A novel spatial and stochastic model to evaluate the within and between farm transmission of classical swine fever virus: II Validation of the model.

Martínez-López, B.[a*]; Ivorra, B.[b]; Ngom, D.[c]; Ramos, A.M.[b]; Sánchez-Vizcaíno, J. M.[a]

[a] VISAVET Center and Animal Health Department. Veterinary School. Complutense University of Madrid. Av. Puerta de Hierro s/n. 28040, Madrid, Spain.
[b] MOMAT research group and IMI institute. Applied Mathematics Department. Mathematics School. Complutense University of Madrid. Plaza de Ciencias, 3, 28040, Madrid, Spain.
[c] Mathematics Department. University of Ziguinchor. Bp: 523, Ziguinchor, Senegal.
& LANI (Gaston Berger University). Bp:234,, Saint Louis, Senegal
[*] Corresponding author: Tel: +34 91 394 37 02; Fax: +34 91 394 39 08; E-mail address: beatriz@sanidadanimal.info; Av. Puerta de Hierro s/n. 28040, Madrid, Spain.

Abstract

A new, recently published, stochastic and spatial model for the evaluation of classical swine fever virus (CSFV) spread into Spain has been validated by using several methods. Internal validity, sensitivity analysis, validation using historical data, comparison with other models and experiments on data validity were used to evaluate the overall reliability and robustness of the model. More than 100 modifications in input data and parameters were evaluated. Outputs were obtained after 1000 iterations for each new scenario of the model. As a result, the model was shown to be robust, being the probability of infection by local spread, the time from infectious to clinical signs state, the probability of detection based on clinical signs at day t after detection of the index case outside the control and surveillance zones and the maximum number of farms to be depopulated at day t the parameters that more influence (>10% of change) on the magnitude and duration of the epidemic. The combination of a within- and between- farm spread model was also shown to give significantly different results than using a purely between-farm spread model. Methods and results presented here were intended to be useful to better understand and apply the model, to identify key parameters for which it will be critical to have good estimates and to provide better support for prevention and control of future CSFV outbreaks.

1. Introduction

A spatial and stochastic model to simulate the spread of within- and between- farm transmission of Classical swine fever virus (CSFV), referred to as Be-FAST (Between Farm Animal Spread Transmission), has recently been described in Martinez-López et al., (2011). Model parameters and assumptions were provided and an illustration of the model results was performed by using available data from Spanish region of Segovia. The aim of this new model was to quantify the magnitude and duration of potential CSFV epidemics and, ultimately, to provide support for the decision making process in future CSFV outbreaks. However, an important requirement to evaluate and fully understand the behaviour and

performance of any new model, before using it for decision making, is to ensure that the model structure is "correct" (results are consistent with experimental ones) and "robust" (results variation due to small perturbations of the input data or, as our model is stochastic, from one run to another one is ''small'') . This only can be done by performing an extensive verification and validation process of the model.

Verification and validation processes allow to verify that the model is correctly formulated and implemented to satisfy the intended objectives and that provides a satisfactory range of accuracy about the system being modelled (Sargent, 1998, 2001). Specifically, model verification is described as the procedure implemented to ensure that the programming code and its implementation are correct. Model validation is the process of determining the degree to which a model gives an accurate representation of the real world from the perspective of the desired uses of the model (AIAA, 1998). An intensive verification of the programming code (implemented in MATLAB) was performed before publication of Martinez-López et al. (2011) and Ivorra et al., (2011, submitted) to verify the correctness and appropriateness of the code before obtaining the outputs. In the manuscript presented here we will focus on the validation process.

There are many validation techniques that can be either subjective (base on graphs), or objective (based on statistical tests or mathematical methods). An extensive review of the techniques and methods that can be used for model validation has been provided elsewhere (Sargent, 1998, 2001; Thacker et al., 2004; Kopec, J.A. et al, 2010). However, we will briefly describe the methods that are most commonly used for validation of stochastic disease-spread models in order to proper understand the validation process applied in this manuscript. The methods described are internal validity, sensitivity analysis, historical data validation and comparison with other models.

Internal validity is a process intended to assess the consistency of the results after several runs of the stochastic model. This method consists on running several replications of the stochastic model to quantify the (internal) variability and the robustness of the model results. If model outcomes have a high variance, the model will neither be reliable nor useful for decision making.

One of the most commonly ways to validate a model is the parameter validity or more commonly referred to as sensitivity analysis. This technique consists on evaluating the influence that variation (changes) in the values of input parameters have on the model outcomes. The use of sensitivity analysis in model evaluation will help to identify the input parameters that more influence has on the model results, and for which good (realistic) estimates are highly recommended. Sensitivity analysis is one of the validation method most frequently used for disease-spread models and many examples are found in literature (Ezanno et al., 2007; Karsten et al., 2005b; Saatkamp et al., 1996; Jalvingh et al., 1999; Chitnis, et al., 2008; Hess et al., 2008).

The historical data validation is a method that uses historical information to determine if the model behaves as the real system does. In this case, the outputs generated by the model are compared with data from real outbreaks. Some examples of this type of validation can be found in Jalvingh et al. (1999) and Saatkamp et al. (1996).

