Progress in Nuclear Energy 69 (2013) 35–43

Contents lists available at SciVerse ScienceDirect

Progress in Nuclear Energy

journal homepage: www.elsevier.com/locate/pnucene

New insights on modeling news dissemination on nuclear issues

José S.B. Reis Junior [a,b,*], Antonio C.O. Barroso [a], Mario O. Menezes [a]

[a] Nuclear and Energy Research Institute, IPEN–CNEN/SP, Av. Prof. Lineu Prestes, 2242, Cidade Universitária, CEP 05508-000 São Paulo, SP, Brazil
[b] Fellowship from the National Council for Scientific and Technological Development (CNPq), Brazil

ARTICLE INFO

Article history:
Received 9 August 2012
Accepted 26 May 2013

Keywords:
News dissemination
Epidemiological model
Nuclear accident repercussion
Embedded simulation optimization

ABSTRACT

Using a modified epidemiological model, the dissemination of news by media agents after the occurrence of large scale disasters was studied. A modified compartmented model was developed in a previous paper presented at INAC 2007 in which, the Chernobyl nuclear accident (1986) and the Concorde airplane crash (2000) were used as a base for the study. Now the model has been applied to a larger and more diverse group of events – nuclear, non-nuclear and naturally caused disasters. To be inclusive, old and recent events from various regions of the world were selected. A more robust news repository was used, and improved search techniques were developed to ensure that the scripts would not contain false positive news. The same model was used but with improved non-linear embedded simulation optimization algorithms to generate the parameters of interest for our model. Individual parameters and some specific combination of these allow a number of interesting perceptions on how the nature of the accident/disaster gives rise to different profiles of growth and decay of the news. In our studies, events involving nuclear causes generate news repercussion with more explosive/robust surge profiles and longer decaying tails than those of other nature. As a consequence of these differences, public opinion and policy makers are also much more sensitive to some issues than to others. The model, through its epidemiological parameters, shows in quantitative manner how "nervous" the media content generators are with respect to nuclear installations and how resilient this negative feelings toward nuclear can be.

1. Introduction

News on high impact subjects such as climate changing, global warming, or major disasters, be they of natural causes or not, always grabs people's attention. This makes plausible the conjecture that this interaction is responsible for shaping up an important part of our collective common sense. In a previous work, an adapted epidemiological model was used to study, in the international media, the magnitude and the longevity of the repercussion of two accidents – the Chernobyl nuclear accident, Ukraine (1986) and the crash of the Concorde airplane near Paris (2000). Unfortunately that research could not be extended or deepened using the original news database (Google, 2007b) due to an inconsistency found in the news archival methodology. The repository owner (Google, 2007a) changed the manner in which news was indexed and archived making the accuracy of the news publishing date unreliable and therefore unsuitable for the purpose of the research. Three years later, another repository (NewsBank, 2011b) was found with adequate depth, breadth characteristics and accurate dating procedures. The studies were restarted, using a larger and more diverse group of events: for the present work, old and recent events were chosen, from several regions of the world: Bhopal (1984), Chernobyl (1986), Deepwater Horizon's rig (2010), Haiti's earthquake (2010), Japan's earthquake (2011) and Fukushima/Daiichi nuclear plants events (2011).

This paper is divided into six sections. Following the Introduction, a summary of the pertinent literature is presented. This is succeeded by problem formulation and model description. Afterward are the sections on data collection and treatment, results and discussions as well as the conclusions.

2. The literature and clues for our model

The public acceptance of nuclear projects became, notably since the eighties, the object of special attention from managers and scientists. Nowadays, "it is consensus that the public participation on the decision process is essential to the success of a new project" (Rocca, 2002). Indeed, great industrial accidents at the end of the seventies and throughout the eighties have spiked people's interest in debating the benefits of facing the risks of complex technologies,

* Corresponding author. Nuclear and Energy Research Institute, IPEN–CNEN/SP, Av. Prof. Lineu Prestes, 2242, Cidade Universitária, CEP 05508-000 São Paulo, SP, Brazil. Tel.: +55 11 3133 9032.
E-mail addresses: jsreis@ipen.br, jsbrjunior@gmail.com (J.S.B. Reis Junior), barroso@ipen.br (A.C.O. Barroso), mario@ipen.br (M.O. Menezes).

0149-1970/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.pnucene.2013.05.008

like nuclear power production. Sauer and Oliveira Neto (1999) pointed that disasters such as the Three Mile Island (1979), Bhopal (1984), Chernobyl (1986), the explosion of the space shuttle Challenger (1986) and the Piper Alpha accident (1988) have promoted public opinion discrediting the government and industry's technical and political competence on securely managing the process related to highly impacting technologies. However, it appears that public opinion perceives risk in a different way than experts. "The divergences between public and experts on what an acceptable risk is have promoted the study about two important aspects of risk management: the public risk perception and the risk communication" (Rocca, 2002). It is reasonable that, for different individuals or different social groups, risk has different meanings. The media plays a single relevant role in this context: it is an important risk information communicator. Wahlberg and Sjoberg (2000) concluded that media truly influences our risk perception, granting that it is only one factor among many others; as well, it is a somewhat biased factor considering the media tends to focus on dramatic, controversial events that cause social upheaval.

News about industrial processes not rarely are more obscuring than enlightening to people searching for information about their associated risks: many news "are concentrated on potentially catastrophic effects and on the risks of diseases, deaths and injury for the next generations" (Sauer and Oliveira Neto, 1999). The most fundamental ways in which media contributes to distort the public's risk perception are the numbers and the spectacular tone of news concerning some subjects (Wahlberg and Sjoberg, 2000). This fact has motivated many authors to search for a quantitative model for the diffusion of news concerning a risk agent. We looked for such models within the famous book entitled *Diffusion of innovations*, by Rogers (1962), which presents some approaches on modeling the diffusion of innovations. His research and work became widely accepted in communications and technology adoption studies, on the other hand, the presented models make assumptions that are not applicable to news dissemination. In 1972, Funkhouser (1972) published *Predicting the diffusion of information to mass audiences*, whereupon he modeled diffusion of information through probabilistic approaches. Stochastic process based tools were used by Karmeshu and Pathria (1980) and Allen (1982) to construct models on the diffusion of information. Although interesting and useful, these models didn't fit our needs. Dodds and Watts (2005) developed a generalized model for social contagion, based on epidemiological models.

