratio W/W^* calculated for the solutions obtained in a random network (left) and a small-world network (right) with 100 agents without the solution guaranteed

Acknowledgments

This work is supported by Spanish DGI grant MTM2010-18674, Consolider Ingenio CSD2007-00022, PROMETEO 2008/051, OVAMAH TIN2009-13839-C03-01, and PAID-06-11-2084.

References

[1] Jerome K. et al. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proc. of SDM*, pages 559–559, 2010.

[2] R. Olfati-Saber et al. Consensus and Cooperation in Networked Multi–Agent Systems. *Proceedings of the IEEE*, 95(1):215–233, 2007.

[3] P. Thagard. *Coherence in Tought and Action.* MIT Press, 2000.

Economic evaluation of computed tomography angiography (CTA) versus conventional angiography (CA) to diagnose Coronary Ischemia.

Francisco Reyes-Santías *, Marta dos Anjos Martins-Ramos**, David Vivas-Consuelo**, Carla Sancho Mestre**, Jose M Carreira***

** Unit of Epidemiology and Clinical Research, University Hospital of Santiago. Santiago de Compostela. Spain. Business Organization and Marketing Department. Universidad de Vigo. Spain.*

*** Research Centre for Health Care Economic and Management, Universitat Politècnica de València. Spain*

**** Radiology Department, University Hospital of Santiago. Santiago de Compostela. Spain. Psychiatry, Radiology and Public Health Department, Universidad de Santiago de Compostela. Spain*

David Vivas Consuelo
dvivas@upvnet.upv.es
Gestión de la Salud (CIEGS) de la Universidad Politécnica.
Facultad de Administración y Dirección de Empresas.
Camino de Vera S/N.
46022 Valencia

Economic evaluation of computed tomography angiography (CTA) versus conventional angiography (CA) to diagnose Coronary Ischemia.

BACKGROUND: Health technology assessment is a good manner in which clinical and cost-effectiveness data can give results to make decisions taking in to account various variables.

Aims: In this study three decision analysis models are compared to evaluate the cost-effectiveness of these two imaging technologies used in coronary ischemia diagnose, the CTA (computed tomography angiography) and the conventional angiography (hemodynamics).

Methods: Data was obtained from a population of 81 individuals, both men and women, with ages between 35 and 88 years old. The variables collected for this study were patient age, sex, imaging procedure used and its costs. Four decision analysis models were used: Discriminant Analysis and Regression analysis – Probability Linear, Logit and Probit models, comparing to the classic Decision tree model.

Results: The model that seems more efficient for these case and variables is the

Conclusion: Despite the try to avoid invasive treatments to achieve the diagnosis, in this study is concluded that this is not always the best option in terms of costs and effectiveness, considering the results from the analysis model.

Keywords: Mathematical Modelling in Medicine, Economic evaluation; Health technology, Coronary Ischemia, Medical imaging

Introduction

Health technology assessment [1] is becoming more and more important, especially given the economic environment affecting society nowadays. One of the areas that have seen more progress is the technology in diagnosis, such as imaging techniques that can help diagnose and treat a patient's condition. Most of the times the decision between one procedure and another can be affected by its costs. Nevertheless, dealing with public health the cost cannot be weighed alone, as we have other aspects

that are of major importance, such as the diagnose and resolution of the condition, and the patient's comfort and quality of life. In a first approach, cost-effectiveness is an appropriate factor in making a decision that can take in to account the interests of institutions, physicians and patients.

In the case of diagnosis of coronary ischemia we need to choose between multidetector computed tomography (CTA) and conventional angiography (CA). Multidetector computed tomography for coronary angiography [2] is non-invasive and allows visualization of both the lumen and vessel walls of coronary arteries and to discriminate between calcified and noncalcified atherosclerotic plaque. Conventional angiography [3,-5] by catheter is an invasive technique which only allows visualisation of the coronary lumen, but does allow a therapeutic invasive procedure.

The aim of this article is to assess four classification statistical techniques: comparing discriminant analysis, Probability Linear Model (PLM), and logistic regression with the gold standard decision tree to identify the best technology option for detecting relevant coronary artery lesions.

Materials and Methods

Data was obtained from a population of 81 individuals, aged between 35 and 88 years old. The variables collected for this study were patient age, sex, imaging procedure used and its cost.

Non-invasive studies were performed on 41 with a helical computer tomography system. A contrast agent was injected into a peripheral vein, and cross-sectional images were reconstructed with a slice thickness of 0.5 mm or 1.0 mm. The other 40 studies were performed with Invasive coronary angiography. Findings from both techniques

were analyzed according to a predetermined segmented anatomical model of the coronary artery. The detection and relevance of coronary artery lesions were evaluated.

Cost was estimated using four scenarios each combining alternative allocation of human resources. For scenario one the time assigned was 30, 60, 120 and 90 minutes for the specialist, radiologist, nurse and technician respectively. Time for scenario two was 60 for each professional. The third scenario considerers the following times 120, 60, 120 and 120. Finally the fourth scenario supposed time of 90, 60, 90 and 90.

The total cost calculation was made assigning the human resources cost plus the cost for medicines material and technologies to each scene.

Decision tree was taken as gold standard to compare with three decision analysis models: Discriminant Analysis, Probability Linear Model, Logistic regression and Probit.

Decision tree [6] is a flow diagram depicting the logical structure of a choice under conditions of uncertainty.

It was calculated from the following expression,

$$p(E \mid R) = \frac{p(E)p(R \mid E)}{p(E)p(R \mid E) + p(\overline{E})p(R \mid E)} \qquad (1)$$

$$E = event;\ \ \overline{E} = complement;\ \ R = \mathrm{Re}\,sult\ \ of\ \ research$$

Discriminant Fisher Function

$$D = \mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_k x_k \qquad (2)$$

The next method used was Probability Linear Model. This model assumes that, for a binary outcome, *Y*, and its associated vector of explanatory variables, *X*

$$\Pr(Y = 1 | X = x) = x\beta$$

An explanation of logistic regression begins with an explanation of the logistic function, which, like probabilities, always takes on values between zero and one:

$$\Pi(x) = \frac{e^{(\beta 0 + \beta 1 X)}}{e^{(\beta 0 + \beta 1 X)} + 1} = \frac{1}{e^{-(\beta 0 + \beta 1 X)} + 1'}$$

and

$$g(x) = \ln \frac{\Pi^{(X)}}{1 - \Pi^{(X)}} = \beta_0 + \beta_1 X$$

and

$$\frac{\Pi(x)}{1 - \Pi(x)} = e^{(\beta_0 + \beta_1 X)}$$

Results

Decision tree

The decision tree (table 1) shows the results for the four scenarios. The total cost for all scenarios is over 2,000,000,000 for computed tomography, and rising from 1,566,438,844 for conventional angiography for scenario one.

We are faced with two alternatives in the diagnosis of Coronary Ischemia. The first is the technology of computed tomography angiography (CTA), the second possibility, conventional angiography (CA).

The problem arises when choosing the most suitable technology. Using conventional angiography we can be sure that this is the technology of finally use, the probability is 100%. The results for the four scenarios were 156, 191, 206 and 225 euros respectively.

We compare this alternative to computed tomography angiography of, but in this case the probability of using this alternative ultimately not be 100%, but 79% and will have a 21% chance of conducting a conventional angiography additionally ie, for a

given patient, there will be two interventions, whereby the cost for this case is calculated as the sum of conducting both interventions. The results show that for all scenarios, it is more advantageous to make a conventional angiography from the start, as we incur lower costs.

Discriminat analysis

The discriminant analysis shows the variable age as significant while the variable costs is not significant, and 65,2 % of the cases being obtained correctly classified.

Regression PLM (Probability Linear Model)

The regression PLM nevertheless, shows significance for variable cost with a positive coefficient in using the CT angiography. The correct classification rate with this method would be 63%.

Logistic regression

With this model we can estimate the probability of an event based on a set of predictor variables, which can er qualitative or quantitative. In our case, sex and age are qualitative while the costs would be the only quantitative variable.

The results of the logistic regression shows significance for the variable age and not significance for variable sex and costs, being positive the coefficients of the significant variable. The percentage of correct classification was 63.8% of cases.

Probit Model

With regard to this methodology, the age again becomes the only significant variables in determining technology, age as occurs in the previous case. The correct classification rate is again 63.8% of cases.

Discussion

According to the decision tree technology choose the conventional angiography to diagnose Coronary Ischemia is advantageous for all scenarios.

Regression PLM only variable results as significant cost, while for the rest of methodologies, age significant variable. As for the percentage of correctly classified cases, the methodology achieves a higher percentage is the discriminant, with 65.2% followed by the logit and probit, both with 63.8% of cases correctly classified. Like other authors [7-9], we made a comparison between different methods for the selection of an alternative. The results demonstrate that while for one of the methodologies (PLM Regression) the cost is significant for the rest of them is the age of the patient. Perhaps this discrepancy in results is because there are not enough variables, or we do not have enough observations.

Conclusions

Sensitivity of 64-multislice CT is high enough to rule out significant coronary artery disease in patients with low probability pre-test of coronary artery disease though the Positive Predicted Value is low.

Diagnosis power of Non-invasive coronary artery angiography by multidetector-row spiral computed tomography although is good it is still inferior to invasive coronary angiography.

Moreover, each methodology used in this study provides different significant variables for the detection of the Coronary Ischemia.

[1] Berger ML, Bingefors K, Hedblom EC, Pashos CL, Torrance GW. *Health Care Cost, Quality, and Outcomes: ISPOR Book of Terms*. Lawrenceville, NJ: ISPOR, 2003.

[2] Sones MS, Shirey EK. *Cine coronary arteriography. Med Concepts Cardiovasc* Dis. 1962;31:735.

[3] Rodríguez-Palomares JF, Cuellar H, Martí G, García B, González-Alujas MT, Mahia P, et al. *Coronariografía mediante tomografía computarizada de 16 detectores antes de la cirugía de recambio valvular*. Rev Esp Cardiol. 2011;64:269-76.

[4] Mendoza-Rodríguez V, Llerena LR, Milián-García V, Linares-Machado R, Hernández-Cañero A, Llerena LD, et al. *Precisión de la tomografía de 64 cortes en el diagnóstico de la cardiopatía isquémica*. Arch Cardiol Mex. 2008;78:162-70.

[5] Leta, R, Garcia Picart, J, Carreras, F, et al. (2004). *Non-invasive coronary angiography with 16 multidetector-row spiral computed tomography: A comparative study with invasive coronary angiography.* Revista española de cardiología, 57(3), 217-24.

[6] Bedding, A. and Lilly, E. (2005), *Decision making in health and medicine Hunink M, Glasziou P, Siegel J, Weeks J, Pliskin J, Elstein A, Weinstein MC* (2001) ISBN 0521770297; 404 pages; £38; $60 Cambridge University Press

[7] Richard's, Maria Marta, et al. *Classification statistical techniques: An applied and comparative study*."Psicothema Revista De Psicologia 20.4 (2008):863-871.

[8] Worth, AP, & Cronin, M. (2003). *The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects*. Journal of molecular structure. Theochem, 622(1-2), 97-111.

[9] Harrell, F.E y Lee, K.L. *A comparision of the discrimination of discriminant analysis and logistic regression under multivariate normality.* En P.K. Sen(Ed.). Biostatistics in biomedical: Public Health and Enviromental Sciences (pp. 333-343). North-Holland: Elsevier Science Publishers.

A Macroscopic Model for Signal Detection in Swarm Robotics

Fidel Aznar*, Mar Pujol*, Francisco Pujol†, Mireia Sempere*, Maria José Pujol‡, and Ramón Rizo*

(*) Department of Computer Science and Artificial Intelligence,

(†) Department of Computing Technology,

(‡) Department of Applied Mathematics,

University of Alicante, San Vicent del Raspeig, Alicante (E-03080). Spain.

November 30, 2012

1 Introduction

Swarm robotics is an approach to solve problems inspired by the collective behaviors of social animals and it is focused on the interaction of multiple robots. Based on this metaphor of social insects, swarm robotics emphasizes aspects like decentralized control, limited communication between agents, local information, emergence of a global behavior and robustness. Multi-swarm robotic systems differ from other multi-robotic systems because [4]: 1) robots in a swarm are autonomous robots located in a certain environment, 2) the swarm has a large number of robots, 3) the swarm is composed of small groups of homogeneous robots, 4) robots will be relatively simple, and 5) the robots will have local sensors and their communication skills will be limited. These features ensure that the coordination between the robots will be distributed and that the system will be fault tolerant, since due to the redundancy of robots each of the agents in the system is not essential and can

*Contact author mail: fidel@dccia.ua.es. This work has been supported by the Spanish Ministerio de Ciencia e Innovación, project TIN2009-10581

be replaced by another agent. Thus, the system is easily scalable, allowing more agents to be added or deleted according to task demands. Because of these features, it is difficult to develop an architecture that correctly models a swarm and, on the other hand, that coordinates the swarm to perform complex tasks.

1.1 Presence of RF signals in an urban environment

In urban environments there are many low- and high-frequency electromagnetic waves, where low-frequency signals correspond to transformers, transmission lines, electrical wiring, electrical appliances, etc., and high-frequency electromagnetic waves belong to mobile phone networks, wireless networks, radio and television, among others. Although radiofrequency (RF) signals operate in different wavelengths, the fact of using multiple wireless technologies at the same time, such as Bluetooth, GSM/GPRS or IEEE 802.11b, makes the environment to have interferences between the different signals, slowing down, as a result, the transmission speed. Moreover, there is a loss of performance when multiple users use the wireless network at the same time, important because the medium access protocol becomes inefficient. The quality of service for RF signals is influenced by different effects (such as geometric attenuation, atmospheric conditions, buildings, etc.); there are different kind of models to predict the attenuation. In all cases, the theoretical models for chosing antennas (their locations and characteristics), must be validated by using field tests, in order to evaluate these undesired effects [1]. In general, we found that RF networks are exposed to a great source of noise produced by different elements, and these elements can have a significant impact on the interference level in the network [3]. Consequently, to keep the RF network performance certain power levels are used to ensure signal links, but sometimes they generate more power than it is needed. In this paper, we are interested in determining the quality of mobile RF signal coverage in an urban area, identifying the areas where the greatest power of RF signal is received. From these results, we will determine if the intensity level is excessive or not in the identified areas.