Other very common method to validate a model is the comparison to other (valid) models or "docking". This method consists in comparing methods and results of the proposed (new) model with other models that have been validated. If the models compared produce similar results, even if they were developed independently or with different

methods, the confidence and credibility on the model increase. Some examples of docking have been presented in Dubé et al. (2007).

Finally, we will describe other important step in the process of model evaluation, which is data validity. Data validity is the process that ensures that the data necessary for model building and implementation is complete and correct. Although data validity is usually not included in the model validation, it is importatn in order to ensure the correctness of the model performance. In this manuscript we have perturbed the input data to assess the impact that the use of incomplete or not updated information may have on model results.

In general, it is costly and time consuming to determine if a model is absolutely "valid". However, the use of one of more of the validation methods described above can help to assess the model behaviour and credibility. Most of the previously published stochastic models have used one or two of those validation methods to verify the soundness of the model. In this study we have used all the methods described above to provide a clear understanding of the performance and robustness of our model. Methods and results of this study are intended to guide decision makers in the better application and interpretation of the Be-FAST model, which ultimately will help to improve the prevention and control of future CSFV outbreaks. The procedure presented here may also be useful for the evaluation of other stochastic disease spread models.

2. Materials and methods

2.1. Brief description of the Be-FAST model

In this section, we briefly recall the main characteristics of the Be-FAST model, which has been described in detail in Martínez-López et al., (2011) and Ivorra et al., (submitted). The main objective of the Be-FAST model is to evaluate the daily spread of CSFV within- and between- farms into a specific region.

At the beginning of the simulation, the model parameters used to simulate the CSFV spread, the detection and control of the disease are set by the user (see Table 1).

From a general point of view, Be-FAST is based on a Monte Carlo approach that generates *NS* possible epidemic scenarios. More precisely, at the beginning of each scenario (i.e., at time t=0), all farms are in the susceptible state (i.e., pigs are free of CSFV) except one randomly selected farm, which is assumed to have one infectious pig and is classified as 'Infectious'. Then, during a time interval [0, T], T being the last simulated day, a within- and between-farm daily spread process (and control, after the detection of the index case) is applied.

The daily CSFV spread within a particular farm *i* is modelled by using a discrete time stochastic Susceptible-Infected model (see, for example, Brauer, F. et al., 2001 and Klinkenberg, D. et al., 2002), where the pigs in each farm are considered to be in one of the two possible states: 'Susceptible' or 'Infected'.

CSFV spread between farms is modelled by using a spatial stochastic individual based model (DeAngelis, D.L. et al., 1991; Karsten, S. et al., 2005a). In this model, farms are assumed to be in one of four possible states: 'Susceptible', 'Infected', 'Infectious' or 'Clinical signs'. The daily transition from 'Susceptible' to 'Infected' state is modelled by considering direct contacts (i.e. the movement of infected pigs between farms) or indirect contacts (i.e.

local spread, movement of vehicles transporting pigs or products or movement of people). The transition from 'Infected' to 'Infectious' and from 'Infectious' to 'Clinical Signs' states are modelled by using the latent and incubation period of CSFV, respectively.

A daily process simulating the detection of infected farms by Animal Health Authorities and the application of control measures dictated by the European Union legislation (Council Directive 2001/89/EC), which are based on zoning and movement restriction, tracing and stamping-out, is also implemented.

If at the end of a simulation day, the CSFV epidemic disappears, the scenario is stopped and we start the simulation of the next scenario.

At the end of the simulation (i.e., when the scenario number *NS* is finished) for each scenario, various outputs, denoted by O_k, are generated and analyzed. The outputs considered here are: the number of infected farms (O_1); the duration of the epidemic in days (O_2); the percentage of infections due to local spread (O_3), to animal transport (O_4), to movement of people (O_5) and, to the integrator group (i.e. vehicles, materials and other fomites) (O_6); the percentage of detection of infected farms due to zoning (O_7), to observation of clinical signs (O_8) or to tracing (O_9); the number of farms included into the control or surveillance zones (O_{10}); the number of traced farms (O_{11}). For all those outputs, we compute their mean valueand their 95% prediction interval (denoted by 95% PI) obtained considering all scenarios.

Furthermore we compute the basic reproduction ratio of each farm *i*, denoted by R0(*i*), which is defined as the number of times that a farm *i* infects another farm in 'Susceptible' state considering all scenarios; and the risk of CSFV introduction into each farm *i*, denoted by RI(*i*), which is defined as the number of times that farm *i* becomes infected considering all scenarios (Anderson R.M., et al., 1979; 1991). The values of MR0 (O_{12}) and MRI (O_{13}), denote the mean R0 and RI values considering all farms.. In addition, we generate the spatial distribution of R0 and RI in the considered region, by interpolating the R0(*i*) and RI(*i*) values obtained for each farm *i*.

These results obtained for the model without considering perturbation in the parameters , which will be referred to as "reference scenario", will be compared to those obtained by each experiments performed during the model validation process.