A paper by Bettencourt et al. (2006), entitled *The power of a good idea: quantitative modeling of the spread of ideas from epidemiological models*, has hinted the great potential of compartmented epidemiological models to deal with the diffusion of news. From this we then created a modified epidemiological model for news generation, having its application to two big industrial disasters as well as its construction presented at the 2007 International Nuclear Atlantic Conference – INAC 2007 (Reis Junior et al., 2007). In 2009, Leskovec et al. (2009) developed a framework for tracking memes on news media, and used said tool to make a representation of the news cycle; in their paper, they indicate that "one can give an argument for the characteristic shape of thread volume (…) through an approximation using differential equations" (Leskovec et al., 2009) (Note: each thread consists of all news articles and blog posts containing a textual variant of a particular quoted phrase). These findings and arguments reinforce our previous findings that a conveniently modified epidemiological model, based on ordinary differential equations, is able to justify the thread shape that they have found. This paper adds some confirmation of the conjecture by Leskovec et al. (2009) and is also a natural outspread of the one we have presented at INAC 2007 (Reis Junior et al., 2007). So to a certain extent, our deterministic approach to model news generation seems to constitute an innovative approach to this issue.

3. Problem formulation and model description

It seems appealing to apply an epidemiological model to the news population, once the available evidences were precisely the amount of published news about each catastrophe. However, there is an immediate problem concerning this approach – news on newspapers and magazines, once they are published, do not change, so they cannot undergo state transitions. On the other hand, journalists can change their state of willingness to produce news about a subject. So the question presented was: would it be possible to assess information about this population (journalists) from available data on published news? Assuming, a priori, yes, the search for an appropriate epidemiological model for the journalists' population was elaborated. Such a model should describe the possible states for the members of the population and should also describe the transitions between these states.

3.1. Compartmental epidemiological model for news dissemination

Once an accident occurs, journalists producing news about the accident are considered as influenced (state I). Those who have yet to publish the news are considered as susceptible (state S). Only these two states are considered in the model – a simple epidemiological model – the *SIS* model. In fact there is no need to consider different states for other situations, like the incubation period (state E), recovered individuals (state R) and non-interested (or immune) individuals (state Z), for example. It is assumed that journalists that have yet to publish on subjects related to the accident can still do so at any time – in other words, it is conservatively assumed that no one is immune. Also journalists that have stopped writing about the incident can still publish on that subject in the future – so there is no need to consider the immunized recovered individuals. Finally, there is no reason to distinguish between susceptible individuals (S) and those in the incubation period (E) considering that every non-influenced journalist is permanently exposed to news overall and are also in direct contact with their colleagues.

The total population of journalists is called J; so, we get $J = I + S$. The standard SIS epidemiological model is set by the following equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}S = -\beta S\frac{I}{J} + \gamma I \qquad (3.1)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}I = \beta S\frac{I}{J} - \gamma I \qquad (3.2)$$

where J is the total of individuals, and is always constant in our model; β is the influence rate, and always assumes non-negative values. When multiplied by I/J, it becomes the influence probability; and γ is the recovery rate, also assuming only non-negative values. S, I and J are, obviously, time functions.

It is appropriate to note that there is no reason for β to be constant. It is reasonable to expect the influence rate to decrease over time, considering that, as time goes by, the reader's interest on the topic will decrease and with it, the amount of space allotted for that fact-related type of news. Therefore, the rate at which journalists become influenced should likewise decrease. To be consistent, the model will treat β as a function of time. It's natural to think of a function that has a peak soon after the event and decays with time until it reaches a constant low with little variation.

The function $B = \beta\cdot(I/J)$ corresponds to the probability of a journalist to become influenced after being exposed:

$$B := \begin{cases} \left(\beta_0 + \beta_1 e^{-\xi(t-t_0)}\right)\frac{I}{J}, & \text{for } t \geq t_0, \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

Table 1 introduces the parameters of interest and their possible interpretation according to their use in the construction of the function $B(t)$.

However the differential equations that rule the state transitions of our system depend, at this point, on functions S, I and J, which refer, respectively, to the number of susceptible, influenced and all journalists at a given moment. All these variables refer to data that is not available, for this reason some transformations should be applied to the variables.

Some plausible hypotheses had to be taken into account to relate the above mentioned quantities with those for which data is available. Let the number of news that cites the accident published at time t be denoted by $Q(t)$; the number of news that don't cite the accident published at time t by $P(t)$; and the total of published news at time t be denoted by $N(t)$. It's obvious that $N = Q + P$ holds. Thereafter let ψ be the average productivity for all journalists – in other words, the mean number of news a journalist publishes in one day. Indeed it's reasonable to think ψ shouldn't vary when the studied period of time is not very long. Then the following equations are valid to relate the number of news – citing and not citing the accident – and the population of journalists:

$$\begin{aligned} S(t) &= \psi P(t); \\ I(t) &= \psi Q(t); \\ J(t) &= \psi N(t) \end{aligned} \tag{3.4}$$

Functions are then normalized with respect to their values at time $t_0 = 0$; this day corresponds to the first day the news on the event appear. Adopting the notation $S_0{:=}S(0), I_0{:=}I(0), J_0{:=}J(0)$, the following definitions can be introduced:

$$\begin{aligned} \widehat{S}(t) &:= \tfrac{S(t)}{S_0} = \tfrac{P(t)}{P_0}; \\ \widehat{I}(t) &:= \tfrac{I(t)}{I_0} = \tfrac{Q(t)}{Q_0}; \\ \widehat{J}(t) &:= \tfrac{J(t)}{J_0} = \tfrac{N(t)}{N_0} \end{aligned} \tag{3.5}$$

This normalization brings, for example, the following premise: being the population J considered constant over the observation period, we get $\widehat{J}(t)$ constant equal to 1. Therefore, we have so far, an epidemiological model defined by the following equation system:

$$\frac{d}{dt}S = -BS + \gamma I \tag{3.6}$$

$$\frac{d}{dt}I = BS - \gamma I \tag{3.7}$$

Dividing (3.6) by S_0 and (3.7) by I_0 we have:

$$\frac{d}{dt}\widehat{S} = -B\widehat{S} + \gamma\frac{I}{S_0} \tag{3.8}$$

Table 1
Parameters of interest to the function $B(t)$.

