INVESTIGACIÓN
e INNOVACIÓN
en INGENERÍAS
ISSN2344-8652

Modeling the Behavior of COVID-19-related Deaths using Two Mathematical Epidemiological Approaches

Dos enfoques matemáticos epidemiológicos para modelar el comportamiento de los decesos causados por el COVID-19

Jeinny Peralta Polo **Miguel Caro Candezano**

Leonardo Niebles Núñez
Universidad del Atlántico, Colombia

Recibido: 01/06/2020
Aceptado: 29/06/2020
Publicado: 28/07/2020

Correspondencia de autores:
jeinny.peralta@unad.edu.co

Abstract

Objective: This study compares two deterministic mathematical epidemiological models from the literature to simulate the curve of COVID-19-related deaths in the Colombian department of Atlántico. **Methodology:** The Tappe model was first proposed to simulate COVID-19-related deaths. This approach was based on the exponential behavior of the COVID-19 death curve, using the number of deaths reported in China. The second model proposed was a susceptible–infectious–recovered–dead (SIRD) model, an extension of the susceptible–infectious–recovered (SIR) model, which sorts the population into susceptible, infected, dead, and recovered. **Results:** The specified date ranges revealed that both models describe a relatively similar behavior pattern to the data provided by the National Colombian Institute of Health (INS). However, the SIRD model estimates higher curve growth than the Tappe model when performing a 90-day projection, thirty days as of the final assessment date (May 26). This is probably because the SIRD model used a larger number of variables. **Conclusions:** SIRD projections are more reliable than those generated by the Tappe model because SIRD is a complete model with a larger number of representative population variables. The authors will use the susceptible–exposed–infected–recovered–deceased (SEIRD) model in future studies to forecast the number of possible deaths in the Colombian department of Atlántico.

Keywords: COVID-19, coronavirus, epidemiological models, SIRD model.

Resumen

Objetivo: Comparar dos modelos epidemiológicos matemáticos determinísticos de la literatura, para simular la curva de decesos en el departamento del Atlántico causados por el COVID-19. **Metodología:** El primer modelo propuesto para simular el número de decesos por el COVID-19 es el de Tappe. Este enfoque se basa en el comportamiento exponencial de la curva del número de decesos, e inicialmente fue usado por el autor con los datos disponibles del número de muertos en China. El otro modelo propuesto es el SIRD, una extensión del modelo SIR, que divide la población entre susceptibles, infectados, muertos y recuperados. **Resultados:** Los resultados obtenidos a partir de los dos modelos, en las fechas estipuladas, mostraron que, comparados con los datos tomados del INS, ambos describen un comportamiento relativamente similar. Sin embargo, al analizar una proyección realizada para noventa días, treinta días después de la fecha final de análisis (26 de mayo), se observa que el modelo SIRD describe una curva de crecimiento mayor que la del modelo de Tappe, esto se debe, probablemente, a la inserción de más variables en el modelo. **Conclusiones:** Al ser SIRD un modelo más completo, con mayor número de variables representativas de la población, la proyección realizada con éste es más confiable que la realizada con el modelo de Tappe. Para estudios futuros se pretende incorporar la población de los expuestos para describir el número de decesos, en un modelo SEIRD, en este departamento.

Palabras clave: COVID-19, coronavirus, modelos epidemiológicos, modelo SIRD.

Como citar (IEEE): J. Peralta-Polo., M. Caro-Candezano., y L. Caro-Candezano. “Dos enfoques matemáticos epidemiológicos para modelar el comportamiento de los decesos causados por el COVID-19”. *Investigación e Innovación en Ingenierías*, vol. 8, n°2, 75-86, 2020. https://doi.org/10.17081/invinno.8.2.4257

Artículos resultados del evento XV Jornadas iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento - JIISIC 2020

Introduction

The coronavirus (COVID-19) is a virus that causes respiratory infections among humans. This virus spreads mostly through coughing, sneezing, and talking. That is when infected people's saliva droplets spread into the environment [1]. Several scientific debates have focused on whether the virus can be transmitted by touching objects previously contaminated by infected people and whether asymptomatic infected people may serve as transmission vectors. In December 2019, the first COVID-19 case was reported in Wuhan, China. It swiftly spread throughout the country and other parts of the world, including Italy and Spain, before reaching the United States. See [2]. In March 2020, the World Health Organization (WHO) declared COVID-19 a pandemic.