2.2. Internal validity

Because the Be-FAST model is based on the combination of various stochastic processes, we were firstly interested in studying the variation of the output values from one execution to another. From a general point of view, as our model is based on a Monte-Carlo approach, a large number of scenarios (i.e., a high value for *NS*) should be considered to ensure a good stability of the outputs between two different runs (Ivorra, B. et al., 2009) However, the larger the value of *M* the higher the resources and computational time required to obtain results, mainly if the model is complex. Thus, for the experiments presented here, we have considered an intermediate number of scenarios (*NS*=1000), which was considered to guaranty a certain stability of the outputs while requiring a reasonable computational time.

In order to check the robustness of the BE-FAST outputs, we ran 10 times the model keeping the same input values for the parameters as described in Martinez-López et al. (2011). Then, for each output O_k, we compute its maximum ($MA(O_k)$), minimum ($MI(O_k)$)

and mean ($M(O_k)$) values obtained during all those runs. We have also defined a mean relative error measure, denoted by ME, expressed in percentage respecting to $M(O_k)$ (which is taken as the reference value of the output O_k), by considering for all outputs, the following formula (Infante, J.A. et al., 2009):

$$\mathrm{ME}(O_k) = \frac{1}{10}\sum_{i=1}^{10}\frac{|\,\mathrm{V_i}(O_k) - \mathrm{M}(O_k)\,|}{\mathrm{M}(O_k)}, \qquad (1)$$

where V_i (O_k) is the mean value of the output O_k obtained in the i^{th} run of the model. by

Finally, the R0 and RI distribution maps obtained during the 10 experiments are compared to the reference R0 and RI maps (obtained by interpolating the mean value of R0(i) and RI(i) considering the 10 experiments) by using thePearson correlation coefficient, R^2 (Aitken, A.C., 1957).

2.3. Data validity

Information used to feed the model (i.e. farm type, geographical position, number of animals, ADS, integrator groups and pigs movements), which was described by Martinez-Lopez et al. (2011), was provided by the Regional Government of Castile and Leon Region of Spain and was considered to be complete, updated and reliable. However, we intended to evaluate the impact that incomplete or not updated information regarding the farm demographics and characteristics (i.e. number and type of farms, incoming and outgoing movements of pigs and number of pigs per farm) has on model outcomes. To do so, we have performed 10 experiments considering the information regarding the farm demographics and characteristics from two different years (2005 and 2008). More precisely, for each experiment, we first have generated a number of farms of the order of the 2008´s one (1401 farms) by considering a Poisson distribution with mean 1401. Then, we have chosen 10% of the farms in the 2005 database the remaining farms in the 2008 database. At the end of those 10 experiments, the mean values of the outputs O_k were compared to their respective reference values $M(O_k)$, obtained during experiments presented in Section 2.2, by considering the error formula (Eq. 1). The obtained R0 and RI maps are also compared to the reference R0 and RI maps. The objective was not only to assess the impact of using “old” information on model outcomes but also the impact that changes in farm demographics and characteristics has in the spread of CSFV.

2.4. Sensitivity analysis

The input values for the 33 parameters used to simulate the within- and between-farms transmission processes and the detection and control of CSFV were obtained either from literature review or from expert opinion (Table 1). Whereas some of these inputs (from 1S to 11M and 22M) are well documented and used in other published models; other input values (from 12M to 33M, except 22M) are either not so well documented or based on potential subjective opinions (i.e., expert opinion). In any case, both (well documented or not) input values are likely to impact results and, therefore, should be carefully evaluated.

In this section, we used sensitivity analysis (SA) to quantify the amount of change on outcomes when varying the input values used in the model. Specifically, we evaluate three aspects of the model: (1) the global behaviour of the model when perturbing the whole set of parameters; (2) the impact of changes on each of the 33 parameters used in the model; and (3) the impact of deactivation of one infection route or one control measure (i.e. deactivation of group of parameters). Next we give the details of these three cases under study.

a) Sensitivity analysis of all parameters using a random perturbation of 10%

Firstly, we aim to study the model behaviour when the whole set of parameters was randomly perturbed. To do so, we ran the Be-FAST model perturbing randomly all the model parameters by a variation between [-10%,+10%] of their reference value. This experiment was repeated 10 times, and the mean values of the output O_k was compared to ME(O_k) by considering the error formula (Eq. 1). Finally, the R0 and RI maps are compared to the reference maps.

b) Sensitivity analysis of the individual parameters used in the model

In this section we intended to identify the most influential parameters in the model. This sensitivity analysis was performed by perturbing every single parameter +/-80% their initial values. The mean values obtained for each O_k were compared with the ones obtained in the reference model by considering:

$$\mathrm{E}(O_k) = \frac{|\,\mathrm{V}(O_k) - \mathrm{M}(O_k)\,|}{\mathrm{M}(O_k)}, \qquad (2)$$

where V (O_k) is the mean value of the output O_k obtained in the considered run of the model.

c) Sensitivity analysis of a set of parameters involved in the within- and between-farm transmission, detection and control processes of CSFV

Here we evaluate the evolution of the epidemic when one of the infection routes or one of the control measures was neglected. Specifically we perform seven experiments: deactivation of transmission (1) by local spread (i.e parameter 7S set to 0), (2) by animal transport (2S set to 0), (3) by contact by persons (5S, 6S and 29M set to 0), (4) by contact with vehicles (3S, 4S and 30M set to 0), and deactivation of measure of (5) zoning (parameters 31M and 33M were set to 0), (6) restriction of movements (15M to 21M and 24M set to 0); and (7) tracing (25M-28M and 32M to 0). In addition, we also evaluated the potential impact of not considering the within-farm transmission component, which was equivalent to consider a pure between-farm transmission model. This was done by considering all animals to be infected as soon as a farm becomes infected (parameter 1S set to $+\infty$). The obtained mean outputs were compared with their reference value using the error formula (Eq. 2).