β_0	Coefficient of the long-term news persistence of the event
β_1	Coefficient of the news outbreak of the event
ξ	Decay constant of the news outbreak of the event

$$\frac{d}{dt}\widehat{I} = B\frac{S}{I_0} - \gamma\widehat{I} \tag{3.9}$$

As a starting point, the above equation (3.9) will be used. Since $\widehat{I}$ is a normalized function, it would be interesting to write the equation only in terms of normalized functions. Of course, by the definition of $\widehat{S}$, we have

$$S = \widehat{S}S_0$$

So, one can write the equation (3.9) as follows:

$$\frac{d}{dt}\widehat{I} = B\frac{\widehat{S}S_0}{I_0} - \gamma\widehat{I} \tag{3.10}$$

Once $P_0 = \psi S_0$ and $Q_0 = \psi I_0$, the following equality holds:

$$\frac{d}{dt}\widehat{I} = B\frac{\widehat{S}\psi S_0}{\psi I_0} - \gamma\widehat{I} = B\frac{\widehat{S}P_0}{Q_0} - \gamma\widehat{I} \tag{3.11}$$

It would be interesting if the differential equation depended only on the function $\widehat{I}$, since that would make its solution easier. So one can note that

$$\widehat{S} := \frac{S}{S_0} = \frac{P}{P_0} = \frac{N-Q}{P_0} \tag{3.12}$$

As N is a constant function, $N \equiv N_0$ holds. Moreover we have $Q = \widehat{I}Q_0$. So the following equality is true:

$$\widehat{S} = \frac{N_0 - \widehat{I}Q_0}{P_0} \tag{3.13}$$

Thus, the equation (3.11) can be written as follows:

$$\frac{d}{dt}\widehat{I} = B\frac{N_0 - \widehat{I}Q_0}{Q_0} - \gamma\widehat{I} = B\left(\frac{N_o}{Q_0} - \widehat{I}\right) - \gamma\widehat{I} \tag{3.14}$$

Having in mind the definition of function B, one can observe that $I/J = \psi I/\psi J = Q/N$, and using $N \equiv N_0$ and $Q = \widehat{I}Q_0$, one can conclude the following formula:

$$B := \begin{cases} \left(\beta_0 + \beta_1 e^{-\xi(t-t_0)}\right)\frac{\widehat{I}Q_0}{N_0}, & \text{for } t \geq t_0, \\ 0, & \text{otherwise} \end{cases} \tag{3.15}$$

Given this formula of B, we can finally write equation (3.14) in such a way that it depends only on $\widehat{I}$. We then get a *second order Bernoulli differential equation*, which is given by the following formula:

$$\begin{aligned} \frac{d}{dt}\widehat{I} &= \left(\beta_0 + \beta_1 e^{-\xi(t-t_0)}\right)\frac{\widehat{I}Q_0}{N_0}\left(\frac{N_0}{Q_0} - \widehat{I}\right) - \gamma\widehat{I} \\ &= -\left(\beta_0 + \beta_1 e^{-\xi(t-t_0)}\right)\frac{Q_0}{N_0}\left(\widehat{I}\right)^2 - \left[\gamma - \left(\beta_0 + \beta_1 e^{-\xi(t-t_0)}\right)\right]\widehat{I} \end{aligned} \tag{3.16}$$

The equation (3.16) governs, accordingly, in our model, the quantitative variation of the influenced parcel of the population of journalists. From there we'll seek to identify the parameters of interest β_0, β_1, ξ and γ for each case – for each industrial or natural catastrophe. These parameters, once given the utilization of each one at the model's construction, can be interpreted according to Table 2.

Table 2
Parameters of interest of the developed model.

β_0	Coefficient of the long-term news persistence of the event
β_1	Coefficient of the news outbreak of the event
ξ	Decay constant of the news outbreak of the event
γ	Decay constant of influence (recovery rate from influenced state)

3.2. System behavior for large time values

Recalling the definition of B – the probability of publishing influenced news after being exposed to this type of news:

$$B := \beta\frac{I}{J} = \begin{cases} \left(\beta_0 + \beta_1 e^{-\xi(t-t_0)}\right)\frac{I}{J}, & \text{for } t \geq t_0, \\ 0, & \text{otherwise} \end{cases} \tag{3.17}$$

$$B := \beta\frac{I}{J} = \begin{cases} \left(\beta_0 + \beta_1 e^{-\xi(t-t_0)}\right)\frac{\widehat{I}Q_0}{N_0}, & \text{for } t \geq t_0, \\ 0, & \text{otherwise} \end{cases} \tag{3.18}$$

So we are able to see the equality

$$\beta = \beta_0 + \beta_1 \exp[-\xi(t-t_0)] \tag{3.19}$$

involving the influence rate β.

When $t \to \infty$, we will get:

$$\lim_{t\to\infty} \exp[-\xi(t-t_0)] = 0 \Rightarrow \tag{3.20}$$

$$\Rightarrow \lim_{t\to\infty}(\beta_1)\exp[-\xi(t-t_0)] = 0 \tag{3.21}$$

$$\Rightarrow \lim_{t\to\infty}\beta = \beta_0 \tag{3.22}$$

Thus, for large enough values of t, the influence rate β is arbitrarily close to β_0.