2 Design of the swarm

An example of using extremely simple agents in swarm systems is described in the emergent game proposed in (`http://icosystem.com/game.htm`). In this game, an agent can develop three types of behavior (refugee or evasive, defender and attacker), from which complex behaviors emerge at the group level. These behaviors are discussed in [2], which shows that from these simple behaviors, three different behaviors of the swarm can be obtained: expansion, cycle or aggregation. This kind of behaviors does not allow exploration tasks or utilization of resources.

Inspired by these studies we propose to extend some of the behaviors outlined. For our purposes we will use three behaviors: the aggressive agents, who pursue other agents; the recognition agents, that wander around the environment; and the elusive agents, whose main purpose is to escape from their attackers. As discussed in [2], if each agent pursues another agent, then the swarm tends to cluster at a specific point in the space. On the contrary, the behavior of elusive agents makes the swarm to expand infinitely.

3 Experimentation

Given an area with radio frequency (RF) signal coverage, there will usually be an array of antennas to receive the signal, having different intensities and overlapping signal areas; there will be added elements, as well, either amplifying signals or attenuating them.

Regarding the test area that has been used to validate our model, we must remark the following issues. First of all, the area is located in an area centered on the main Campus of the University of Alicante. Second of all, the RF signal coverage was generated using the program cloudrf (`http://cloudrf.com/`), having three antennas located on the perimeter of the Campus. These antennas generate different signal intensities over the Campus.

As indicated in the swarm robot model, it is essential that robots will have a reduced cost for economy, robustness and viability reasons. In this case, we would use low cost quadricopters to fly over the exploration area at a preset altitude high enough to avoid colliding with the buildings. They will also have the necessary means for autonomous movement and a RF sensor to measure the signal strength.

For the simulation of the system, the powerful and versatile MASON simulator for multi-agent systems has been used. Based on the MASON simulator, we have developed a continuous 2D environment, which contains discrete resource cells, and each cell has the same size as an agent (1x1m approx.). The amount of resources of a cell (RF signal strength) is described from black (no resource) to white (maximum amount of resources). In the simulation model, we consider different swarm sizes with (10,50,100,500) individuals, whose objective is to locate the presence of signal and to indicate the areas of maximum coverage in a collective way, according to the behaviors defined by our model.

4 Conclusions and future works

Once the process of locating ends, the individuals of the swarm are located on sites that have the highest RF intensities. This allows engineers to determine how to distribute the signal more efficiently by means of two alternatives: firstly, by modifying the transmission power of the antennas and, secondly, by relocating some of them. In any case, the proposed method verifies that the theoretical model of RF signals in an urban area can be validated using the swarm model proposed in this paper. We therefore consider that the data provided demonstrated empirically that the swarm will always aggregate in one of the highest RF signal locations. This application of swarm systems can establish new ways to improve signal coverage and to reduce the signal strength in areas with enough intensity levels.

References

[1] M. Chabane M. Alnaboulsi H. Sizun Bouchet, T. Marquis. Fso and quality of service software prediction. *Proc. SPIE 5892, 589204*, 2005.

[2] Ian A. Gravagne and Robert J. Marks. Emergent behaviors of protector, refugee, and aggressor swarms. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 37(2):471–476, 2007.

[3] Aniket Mahanti, Niklas Carlsson, Carey Williamson, and Martin Arlitt. Ambient interference effects in wi-fi networks, 2010.

[4] Erol Şahin. Swarm robotics: From sources of inspiration to domains of application, 2005.

Computational study of the influence of the needle eccentricity on the internal flow in diesel injector nozzles

F.J. Salvador[†,*], J. Martínez-López[†], J.-V. Romero[‡] and M.-D. Roselló[‡]

([†]) CMT-Motores Térmicos, Universitat Politècnica de València,

Camino de Vera s/n, Edificio 6D, 46022 Valencia, Spain,

([‡]) Instituto de Matemática Multidisciplinar, Universitat Politècnica de València,

Camino de Vera s/n, Edificio 8G, 2°, 46022 Valencia, Spain

November 30, 2012

1 Introduction

It is well known that during the opening and closing process of a diesel injector, the fuel characteristics at the nozzle exit change significantly as a consequence of the needle movement [7, 2]. This change of fluid properties at the exit of the discharge orifices strongly affects the spray pattern and the air-fuel mixing process, and therefore its subsequent combustion [5, 6].

Nevertheless, despite most investigations focus only on the vertical motion of the needle, the internal flow and spray characteristics can be also affected by an eccentric location of the needle [3, 1]. This phenomenon, produced by random oscillations of the needle in the transverse direction during the opening or closing of the injector, makes difficult the study of the internal flow, especially at cavitating conditions.

[*]e-mail: fsalvado@mot.upv.es

2 Simulations description

In order to study the influence of the needle eccentricity on the internal flow, two different geometries have been simulated: one nozzle with a needle perfectly centered and a second one with an eccentric position of the needle. Both geometries belong to the same diesel injector nozzle, which has 6 cylindrical holes with a length of 1 mm, a diameter of 170 μm and a curvature inlet radius of 13 μm.

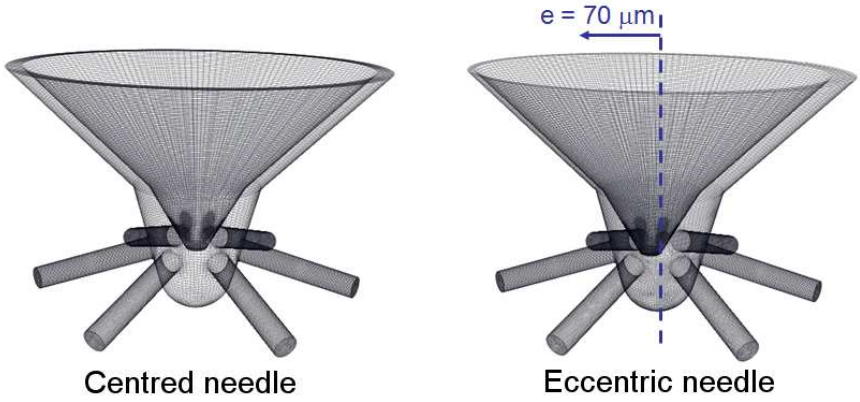


Figure 1: Geometries simulated: nozzle with a needle perfectly centred (left) and with an eccentric position (right).

The geometries, depicted in Figure 1, have been modeled at fully needle lift conditions (250 μm) and they have been discretized in a hybrid mesh of around 600000 hexahedral cells. The cell size ranges from 7.5 μm in the orifice core to a minimum of 2 μm in the orifice wall. For the rest of the geometry (sac, needle seat, etc.) the cell size ranges from 7.5 μm to 22.5 μm depending on the distance between the needle and the nozzle wall.

As far as the boundary conditions are concerned, five pressure conditions have been tested. The injection pressure, defined as the pressure existing in the common rail, was set with a constant value of 30 MPa, whereas the backpressures simulated were 1, 3, 5, 7 and 9 MPa.

3 Influence on the internal flow

3.1 Cavitation appearance

Analyzing the cavitation appearance, there are not hole to hole differences for the nozzle with the needle placed in the center. As a sample, Figure 2 shows how cavitation starts in the curvature radius of the orifice inlet and grows mainly along the upper part of the hole until reaching the nozzle exit. This cavitation field is exactly the same for the rest of orifices.

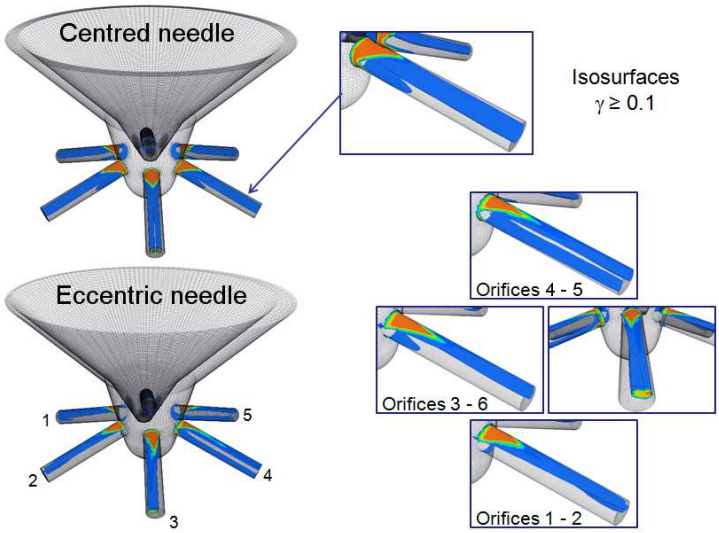


Figure 2: Cavitation fields at $P_{inj} = 30$ MPa $- P_{back} = 1$ MPa.

For the case with an eccentric position of the needle, the fact of having a variable pressure field in the nozzle depending on the needle position affects the cavitation development. Indeed, if the isosurfaces for $\gamma \geq 0.1$ are analyzed for each orifice, it is clear that the vapour phase distribution is completely different. For holes 4 and 5 the entry of fuel from the upstream area and from the sac, forces that cavitation reaches the outlet from the upper and lower part of the hole. However, as the entry of fuel in the orifices 1 and 2 from the sac is lower, the cavitation developed in the lower part of the orifice is almost negligible.

Once understood the cavitation development for orifices 1–2 and 4–5, it is quite easy to understand the vapour phase developed in the orifices 3

and 6. The amount of fuel coming in these orifices by their bottom part is lower than the orifices 4–5 and higher than the orifices 1–2. For that reason, the cavitation length in the lower part of the orifice is between the length observed in the orifices 1–2 and 4–5.

3.2 Hole to hole deviation

Obviously, for the case where the needle is placed in the center of the nozzle the values of mass flow, momentum flux and effective velocity are the same for all the orifices. Nevertheless, as can be seen in Figure 3 corresponding to the eccentric needle case at $P_{back} = 1$ MPa, the mass flow and momentum flux values strongly depend on the needle position.

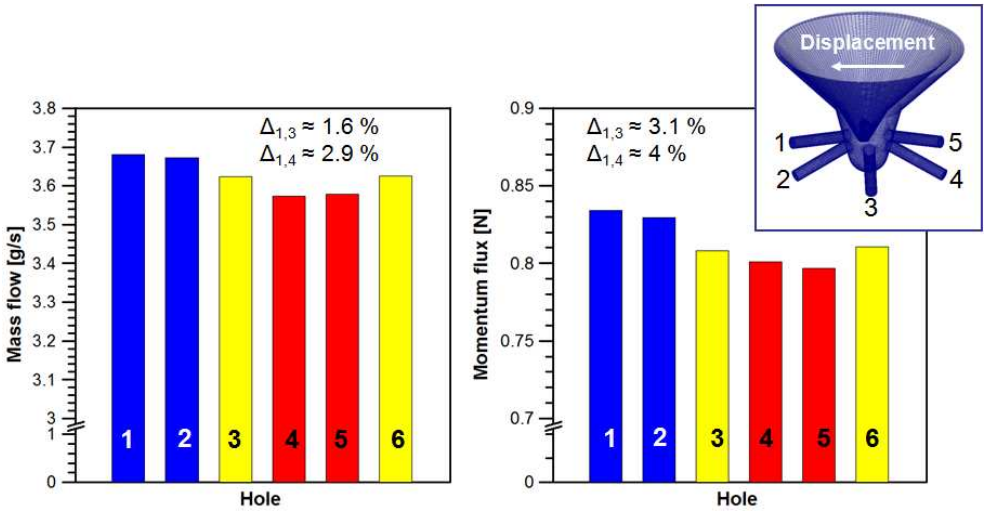


Figure 3: Mass flow and momentum flux values for each orifice for the case with the eccentric needle. $P_{inj} = 30$ MPa $- P_{back} = 1$ MPa.

On the upper right part of Figure 3 there is a nozzle scheme with all the orifices labeled for making easier the explanation of the results, being the orifices number 1 and 2 the orifices closer to the needle and the orifices 4 and 5 the farthest ones. Surprisingly, the mass flow and momentum flux in the orifices closer to the needle are higher, being the orifices 4 and 5, which are the farthest from the needle, the holes with less momentum and mass flow. This behavior can be explained remembering the cavitation field seen

in Figure 2, since the existence of vapor in the upper and lower part of the orifices 4 and 5 strongly reduces the amount of fuel injected.

3.3 Total/averaged flow properties

Figure 4 shows the total mass flow (considering all the orifices of the nozzle) and the averaged momentum flux and effective velocity at the nozzle outlet as a function of the square root of pressure drop, defined as the difference between the injection pressure and the discharge pressure.