Government entities have proposed strategies to contain the virus, including lockdown, social distancing, and social health education. As time progressed, more severe restriction measures were imposed, but they were later partially lifted to allow the population to return to work and guarantee sustenance. Some studies, such as [3], have already highlighted the importance of social distancing in reducing transmission rates in this context.

Hence, to help foster decision-making by government health agencies that must enact actions to mitigate the effects of the pandemic on the population, mathematical models in epidemiology have been implemented. These models play a significant role because they naturally predict future disease behavior while considering their corresponding margins of error and simplifications.

Throughout these first five months of the year , countless studies have proposed several mathematical epidemiological models for COVID-19. These models use different mathematical approaches, including deterministic, probabilistic, machine learning, artificial intelligence, and big data elements. Some of thesestudies are [4, 5, 6, 7, 8, 9]. For example, [10] discussed the importance and rigor of mathematical models in decision-making to reduce deaths and infections during a pandemic. The authors of [11] examined mathematical models applied to COVID-19. In [12], the authors evaluated and compared the seasonal dynamics of the COVID-19 pandemic in China, Italy, and France. Among other studies, [13] used a combination of Bayesian methods and a dynamic meta-population model of the pandemic.

Similarly, different studies have been found in the scientific literature describing the epidemiological behavior of COVID-19 in Colombia. For example, [14, 15]. In both studies, the susceptible–infectious–recovered (SIR) method by Kermack and Mackendrick was implemented [16] to provide a forecast model of the pandemic in Colombia.

This study compares two epidemiological approaches based on deterministic mathematical models that describe curve behavior for the number of COVID-19-related deaths in the Colombian department of Atlántico. Data from the National Colombian Health Institute (INS) on May 26, 2020, for the city of Barranquilla and the entire department of Atlántico [17] were used. The datasets for the results were collected as of March 27, 2020—the date on which the first COVID-19-related death was reported in the department of Atlántico. This study adds to the analysis of the COVID-19 pandemic and its development, particularly in the Colombian department of Atlántico. MATLAB mathematics software was used to perform all numerical calculations in this study.

Methodology

Mathematical Model Description

The first model used in this study was an approach derived from the empirical values proposed by Tappe [18]. This model is easy to implement and does not involve complex mathematical elements. Instead, it simply uses elements typical of a high school course, making it easy to understand by any average undergraduate Engineering or Basic Science student.

In [18], the number of deaths is defined as follows:

$$D(t) = \exp\big(L(t)\big), \ [t_1, \infty). \tag{1}$$

Where $L(t)\colon\ [t_1, \infty) \rightarrow \mathbb{R}^+$ is a piecewise function, as defined in (2).

Here, $T_1 \in (t_1, \infty)$ is the number of days after which the effects of the government's measures implemented can be observed. This study adopted a T1 = 17 value, just like in [18]. Tappe's proposal consists of calculating L(t) values considering that, as of the T1 day, the growth curve for the number of deaths changes from convex to concave, as shown in Figure 1. The foregoing reasoning is valid since the death curve slows down as the government's restriction measures become effective.

A dataset of reported deaths *(d_1, d_2, ..., d_n)* or each time value *(t_1, t_2, ..., t_n)*was used to construct the model. *Additionally,* l_i was considered as $l_i = \ln(d_i)$, for $i = 1, \ldots, n$, *and* $t_n \leq T_1$. Taking *n = 14 in this study*, the *L(t)* function is defined as

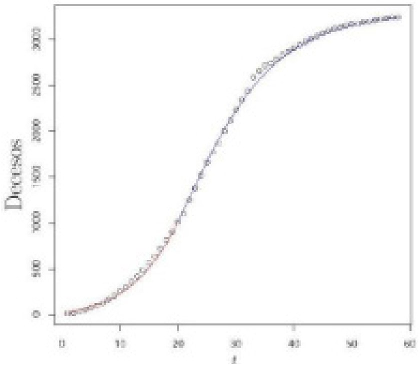
$$L(t) = \begin{cases} l_1 + (l_n - l_1)\left(\frac{t-t_1}{t_n-t_1}\right)^{\beta}, & \text{si } t \in (t_1, T_1]; \\ L(T_1) + \lambda\big(1 - \exp\big(-\nu(t - T_1)\big)\big), & \text{si } t \in (T_1, \infty), \end{cases} \tag{2}$$