This fact sets a scenario where sometimes the system, depending on β_0, β_1, ξ and γ, can present an interesting growing trend of influence on a population. In other words: some events would observe a distal turning point from which the proportion of influenced journalists I would start re-growing, overturning it's decay due to passing of time since the accident. Indeed, given the equation (3.19) and the definition

$$K_0 := \frac{Q_0}{N_0} \tag{3.23}$$

we are able to write the ordinary differential equation (3.16) in the following form:

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{I} = -\beta K_0 \widehat{I}^2 - \gamma\widehat{I} + \beta\widehat{I} \tag{3.24}$$

Such equation informs us that the function $\widehat{I}$ has critical points if and only if

$$-\beta K_0 \widehat{I}^2 - \gamma\widehat{I} + \beta\widehat{I} = 0 \Leftrightarrow \tag{3.25}$$

$$\Leftrightarrow \left(\widehat{I}\right)\left(-\beta K_0\widehat{I} - \gamma + \beta\right) = 0 \tag{3.26}$$

If $\widehat{I} = 0$, then $(\mathrm{d}/\mathrm{d}t)\widehat{I} = 0$ and $\widehat{I}$ would remain zero forever – in other words, there would not be any influenced journalist, and so a re-infection process would not happen. Thus, let's see the case in which $\widehat{I} \neq 0$.

Indeed it is always true that

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{I} \leq 0 \tag{3.27}$$

and so $\widehat{I}$ will be continuously approaching zero. However using equality (3.22), when t is large enough, the critical point at which $\widehat{I}$ starts to increase again can be calculated as the time t for which $\widehat{I}$ reaches the value

$$\widehat{I} = \frac{\beta_0 - \gamma}{\beta_0 K_0} \tag{3.28}$$

so that we get

$$-\beta K_0\widehat{I} - \gamma + \beta = 0 \tag{3.29}$$

as we wished. Therefore, from equation (3.28) one can conclude that if and only if $\beta_0 > \gamma$ the critical point can be reached, because otherwise $\widehat{I}$ will reach zero at an earlier time and $(\mathrm{d}/\mathrm{d}t)\widehat{I} \leq 0$ at the same point making $\widehat{I} = 0$ to become a permanent condition.

For processes where $\beta_0 > \gamma$ there will be a bounded oscillatory behavior – because as soon as $\widehat{I} > 0$ its derivative will become less than zero pushing it back again – that leads to a steady state commonly designated *endemic equilibrium* by epidemilogists: in this case the infection should establish itself (Brauer and Castillo-Chávez, 2001). Note that we can define the tendency – or not – of the system to an endemic equilibrium just by looking at the coefficient β_0/γ, that is usually denoted by R_0 and called the *basic reproductive number*. "In studying an infectious disease, the determination of the basic reproductive number is invariably a vital first step" (Brauer and Castillo-Chávez, 2001). Indeed if $R_0 > 1$ the infection must be persistent throughout the population while if $R_0 < 1$ we will get, for large enough time values, a constant condition $\widehat{I} = 0$, what is commonly called the *disease-free equilibrium*. "The value one for the basic reproductive number defines a threshhold at which the course of the infection changes between disappearance and persistence" (Brauer and Castillo-Chávez, 2001). So the first feature we need to measure for each event is the basic reproductive number R_0.

4. Data collection and processing

Once we have the ordinary differential equation:

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t}\widehat{I} = {} & \left(\beta_0 + \beta_1 e^{-\xi(t-t_0)}\right) \tfrac{Q_0}{N_0}\left(\widehat{I}\right)^2 + \\ & -\left[\gamma - \left(\beta_0 + \beta_1 e^{-\xi(t-t_0)}\right)\right]\widehat{I} \end{aligned} \tag{4.1}$$

our main objective is to identify the parameters β_0, β_1, ξ and γ that provide the best fit of equation (4.1) to the reality of the collected data from the published news. We should find a point $(\beta_0,\beta_1,\xi,\gamma)$ in $\mathbb{R}^4$ that fits as perfectly as possible the solution of this equation to the collected data of published news about each disaster of interest; so we should get one optimal point $(\beta_0,\beta_1,\xi,\gamma)$ for each catastrophe.

4.1. Data collection

Data was collected from the News Library repository (NewsBank, 2011b), provided by NewsBank Inc. (2011a). This has shown to be the most robust, reliable and accessible repository available on the web at this time. In a previous work (Reis Junior et al., 2007), we used the Google News Archive Search; but this tool often replaced the date in which an article was published with

Table 3
Analyzed period and correspondent *day W* for all studied events.

Event	Occurrence	Analyzed period	*Day W*
Chernobyl	April 26, 1896	2 years	729
Bhopal	December 3, 1984	2 years	729
Challenger	January 28, 1986	2 years	729
Fukushima	March 11, 2011	1½ month	44
Deepwater	April 20, 2010	1 year	364
Japan	March 11, 2011	1½ month	44
Haiti	January 12, 2010	1 year 3 months	454

dates cited in the article's body, resulting in wrong data. We communicated this discrepancy to Google but they insisted it was not a flaw, but a resource to identify the time periods that are likely to be relevant to each query. The decision was then made to instead make use of the News Library repository. The search took place between May and June of 2011. For each disaster, we collected the number of articles published per day, since the first day in which there were articles citing the event – *day 0* – until a certain time that was called *day W*, which depends on the case, as shown in Table 3. Thus, for each integer t, $0 \le t \le W$, we obtained a number $I_R(t)$ that corresponds to the quantity of articles published by journalists that day. The graph of the function $I_R(t)$ for all studied events is represented by Fig. 1.

4.2. Data smoothing

Once the data collection for each disaster was completed, the data required smoothing prior to being approximated as it would be approximated by a continuous function – a solution of equation (4.1).