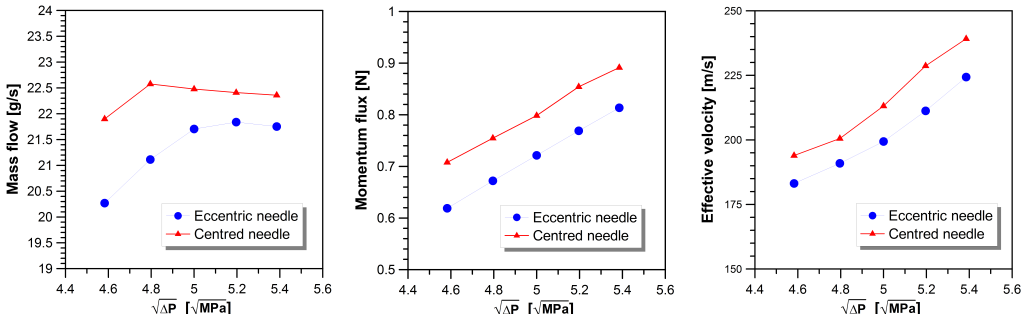


Figure 4: Mass flow, momentum flux and effective velocity results.

Attending to the mass flow graph, the amount of fuel injected in the combustion chamber for the case where the needle is perfectly placed increases as the backpressure decreases from 9 to 5 MPa. Once the backpressure arrives to 5 MPa, the mass flow remains chocked or collapsed, and despite decreasing the discharge pressure the mass flow remains invariable. This phenomenon is a consequence of phase change of the fuel from liquid to vapour due to cavitation [4].

For the nozzle with an eccentric position of the needle, the critical cavitation conditions or the conditions where the mass flow starts to be constant is achieved earlier, at 7 MPa. However, not only the evolution of the flow is different between both cases, but also the values obtained. Indeed for all the pressure conditions the amount of fuel injected is higher when the needle is placed in the center of the nozzle, with a maximum difference of around 7.4% for the highest backpressures.

With regard to the averaged momentum flux at the nozzle outlet, the trend of this parameter is similar in both cases with a continuous increase

of the force of the spray with the pressure drop. As happened with the mass flow results, the comparison of momentum flux values between both cases shows important differences, being the averaged momentum flux of the orifices when the needle is centered 12% higher.

This tendency can be seen also analyzing the effective velocity of the flow in the outlet section of the orifices, since for all the pressure conditions the velocity when the needle is displaced is 7% lower. From this graph it is also interesting to notice the increase of the slope when cavitation takes place (at 7 MPa for the red line and 5 MPa for the blue line). This fact has two reasons: on the one hand, as there are vapour bubbles in the flow, the cross section for the pass of fluid is lower; on the other hand, as the viscosity of the vapour phase is lower, the friction losses decreases [4].

4 Conclusions

From this analysis, the following conclusions can be drawn:

- When the needle is perfectly centered, the cavitation appearance is similar in all the orifices, whereas if the needle is displaced there are strong variations of the vapour phase between orifices.
- The eccentricity of the needle produces strong hole to hole differences in terms of mass flow, momentum flux and effective velocity.
- An eccentric position of the needle produces lower values of mass flow, momentum flux and effective velocity.

Acknowledgements

This work was partly sponsored by "*Vicerrectorado de Investigación, Desarrollo e Innovación*" of the "*Universitat Politècnica de València*" in the frame of the project "*Estudio de la influencia del uso de combustibles alternativos sobre el proceso de inyección mediante GRID computing (FUEL-GRID)*", and by "*Ministerio de Ciencia e Innovación*" in the frame of the project "*Estudio teórico-experimental sobre la influencia del tipo de combustible en los procesos de atomización y evaporación del chorro Diesel (PRO-FUEL)*", reference TRA2011–26293. This support is gratefully acknowledged by the authors. The authors would also like to thank the computer resources

and assistance provided by the Universidad de Valencia in the use of the supercomputer "Tirant".

References

[1] O. Chiavola and F. Palmieri. Modeling needle motion influence on nozzle flow in high pressure injection systems. SAE Paper 2007 00–0250.

[2] J. W. Lee, K. D. Min, K. Y. Kang, C. S. Bae, E. Giannadakis, M. Gavaises, and C. Arcoumanis. Effect of piezo-driven and solenoid-driven needle opening of common rail diesel injectors on internal nozzle flow and spray development. *International Journal of Engine Research*, 7(6):489–502, 2006.

[3] T. Oda, M. Hiratsuka, Y. Goda, S. Kanaike, and K. Ohsawa. Experimental and numerical investigation about internal cavitating flow and primary atomization of a large-scaled vco diesel injector with eccentric needle. ILASS-Europe 2010, 23rd Annual Conference on Liquid Atomization and Spray Systems, Brno, Czech Republic, September 2010.

[4] F. Payri, R. Payri, F. J. Salvador, and J. Martínez-López. A contribution to the understanding of cavitation effects in diesel injector nozzles through a combined experimental and computational investigation. *Computers & Fluids*, 58:88–101, 2012.

[5] R. Payri, J. M. García, F. J. Salvador, and J. Gimeno. Using spray momentum flux measurements to understand the influence of diesel nozzle geometry on spray characteristics. *Fuel*, 84:551–561, 2005.

[6] R. Payri, F. J. Salvador, J. Gimeno, and J. De la Morena. Effects of nozzle geometry on direct injection diesel engine combustion process. *Applied Termal Engineering*, 29:2051–2060, 2009.

[7] R. Payri, F. J. Salvador, P. Martí-Aldaraví, and J. Martínez-López. Using one-dimensional modeling to analyse the influence of the use of biodiesels on the dynamic behavior of solenoid-operated injectors in common rail systems: Detailed injection system model. *Energy Conversion and Management*, 54:90–99, 2012.

Model selection to study the dynamics of the cocaine consumption in Spain using a bayesian approach

F. Guerrero [*], F.-J. Santonja[†], M. Rubio[‡], R.-J. Villanueva[‡] and J.-C. Cortés[‡]

(*) Departamento de Matemática Aplicada. Universidad de Valencia,

Dr. Moliner 50, 46100 Burjassot, Spain.

(†) Departamento de Estadística e Investigación Operativa. Universidad de Valencia,

Dr. Moliner 50, 46100 Burjassot, Spain.

(‡) Instituto de Matemática Multidisciplinar,

Camino de vera s/n, Universidad Politénica de Valencia, 46022 Valencia, Spain.

1 Introduction

In this work we study the dynamics of the cocaine consumption in Spain. Taking into account the possibility to define this dynamics by ordinary differential equations, three possible scenarios (three mathematical models) are presented and using an Approximate Bayesian Computation (ABC) technique we will select the model that best matches the Spanish situation.

2 Mathematical models

In order to build model 1 and model 2, the 15–64 years old Spanish population is considered and divided into four subpopulations (following the division proposed by the National Drug Observatory of Spain [1]): Non-consumers

[*]e-mail:guecor@uv.es

(N(t)), individuals who have never consumed cocaine; occasional consumers ($C_o(t)$), individuals who have consumed sometimes in their life; regular consumers ($C_r(t)$), individuals who have consumed in the last year; and habitual consumers ($C_b(t)$), individuals who have consumed in the last month.

Table 1 shows the prevalence rates of cocaine consumption in Spain for the last years.

Table 1: Evolution of the proportions of the subpopulations for different years. The data have been obtained from the Drug National Observatory Reports (Spanish Ministry of Health)[1]. The subpopulation T (individuas on therapy) is only considered in model 3. This subpopulation is calculated taking into account [1, 2].

	N	Co	Cr	Cb	T
1997	0.947953	0.032	0.015	0.005	0.0000466
1999	0.947887	0.031	0.015	0.005	0.0001127
2001	0.910788	0.049	0.026	0.014	0.0002116
2003	0.902585	0.059	0.027	0.011	0.0004154
2005	0.883483	0.070	0.030	0.016	0.0005174
2007	0.873509	0.080	0.030	0.016	0.0004907
2009	0.859545	0.102	0.026	0.012	0.0004546

According to this, we propose three models to study the evolution over the time of the subpopulations described above. The parameters of the models and their definition intervals are defined in table 2.

The transitions between these subpopulations, according to the model 1 and model 2 are shown in Figure 1 and are described by the equations (1)-(5). In bold, the new term introduced, corresponding to model 2.

$$\frac{dN(t)}{dt} = \mu P(t) - d_N N(t) - \beta \frac{N(t)(C_o(t) + C_r(t) + C_b(t))}{P(t)} + \varepsilon C_b(t) \quad (1)$$

$$\frac{dC_o(t)}{dt} = \beta \frac{N(t)(C_o(t) + C_r(t) + C_b(t))}{P(t)} - d_{c_1} C_o(t) - \gamma C_o(t) + \alpha \mathbf{C_r(t)} \quad (2)$$

$$\frac{dC_r(t)}{dt} = \gamma C_o(t) - d_{c_2} C_r(t) - \sigma C_r(t) - \alpha \mathbf{C_r(t)} \quad (3)$$

$$\frac{dC_b(t)}{dt} = \sigma C_r(t) - d_{c_3} C_b(t) - \varepsilon C_b(t) \quad (4)$$

$$P(t) = N(t) + C_o(t) + C_r(t) + C_b(t) \quad (5)$$

Table 2: Prior definition for the parameters of the models 1 and 2. The minimum and maximum values for the parameters μ, d_N, dc_1, dc_2 and dc_3 have been estimated taking into account [3, 4]. The parameters β, γ, σ and α have been considered with maximum range of variation. For ε minimum and maximum values have been estimated considering [5] and unbiased maximum likelihood estimation. The parameter α corresponds to model 2 and the definition of ε is the one used in model 1.

	Definition	Min	Max
μ	birth rate in Spain	0.008343	0.009228
d_N	death rate in Spain	0.009227	0.010944
dc_1	augmented death rate due to occasional consumption	0.0369	0.087
dc_2	augmented death rate due to regular consumption	0.0369	0.087
dc_3	augmented death rate due to habitual consumption	0.0369	0.087
β	transmission rate due to social pressure to consume cocaine	0.0	1.0
γ	rate at which an occasional consumer transits to the regular consumption subpopulation	0.0	1.0
σ	rate at which a regular consumer transits to the habitual consumption subpopulation	0.0	1.0
α	rate at which a regular consumer becomes an occasional consumer by decreasing his frequency of consumption	0.0	1.0
ε	rate at which a habitual consumer becomes a non-consumer by therapy	0.0	0.00009

Model 2 is defined adding a new transition, C_r to C_o, that it is modeled by αC_r. The introduction of this new transition is in order to test the hypothesis that a non-problematic consumption can be controlled. That is, we consider the possibility that a regular consumer can decrease, or control, his/her cocaine consumption (without therapy) and he/she can become an occasional consumer. In the first model, we admit that the only possibility to decrease cocaine consumption is by therapy (in habitual consumers).

To define model 3 we consider a new subpopulation, T, defined by habitual cocaine consumers who decide to stop consumption and go into therapy. Model 3 has two additional parameters (ϕ, rate at which people in therapy decide to leave therapy and return to habitual consumption, and d_{c_4}, death rate of people in therapy) and a different definition for ε as rate at which habitual consumers enter into therapy. This model is defined by the equations (1), without the last term $\varepsilon C_b(t)$; (2) and (3), without the last term $\alpha C_r(t)$; (4), with a new term, $\phi T(t)$; a new equation corresponding to the new subpopulation T ($\frac{dT(t)}{dt} = \varepsilon C_b(t) - \phi T(t) - d_{c_4} T(t)$); and (5), with a new

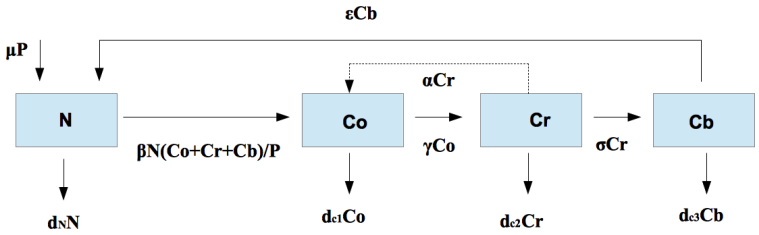


Figure 1: Flow diagram for the evolution dynamics of cocaine consumption in Spain. The boxes represent the subpopulations and the arrows represent the transmission between the subpopulations. Arrows are labeled by their corresponding model transition terms. The dashed arrow is the additional one corresponding to model 2.

term, $T(t)$.

3 Model selection and parameter estimation

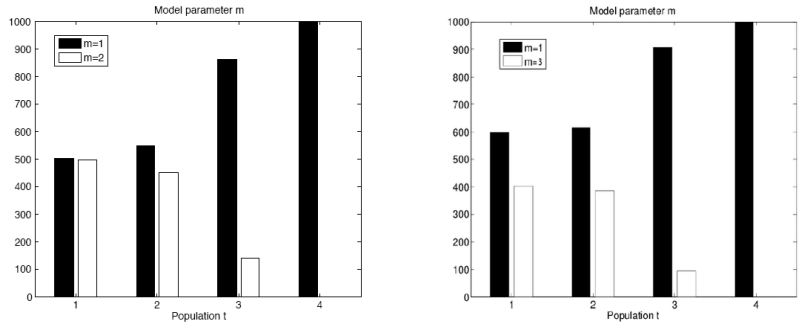
We have already presented the three models and now we have to decide which one is the best to describe the evolution of the different subpopulations. To do this, we are going to use the approximate bayesian computation sequential Monte-Carlo approach (ABC SMC) proposed by T. Toni in [6]. This bayesian approach is based on the study of the evidence provided by the data (D, Table 1) in favour of one model over other one. The objective is to obtain a set of N parameter vectors $\theta(m)$, with m = 1, 2, divided between two models, that satisfy the final condition $d(x^*, D) \leq \epsilon_T$. This condition means that the prediction x^* given by model m with values of the parameters $\theta(m)$ has a distance less than ϵ_T from the observed data. At the end of the process, we select the model having the highest number of $\theta(m)$.

In a first step, we compared model 1 with model 2 and then, we will compare the best model obtained (between this two first models) with the model 3.