With,

$$\nu = \frac{\beta}{\lambda} \cdot \frac{l_n - l_1}{t_n - t_1}\left(\frac{T_1 - t_1}{t_n - t_1}\right) \tag{3}$$

The β ∈ (0,1] and λ > 0 parameters determine the concavity of L(t) and the rigorousness of the measures adopted by the government, respectively. Therefore, lower values of λ indicate that the measures are more rigorous, while higher values indicate greater relaxation of government measures. According to calculations, D(t) will become concave if λ is greater than 1.

Figure 1. Number of Deaths in China from January 22 to March 19, 2020, calculated using the D(t) function



Source: Adapted from [18]

In this study, β and λ parameters were calibrated using the least squares method with positive restrictions [19]. Given the F(D, p) function, which calculates the numerical solutions from the D(t) equation given by (1), with p = (β, λ), and the mortality data provided by [17] for the city of Barranquilla and the Colombian department of Atlántico from March 27, 2020, the date when the first COVID-19-related death was reported through May 26, 2020, these parameters were calibrated by solving the following minimization problem.

$$\min_{p} \frac{1}{2} \|F(D,p) - y\|_2^2 \qquad (4)$$

$$0 < \beta \leq 1$$
$$\lambda > 0$$

The second mathematical model used was susceptible–infectious–recovered–dead (SIRD), an extension of the SIR model developed by Kermack and Mackendrick [16]. Here, the population is divided into three groups: susceptible, infected, and recovered. This extension adds one more category: dead. The model is described by a system of ordinary differential equations, given by the following:

$$\frac{dS(t)}{dt} = -\frac{\alpha}{N}SI, \frac{dI(t)}{dt} = \frac{\alpha}{N}SI - (\gamma_R + \gamma_D)I, \qquad (5)$$
$$\frac{dR(t)}{dt} = \gamma_R I,$$
$$\frac{dD(t)}{dt} = \gamma_D I,$$

where α, γR, and γD are the parameters that determine the infection, recovery, and death rates, respectively. In this model, the total population, which is N = 2,517,897 for the department of Atlántico [17], must meet the N = S + I + R + D condition.

Similar to the first model, the α, γR, and γD parameters were calibrated by solving a nonlinear least squares problem. The SIRD model, for example, defines the u(t) = (S(t), I(t), R(t), D(t)) function, which is determined by the q = (α, γ_R, γ_D) parameters and the y matrix. In this matrix, each row corresponds to the susceptible, infected, recovered, and dead data available for the t day, in which t runs from March 27, 2020, to May 26, 2020, as in the first model. Hence, given the F(u, q) function, which calculates the numerical solution u for the differential equation system (5), the parameters are calibrated by solving the minimization problem.

$$\min_{q} \frac{1}{2}\|F(u,q)-y\|_2^2 \tag{6}$$

$$q \geq 0$$

For both models, the optimization problem was solved using the trust-region method, which is used in the lsqnonlin function in MATLAB. Please refer to [19, 20] for further details on this method.

Results

For the numerical solution of both models, the parameters were calibrated from the data provided by [17] from March 27, 2020, to May 26, 2020. The least squares method was used to solve (4) and (6) for the Tappe and the SIRD models, respectively. The calibration yielded the parameters presented in Table 1.

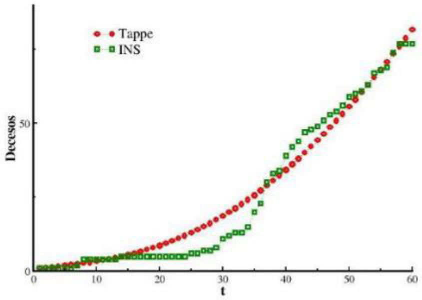
TABLE 1. CALIBRATED PARAMETERS FOR THE TAPPE AND SIRD MODELS

Tappe		SIRD		
β	λ	α	γR	γD
0.7778	3.997	0.154	0.083	0.027

Source: Prepared by the Authors

The results for the Tappe model were compared to the data provided by the National Colombian Health Institute (INS) [17]. Note that the number of days in which deaths occurred in the Atlantic Department used to implement Tappe's methodology was n = 14 because values of n greater than 14 to 17 do not show significant differences, and taking the lowest value is more convenient for calibration purposes.