The procedure performed to smooth the data was as follows:

- From *day 0* to day when $I_R(t)$ reaches its maximum – *day M* – there is no change. That is, for every t, it remains the correspondent $I_R(t)$.
- From the first day after the *day M*, we obtain the mean for each group of three days – one day cannot belong to two different groups – and each day in the group receives as the value of $I_R(t)$, the average of their group. For example, in the case of Chernobyl, the *day M* occurs at $t = 2$. The first group of three days will then be the group for $t = 3,4,5$. Then, as the average of published articles per day in this group is 184, we have $I_R(3) = I_R(4) = I_R(5) = 184$. However, we used a minimum score: if the sum of the $I_R(t)$ values of the group is less than a certain C_{MIN}, then we keep on adding one more day until it reaches the minimum score. Hence, each day in the group receives as $I_R(t)$ value the average value of the group.

For each event, the corresponding minimum score C_{MIN} was set as the average of the number of articles published per day in the first half of the period taken into account (which ends at *day W*):

$$C_{MIN} = \frac{2\sum_{t=0}^{\frac{W-1}{2}} I_R(t)}{W+1} \tag{4.2}$$

We have used rounding by truncation in all operations for the calculation of C_{MIN}. One should note the following: as consequence of imposition of the minimum score, the set of values t can assume may reduce. From a given time t', it is possible for the function $I_R(t)$ to be so small that it no longer can reach the minimum score until $t = T$. The set of values that t can assume now goes from 0 to the last day of the last group that achieved the minimum score, which we will call *day V*, which depends on each specific case, as shown in Table 4. Once the data was smoothed, it was normalized with respect to $I_R(0)$, given the very definition of function $\widehat{I}$.

4.3. Parameters identification

With the data smoothed and normalized, we proceeded to identify the parameters of interest. We have chosen to use the *Sequential Least Squares Programming* technique (Kraft, 1988) to find the point that minimizes our function with non-linear

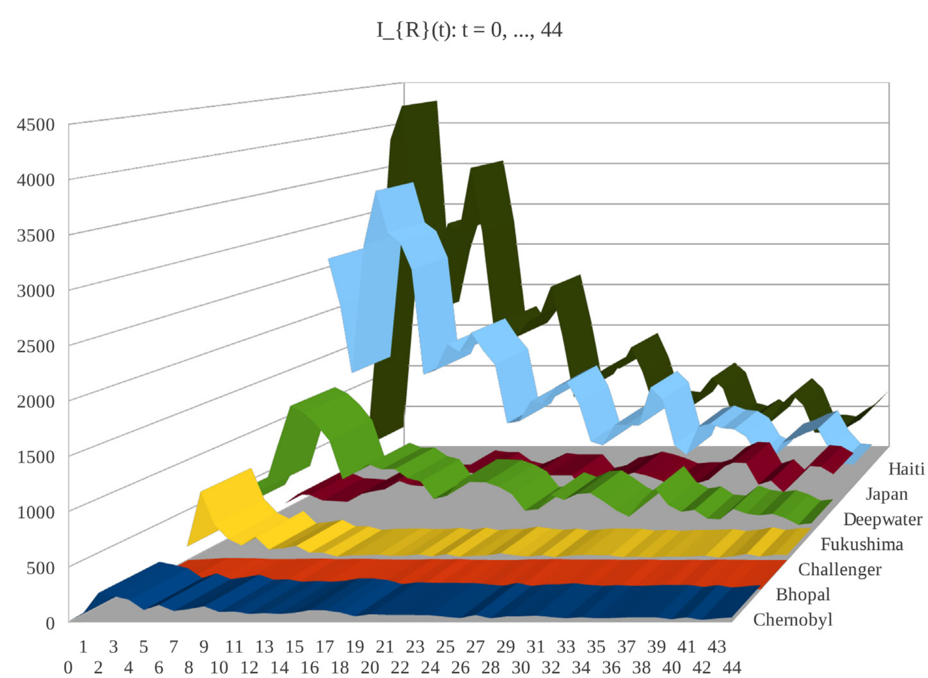

Fig. 1. Graph of the function $I_R(t)$, $t = 0,\ldots,44$, for all studied events.

Table 4
C_{MIN} and consequent *day V* for all studied events.

Event	C_{MIN}	*Day V*
Chernobyl	20	728
Bhopal	3	729
Challenger	26	727
Fukushima	659	42
Deepwater	217	363
Japan	1698	40
Haiti	404	449

Table 6
The basic reproductive number R_0 and early force of infection $\overline{\lambda}$.

Event	R_0	$\overline{\lambda}$
Challenger	1.0034991956	0.1956164989
Fukushima	1.0022138892	0.1345564158
Japan	0.9997718282	0.0861042523
Chernobyl	1.0017771013	0.0707322066
Haiti	0.9980584476	0.0589169629
Bhopal	1.0022176719	0.0565795673
Deepwater	0.999308038	0.0193664885

constraints. Indeed, we have tried many non-linear optimization methods: downhill simplex algorithm, modified Powell's method, conjugate gradient algorithm, BFGS algorithm, Newton-Conjugate Gradient method, simulated annealing, and brute force technique; a few of them provided noticeably bad results – the simulation of the equation of interest using such results got quite far from the reality of the data. As for the others, they were unable to complete optimization in an acceptable amount of time. The fastest and therefore best choice was, undoubtedly, the Sequential Least Squares algorithm.

Hence, we proceeded to define the function to be minimized: let $x = (\beta_0, \beta_1, \xi, \gamma)$ be as defined before. We called $I_S(x)$ the simulation – performed via the fourth-order Runge–Kutta method – of the ordinary differential equation (4.1), with initial conditions $t_0 = 0$, $\widehat{I}(t_0) = 1$ (due to the fact that $\widehat{I}$ is the result of a normalization of function $I(t)$). We then called $I_S(x,t)$ the value of such simulation at time t, with $t = 0,1,2,\ldots,V$, where V is the number corresponding to the day V. The function given to the minimization algorithm was as follows:

$$E(x) = \sqrt{\sum_{t=0}^{V} [I_s(x,t) - I_R(t)]^2} \tag{4.3}$$

In other words: the goal was to minimize, with respect to x, the Euclidean norm of errors between the simulation and data collected from the News Library repository. One can note that, in each iteration of the optimization process, a new simulation was performed to calculate the value of the objective function. The optimization program was executed, for all events, from the initial point (1,1,1,1) – the canonical vector example with non-zero entries in $\mathbb{R}^4$. Table 5 contains the results obtained by performing this optmization process.