4 Results

The values of ϵ_t that we have used to compare model 1 with model 2 are $\epsilon_1 = 0.0160$, $\epsilon_2 = 0.0090$, $\epsilon_3 = 0.0070$ and $\epsilon_4 = 0.0066$; and to compare model 1 with model 3 the considered values are $\epsilon_1 = 1.0$, $\epsilon_2 = 0.5$, $\epsilon_3 = 0.3$

and $\epsilon_4 = 0.13$. These values are defined considering the deterministic fitting of the models in the least square sense.



(a) Model 1 with model 2 (b) Model 1 with model 3

Figure 2: Evolution of the number of parameters vectors, $\theta(m)$, corresponding to each model in each population $t = 1, 2, 3, 4$. We take T = 4 and we have considered N = 1000.

Figure 2 shows the distributions obtained for the parameter $\theta(m)$ for each iteration $t = 1, 2, 3, 4$ according to the four values of ϵ_t for the two selection procedures. We can see how the number of times that the algorithm selects the model 2 is decreasing as ϵ_t is decreasing. In the same way, the number of times that the algorithm selects the model 3 is decreasing as ϵ_t is decreasing. Finally, only model 1 is selected. Thus, we can conclude that model 1 is the one that best describes the evolution of the subpopulations.

In addition, the ABC scheme provides us an approximation to the posterior probability distributions for the parameters of the selected model, in this case, model 1. Table 3 shows the 95% credible interval defined by percentile 5 and 95.

This fact allows us to predict the evolution of the subpopulations in the next few years for the Spanish population. Considering the solution of the selected model –1000 solutions, one solution for each value of $\theta(m_1)$– we will be able to calculate the prediction for the proportions of non-consumer, occasional consumers, regular consumer and habitual consumers by credible intervals. The predictions for cocaine consumption in Spain are shown in Figure 3. We note a decreasing trend in non-consumer subpopulation. Also, there is an increasing trend in all the populations of cocaine consumers, i.e, occasional, regular and habitual consumers. That is, if there are not changes in current policies cocaine consumption in Spain will be increase.

Table 3: Posterior probability distributions for the parameters of model 1.

Parameter	Median	95% Credible Interval
μ	0.008791	(0.008392; 0.009200)
d_N	0.010068	(0.009287; 0.010870)
dc_1	0.047139	(0.037407; 0.067030)
dc_2	0.062998	(0.045338; 0.079426)
dc_3	0.076774	(0.057828; 0.086483)
β	0.057903	(0.030229; 0.073108)
γ	0.021427	(0.016985; 0.024264)
σ	0.000716	(0.000081; 0.001913)
ϵ	0.031593	(0.005467; 0.048865)

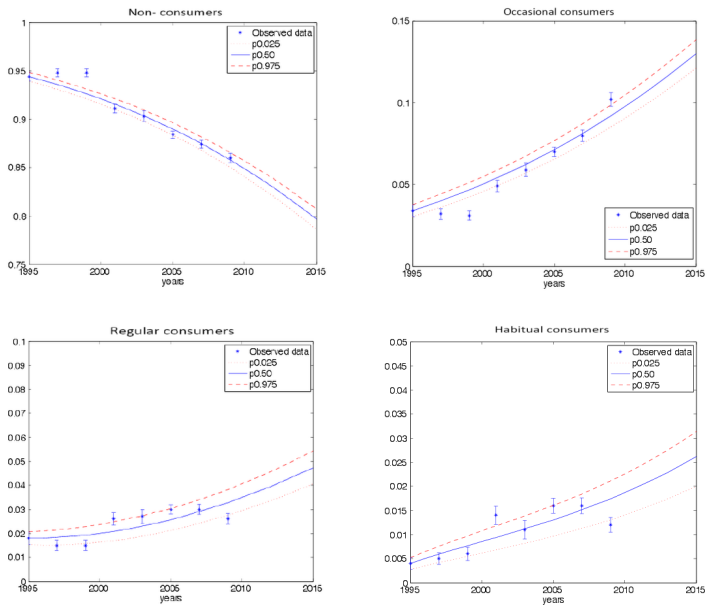


Figure 3: Probabilistic projections of cocaine consumption population in Spain.

References

[1] NPD. Spanish Ministry of Health. Spanish Drugs Survey 2009. Retrieved from http://www.pnsd.msc.es/Categoria2/observa/estudios/home.htm, 2009.

[2] L. Dutra, G. Stathopoulou, S.L. Basdem, T.M. Leyro, M.B. Powers, and M.W. Otto. A meta-analytic review of psychosocial interventions for substance use disorders. *American Journal of Psychiatry*, Volume(165):179–187, 2008.

[3] INE, Spanish Statistic Institute. Retrieved from http://www.ine.es, 2008.

[4] Spanish Health Ministry. Spanish drugs survey 2007-2008. Retrieved from:
http://www.msc.es/gabinetePrensa/notaPrensa/pdf/EncuestaDomiciliariaDrogasAlcohol(EDADES)MINISTRO.pdf

[5] E. Sánchez, R. J. Villanueva, F.J. Santonja and M. Rubio. Predicting cocaine consumption in Spain: a mathematical modeling approach. *Drugs: Education, Prevention and Policy*, Volume(18(2)): 108–115, 2011.

[6] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M.P.H. Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, Volume(6): 187–202, 2009.

A general reference rear-muffler model for exhaust system pre-design

F. Payri[†], A. J. Torregrosa[†] [*], A. Broatch[†], and J.-P. Brunel[‡]

(†) CMT Motores Térmicos, Universitat Politècnica de València,

Camino de Vera s/n, 46022–Valencia, Spain.

(‡) Faurecia Emissions Control Technologies,

Bois sur Pres, 25550 Bavans, France.

November 30, 2012

1 Introduction

At the first stages of muffler design, it may be convenient to dispose of some muffler model in which, for a given total volume and without precise knowledge of the internal structure of the muffler, one may put relevant information regarding very general characteristics of the muffler, such as dissipation, required attenuation, etc.

This general model would be sufficient for exhaust pre-design, since the overall shape of the attenuation, as well as the interactions between exhaust elements, would be accounted for in an approximate but sufficient way.

The model chosen must comply with general conservation principles, so that it is not an arbitrary black-box, but a significant reference so as to determine if design requirements may be affordable when a real muffler is used.

The development of such a model is attempted in the present work. First, the model chosen is described and the way to define its parameters is discussed. Then, some illustrative results are given. Finally, possibilities for further developments are outlined.

[*]e-mail: atorreg@mot.upv.es

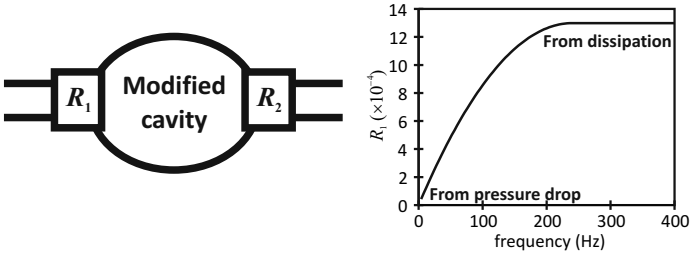


Figure 1: Model scheme (a), and typical inlet resistance (b).

2 Description of the model

The model was developed starting from the ideal cavity, but modifying its behaviour so that the low frequenciy attenuation associated with the volume is preserved, but at higher frequencies one has a constant attenuation (to be fixed as a design criterion). This implies, that the cut-off frequency above which the attenuation remains unchanged is uniquely determined.

Dissipation was incorporated through upstream and downstream resistances, as shown in Fig. 1(a). For equal inlet and outlet diameters, dissipation may be computed as $\eta = 1 - (|T|^2 + |R|^2)$, where T and R are, respectively, the transmission and reflection coefficients, which can be obtained from the transfer matrix elements [1]. The final result is

$$\eta = \frac{4(\tilde{R}_1 + \tilde{R}_2) + 4(1 + \tilde{R}_2)^2 \tilde{R}_1 \varphi^2/\psi^2}{(2 + \tilde{R}_1 + \tilde{R}_2)^2 + (1 + \tilde{R}_2)^2(1 + \tilde{R}_1)^2 \varphi^2/\psi^2} \tag{1}$$

where $\psi = d_e^2/8V$, d_e is the inlet diameter, $\tilde{R}$ is a non-dimensional resistance, and $\varphi = f/a_0$. Taking the limits $\varphi \to 0$ and $\varphi \to \infty$, one has

$$\eta(\varphi \to 0) = \frac{4(\tilde{R}_1 + \tilde{R}_2)}{(2 + \tilde{R}_1 + \tilde{R}_2)^2} \quad ; \quad \eta(\varphi \to \infty) = \frac{4\tilde{R}_1}{(1 + \tilde{R}_1)^2} \tag{2}$$

Therefore, for high frequencies dissipation depends only on $\tilde{R}_1$, which can thus be fixed, , whereas at low frequencies it depends only on $\tilde{R}_1 + \tilde{R}_2$, so that $\tilde{R}_2$ may be estimated from $\tilde{R}_1 + \tilde{R}_2$ and the low frequency dissipation.

Dissipation in usual mufflers suggests that $\tilde{R}_1$ must depend on frequency. The best strategy is then to fix first the value of R_2 so that the first resonance of the system is properly damped, and then fix R_1 at low frequencies so that

$\tilde{R}_1 + \tilde{R}_2$ gives the desired dissipation. The resulting total resistance at zero frequency is $\tilde{R} \sim KM$, with K the steady loss coefficient and M the Mach number, thus linking pressure losses and acoustics (see Fig. 1(b)).

Summarizing, the steps required are: (i) Define volume availability, and requirements in terms of attenuation and dissipation; (ii) From the attenuation requirements, define the cut-off frequency for the modified volume; (iii) From the desired damping of tailpipe resonances, define the constant value of R_2; and (iv) From the desired dissipation, and recalling the value of R_2, define the frequency dependence for R_1.

3 Discussion of results

In order to check the feasibility of the approach presented, a typical rear muffler was considered. The requirements in terms of attenuation and dissipation were obtained from the experimental characterization of the muffler [2] and its geometrical characteristics. The results obtained are shown in Fig. 2 for the heuristic magnitudes of the muffler alone (transmission loss and dissipation) and in Fig 3. for the parameters representing the performance of the muffler and the tailpipe (noise reduction and resistive impedance). It can be observed that, in all the cases, a suitable approach to the results of this precise muffler is obtained. Nevertheless, it should not be forgotten that here the purpose is not to model the muffler, but to obtain some kind of reference.

Nevertheless, some of the discrepancies observed may be of a certain importance, namely those referred to the attenuation shape of the muffler. The low frequency behaviour in most real cases exhibits a sharp rise in attenuation, which comes usually from some resonance effect.

4 Conclusions

The convenience to dispose of a formal approach to muffler design has been justified, so that one may dispose of a simple model for the identification of the main requirements to be fulfilled by a muffler for a certain application, before entering into the precise geometrical details.

A simple model which gives a first approximation to the problem has been presented and a direct procedure for the identification of the model parameters from the actual requirements has been given. Comparison with

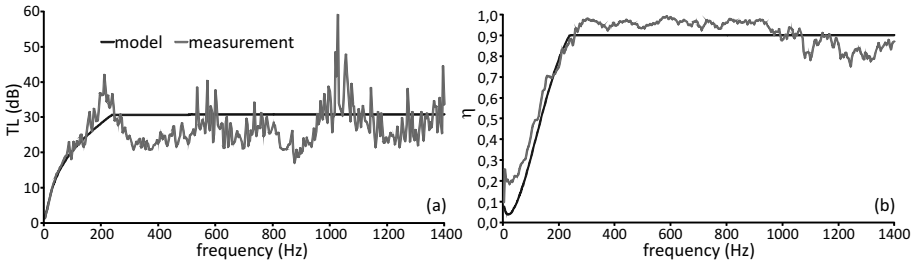


Figure 2: Transmission loss (a), and dissipation (b).

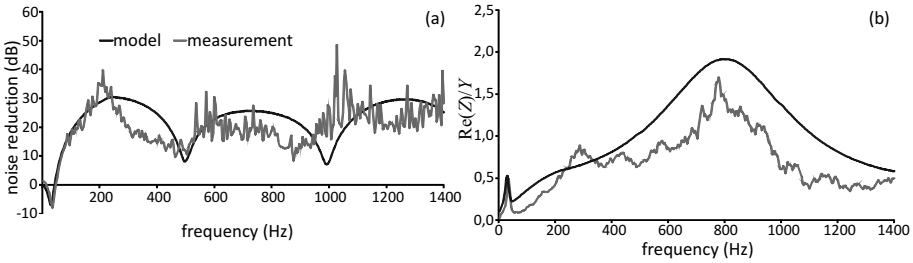


Figure 3: Noise reduction (a), and resistive impedance (b).

experimental results indicate the feasibility of the procedure.

A future developments could be to explore different possibilities to include the resonant effect mentioned above, trying to avoid any unnecessary arbitrariness in the choice of the model parameters.

References

[1] Payri F, Desantes JM and Torregrosa AJ. Acoustic boundary condition for unsteady one-dimensional flow calculations. *Journal of Sound and Vibration*, 188(1): 85–110, 1995.

[2] Payri F, Desantes JM and Broatch A. Modified impulse method for the measurement of the frequency response of acoustic filters to weakly nonlinear transient excitations. *Journal of the Acoustical Society of America*, 107(2): 731–738, 2000.

Modelling Metamaterial Acoustics on Spacetime Manifolds

Michael M. Tung*
Instituto de Matemática Multidisciplinar
Universitat Politècnica de València
Camino de Vera s/n, E-46022 Valencia

November 30, 2012

1 Introduction

Acoustics as the interdisciplinary science that studies the effects and properties of mechanical waves in various media has come a long way from the initial attempts of the ancient Greek philosopher Pythagoras in the 6th century BC to explain the harmonic overtone series on a string to the consolidation of the field of mathematical acoustics in the 19th century mainly pioneered by Helmholtz and Rayleigh.