Figure 2. Number of Deaths in the Department of Atlántico from March 27 to May 26, obtained using the Tappe Strategy with Parameter Calibration and Compared to Actual Data from Instituto Nacional de Salud (INS).



Source: Prepared by the Authors

Figure 2 shows that the curve approximated using the Tappe model is consistent and very close to the actual values. In fact, a concavity change in the curve when T1 = 17 or April 12 was expected. However, this change is not clearly evident.

Concavity changes at several points in the curve from INS data were observed due to the dynamic behavior of virus spread rates. This dynamic behavior, combined with the fact that some measures were either not strict enough or were not rigorously followed by the population, may have a significant impact. For this reason, a clear concavity change could not be observed in the Tappe curve when $T_1 = 17$. However, it must be considered that, in the department of Atlántico and in general in Colombia at a national level, the strictest measures were imposed on March 25, when no deaths had yet been reported. Furthermore, these measures were relaxed after the first lockdown, which may explain why the death curve continued to grow, especially after April 29, when $t = 34$ (17 days after the strict measures had been partially lifted).

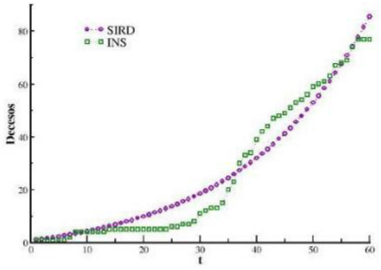
Figure 3 shows the results from the SIRD model, with the parameters presented in Table 1. When comparing these results to the data provided by INS, a consistent growth behavior in the curve was observed.

The SIRD model is more complete when expressed as a system of differential equations because it consists of four equations with four time-dependent unknowns that describe the behavior of the number of susceptible, actively infected, recovered, and dead patients. These equations are also configured based on infection, recovery, and mortality rates, which implies the calibration of more parameters. However, the Tappe model is a single equation with a single variable and two parameters that describe the death curve based on the measures adopted by government entities.

Figure 4 compares the two models to the data provided by INS. According to this figure, both models provide an adequate approximation. In fact, the curves are very close to each other.

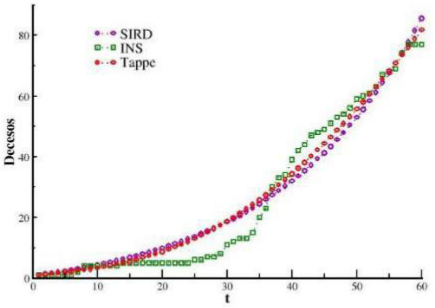
Figure 5 shows a 30-day projection after the last approximation is performed, where the SIRD death curve has a faster exponential growth than the Tappe curve. However, the SIRD model projection is based on the behavior exhibited by the number of infected and recovered patients in addition to deaths, whereas the Tappe model projection is based only on the behavior exhibited by the number of deaths. Hence, this model adds more input to obtain more reliable results. According to the dataset from June 14, 2020, $t = 80$, as provided by INS [11], the number of deaths in Barranquilla and the department of Atlántico reached 397 (in subsequent data, this value rose to 440). This number exceeds the projections obtained using both models and is closer to the projections obtained from the SIRD model.

Figure 3. Number of Deaths in the Department of Atlántico from March 27 to May 26, obtained using the SIRD Model with Parameter Calibration and Compared to Actual Data from INS.



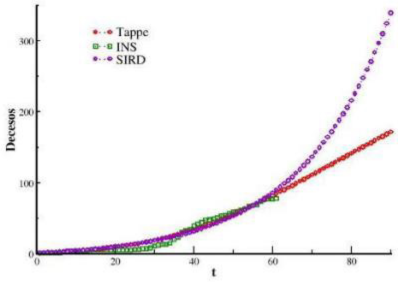
Source: Prepared by the Authors

Figure 4. Comparison between the Number of Deaths Projected for the Department of Atlántico from March 27 to May 26 using the Tappe and SIRD models and the Actual Data from INS.