Once Table 5 was achieved, in order to assess the influence process behavior at the beginning of the analyzed period, we then calculated, for each disaster, the *early force of infection* $\overline{\lambda}$, which is defined as follows:

$$\overline{\lambda} = \frac{1}{30}\sum_{t=0}^{29}[\beta_0 + \beta_1 exp(-\xi t)]\frac{Q_0}{N_0}\widehat{I}(t) \tag{4.4}$$

Note that $\overline{\lambda}$ is the average force of infection (Sutton et al., 2006) for the first thirty days after the occurrence of the disaster. In Table 6 we sorted the events by decreasing values of $\overline{\lambda}$, making clear which events have been more "infectious" at the beginning. Probably those were the ones that caused more comotion during this time. Our experience has shown that 30 days is a reasonable period for calculating $\overline{\lambda}$, but we did not attempt to find an optimum number of days.

As discussed in subsection 3.2, events that present $R_0 > 1$ have long-term reinfection pressure. From Table 6 one can conclude that the only events that present $R_0 > 1$ are Chernobyl, Bhopal, Challenger and Fukushima. It's interesting to note that the only recent disaster in this group is Fukushima – the only event which occurred in a nuclear facility. Indeed, it seems that media will quickly forget the Japan earthquake itself – which tragically resulted in many deaths – but will not forget as quickly the Fukushima nuclear power plant accident – which has no deaths associated.

We have also calculated, for each disaster, the *mean infective period* (Brauer and Castillo-Chávez, 2011), which is defined as the inverse of the parameter γ. A larger $1/\gamma$ value can point to a propensity of the event repercussion to become resilient, i.e., the influenced individuals have more time to influence others, although this process doesn't trigger a new outbreak or epidemic. Such measure may be interpreted as a kind of persistence of the event to become the subject of new articles lasting long after the disaster's occurrence. Table 7 shows all events ordered by calculated $1/\gamma$ values.

5. Discussing the possibility of improving the developed model

Our expression for the influence rate β has the inconvenience of assuming a peak value at $t = 0$. As a result, we decided to test a more flexible expression since data has shown that this is not always true. For instance, the Deepwater Horizon rig disaster, was gradually spiking the media's interest as the oil spill in the ocean became more and more serious. So much so that, for this event, the peak of the curve of infected agents took place in the 57th day after the accident, nearly two months later. This fact makes it natural to search for a form to the influence rate that admits a peak even after the *day 0*. The function typically used to describe curves with movable peaks is the *gaussian function*, also known as "bell curve",

Table 5
Identified parameters and objective function value for all studied events.

Event	β_0	β_1	ξ	γ	$E(x)$
Chernobyl	1.90161633	5.64453851	1.43305041	1.89824296	23.23606580
Bhopal	5.5789935	2.74049343	1.06153285	5.5666485	37.83153979
Challenger	4.22657173	45.71444984	8.70693565	4.2118337	2.16311943
Fukushima	19.5813878	0.81401107	0.29303129	19.53813254	2.45628832
Deepwater	5.95316998	0.13630052	0.0355955	5.9572922	16.44219916
Japan	4.95570448	0.21447007	0.18462508	4.95683549	0.86033893
Haiti	2.23739491	4.1986905	1.35348209	2.24174738	11.00282360

Table 7
Events sorted by $1/\gamma$ – the mean infective period.

Event	$1/\gamma$
Chernobyl	0.5268029547
Haiti	0.4460805927
Challenger	0.2374262783
Japan	0.2017416156
Bhopal	0.1796413048
Deepwater	0.1678614992
Fukushima	0.0511819642

Table 10
Gaussian model: objective function value and the mean infective period $1/\gamma$.

Event	$E(x)$	$1/\gamma$
Chernobyl	21.3107563506	0.5446038119
Haiti	10.7233189612	0.4664834727
Challenger	2.16278972205	0.2375836701
Bhopal	37.792010014	0.1830453495
Japan	0.538529311427	0.1750013881
Fukushima	1.91311339216	0.0836216697
Deepwater	16.8792555173	0.0105078827

Table 8
Parameters of interest of the gaussian model.

β_0	Coefficient of the long-term news persistence of the event
β_1	Coefficient of the news outbreak of the event
η	Position of the centre of the infectivity's peak of the event
σ	Coefficient of the amplitude (over time) of the news outbreak
γ	Decay constant of influence (recovery rate from influenced state)

because of its characteristic shape. We have then set to B (and thus to β, once $B = \beta\cdot(I/J)$) the following formula:

$$B := \begin{cases} \left(\beta_0 + \beta_1 e^{-(t-\eta)^2/2\sigma^2}\right)\frac{I}{J}, & \text{for } t \geq t_0, \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

which provided us, after redoing the calculations exposed in subsection 3.1, with a new epidemiological model for news dissemination. From now on, to avoid ambiguity, we will call it the *gaussian model*. Table 8 presents the parameters of interest and their possible interpretation according to their use in the construction of the new model.

This gaussian model was applied to the same seven events studied with the previous model. The steps presented in Section 4 were repeated for this new model, which gave us a new set of identified parameters for each studied disaster. Note that the gaussian model has five parameters to be identified for each event – β_0, β_1, η, σ, and γ – while the previous model had only four. The results of the procedure for parameter identification are demonstrated in Table 9. Table 10 shows the objective function values after carrying out the optimization process for the gaussian model and also the calculation of the new mean infective period for all studied disasters.