2.5. Validation using historical data

Validation of the Be-FAST model was performed by using data of the 1997-1998 CSFV epidemic in Segovia, which was provided by the Regional Goverment of Castile and Leon region of Spain. Information consist on the unique identifier code of the CSFV infected farm, location of the farm (i.e. latitude and longitude of the farm centroid), production type of farm (i.e. farrowing, fattening or farow- to- finish), number of pigs on farm (i.e. farm size), day of confirmation of the CSFV infection on farm by the Official Laboratory and, day of farm depopulation. Validation of the model was performed by comparing the magnitude, duration and geographical location of the real epidemic with the simulated results obtained with the Be-FAST model.

2.6. Comparison with other models

Methods and results were compared with other three published CSFV spread models. These three models were developed for Germany (Karsten et al., 2005a and b.), The Netherlands (Jalvingh et al. 1999) and Belgium (Saatkamp et al. 1996), which were countries where the pig demographics and epidemiological conditions were assumed to be similar to the ones observed in Spain. Because all those are stochastic models that simulate only the between farm spread process, we also compare the output of those models with the one obtained by our model when neglecting the within-farm transmission component.

3. Results

3.1. Internal validity

The mean error ME(O_k) value obtained for each considered output O_k after the ten runs of the model was about 3% (Table 2). The highest mean error was obtained for the proportion of infections due to people (5.5%) and the RIvalue (5.2%). The range (max-min) of the mean value for each output O_k was very small, with a mean value of 3.1% (Figure 1). The distribution of mean error for the RI and R0 values was mainly concentrated in the areas with high pig density but, in general, was similar to the reference values with values for the Pearson correlation coefficient of R^2=0.97 and R^2=0.99, respectively (Figure 2).

3.2. Data validity

Results for the ten experiments described in Section 2.3 are presented in Table 2. The mean error value was of 13.54%. The highest mean error value was found in the proportion of infections due to people (27.1%), the proportion of detections due to tracing (19.9%) and the proportion of infections due to animal movements (19.6%). The range (max-min) of the mean value for each output O_k was of 11.5% (Figure 1), with the maximum ranges found in the MR0 (Range=32%) and the MRI (Range=31%). The distribution of the Risk and R0 mean error values was only in part comparable to the reference values, with a R^2 = 0.52 and R^2 = 0.48, respectively (Figure 2).

3.3. Sensitivity analysis

a) Sensitivity analysis of all parameters using a random perturbation of 10%

All results are presented in Table 2. The mean error obtained after the perturbation of randomly selected parameters was 6.54%, which was of the order of the parameter perturbations (10%). The mean range (max-min) for each output O_k was of 10.3% (Figure 1), with the maximum ranges found in the number of farms affected by zoning (Range = 26%) and the number of traced farms (Range = 25%). The distribution of the mean error for the Risk and R0 values was not meaningfully different from the reference scenario (R^2 = 0.94 for both) (Figure 2).

b) Sensitivity analysis of the individual parameters used in the model

The most influencing parameters in the whole set of outputs were the probability of infection by local spread (7S), the time from infectious to clinical signs state (9S), the probability of detection based on clinical signs at day *t* after detection of the index case outside the control and surveillance zones (11M), the maximum number of farms to be depopulated at day *t* (23M), the probability of detection based on clinical signs at day *t* before detection of the index case (10M) and the latent period (8S) (Figure 3).

Specifically, the magnitude and duration of the CSFV epidemic was mostly impacted (>10% of change) by the probability of infection by local spread at day t (7S), the transition from infectious to clinical signs state (9S) and the probability of detection based on clinical signs at day t outside the control and surveillance zones (11M) (Figure 4). The maximum number of farms to be depopulated at day t (23M) was also very influential on the duration of the epidemic (Figure 4).

The proportion of infections due to local spread, pig movements, people and other fomites were mostly sensitive to the number of contacts with vehicles transporting products per farm at day t (3S), probability of infection by contacts with vehicles transporting products (4S), the probability of infection by local spread at day t (7S) and the number of farms visited by a person (29M) and by a vehicle (30M) during one trip (Figure 4).

The proportion of detections by zoning, clinical signs and tracing were sensitive to the probability of infection by local spread at day t (7S), the transition from infectious to clinical signs state (9S), the probability of detection based on clinical signs after detection of the index case (10M and 11M) and the number of farms visited by a person during one trip (29M) (Figure 5).

The most important parameters regarding the number of traced farms and farms affected by zoning were the probability of infection by local spread at day t (7S), the transition from infectious to clinical signs state (9S), the number of farms visited by a vehicle during one trip (30M) and the radius (km) applied for control and surveillance zones (31M) (Figure 5).