Source: Prepared by the Authors

Figure 5. Number of Deaths Projection for the Following 30 days after May 26.



Source: Prepared by the Authors

A concavity change in the death curve, i.e., a decrease in the growth rate of the number of deaths, mostly depends on maintaining social distancing measures, introducing more rigorous quarantine measures, and, most importantly, on the degree of compliance from the community with these measures.

Table 2 presents the projected death values from both the Tappe and the SIRD methods for the following 30 days after May 26, as well as actual data for these dates as provided by INS in its July 22 report. Here, as the number of days increases, the projected values are moved further away, which clearly reflects a change in the parameters obtained and acceleration in virus spread rates.

Conclusions

This study evaluated the number of deaths in the department of Atlántico using two different mathematical epidemiological models: Tappe and SIRD, which is an extension of the SIR model. These two models were simulated between March 27 and May 26. The results revealed that, compared to the actual data obtained by INS, both models describe similar behavior on these dates. However, the SIRD model estimates higher curve growth than the Tappe model when conducting a 90-day projection, thirty days as of the final assessment date (May 26). The SIRD model introduces more information into the model, i.e., three unknowns: susceptible, infected, and recovered patients, which influence the development of the death curve. Therefore, the SIRD model better describes the behavior of the pandemic. It could be thought that to project the number of

COVID-related deaths, Tappe's empirical model, whose mathematical simplicity is well-known, can be implemented, thus avoiding the complexity of solving differential equation systems. However, note that this model does not consider the behavior of the rest of the population, which influences the development of the pandemic.

Table 2. Projected Numbers of Deaths from the Tappe and SIRD Model, and the Data Provided by INS

Date	Tappe	SIRD	INS	Date	Tappe	SIRD	INS	Date	Tappe	SIRD	INS
27/05	87	94	83	06/06	117	149	231	16/06	147	236	508
28/05	90	98	92	07/06	120	156	255	17/06	150	247	541
29/05	93	103	98	08/06	123	163	274	18/06	153	259	580
30/05	96	108	111	09/06	126	171	302	19/06	156	271	611
31/05	99	113	120	10/06	129	180	324	20/06	159	283	643
01/06	102	119	134	11/06	132	188	353	21/06	162	296	675
02/06	105	124	151	12/06	135	197	381	22/06	165	309	710
03/06	108	130	179	13/06	138	206	407	23/06	168	324	748
04/06	111	136	193	14/06	141	216	440	24/06	171	339	781
05/06	114	142	212	15/06	144	226	468	25/06	174	355	818

Source: Prepared by the Authors

In addition to this assessment, no changes were observed in the concavity from the data produced by both simulations. This is probably due to the dynamic nature of virus spread rates and the early implementation of measures such as lockdown, mobility restrictions, and business restrictions before the first COVID-related deaths were reported. Additionally, these restrictions were perhaps not strictly enforced by the population. The relaxation of measures over time, coupled with social indiscipline, led to increased infections and, therefore, deaths.

Here, as the number of days increases, the projected values are moved further away, which clearly reflects a change in the parameters obtained and acceleration in virus spread rates.

When comparing the projections from both methodologies to the data provided by INS, it was observed that, as the days increase, these projections further deviate from one another. Because of the dynamic nature of virus propagation rates, this concludes that projections generated using calibrated parameters are only reliable for a limited number of days.

Virus spread rates and, thus, the growth of the death curve depend directly on the measures adopted. These measures must be more drastic and require strict compliance with quarantine and physical-distancing restrictions, as well as raising awareness of health care measures to prevent contagion.