Recent successes in analogue models of gravity for optics, known as *transformation optics* (see [1] and references therein), have inspired similar frameworks for other physical systems such as acoustic phenomena in acoustic metamaterials [2–4]. Here, *transformation acoustics* allows for deformations of the acoustic propagation space, which mathematically resembles spacetime of general relativity in the presence of gravitating matter or energy—the underlying mathematical structure is described by pseudo-Riemannian manifolds. Consequently, transformation acoustics employs sophisticated differential-geometric tools to design and manufacture advanced acoustic devices.

*e-mail: mtung@imm.upv.es

Acoustic devices are composed of acoustic metamaterials, *i.e.*, materials which are mostly artificially produced with suitable and highly unusual properties that are not found in nature [2]. The technical and industrial applications are innumerous, ranging from the acoustic enhancement of concert halls to the design of submarines undetectable for sonar probes. More interesting details on *acoustic cloaking* may be found in Ref. [4].

2 Covariant acoustic framework

In order to derive an efficient framework for controlling the propagation of acoustic waves in industrial applications, we have postulated a variational principle on a Riemannian manifold for the acoustic scalar potential $\phi(x^\mu)$, where x^μ represents the contravariant spacetime components of an event p on the manifold M endowed with a non-degenerate, smooth, symmetric metric $\mathbf{g} : T_pM \times T_pM \to \mathbb{R}$ of positive signature. For the fundamental description of transformation acoustics (without explicit sources) it then suffices to require that the following functional derivative vanishes [5]:

$$\frac{\delta}{\delta\phi} \int\limits_{\Omega} d^4x \left(\tfrac{1}{2}\sqrt{-g}\, g^{\mu\nu}\phi_{,\mu}\phi_{,\nu} \right) = 0, \tag{1}$$

where as usual the comma notation in the index refers to a partial derivative of the respective spacetime component, and $g^{\mu\nu}$ is the inverse metric tensor in a particular coordinate frame. The invariant volume element $d^4x\sqrt{-g} = dx^0dx^1dx^2dx^3\sqrt{-g}$ contains the determinant $g = \det(g_{\mu\nu})$ and the integration is carried out over the closed and bounded spacetime domain $\Omega \subset M$. The bracket expression just represents the kinetic term of the field in covariant four-dimensional form.

This formalism automatically guarantees the essential condition that the corresponding acoustic wave equation is invariant under certain coordinate transformations. This necessarily leads to the constitutive relations of the acoustic parameters (bulk modulus and mass-density tensor) which link together the *virtual space* (flat space with known wave propagation) and *physical space* (curved, transformed space with the desirable acoustic properties) under consideration [5].

3 Acoustic analogue of gravitational redshift

The variational principle, Eq. (1), implies in combination with the standard differential-geometric methods that for every predefined spacetime metric one obtains via the *Euler-Lagrange equations* the corresponding acoustic wave equation in curved spacetime

$$\Delta_M \phi = \frac{1}{\sqrt{-g}} \left(\sqrt{-g}\, g^{\mu\nu} \phi_{,\mu} \right)_{,\nu} = 0, \tag{2}$$

where $\Delta_M \phi$ is the Laplace-Beltrami operator on the pseudo-Riemannian manifold M. The acoustic wave equation, Eq. (2), fully determines the propagation of the acoustic wave in the prefabricated metamaterial. Hence, every concrete physical spacetime geometry has its acoustic equivalent which may so be made subject to further investigation—in full agreement with the *gravity analogue programme* [6].

Of particular interest is the flat spacetime metric proposed by Desloge [7] to study the gravitational Doppler effect according to which the wavelength of electromagnetic radiation varies in a gravitational field, especially the measurements between the emitter and receiver of the signal. In today's satellite-based global positioning system (GPS) such corrections (in addition to the contributions which arise from special relativity) are essential for a proper functioning.

Suppressing the third space component, the metric which corresponds to a uniformly accelerating rigid frame (UAF) in field-free space is (see also Ref. [8])

$$g_{\mu\nu} = \begin{pmatrix} -\left(1 + g_0 y/c^2\right)^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{3}$$

where $g_0 > 0$ is the proper acceleration in y-direction for an observer located at the origin. Now it is straightforward to fine-tune the parameters of the metamaterial to implement the acoustic UAF model.

A detailed analysis of the corresponding solutions of the wave equation, Eq. (2), shows that the y-dependence of the acoustic potential $\phi(t, x, y) = \phi_0(t)\phi_1(x)\phi_2(y)$ for large y is dictated by the differential equation

$$\phi_2''(y) + \frac{1}{y}\phi_2'(y) - \frac{C}{y^2}\phi_2(y) = 0, \tag{4}$$

where C is an appropriate constant comprising the integration constant $\phi_0''(t)/\phi_0(t) > 0$ and the physical constants g_0 and c. The analytical solutions in this approximation are well-defined in terms of oscillatory trigonometric functions or damped hyperbolic functions, depending on the sign of C. However, as y approaches $y_0 = -g_0/c^2$ the underlying metric, Eq. (3), displays a singularity and corresponds to a physical domain within the reference frame where time dilation becomes more and more relevant. This means that any acoustic wave will be trapped in either one of the semi-finite regions $y < y_0$ and $y > y_0$, where no communication to the outside of each respective region will be possible. In practice, the value and sign of g_0 can be chosen such that this effect may be detected at a reasonably fine-tuned position. The boundary where $y = y_0$ demarcates the region at which the accelerational pull becomes so great as to make the escape of an acoustic signal impossible, that is, it constitutes an acoustic *event horizon*.

Especially noteworthy is that the acoustic UAF model with its linear event horizon corresponds in practice to a wall with a complete sound barrier, independent of the frequency of the incoming sound waves. This salient feature makes the model very attractive for future use in industrial devices.

Acknowledgements

The author wishes to thank for support by the Universidad Politécnica de Valencia under grant PAID-06-11-2020.

References

[1] C. García-Meca and M. M. Tung. The variational principle in transformation optics engineering and some applications, *Math. Comput. Model.*, doi:10.1016/j.mcm.2011.11.035, 2012 (available online, in press).

[2] A. Norris. Acoustic metafluids, *J. Acoust. Soc. Am.*, 125:839–849, 2009.

[3] S. A. Cummer and D. Schurig. One path to acoustic cloaking, *New J. Phys.*, 9(3):45–52, 2007.

[4] H. Y. Chen and C. T. Chan. Acoustic cloaking and transformation acoustics, *J. Phys. D*, 43(11):113001-113014, 2010.

[5] M. M. Tung, A fundamental Lagrangian approach to transformation acoustics and spherical spacetime cloaking, *Europhys. Lett.*, 98:34002–34006, 2012.

[6] M. Visser, C. Barceló und S. Liberati, Analogue models of and for gravity, *Gen. Rel. Grav.*, 34:1719–1734, 2002.

[7] E. A. Desloge, The gravitational red shift in a uniform field, *Am. J. Phys.*, 58(9):856–858, 1989.

[8] L. Acedo and M. M. Tung, Electromagnetic waves in a uniform gravitational field and Planck's postulate, *Eur. J. Phys.*, 33:1073–1082, 2012.

On generalized cooperative systems and the computation of their solution envelopes

Diego de Pereda[†], Sergio Romero-Vivo[‡], Beatriz Ricarte[‡] and Jorge Bondia[†, *]

(†) Institut Universitari d'Automàtica i Informàtica Industrial,

Universitat Politècnica de València, Spain.

(‡) Institut Universitari de Matemàtica Multidisciplinar,

Universitat Politècnica de València, Spain.

November 30, 2012

1 Introduction

Mathematical modelling is widely used to mimic different real life processes. However, models are usually a simplified version of the actual process, producing non-modelled dynamics. Furthermore, a common characteristic of real processes is variability, leading to parametric uncertainty. Under these circumstances, the exact values of the parameters and the initial conditions of the model are unknown, but they can be bounded by intervals. Under known parameters and initial condition, there is a single possible solution for the model. However, as interval uncertainty is considered, a set of different possible solutions is achieved. The aim of this work is to compute solution bounds that guarantee the inclusion of all possible solutions, and minimize the overestimation performed.

Monotonicity approaches allow us to compute a guaranteed solution envelope by taking into account the ordering induced by an orthant [1]. As

*email: jbondia@isa.upv.es

seen in Figure 1, just two simulations have to be performed to compute the solution bounds for monotone states:

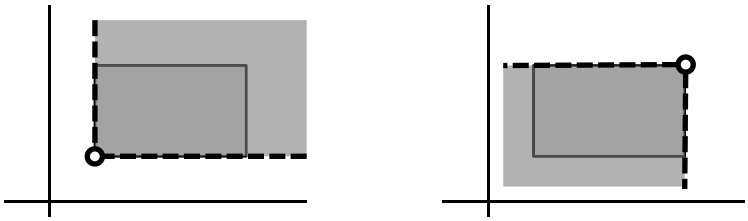


Figure 1: *Possible values for the input space to compute the lower and the upper bounds in the case of orthant monotonicity.*

In order to reduce the overestimation due to the non-monotone states, the input space can be divided into smaller input spaces, shown in Figure 2. Assuming the infimum (supremum) of all the smaller input space simulations, the lower (upper) bound of the solution is computed.

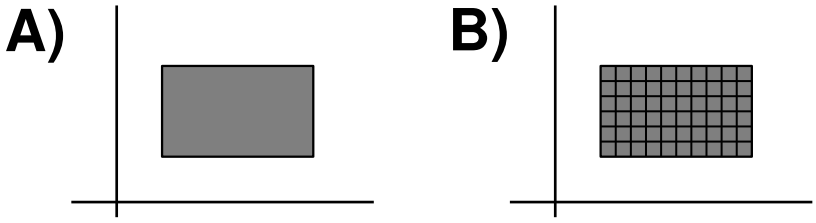


Figure 2: *A division of the non-monotone states of the input space. A) Input space. B) Input space divided.*

2 Generalized cooperative systems

Without orthant monotonicity among the system states, an overestimation occurs when computing a solution envelope. For those systems with non-monotone states or parameters, we propose to analyse the monotonicity induced by a cone P [2], instead of an orthant. This new class of monotone systems with respect to a cone is name generalized cooperative systems.

Now, simulations considering the boundary of the input space are needed to embrace all the possible solutions, as seen in Figure 3.

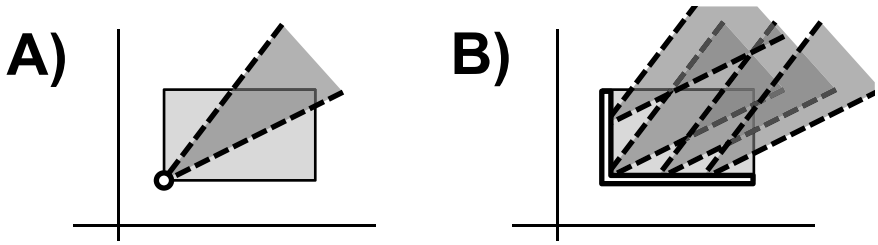


Figure 3: *An example of cone monotonicity. A) A simulation considering a finite number of points does not embrace all the input space. B) Simulations considering the entire input space boundary embrace all the input space.*

Moreover, in order to further reduce the overestimation committed, the boundary of the input space can be divided into smaller input spaces, as seen in Figure 4. The infimum (supremum) of all the smaller input spaces determines the lower (upper) bound of the solution envelope.

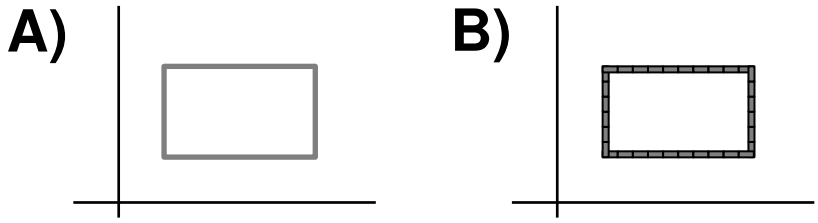


Figure 4: *A division of the input space boundary under cone monotonicity. A) Input space boundary. B) Input space boundary division.*

3 Non-linear example

A chemostat is an important laboratory device used in microbial ecology for the continuous culture of microorganisms [3], given by the following equations:

$$\dot{x}_1(t) = r \cdot (k - x_1) - \beta \cdot x_1 \cdot x_2$$
$$\dot{x}_2(t) = -r \cdot x_2 + x_1 \cdot x_2$$

where $x_1 \geq 0$ denotes the substrate concentration and $x_2 \geq 0$ the microbes concentration. The parameter $r \geq 0$ stands for the elimination rates, while the parameter $k \geq 0$ represents the input rate. Finally, the parameter $\beta \geq 0$ denotes the consumption of the substrate.