The mean Risk was sensitive mainly to the probability of infection by local spread at day *t* (7S) and the transition from infectious to clinical signs state (9S) (Figure5).

c) Sensitivity analysis of a set of parameters involved in the within- and between-farm transmission or detection and control of CSFV

The error values of the output generated by the Be-FAST model by deactivating one by one the ways of CSFV transmission or control measures are presented in Table 3.

The deactivation of the local spread transmission was the experiment that most affected the model outputs with a mean error of 43.38%. In particular, the magnitude and duration of the epidemic omitting this route reduces the number of infections and the epidemic length by 51.0% and 13.5%, respectively. Other ways of transmission such as animal movements or contacts with people or vehicles produced lower impact on the output with a mean error around 10%.

Similarly, zoning and tracing were the control measures that mostly impact model outcomes (ME ≈ 22%), producing an increment in the magnitude and duration of the epidemic from 15% to 20%.

Finally, the suppression of the set of parameters involved in the within-farm transmission lead to the most important increase in the magnitude (100.5%) of the CSFV epidemic, with an important increase also in the epidemic duration (18.2%). In this case, the mean error on the output was about 58%.

3.4. Validation using historical data

In the 1997-1998 the CSFV epidemic in the province of Segovia lasted 50 days, with a total of 22 farms infected and 29 indirectly affected by pre-emptive depopulation (Del Pozo, 2006; Martínez-López, B et al., 2007).

When comparing the real epidemic with the simulated epidemic, we observed that most (93%) of the confirmed outbreaks in 1997-1998 were located in areas identified as medium (Risk = 4-7) or high risk (Risk > 7) areas for CSFV introduction by the model (Figure 6). Specifically, 61.4% of the confirmed outbreaks in 1997-1998 were allocated in areas estimated to be at high risk of CSFV introduction, 32% in areas at medium risk and 7% in areas at low risk.

3.5. Comparison with other models

Comparison of the methods and main results of the Be-FAST model with the three models for The Netherlands, Belgium and Germany are presented in Table 4. The magnitude and duration of the CSF simulated epidemic was only comparable with the model presented by Karsten et al., (2005a and b). The other models presented much higher number of infected farms and epidemic duration. The infection due to local spread was the main route of infection, similar to results presented by Karsten et al., (2005b) and Jalvingh et al. (1999).

4. Discussion

The exhaustive validation process conducted in this study aimed to provide a better understanding of the performance of the Be-FAST model. The five methods used for the validation process, which were internal validity, data validity, sensitivity analysis, validation using historical data and model comparison, were intended not only to assess the

robustness and reliability of the Be-FAST model but also to identify the most influential parameters for which good (realistic) estimates are highly recommended.

4.1. Internal validity

The internal validity allowed to verify the consistency of the stochastic model after different runs. The small mean error value obtained (ME=3%) was found to be reasonably low considering the reduced number of Monte-Carlo simulations used [M=1000] and shows that this value of M gives a good ratio between computational complexity and output precision.

4.2. Data validity

As expected, the higher variation on model outputs was obtained after altering the input data used to feed the model (ME=13.54%) (Table 2). Moreover, the distribution of the areas at risk of introducing (High risk value) or spreading (High R0 value) the disease were importantly modified (almost 50%), compared with the reference scenario (Figure 1). These results highlight the importance of using updated and complete information regarding the area of study to obtain realistic and useful results for the decision making process.

4.3. Sensitivity analysis

a) Sensitivity analysis of all parameters using a random perturbation of 10%

The impact that variations (uncertainty) in the input values has on the model results was assessed by sensitivity analysis (Saltelli et al., 2008). The low mean error (6.54%) obtained after the 10% random perturbation of all parameters confirmed the robustness of the model to general variations on the input parameters.

b) Sensitivity analysis of the individual parameters used in the model

The sensitivity analysis performed by doing a strong perturbation (+/-80% the initial values) of the 33 input parameters intended to identify the most influential parameters in the model (Figure 3). As a result, the model results were found to be mostly sensitive (>10% of change) to six parameters: the probability of infection by local spread (7S), the time from infectious to clinical signs state (9S), the probability of detection based on clinical signs at day *t* after detection of the index case outside the control and surveillance zones (11M), the maximum number of farms to be depopulated at day *t* (23M), the probability of detection based on clinical signs at day *t* before detection of the index case (10M) and the latent period (8S). Results are consistent with previous studies if we consider that those parameters are related with the local spread (7S and 23M) and with the time from infection to detection of a CSFV infected farm, also referred to as the high risk period, (9S, 11M, 10M and 8S), which both has been traditionally identified as key-aspects to determine the magnitude and duration of an CSFV epidemic (Jalvingh et al. 1999; Nielen et al., 1999; Karsten et al., 2005b). Those results highlight the fact that studies that help to quantify the local spread and the time from infection to detection of CSFV in real epidemics, such as the one presented by Stegeman et al. (1999); (2002), are extremely useful to implement realistic

estimates for disease spread models which ultimately will help to better prevent and control future CSFV epidemics.

c) Sensitivity analysis of a set of parameters involved in the within- and between-farm transmission or detection and control of CSFV