References

1. Panamerican Health Organization, www.pho.com
2. Corona Virus Resource Center, Johns Hopkins University & Medicine. Disponilble en: https://coronavirus.jhu.edu/
3. E.M.L. Aquino, I.H. Silveira, J.M. Pescarini, R. Aquino., and J.A. Souza-Filho, “Social distancing measures to control the COVID-19 pandemic: potential impacts and challenges in Brazil. Social distancing measures to control the COVID-19 pandemic: potential impacts and challenges in Brazil”, Ciência & Saúde Coletiva, vol .25, pp. 2423-2446. DOI:https://dx.doi.org/10.1590/1413-81232020256.1.10502020
4. J. Jia, J. Ding, S. Liu, G. Liaq, J. Li, B. Duan, G. Wang., and R. Zhang, “Modeling the control of COVID-19: Impact of policy interventions and meteorological factors,” arXiv, 2020.
5. W. C. Roda, M.B. Varughese, D. H. Michael and Y. Li, “Why is it difficult to accurately predict the COVID-19 epidemic?”, Infectious Disease Modell, vol. 5, pp. 271 – 281, 2020, DOI: https://doi.org/10.1016/j.idm.2020.03.001
6. E.L. Piccolomini and F. Zama, “Preliminary analysis of COVID-19 spread in Italy with an adaptive SEIRD model,” arXiv, 2020.
7. S. Tuli, S. Tuli, R. Tuli., and S. Singh Gill, (2020, may), “Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing”, Internet of Things, vol. 11, pp. 1-16, 2020. DOI: https://doi.org/10.1016/j.iot.2020.100222
8. M. Niazkar., and H.R. Niazkar, “COVID-19 Outbreak: Application of Multi-gene Genetic Programming to Country-based Prediction Models”, Electron J. Gen. Med, vol 17, No. 5., pp. 1 –7, 2020. DOI: https://doi.org/10.29333/ejgm/8232
9. G.R Shinde, A.B Kalamkar, P.N. Mahalle, N. Dey, J. Chaki & A.E. Hassanien, “Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art”, SN COMPUT. SCI. 1, 197, 2020. DOI: https://doi.org/10.1007/s42979-020-00209-9
10. NP. Jewell, JA. Lewnard., and BL. Jewell. “Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections”, JAMA, vol. 323, No.19, pp. 1893–1894, 2020. DOI. https://doi:10.1001/jama.2020.6585
11. A.E.K. Grillo, T.J. Santaella, R. Guerrero. L.E. Bravo, “Modelos matemáticos y el COVID-19” [Mathematical Modelling and COVID-19], Colomb.Med, vol. 51, No. 2. 2020; DOI: https://doi.org/10.25100/cm.v51.i2.4277
12. D. Fanelli, F. Piazza, “Analysis and forecast of COVID-19 spreading in China, Italy and France”, Chaos, Solitons & Fractals, vol. 134, pp. 1 – 5. 2020. DOI: https://doi.org/10.1016/j.chaos.2020.109761.
13. R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, “Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)”, Science, vol. 368, No. 6490, pp. 489-493. 2020. DOI: https://doi.org/10.1126/science.abb3221
14. 14. C. De Castro. “SIR model for COVID-19 calibrated with existing data and projected for Colombia,” arXiv, 2020.
15. F.G. Manrique-Abril, C.A. Agudelo-Calderon, V.M. González-Chordá, O. Gutiérrez-Lesmes, C. F. Téllez-Piñerez y G. Herrera- Amaya, “Modelo SIR de la pandemia de Covid-19 en Colombia” [SIR Model for the COVID-19 Pandemic in Colombia], Rev. Salúd Pública, vol. 22, pp. 1- 9, 2020. DOI: https://doi.org/10.15446/rsap.V22.85977
16. W. O. Kermack, A. G. McKendrick and Gilbert Thomas Walker, “A contribution to the mathematical theory of epidemics”, Proc. R. Soc. Lond. 1927. DOI: http://doi.org/10.1098/rspa.1927.0118

17. Coronavirus (COVID-2019) en Colombia.[Coronavirus (COVID-2019) in Colombia Instituto Nacional de Salud (INS), https://www.ins.gov.co/Noticias/Paginas/Coronavirus.aspx
18. S. Tappe, “A simple mathematical model for the evolution of the corona virus,” arXiv, 2020.
19. F. Zama, D. Frascari, D. Pinelli, A. E. M. Bacca, “Parameter estimation algorithms for kinetic modeling from noisy data”, System Modeling and Optimization, Springer International Publishing, pp. 517–527, 2016.
20. B. Morini, M. Porcelli, “TRESNEI, a Matlab trust-region solver for systems of nonlinear equalities and inequalities”, Comput. Optim. Appl. vol. 51, pp. 27–49 ,2012. DOI: https://doi.org/10.1007/s10589-010-9327-5