It can be proven that the system states are not monotone with respect to an orthant, but there is monotonicity with respect to the ordering induced by the cone P:

$$P = \{x \in \mathbb{R}^2 : -x_1 \cdot (x_1 + \beta \cdot x_2) > 0\}.$$

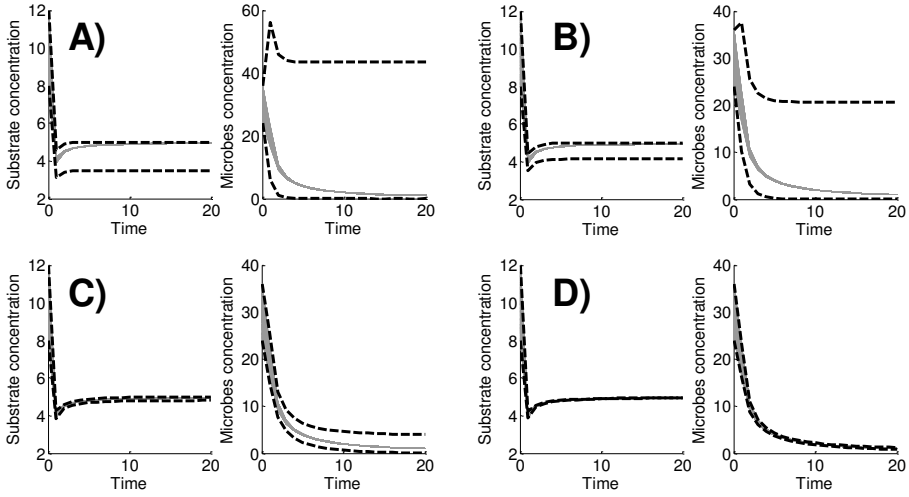


Figure 5: *Computation of solution envelopes for the non-linear chemostat model for parameters* $r = 5$, $k = 5$ *and beta* $= 0.05$, *under 20% uncertainty for the initial conditions. A) Without monotonicity. B) Cone monotonicity. C) Input space division without monotonicity (100 simulations). D) Input space division under cone monotonicity (100 simulations).*

The black dashed lines in Figure 5 display the computed solution bounds for the substrate and the microbe concentrations, and the light grey lines represent several numerical simulations executed for different values of the initial conditions.

4 Discussion and Conclusion

Traditionally, orthant monotonicity has been applied to compute a solution envelope under interval uncertainty. This analysis allows us to compute the solution bounds without overestimation when the system is monotone with respect to the states and parameters of the model. However, if there is at least one state or parameter without monotonicity, it has to be considered as an interval and it will produce an overestimation in the solution envelope computation.

Without orthant monotonicity among the system states, a monotonicity analysis of the model by taking the ordering induced by any arbitrary cone can be performed. If the system is monotone with respect to a cone P then just the boundary of the input space has to be simulated to compute a guaranteed solution envelope, instead of considering all the uncertain input space. As the input space is reduced, the overestimation committed will also be reduced.

Furthermore, the input space can be divided into smaller fragments, as seen in Figure 2 and Figure 4. The overestimation committed is significantly reduced in the computation of a solution envelope for the same computational cost, as seen in Figure 5.

References

[1] Sontag, E.D. Monotone and near-monotone biochemical networks *Systems and Synthetic Biology*, Volume(1):59–87, 2007.

[2] Walcher, S. On cooperative systems with respect to arbitrary orderings *Journal of Mathematical Analysis and Applications*, Volume(263):543–554, 2001.

[3] Smith, H.L. and Waltman, P.E., The theory of the chemostat: dynamics of microbial competition. Cambridge Univ Pr, 1995.

A Multivariate Missing Data Inputation Method Based on Clustering. Application to World Health Organization data.

K. Gibert*

Knowledge Engineering and Machine Learning Group (KEMLG)

& Dep. Statistics and Operations Research

Universitat Politècnica de Catalunya, Barcelona, 08034 (Spain)

L. Salvador-Carulla

Professor of Psychiatry. Department of Neurosciences.

University of Cadiz. Plaza Falla 9 11003 Cadiz (Spain)

Working Group on Clinical Management of the Spanish Society of Psychiatry (GClin-SEP)

J. Morris, S. Saxena

Department of Mental Health and Substance Abuse,

World Health Organization, 20 Ave. Appia, Geneva 1211, Switzerland

November 30, 2012

1 Introduction

In real applications, it is quite usual to find important rates of missing data that have to be preprocessed before the analysis. The literature for missing imputation is abundant [1]. However, the most precise imputation methods require quite a long time, and sometimes specific software. This implies a significant delay to get a complete data matrix suitable for the analysis. In many real applications a listwise deletion or complete case analysis is used as the most simple, and automatic, way to overcome the missing data analysis.

This, of course, can drastically reduce the number of cases to be considered in the analysis, with the corresponding loose of useful information. In the worst cases, this can imply significant biases reducing the representativeness of the analyzed data. As a simple and quick solution, substitution by the variable means is very used in the practice, while

*e-mail: karina.gibert@upc.edu

it is known that this reduces the intrinsic variability of statistics like variability and also diminishes the degree of correlation between variables.

In this work, we propose to introduce clustering in the process of missing imputation, in order to get a good trade-off between precision and required time to prepare data for the analysis. The proposal is applied in the context of better understanding the Mental Health Systems in Low and Middle Income Countries, project leaded by the Mental Health Department of the World Health Organization.

2 The MIMMI method

The *Mixed Intelligent-Multivariate Missing Inputation (MIMMI) method* is presented in this work. It is a non parametric method that uses the conditional mean for imputation according to the underlying structure of the dataset itself, i. e. the means local to the classes discovered by the clustering.

The main idea is to identify, together with the experts, a first subset of quasi-full relevant variables. Since this reduced set of variables contain a small set of missing data, it is possible to perform the imputation of this missing cells manually, by using the background knowledge of the experts. For this purpose, experts use their knowledge on the domain and identify either variables related with the missing one for a given individual or other related individuals with similar values in the variables to be taken as a reference and get a consensus imputation value for the missing.

The experience says that this process can be done in short time because the set of missing cells to be discussed is small and the set of variables considered is also reduced. The imputed data matrix is clustered and an auxiliary partition of data is found. The class identifier is used to find conditional means for all remaining variables in the dataset.

MIMMI is a method combining the prior expert knowledge with multivariate analysis to find imputation values that take into account the joint distribution of all variables and can be completed in a relatively short time, without requiring assumptions on the probabilistic models of the variables (normality, exponentiality, etc).

3 Application

Real applications shown a good performance in both quality of results and required time. In particular, this method has been used with the WHO-AIMS database, a database collected by the WHO [2] about the situation of Mental Health Systems in Low and Middle Income Countries.

The database had a 21,43% of missing data from a total of 256 variables. Applying the MIMMI methodology a set of 8 missing data had to be discussed among the experts to manually propose an imputation value, while the remaining missing data could be substituted by using the local means of a clustering method. Applying MIMMI to the missing data treatment, significantly enshorted the time devoted to data cleaning, which

is assumed to consume a 70% or more of the time of a data mining project [3], and the resulting imputation values where retained sufficiently accurate by the experts.

Acknowledgements

The author gratefully acknowledges Juan Carlos Martín Sánchez for helping with data processing.

References

[1] Schafer, Joseph L.; Graham, John W. (2002): Missing data: Our view of the state of the art. Psychological Methods, Vol 7(2): 147-177.

[2] Saxena S, Lora A, van Emmeren M, Barrett T, Morris J, Saraceno B (2007) WHO's Assessment Instrument for Mental Health Systems: Collecting essential information ofr policy and service delivery. Psychiatric Services 58(6): 816-21

[3] http://www.kdnuggets.com/polls/2003/data_preparation.htm

A comparison of the ARMA-GARCH-M and the Back-propagation Neural Network in estimating returns and conditional volatility: application to the Ibex-35 Spanish stock market index

Fernando García

Francisco Guijarro (*)

Ismael Moya

Javier Oliver

Affiliation: Department of Economic and Social Sciences, Universidad Politécnica de Valencia, Valencia, Spain

Address: Universidad Politécnica de Valencia, Departamento de Economía y Ciencias Sociales. Camí de Vera s/n, 46022 – Valencia, Spain. Corresponding author e-mail: Corresponding author's e-mail: fraguima@upvnet.upv.es (*)

Keywords: conditional volatility; GARCH; neural network; return equation; ibex-35

Subject classification codes: 97M30 68T05

1. Introduction

It could be said that financial market volatility has risen considerably with the passage of time [6], especially in recent years. Examples of this phenomenon include the 1987 stock market collapse, the crisis of the European monetary system in 1992-93, the bursting of the dot-com business bubble in 2000, or the 2008 financial crisis. It was precisely after the 1987 collapse that studies on stock market volatility and its modelling began to multiply and became a fundamental element in risk management.

Quantitative research tries to identify investment opportunities that maximize return while assessing any risks involved by measuring the volatility of returns, an aspect to which investors give great importance. Modelling has thus become a primary

research field [1,2,3], using the capacity of econometric models to estimate the returns on investments, volatility of returns, and the relationship between these two variables [5,7,10].

This paper describes a comparison of one of the econometric models most widely used in risk simulation, the ARMA-GARCH-M, with a model based on artificial intelligence, the Backpropagation neural network. The rest of the paper is laid out as follows: the following section deals with a brief description of the methods used to estimate returns on investments and conditional volatility by econometric models and neural networks. In Section 3 the performance of ARMA-GARCH-M and the Backpropagation neural network are assessed by processing a historical series of the Spanish Ibex-35 closing prices and the results are compared by means of different error statistics. The main conclusions drawn from the work are presented in the final section.

2. Methodology

This section describes the two models whose performance in explaining the returns and conditional volatility is compared: the ARMA-GARCH-M model and the Backpropagation neural network.

2.1 GARCH model

One of the variants of the GARCH models proposed by [1] and the ARCH-M proposed by [4] is the GARCH-M or GARCH-in-Mean. This model proposes incorporating conditional variance into the return equation; in other words, the expected returns will also depend on their conditional variance. The analytical expression of the GARCH-M model is given in the equations below, in which (1) expresses the conditional variance equation and (2) expresses the return equation.

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} = \alpha_0 + A(L)\varepsilon_t^2 + B(L)h_t \qquad (1)$$

$$\mathrm{R_t} = \delta + \gamma \mathrm{h_t} + \mathrm{h_t}^{1/2} \varepsilon_t \qquad (2)$$

with $p \geq 0$, $q > 0$, $\alpha_0 > 0$, $\alpha_i \geq 0$ $(i = 1..q)$, $\beta_j \geq 0$ $(j = 1..p)$ and $\varepsilon_t \approx N(0,1)$.

The influence of conditional volatility on performance can be expressed in different ways: as variance h_t, as the logarithm of variance $\log h_t$ or standard deviation $h_t^{1/2}$. The last option is the one most widely used in empirical studies and appears in expression (2).

The relationship between conditional volatility and trading volume has also been considered. To study this relationship we followed the method used by [8]. These authors consider the GARCH (1,1) model to be the most suitable for modelling the performance of returns, as it usually has a leptokurtic distribution and presents long persistence in the variance of distribution. This model appears in (3):

$$\mathrm{y_t} = \mu + \varepsilon_t$$

$$\varepsilon_t | \phi_{t-1} \sim \mathrm{N}(0, \mathrm{h_t})$$

$$\mathrm{h_t} = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \mathrm{h_{t-1}} + \delta \mathrm{V_t} \qquad (3)$$

where y_t is the return on shares $ln(close_t/close_{t-1})$ and V_t is the volume. The use of the logarithm to calculate performance is due to any non-seasonal influences the series may show.

2.2 The Backpropagation neural network

One of the options for estimating returns and conditional volatility by neural networks is

the Backpropagation variant [9], which uses delta rule-based supervised learning, or *error backpropagation*. In the case of this network, the learning algorithm is generalized so that it can be used with networks of more than two layers. Operations are carried out in two phases. The information initially enters the first-layer neurons and generates an association of input-output data pairs. In the second phase, the information is propagated to the rest of the neurons in the rest of the layers and the different neuron outputs are compared to the desired output, after which the learning error is calculated. The errors from each neuron are then transmitted backwards from the output neurons in order to determine the contribution of each neuron to the total error. With this new information the weights of the neurons are varied until a certain error threshold is reached.

3. A comparative study of the ARMA-GARCH-M and the Backpropagation Neural Network: Application to the Ibex-35 Index.

The series of Ibex-35 daily closing prices chosen for the comparative study ranged from 3 January 2000 until 14 July 2010 and contained a total of 2,658 observations. The series included periods of both rising and falling price trends and high and low volatility.

3.1 ARMA-GARCH-M Model

Before constructing the GARCH-M model, a preliminary step must be carried out to analyse the seasonality of the sample. The augmented Dickey-Fuller test was used to confirm series seasonality, leaving out the constant and trend, which turned out to be non significant. The Schwarz criterion was used to select lags in the returns: ARMA (1,1) model was finally selected, whose terms were all significant.

Besides selecting the returns series-generating model, the possible heteroskedasticity of the residues generated by this model was analysed to detect any possible relationship between the series residues at time t and residues from earlier times: $t-1$, $t-2$, etc. Several coefficients are statistically significant, which confirms the existence of ARCH processes. The use of the GARCH family model is therefore justified for modelling volatility. Since the aim was to analyse both stock yields and index volatility together, the GARCH-M was selected, as it permits the joint analysis of these variables by including conditional variance in the return equation.

The GARCH model estimation was performed considering different delays. Following Najad & Yung (1991), besides conditional volatility, volume was included as an explanatory variable in the model in three different ways: (1) incorporating volume in the variance equation, (2) incorporating lagged volume, (3) including the lagged logarithmic form.

The model was chosen using the Schwarz and Hannan-Quinn criteria. Both criteria select the same model: ARMA(1,1)-GARCH-M(2,1). The values of both criteria are shown in Table 1, modelling the conditional variance equation in its three possible forms: variance (GARCH), logarithm of variance (LN GARCH) and standard deviation (SDEV GARCH). Also considered was the possibility of including the logarithmic form of the lagged volume.

Table 1. Selection of the ARMA(1,1)-GARCH-M(2,1) model.

Return Equation	GARCH		LN GARCH		SDEV GARCH	
Variance Equation	Vol(-1)	LnVol(-1)	Vol(-1)	LnVol(-1)	Vol(-1)	LnVol(-1)
Schwarz Criterion	-5.9188	-5.9185	-5.9211	-5.9205	-5.9204	-5.9175
Hannan-Quinn	-5.9303	-5.9299	-5.9326	-5.9320	-5.9319	-5.9290

Criterion						

According to the results given in Table 1, the model with the best scores for both criteria is the one that expresses the conditional variance equation in the form of standard deviation and includes delayed volume in its logarithmic form (Table 1, last column).