The aim of the deactivation of a set of parameters involved in the spread, the detection and the control of CSFV was, firstly, to identify the most important routes of disease spread and, secondly, the most effective detection and control measures to be applied during a CSFV epidemic. As a result, we found that local spread was the most important way of CSFV transmission in Segovia region. Also, the elimination of the within-farm transmission component was found to produce a much larger epidemic. This was an expected result as other influent parameters (such as 7S) directly depend on the number of infected animals in the farm. Moreover, deactivation of the SIR component implies that all animals in a farm become infected immediately (at time 0) after the infection of the farm, which directly increases the probability of CSFV transmission by any route from this farm to any other farms. This result also reveals that the simplification of the model to a purely between-farm transmission model, will lead to an overestimation of the epidemic size and duration. In fact, the use of a combined within- and between- farm spread model will produce two times smaller epidemics than a simple between-farm transmission model. Policy makers should be aware of this potential overestimation of the simple between-farms spread models before interpreting and using the model for allocation of preventive and control measures.

On the other hand, tracing was considered the most effective measure to control the disease spread, because its suppression leaded to the major increase (+20%) in the magnitude and length of the epidemic. It is important to note that other the control measures were also important for the disease control, as their deactivation imply from 10% to 15% larger epidemics, but the role of tracing was crucial. This result reveals that the capabilities of the Animal Health Services to implement timing and effective tracing are extremely important to control disease spread in the CSFV infected regions and may certainly determine the final sanitary and economical consequences of a CSFV epidemic.

4.4. Historical validation

Not many differences in the areas at risk for introduction or spread of disease were found when comparing the simulated epidemic with the real epidemic of 1997-1998 in Segovia region. The difference could be explained at least in part by the differences in the number of farms in Segovia region from 1997 to 2008. In fact, in 1997 the number of farms was around 2.205 and in 2008 this number was only 1.401. Moreover, the epidemiological conditions and resources for tracing, control and depopulation may have changed a lot in the last ten years. In fact, there have been a dramatic changes of integrator groups in the last ten years, associated with the decline of '*Proinserga*' (Official Journal of the European Union, 21,8,2010; pigmeat.blog.com, 28-07-2008), an enterprise in Segovia, which produced not only a decrease on the number of farms but also a change in the structure of pig trade in Segovia (Official Journal of the European Union, 21,8,2010). As an expected consequence, the distribution of Risk and R0 values have been widely modified (almost 50%) during this time period.

4.5. Comparison with other models

Some agreements as well as differences were found when comparing our model with other available models. The work presented by Karsten et al., 2005a and b is the one most similar in methods and results to the Be-FAST model. In contrast, our outputs are quite different from the one described by Saatkamp et al. (1996) and Jalvingh et al. (1999). Those results can be explained by several reasons. Firstly, the values of the parameters that we used were obtained by recently published studies, similar to the values considered by Karsten, whereas in the other models (oldest ones) the coefficients were calibrated using the 1997-1998 epidemic in the Netherlands, which magnitud was dramatically high (Elbers, A., et al. 1999). Nowadays, European Animal Health Authorities are much better prepared to prevent and control CSFV outbreaks, mainly thanks to the evident improvement in tracing capabilities. Therefore, recent epidemics in EU countries have been much smaller compared with the ones occurring in the 90´s (OIE, 2011). Furthermore, from a modelling point of view, whereas the model proposed by Saatkamp et al. 1996 and Jalvingh et al. 1999 are based on the use of black-boxes (Quattro-Pro Spread Sheet and InterSpread), the model presented by Karsten et al., 2005a and b. was a self made C++ code and was the closest in the sense of programming, to our approach.

Specific conditions of Segovia region (i.e. pig density, direct and indirect contacts, etc.) may also explain some of the differences found when comparing outputs of our model to other models. Further analysis should be performed by using other regions than Segovia to fully evaluate the degree of agreements or disagreements of the Be-FAST with other models.

4.6. Future works

The next step would be to apply this model to other regions in order to perform a better comparison with other available models. Moreover, the methodology presented here will be extended by introducing an economical component and alternative control measures (i.e. vaccination, etc.) in order to provide an estimation of direct costs and to evaluate the cost-benefit of alternative measures in future CSFV epidemics. Finally, this model could be adapted to other diseases to provide a more useful and complete disease management and decision support system.

5. Conclusion

The exhaustive validation process presented in this study aimed to provide a better understanding of the behavior and performances of the Be-FAST model and to identify the most influential parameters. As a result, model was found to be robust to general perturbations of input parameters, but sensitive to changes in the input data used to feed the model and to removal of the within-farm spread component. Parameters related with the local spread and to the time from infection to detection of disease were found to be the most influential in the model outputs. This reveals the need to incorporate good estimates for those parameters to get realist results. Methods and results presented here may be useful for decision makers to better prevent and control future CSFV outbreaks.

Acknowledgements

The Project was funded by the Projects MTM2008-04621 and CONS-C6-0356 of the Spanish Ministry of Science and Innovation, the University Complutense of Madrid (Reseach group 910480), the Regional Government of Castile and Leon Region (JCyL) and the Spanish Ministry of Environment and Rural and Marine Affairs (MARM).

References

AIAA (American Institute of Aeronautics and Astronautics), 1998. Guide for the Verification and Validation of Computational Fluid Dynamics Simulations. AIAA-G-077-1998, Reston, VA.