6. Some considerations on the model's predictive ability

It is true that we can use the parameters obtained through the optimization process to extrapolate the curve of influenced agents – the function $\widehat{I}(t)$ – toward the future, trying to observe how the epidemiological process tends to advance over time. Note that we are keen to use the term "tends", once a model, however good, is subject to failure when it comes to covering all the intricacies of the reality which it aims to represent, especially when that reality involves the dynamics of the news being spread among people. In order to evaluate whether the extrapolations based on our model tend to be close to reality, we have performed the following experiment: one year after the accident at the Fukushima/Daiichi power plant, we collected data on the amount of news published per day concerning the event, from the 45th day following the accident (we already had data on the previous days) until the 364th day, covering a period of one year after the disaster. Thereon we smoothed the data and conducted the procedures for parameter identification that we previously exposed. At the same time, we used the parameters already identified for the case of Fukushima – presented in Tables 5 and 9 – to extrapolate the curve of influenced journalists up to the time point of one year, using both original and gaussian models. Accordingly, the predicted curves used only data from the first 45 days to predict the behavior of the infectious process for the full first year after the event, what would be able to display the predictive ability of the developed model. The result of this experiment–the graph of the three curves – is presented on Fig. 2.

It is observed that both models tend to make a conservative estimate, i.e., they tend to foresee an amount of news greater than that which was actually released and so the models give us a kind of upper bound for the actual amount of news producted by media agents. This experiment was repeated for the other events in which the *day W* was less than 364, and in all cases we obtained the same result: a conservative estimate that surmounted the actual number of published news on each day of the remaining period of the first year. These results showed us that, despite the fact that extrapolations based on the developed models are far from providing extremely accurate forecasts, they are able to indicate suitable upper bounds to the course of the epidemic process. However, it is quite clear that the gaussian model seems to have a better predictive capacity. In Fig. 2 we can see that, in addition to following "more closely" the peak of news in the 4th day after the accident, the prediction of the gaussian model "sees" more effectively the downward trend in the amount of news concerning the Fukushima accident during the first year following the tragedy.

We should note that, even after one year, the Fukushima event remains endemic: the new set of found parameters provided a basic reproductive number of 1.000242583 on the original model and of 1.000109825 on the gaussian model; so, both are greater than 1,

Table 9
Gaussian model: identified parameters value for all studied events.

Event	β_0	β_1	η	σ	γ
Chernobyl	1.8393258	6.59120611	0.89350992	0.24679385	1.83619721
Bhopal	5.47520934	1.51123002	0.24602168	1.15177979	5.46312705
Challenger	4.22376904	8.69544872	0.21223581	0.19809338	4.20904349
Fukushima	11.97476727	0.72356926	2.35371666	1.13030623	11.95862273
Deepwater	95.20688091	0.40762454	42.98742773	43.73864172	95.16665033
Japan	5.73327519	1.57458544	3.00453172	0.24093338	5.71424039
Haiti	2.13866199	2.27663888	0.45724272	0.72334571	2.14369867

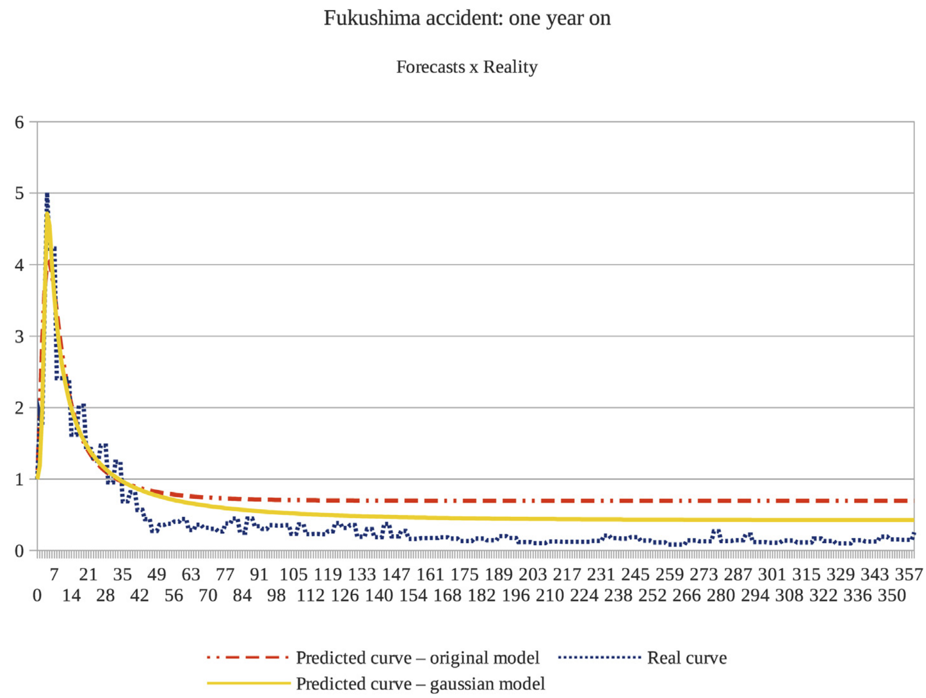

Fig. 2. Real and predicted curves for the first year after Fukushima event.

which indicates that the Fukushima accident tends to persist somewhat as a recurring issue in the media. And, indeed, this is really consistent with what is observed by those who follow newspaper reports presently. Lastly, the new mean infective period for Fukushima is 0.079376519 in the original model and 0.093638482 in the gaussian model, which makes this disaster rank among the last ones in this regard; this fact leads us to conclude that the current "endemic" presence of this event in the media is probably not due to the accident itself, but is the result of a kind of "feedback process" originated by new occurrences involving the consequences of the explosions on the nuclear plant.

7. Conclusions

First it should be noted that the research made use of a news repository based in the USA and, although they collect news from all over the world, the majority of the sources come from the US. So the sample has a natural bias capturing more news items that may be of concern and interest to the American population.

In general, media gives more prominence to disasters involving nuclear facilities than to other types of catastrophes. Comparing the early force of infection after the occurrence, Chernobyl's nuclear accident ranked fourth and Fukushima/Daiichi ranked second, well ahead of Bhopal, Deepwater Horizon and Haiti Earthquake accidents, whose tragic consequences were of much larger proportions. Something that would not occur had the media's attention been distributed in proportion to the seriousness of the event. Perhaps the Chernobyl accident would have shown a greater early force of infection if it were not for the Soviet government's attempt to conceal the accident and its consequences immediately following the event.