After designing the definitive model, its coefficients were estimated from the 2,658 observations in the sample. Table 2 gives different error statistics for the returns and conditional variance equations: MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error), MSE (Mean Squared Error), AMPE (Absolute Mean Percentage Error) and RMSE (Root Mean Squared Error).

Table 2: Error statistics for the return and conditional variance equations in the ARMA-GARCH-M model.

Error	MAPE	MAE	MSE	AMPE	RMSE
Return Equation	1.0983	0.0108	0.0002	0.9715	0.0154
Conditional Variance Equation	2.0463	0.0002	0.0000	1.8903	0.0003

3.2 Backpropagation Neural Network Model

From the different neural network configurations at present available we chose the Backpropagation for its capacity to adapt neuron weights from the errors made during the learning process.

The inputs established for the network learning process were: index returns with one time lag (t-1), conditional variance with one (t-1) and two time lags (t-2), and

volume with one time lag (t-1). The outputs were returns and conditional variance at time t.

The same variables were chosen for both systems in order to make it possible to compare the performance of the neural network with the econometric model. On one hand, network training indicates possible relationships between returns and their time lag (ARMA (1,1)). Conditional variance with one and two time lags establishes the relationship between returns and conditional variance (GARCH-M). Finally, the relationship between delayed volume and conditional variance can be studied. The relationships between inputs and outputs detected by the neural network are somewhat more complex than those identified by the econometric model, so that better results can be expected from it and consequently fewer errors in predicting Ibex-35 stock returns and conditional variance.

Two hidden layers were built into the network, with a maximum of 256 neurons per layer and multiple inter-neuron connections were permitted. The learning rates established for the two hidden layers were from 0.1 to 0.4 and for the output layer were from 0.1 to 0.2. The moment parameter was fixed between 0.1 and 0.3 for the hidden layers and between 0.1 and 0.2 for the output layer. As all these values are expressed in ranges, the network carries out an optimization process that allows it to select the best-fitting values. The initial weighting of the neuron connections were fixed at ±0.3 by default.

Learning was configured to minimize three error measurements: Average Absolute Error (AAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), so that the results of three different networks could be compared. Tables 3 give the error statistics of the three networks considered. It can be seen that there is little difference between the results of the three networks when estimating the return

equation, while the differences are in general somewhat higher when it comes to estimating conditional volatility.

If these results are compared with those from the econometric model, we again find little difference as to the return equation, but more significant ones in estimating conditional volatility: in this case, the results of the Backpropagation neural network are a considerable improvement on those of the ARMA-GARCH-M econometric model.

Table 3: Error statistics for the return equation in the Backpropagation Neural Network Model.

Equation ->	Return equation			Conditional volatility equation		
Error ->	AAE	MSE	RSME	AAE	MSE	RMSE
MAPE	1.1420	1.0711	1.1489	0.1888	0.1483	0.1754
MAE	0.0108	0.0108	0.0108	0.0000	0.0000	0.0000
MSE	0.0002	0.0002	0.0002	0.0000	0.0000	0.0000
AMPE	0.9915	0.9980	1.0247	0.1046	0.0060	0.0898
RMSE	0.0154	0.0154	0.0154	0.0001	0.0001	0.0001

4. Conclusions

This paper presents a comparison of the performance of the GARCH family of econometric models and neural networks in estimating the returns and conditional variance of the Ibex-35 Spanish Stock Exchange Index.

From a comparison of the results of both models it can be concluded that there are no significant differences in their explanations of the return equation, so that one model cannot be said to be better than the other in this respect. However, significant

differences were found in favour of the neural network for its explanation of conditional variance in each of the three networks estimated with different optimized error criteria. It can therefore be concluded that the Backpropagation neural network is better able to explain index volatility than the ARMA-GARCH-M econometric model.

It should be pointed out that we did not consider the most complex GARCH models, such as the EGARCH, which is capable of giving a more precise explanation of past behaviour and of the interrelationships among variables. In order to validate these results it would therefore be advisable to amplify the study to other time horizons and/or other stock exchange indexes or individual stocks.

Bibliography

T. Bollerslev, *Generalized Autorregressive Conditional Heteroskedasticity*, J. Econometrics 31 (1986), pp 307-327.

T. Bollerslev and H. Mikkelsen, *Modeling and Pricing long memory in Stock Market Volatility*, J. Econometrics 73 (1996), pp 151-184.

M. Deo, K. Srinivasan, and K. Devanadhen, *The empirical relationship between Stock Returns, Trading Volume and Volatility*, Eur. J. Econ. 12 (2008), pp. 58-68.

R.F. Engle, D.M. Lilien, and R.P. Robins, *Estimating Time-Varying Risk Premia in the Term Structure: the ARCH-M Model*, Econometrica 59 (1987), pp 391-407.

M. Ghahramani and A. Thavaneswaran, *A note on GARCH model Identification*, Comput. Mat. App. 55 (2008), pp 2469-2475.

P. Kupiec, *Stock market volatility in OECD countries: recent trends. Consequences for the real economy and proposals for reform*, Econom. Stud. 17 (1991), pp. 31-62.

S. Lundbergh and T. Teräsvirta, *Evaluating GARCH models*, J. Econometrics 110 (2002), pp 417-435.

M. Najand and K. Yung, *A GARCH examination of the relationship between volume and price variability in futures markets*, J. Futures Markets 11 (1991), pp. 613-621.

D. Rumelhart, G. Hinton, and R. Williams, *Learning representations by back-propagating errors*, Nature 323 (1986), pp. 533–536.

A.D. Schepper and M.J. Goovaerts, *The GARCH(1,1)-M model: results of densities of the variance and the mean*, Insur. Math. Econ. 24 (1999), pp 83-94.

Video analysis of the bouncing ball system*

J.L. Hueso[#†], E. Martínez[§], and J. Riera[#‡]

[#]Instituto de Matemática Multidisciplinar,

[§]Instituto de Matemática Pura y Aplicada,

Universitat Politècnica de València.

November 30, 2012

1 Introduction

The bouncing ball dynamical system has been extensively studied since its introduction by Fermi [1]. It consists in a ball, moving under the gravity action, that bounces on a membrane vibrating under the action of a periodic force. The model is easy to realize in practice, allowing the comparison between the theoretical dynamics of the system, the numerical simulation and the experimental results.

In Section 2 we introduce the dynamical system modeling the bouncing ball problem, that will be used for the simulations. In Section 3 we describe the experimental setup and the process of recording the video sequence to obtain the optical flow. Section 4 is devoted to compare the experimental results against the simulation results for different observed behaviors of the bouncing ball system.

*This research was supported by Ministerio de Ciencia y Tecnología MTM2011-28636-C02-02 and by Vicerrectorado de Investigación, Universitat Politècnica de València PAID-06-2010-2285

[†]e-mail: jlhueso@mat.upv.es

[‡]Member of the Equipo de Innovación y Calidad Educativa de la UPV, E-MACAFI.

2 Bouncing ball models

Consider an elastic ball, with coefficient of restitution ϵ, which is kept bouncing off a vertically oscillating base which vibrates sinusoidally as $S(t) = A\sin\omega t$. Between two successive collisions, the ball motion is governed by a gravitational field g. If the ball departs from the membrane at time t_i, the time of the next impact t_{i+1} is the smallest solution $t_{i+1} > t_i$ of the discrete-time dynamics map

$$A(\sin\omega t_{i+1} - \sin\omega t_i) = V_i(t_{i+1} - t_i) - \frac{1}{2}g(t_{i+1} - t_i)^2, \tag{1}$$

where V_i is the post-impact velocity, which relates to the pre-impact U_{i+1} velocity at time t_{i+1} through

$$U_{i+1} = V_i - g(t_{i+1} - t_i). \tag{2}$$

As far as the collision is partially elastic, the ball bounces back instantaneously at t_{i+1} with a relative positive velocity

$$V_{i+1} - \dot{S}(t_{i+1}) = -\epsilon[U_{i+1} - \dot{S}(t_{i+1})], \tag{3}$$

where the relative landing velocity $U_{i+1} - \dot{S}(t_{i+1})$ is always negative. Physically, the coefficient ϵ (defined as the ratio of the relative velocities before and after the collision and sometimes called restitution coefficient) gives a measure through the quantity $(1 - \epsilon^2)$ of the energy lost in the collision. Combining equations (1-3) one can simulate the bouncing ball system.

The state variables are the time t_i and the post impact velocity V_i. We will use the amplitude of the vibration, A, as control parameter, keeping fixed the frequency, ω, and the restitution parameter, ϵ.

Our objective is to compare the theoretical model against the measurements obtained from the experimental setup described in the next Section. The main practical difficulty is to precisely detect the times when the ball touches the membrane. In some devices found in the literature, this time is obtained placing a piezoelectric film on the vibrating surface or attaching thin and light metallic wires to the ball an to a Nickel sheet deposited in the base. We have tried a less invasive setup where video image processing is the main tool to obtain the data of the motion of the ball.

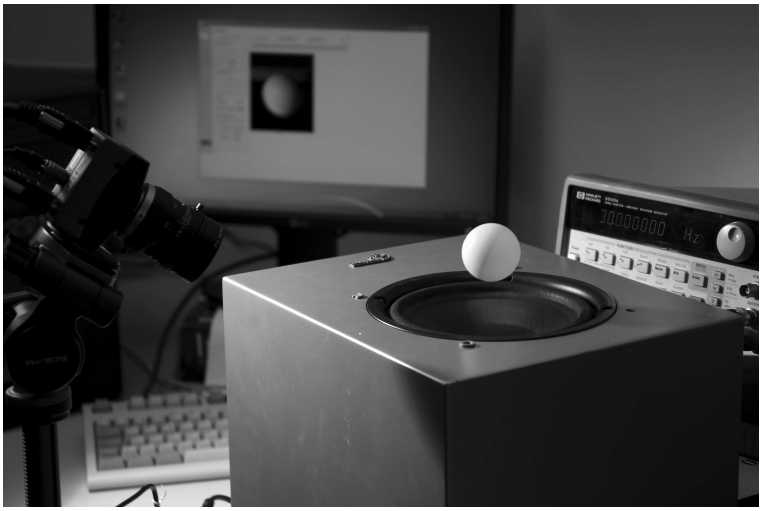


Figure 1: Experimental setup

3 Experimental setup

The bouncing ball model consists in a particle bouncing on a vibrating surface. In our experiment, the particle is a table tennis ball that bounces on the membrane of a loudspeaker. The restitution coefficient of the ball has been estimated as 0.75. The loudspeaker is driven by a sinusoidal voltage with controlled amplitude. This signal is provided by a function generator and amplified by an audio amplifier to excite the vibration (see Figure 1). The vibrating frequency has been fixed to 30 Hz, using the amplitude as control parameter.

We have used a digital camera link Mikrotron Eo Sens-MC1362, that captures up to 1600 frames per second (fps). In the experiments we have registered videos at 430 fps, with frames of 256×380 pixels. The recorded images were sent to a computer running MS Windows 7, throughout a x64 xCelera-CLFull board, inserted in a PCI Expressx4 slot of the computer. The camera has been calibrated using a camera calibration toolbox for MATLAB [2].

The analysis of the images was performed using algorithms implemented in MATLAB R2011b.

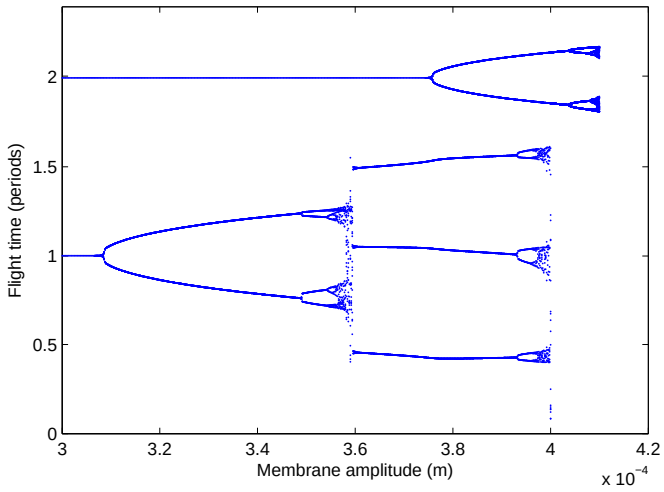


Figure 2: Bifurcation diagram of the bouncing ball

4 Simulation and experimental results

In order to compare the simulation results with the experimental ones, we run the simulation starting from an initial position until the system becomes periodic. Then repeat the simulation with a slight change in the amplitude of the driving force. The behavior of the system can be described by the successive flight times in the periodic state (see Figure 2). Time is measured in periods of the driving vibration, to ease the interpretation of the behavior of the ball. The lower path in the figure corresponds to the evolution of the state where the ball bounces regularly at the same frequency as the membrane. In the upper path the jumps span two periods of the membrane.

With our experimental setup, we have recorded the ball movement with the loudspeaker fed with different voltages and the ball bouncing in different modes. The trajectories obtained by video analysis are matched with the simulation, comparing the time flight in both cases.

The state of the system is analyzed by applying the FFT to the tracked ball position. In most cases, there is a good agreement between the experimental results and the simulation.

For example, in Figure 2 , at amplitudes about 3.9×10^{-4}m, there are two different states that have been recorded in our experiments.