Aitken, A.C, 1957. Statistical mathematics. Oliver & Boyd; 8th Revised edition.

Anderson, R.M., May, R.M., 1979. Population biology of infectious diseases: Part I. Nature. 280: 361-367.

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans. Oxford, U.K.: Oxford University Press.

Brauer, F., Castillo-Chavez, C., 2001. Mathematical Models in Population Biology and Epidemiology. Springer.

Chitnis N., Hyman J.M., Cushing J.M., 2008. Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. Bull Math Biol 2008, 70, 1272-1296

DeAngelis, D.L., Gross L.J. (Editors), 1991. Individual-based Models and Approaches in Ecology. Chapman and Hall, NY.

Del Pozo, M. 2006. Estudio del brote de peste porcina clásica ocurrido en Castilla y León en los años 1997-1998 [Study of the classical swine fever epidemic occurred in 1997-1998 in Castile and Leon]. PhD tesis.

Dubé, C., Stevenson, M., Garner, M., Sanson, R., Corso, B., Harveyl, N., Griffin, J., Wilesmith, J., Estrada. C., 2007. A comparison of predictions made by three simulation models of foot-and-mouth disease. New Zealand Veterinary Journal, 55, 280–288.

Elbers, A.T.W., Stegeman, A., Moser, H., Ekker, H.M., Smak, J.A., Pluimers, H., 1999. The classical swine fever epidemic 1997-1998 in the Netherlands: descriptive epidemiology. Prev. Vet. Med., 42, 157-184.

Ezanno P, Fourichon C, Viet AF, Seegers H., 2007. Sensitivity analysis to identify key-parameters in modelling the spread of bovine viral diarrhoea virus in a dairy herd. Prev Vet Med. 80, 49-64.

Hess G.D., Garner M.G., Yang X. 2008. A sensitivity analysis of an integrated modelling approach to assess the risk of wind-borne spread of foot-and-mouth disease virus from infected premises. Environ. Model. Assess.13, 209–220.

Infante, J.A., Ivorra, B., Ramos, A.M., Rey, J.M., 2009. On the Modelling and Simulation of High Pressure Processes and Inactivation of Enzymes in Food Engineering. Mathematical Models and Methods in Applied Sciences, 19, 2203-2229.

Ivorra, B., Martínez-López, B., Sánchez-Vizcaíno, J.M., Ramos, A.M., 2011 (submitted). Modeling and simulation of Classical Swine Fever Virus spread between and within farms. Annals of Operations Research. In press.

Ivorra, B., Mohammadi, B., Ramos, A.M., 2009. Optimization strategies in credit portfolio management, Journal Of Global Optimization, 43, 415-427.

Jalvingh, A.W., Nielen, M., Maurice, H., Stegeman, A.J., Elbers, A.R.W., Dijkhuizen, A.A., 1999. Spatial and stochastic simulation to evaluate the impact of events and control measures on the 1997-1998 classical swine fever epidemic in The Netherlands. I. Description of simulation model. Prev. Vet. Med, 42, 271-295.

Karsten, S., Rave, G., Krieter, J., 2005a. Monte Carlo simulation of classical swine fever epidemics and control: I. General concepts and description of the model. Vet. Microb., 108, 187-198.

Karsten, S., Rave, G., Krieter, J., 2005b. Monte Carlo simulation of classical swine fever epidemics and control: II. Validation of the model. Vet. Microb., 108, 199-205.

Klinkenberg, D., De Bree, J., Laevens, H., De Jong, M.C.M., 2002. Within and between-pen transmission of Classical Swine Fever Virus: a new method to estimate the basic reproduction ration from transmission experiments. Epidemiol. Infect. 128, 293–299.

Kopec, J.A., Finès, P., Manuel, D.G., Buckeridge, D.L., Flanagan, W.M., Oderkirk, J., Abrahamowick, M., Harper, S., Sharif, B., Okhmatovskaia, A., Sayre, E.C., Rahman, M.M., Wolfson, M., 2010. Validation of population-based disease simulation models: a review of concepts and methods. doi:10.1186/1471-2458-10-710. BMC Public Health, 10:710.

Martínez-López, B., Del Pozo, M., Sánchez-Vizcaíno, J.M., 2007. Spatial analysis and modeling of Classical Swine Fever outbreaks during 1997 in the Spanish Province of Segovia. GisVet proceedings.

Martínez-López, B., Ivorra, B., Ramos, A.M., Sánchez-Vizcaíno, J.M., 2011. A novel spatial and stochastic model to evaluate the within- and between-farm transmission of classical swine fever virus. I. General concepts and description of the model. Vet. Microb., 147, 300–309.

Martínez-López, B., Perez A.M., Sánchez-Vizcaíno J.M., 2010. A simulation model for the potential spread of foot-and-mouth disease in the Castile and Leon region of Spain. Prev Vet Med. 96, 19-29.

Nielen, M., Jalvingh, A.W., Meuwissen, M.P.M, Horst, S.H., Dijkhuizen, A.A., 1999. Spatial and stochastic simulation to evaluate the impact of events and control measures on the 1997-1998 classical swine fever epidemic in The Netherlands. II. Comparison of control strategies. Prev Vet Med 42, 297-317.