Allowing concessions to the bias mentioned previously, one can conclude that nuclear accidents don't always receive the vast majority of the attention of media agents: events that cause more commotion in public, have a greater appeal to emotion, or are associated to the misfortunes of defenseless people, seem to cause an even larger news outbreak. As one can see that the Challenger shuttle explosion has surmounted Fukushima in terms of β_1 (news outbreak coefficient) in both original model and gaussian model – and the devastating Japan earthquake has surmounted Chernobyl in early force of infection.

The accident involving the space shuttle Challenger caused great commotion in the USA as many people were watching the launch live on television, and the explosion was terrifying to the American people. But, after Challenger, the Fukushima accident is the most short-time striking event we have studied. In third place comes the Japan earthquake, another event that deeply affected people, given the scale of the damage – in human and material aspects – immediately caused by an earthquake of unprecedented proportions. Moreover one can note that the Deepwater Horizon rig's accident presents the smallest influence rate value shortly after occurring; this is certainly due to the fact that it was believed that the oil spill would be shortly contained and therefore not attracting people's attention in as large a scale as the other accidents. However the oil spill could not be contained and, as the situation worsened, the media began to give the attention deserved to the accident due to its serious implications.

When we look at the long-term scenario for each event we also realize that disasters involving nuclear issues receive differentiated preference from media agents. Chernobyl's nuclear accident presents the second greatest propensity to become endemic when we analyze the basic reproductive number, just behind the Challenger disaster, and also presents the greatest mean infective period of all studied cases. This seems to indicate that media will keep on coming back to this nuclear accident much more frequently and longer than they do for other types of events. While Chernobyl has the greatest tendency of persistence in media of all catastrophes, the Fukushima plant accident has the greatest propensity of endemic of all its contemporary disasters, given that its basic reproductive number is the only among all the recent events to be greater than 1: even the Japan and Haiti earthquakes are less resilient in media than Fukushima. However, while looking at the low mean infective period of Fukushima, it seems that this recent nuclear event does not tend to be as resilient as Chernobyl and it could even stay less resilient than the Bhopal gas tragedy, as its basic reproductive number is slightly smaller than that of Bhopal. Furthermore, the experiment presented in Section 6 showed us that the production of news concerning the disaster on Fukushima's plant has kept below the forecasts in the first year following its occurrence. From these findings one might conclude that the Fukushima event is not a "new Chernobyl event" regarding dissemination and persistence of news in the media.

Lastly it is clear that there is still much work to be conducted in modeling news dissemination by epidemiological models. The next targeted step involves implementation of our model equipped with other – maybe better – functions to influence rate β. Additionally, the use of other metrics from epidemiological sciences to help the interpretation of the modeled data has to be careful and thoroughly analyzed. An enlargement of the model to include multiple "generating facts" is currently being considered as this would be very useful in modeling further peaks that could be generated by world conferences that are organized to discuss the event one or more years later. Finally research is being conducted to find alternate repositories of periodicals robust and accessible enough from other countries and continents. The purpose is to be able to evaluate the Country bias referred to in the beginning of this section.

Acknowledgment

We acknowledge the National Council for Scientific and Technological Development (CNPq) for the financial support of our research through the National Institute of Science and Technology of Innovative Nuclear Reactors (INCT-RNI)

References

Allen, B., 1982. Stochastic interactive model for the diffusion of information. J. Math. Sociol. 8, 265–281.
Bettencourt, L.M.A., Cintrón-Arias, A., Kaiser, D.I., Castillo-Chávez, C., 2006. The power of a good idea: quantitative modeling of the spread of ideas from epidemiological models. Phys. A 364, 513–536.
Brauer, F., Castillo-Chávez, C., 2001. Mathematical Models in Population Biology. Springer-Verlag, New York, NY.
Dodds, P.S., Watts, D.J., 2005. A generalized model of social and biological contagion. J. Theor. Biol. 232, 587–604.
Funkhouser, G.R., 1972. Predicting the diffusion of information to mass audiences. J. Math. Sociol. 2, 121–130.
Google, 2007a. Google.
Google, 2007b. Google News Archive Search.
Karmeshu, R.K., Pathria, R.K., 1980. Stochastic evolution of a nonlinear model of diffusion of information. J. Math. Sociol. 7, 59–71.
Kraft, D., 1988. A software package for sequential quadratic programming. In: DFVLR-FB 88–28. DLR German Aerospace Center. Institute for Flight Mechanics, Koln, Germany. Tech. Rep.
Leskovec, J., Backstrom, L., Kleinberg, J., 2009. Meme-tracking and the dynamics of the news cycle. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France.
NewsBank, 2011a. Newsbank, Inc. a Premier Information Provider.
NewsBank, 2011b. Newslibrary.com – Newspaper Archive, Clipping Service – Newspapers and Other News Sources.
Reis Junior, J.S.B., Barroso, A.C.O., Imakuma, K., Menezes, M.O., 2007. News and its influence on the viability of nuclear power plants deployment - a modified epidemiological model for news generation. In: 2007 International Nuclear Atlantic Conference – INAC 2007, Santos, Brazil.
Rocca, F.F.T., 2002. A percepcao de risco como subsidio para os processos de gerenciamento ambiental. Nuclear and Energy Research Institute, IPEN–CNEN/SP, Brazil. Ph.D. thesis.
Rogers, E.M., 1962. Diffusion of Innovations. Free Press of Glencoe, New York, NY.
Sauer, M.E.L.J., Oliveira Neto, J.M., 1999. Comunicacao de risco na area nuclear. In: VII Congresso Geral de Energia Nuclear – VII CGEN, Belo Horizonte, Brazil.
Sutton, A.J., Gay, N.J., Edmunds, W.J., Hope, V.D., Gill, O.N., Hickman, M., 2006. Modelling the force of infection for hepatitis b and hepatitis c in injecting drug users in England and wales. BMC Infect. Dis. 6.
Wahlberg, A., Sjoberg, L., 2000. Risk perception and the media. J. Risk Res. 3 (1), 31–50.