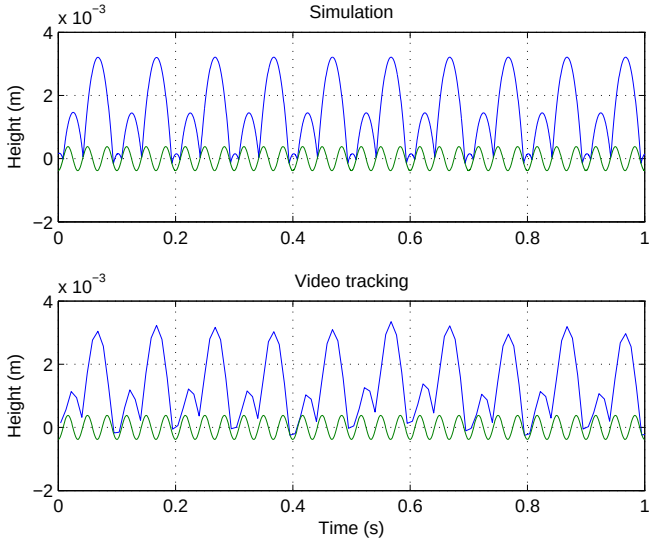


Figure 3: Three different bounces at 2.25 V

In one state, the ball makes three different bounces in three driving periods. The smallest bounce is difficult to distinguish in the tracked trajectory in Figure 3, but the spectrum clearly shows the period triplication (see Fig. 4). This state corresponds to the three branches region in Figure 2.

In other state, the period is quadruple and the ball makes two different bounces spanning two periods each. In this case, we have tracked both the ball an the membrane motions. Figure 5 shows the trajectories of a point in the ball and another in the membrane. The distance between them has been adjusted in order to make the figure more clear, but the amplitude of each motion has not been scaled. This state corresponds to the bifurcation observed in the upper branch on Figure 2.

References

[1] E. Fermi, On the origin of the cosmic radiation, *Phisical Review* 75:1169–1174, 1949.

[2] J. Bouguet, Camera calibration toolbox for matlab, *http://www.vision.caltech.edu/bouguetj/index.html*, April, 2012.

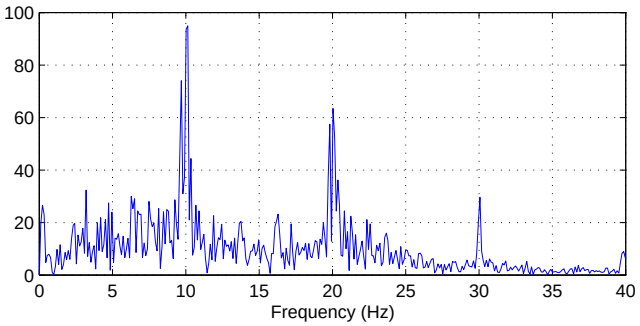


Figure 4: Triple period at 2.25V

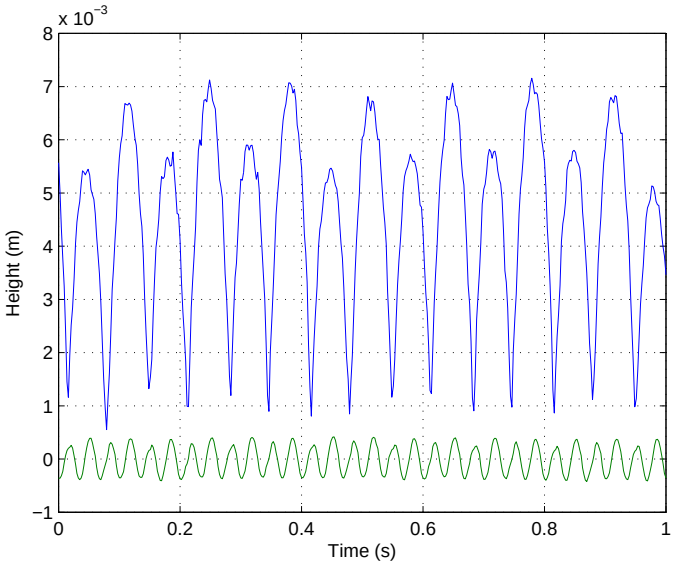


Figure 5: Tracked trajectories of ball and membrane at 2.25 V

Analysis of the multipactor effect in a parallel plate waveguide with multiple modulations *

M. Belloch[(a)], B. Gimeno[(b)], V. E. Boria[(a)],
J. L. Hueso[(c)], E. Martínez[(d)†]

[(a)]Instituto Universitario de Telecomunicación y Aplicaciones Multimedia

[(b)]Institut Universitari de Ciència dels Materials

[(c)]Instituto Universitario de Matemática Multidisciplinar

[(d)]Instituto Universitario de Matemática Pura y Aplicada

[(a,c,d)]Universitat Politècnica de València

[(b)]Universitat de València

Spain

November 30, 2012

1 Introduction

Multipactor is an RF breakdown that may occur in the high power microwave devices working under the vacuum condition. A particular scenario where the multipactor appears is inside a parallel plate waveguide. Between these two plates, it exists an electric field with an electric potential difference which produces the electron movement. Applying the Newton equations to this environment, it is possible to find out the position and the speed of the particle at any time.

*This research was supported by Ministerio de Ciencia y Tecnología MTM2011-28636-C02-02 and by Vicerrectorado de Investigación, Universitat Politècnica de València PAID-06-2010-2285

†Corresponding author, eumarti@mat.upv.es

$$\overrightarrow{F} = m \cdot \vec{a} = -e \cdot \overrightarrow{E_{RF}}. \quad (1)$$

$$\overrightarrow{E_{RF}} = \frac{V}{d} \cdot \cos(w \cdot t + \theta) \quad (2)$$

When an electric field is applied to the electron inside this structure, this electron tends to impact with the plate. One of the objectives is to find out the speed that the electron achieves when the impact is produced. Applying this speed value to the kinetic energy equation, it is possible to model the electron impact in the plate. The consequence of the crash is the appearance of new electrons. Some of them will impact with the opposite plate and so on [**?**, **?**]. Simulating these impacts, we determine when the multipactor effect appears.

The differential equation (1) can be numerically solved using Runge-Kutta's method. The key idea is to determine the impact time for the electron subject to the field given by equation (2). However, when multiple modulations are applied, the difficulty of resolving the equation increases greatly. The purpose of this study is to analyse if multipactor effects appears in this particular environment under different field conditions (multiple carriers and different modulations).

2 Secondary emision model

Secondary emission is a phenomenon where additional electrons, called secondary electrons, are emitted from the surface of a material when an incident particle (often, a charged particle such as electron or ion) impacts the material with sufficient energy. In this case, the number of secondary electrons emitted per incident particle is called Secondary Emission Yield (SEY).

The electric field can eventually drive an electron to the waveguide walls and when this happens, this electron can be absorbed, reflected or can extract secondary electrons from the surface. It is well known that the multipactor effect characteristics strongly depend on the surface properties. This is basically material dependent and is quantitatively considered by means of the secondary electron emission coefficient (SEEC).

At each electron impact the average number of electrons generated is given by the SEEC value for the energy and angle of the colliding electron. A well-known approach to model multipactor is to use effective electrons in

such a way that the real number of electrons to be tracked remains constant throughout the calculation but each electron represents a larger (or smaller) number of electrons depending on the accumulated SEEC values at each impact.

3 Multipactor model

In this system, it is considered infinite parallel plates in the xy plane, with the RF electric field unidirectional in the z coordinate. Border effects are not considered. The approximation corresponds to a narrow gap (relative to the other dimensions). Electrons are modelled individually, assuming that their trajectories are only modified by the electric field and are not affected by other electrons in the system. That is, effects due to space charge are not taken into account. The collision of the electron with a plate can rip zero or more electrons from the wall following a probabilistic model that is described in next section. The newly created electrons are again individually tracked. Given the parameters of a certain material, the SEY depends only on the primary energy, E_p and its incidence angle. Following the energy conservation principle, the total output kinetic energy should be equal or less than the input electron kinetic energy. For the initial conditions, a certain number of free electrons (created during the first period of the electric field) are assigned a normally distributed energy and start at plate $x = 0$.

The simulations show the difference in the trajectory of the electron when the initial phase is different. The higher a is, the longer is time the electron takes to interact with the electric field. The amplitude value (potential generated between plates) of the electric field is different in every simulation and therefore each simulation results are also different. In the first case multipactor does not occur because applying a potential of 30 Volts the electron does not impact the wall with enough power to cause the appearance of new electrons. However, if the electric field has an amplitude about 100 Volts as in simulation 2, this time the electron impacts with enough energy to generate new electrons producing a multipactor discharge.

4 Multipactor effect in modulated signals

Power is a critical issue in telecommunications and narrow-band communication systems require the use of tightly band limited signalling formats. This analysis has a major importance in case of using high amplitude signals due to the ever-growing demand in high spectral efficiency telecommunications systems implying multi-dimensional waveforms considerations (in frequency, time, space, etc.) where parameters like PAPR (Peak Power to Average) are of major concern. Traditional modulation and coding schemes have been designed from the standpoint of minimizing average power but it is also important to look into modulation formats to minimize peak power and retain high spectral efficiency. In this section, the focus is done in single carrier modulations. The main objective of the analysis is to understand the effect of modulated signals in the RF breakdown values. Also it is presented an analysis of different performance parameters such as EVM (Error Vector Magnitude), Bandwidth Efficiency and BER (Bit Error Rate) of modulated signals in transmission systems, which are considered to be useful system metrics for any digital communication system. The purpose consists in the study of the variation in the multipactor threshold for a set of digitally modulated signals and the Multipactor simulator tool is improved to validate a set of different pass-band modulated signals. At the present, the study has been focused mainly in BPSK and QPSK modulation schemes. The reason is that the most links used in satellite telecommunication are employing modulation by means of quadrature phase-shift keying (QPSK). However, to analyse multipactor initiation in QPSK-modulated signals, firstly a digital phase modulation study is included making a review of the fundamentals of binary phase shift keying (BPSK) which is the simplest form of digital phase modulation. The main characteristic of this type of modulation is that the signal envelope stays constant, but the phase switches with regular intervals. It is an important issue to consider that any jump in the RF phase can be treated as a considerable perturbation of the multipactor resonance. The aim of this section is to study the influence of such phase switches on the initiation and dynamics of the multipactor discharge. Finally, one of the purposes is to predict important parameters as the Symbol Period, T_s, we must use to prevent and avoid the growth of the electron avalanche.

We generally regard PSK as a form of constant-envelope modulation, since we are modulating the phase instead of the amplitude of the carrier signal. Phase modulation of a sinusoidal carrier is equivalent to amplitude

modulation of a quadrature carrier. This leads us to wonder whether or not PSK is truly constant-envelope modulation. The answer can be found in the transition diagram. The transition diagram shows the signal in complex signal space, plotted over some period of time. At any instant, the signal amplitude is simply the distance from the origin to a specific point on the transition diagram. In order to have a constant envelope, the signal must always be equally distant from the origin. For unfiltered QPSK, the symbols are simply rectangular pulses, and the transitions between symbols are instantaneous. In this case, the transition diagram is a square. Although it has straight lines between all four symbols, these transitions occur in zero time. Therefore, the signal must be at one of the four corners of the square at all times, resulting in a constant envelope signal. However, when we apply a pulse shaping filter to the symbols, the envelope is no longer constant.

The signal envelope is important on channels which suffer from amplitude distortion, especially channels which are hard limited. If a signal does not have a nearly- constant envelope, it will be severely distorted on such a channel. This can result in bandwidth expansion, intersymbol interference, and quadrature channel crosstalk. If the distortion is severe, it may not be possible for the receiver to recover the modulation.

In order to study what is the minimun symbol period trying to avoid the multipactor discharge, we can choose how many RF periods must take one symbol. For each simulation we can determinate this value, that will be an integer number of RF periods. It is an intuitive way to check how many RF periods must take one symbol to produce multipactor. The dependence of the multipactor behaviour on the moment of the phase switches is being studied by varying the phase of the RF electric field when the phase jump was applied.

Another important factor to take into account is the sequence choice to modulate the signal. In next examples is used a general sequence composed by $[1, 0, 1, 0, 1, 0, 1, \ldots]$. It is easy to understand that this is the worst sequence case because every symbol period the phase suffers an abrupt change from one phase to another (0 to π and vice versa, evidently considering a BPSK modulated signal).

5 Conclusions

To compare the results between the modulated signal and the non-modulated the total number of impacts is a parameter to take into account. In most cases the final number of impacts are quite similar except in the case of applying an amplitude of 30 Volts. For this particular case to apply a BPSK modulation reduces the number of impacts on the device walls. This is probably due to phase change affects more the trajectory of the electron. Also, a curious aspect to consider is that independently the rest of conditions, if $T_s = T$, multipactor discharge does not occur. In most cases, this fast phase change affects the electron changing its trajectory and therefore it does not time to impact against the plates with sufficient energy to generate new electrons. The main issue is that in real cases there are significant limitations to implement such short symbol periods.

According to the results, when the ratio between the phase-switching period and the RF signal period is smaller than then the interval between successive phase switches was enough to change the minimum voltage to initiate the discharge, therefore the multipactor threshold oscillates due to the fast phase switches. The trajectory followed by the electrons is not the same because the signal that excites their movements between the plates is changing due to the different modulation parameters. In this work, one of the main conclusions is that the type of filter applied (rectpulse or RRC) to a BPSK-modulated signals is not a determining factor in the multipactor threshold. In space telecommunication applications, the ratio of the phase-switching interval to the RF period is in the range of several hundreds, which means that the phase switching interval is large enough to essentially correspond to the situation of a monochromatic RF field.

References

[1] R. Vaughan, Secondary emission formulas, *Electron Devices, IEEE Transactions on*, 40(4): 830, 1993.

[2] J. R. M. Vaughan, Multipactor, *IEEE Trans. Electron Devices*, 35(7):1172-1188, 1988.