Saatkamp,H.W., Huirne, R.B.M., Geers, R., Dijkhuizen, A.A., Noordhuizen, J.P.T.M., Goedseels, V., 1996. State-Transition Modelling of Classical Swine Fever to Evaluate National Identification and Recording Systems -General Aspects and Model Description. Agricultural Systems, 51, 215-236.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D. Saisana, M., and Tarantola, S., 2008, Global Sensitivity Analysis. The Primer, John Wiley & Sons.

Sargent, R.G. 1998. Verification and validation of simulation models. Proceedings of the 1998 Winter Simulation Conference.

Sargent, R.G., 2001. Some approaches and paradigms for verifying and validating simulation models. Proceedings of the 2001 Winter Simulation Conference.

Stegeman, J.A., Elbers, A.R.W., Smak, J., De Jong, M.C.M., 1999. Quantification of the transmission of classical swine fever virus between herds during the 1997-1998 epidemic in The Netherlands. Prev. Vet. Med. 42, 219-234

Stegeman, J.A., Elbers, A.R.W., Bouma, A., De Jong, M.C.M., 2002. Rate of inter-farm transmission of classical swine fever virus by different types of contact during the 1997-8 epidemic in The Netherlands. Epidemiol. Infect., 128, 285-291.

Tracker, B.H., Doebling, S.W., Hemez, F.M., Anderson, M.C., Pepin, J.E., Rodriguez, E.A., 2004. Concepts of Model Verification and Validation. Ed. Los Alamos, National Laboratory. LA-14167-MS.

Table 1. Assumptions and parameters used in the Be-FAST model for the within- and between- farm spread (S) and control measures (M) of the CSFV.

ID	Parameter	Initial Value	Reference
1S	Within-farm transmission parameter for farrowing pig farms, fattening pig farms and farrow-to-finish pig farms	β_h = 8.52; β_h = 1.85 and β_h = 5.18, respectively	Klinkenberg et al., 2002
2S	Probability of infection by contact with vehicles transporting infected animals	Bernoulli [0.011]	Stegeman et al., 2002
3S	Number of contacts with vehicles transporting products per farm at day *t*	Poisson [0.4]	Karsten et al., 2005a
4S	Probability of infection by contacts with vehicles transporting products	Bernoulli [0.0068]	Stegeman et al., 2002
5S	Number of contacts with people per farm at day *t*	Poisson [0.3]	Karsten et al., 2005a
6S	Probability of infection by contact with people	Bernoulli [0.0065]	Stegeman et al., 2002
7S	Probability of infection by local spread at day *t*	$\frac{Ip_h(t)}{N_h(t)} * LSM(h,k)$	Karsten et al., 2005b
8S	Latent period (transition from infected to infectious state)	Poisson [7]	Karsten et al., 2005a
9S	Transition from infectious to clinical signs state	Poisson [21]	Karsten et al., 2005a
10M	Probability of detection based on clinical signs at day *t* before detection of the index case	Bernoulli [0.03]	Karsten et al., 2005b
11M	Probability of detection based on clinical signs at day *t* after detection of the index case outside control and surv. zones.	Bernoulli [0.06]	Karsten et al., 2005b
12M	Probability of detection based on clinical signs at day *t* after detection of the index case in the control zone	Bernoulli $[0.98\left(\frac{Ip_h(t)}{N_h(t)}\right)]$	CyL expert opinion, 2008
13M	Probability of detection based on clinical signs at day *t* in the surveillance zone	Bernoulli $[0.95\left(\frac{Ip_h(t)}{N_h(t)}\right)]$	CyL expert opinion, 2008
14M	Probability of detection based on serological test	Bernoulli [0.95]	MAPA, 2006
15M	Probability of restriction of animal movements on the detected as infected farm	Bernoulli [0.99]	CyL expert opinion, 2008
16M	Probability of restriction of vehicle movements on the detected as infected farm	Bernoulli [0.95]	CyL expert opinion, 2008
17M	Probability of restriction of people movements on the detected as infected farm	Bernoulli [0.80]	CyL expert opinion, 2008
18M	Probability of restriction of animal movements within the control and surveillance zone	Bernoulli [0.95]	CyL expert opinion, 2008
19M	Probability of restriction of vehicle movements within the control and surveillance zone	Bernoulli [0.90]	CyL expert opinion, 2008
20M	Probability of restriction of people movements within the control and surveillance zone	Bernoulli [0.70]	CyL expert opinion, 2008
21M	Probability of restriction of movements outside the control and surveillance zones	Bernoulli [0.4]	CyL expert opinion, 2008
22M	Probability to depopulate a detected as infected farm	Table [prob,day]: [0.11,0], [0.58,1], [0.2,2], [0.06,3], [0.04,4], [0.004,5], [0.003,6], [0.0015,7] and [0.0015,8]	Elbers et al., 1999
23M	Maximum number of farms to be depopulated at day *t*	Poisson [20]	CyL expert opinion, 2008
24M	Time to repopulation of a depopulated farm	Poisson [90]	CyL expert opinion, 2008
25M	Maximum number of farms to be traced at day *t*	Poisson [60]	CyL expert opinion, 2008
26M	Probability of tracing an animal movement	Bernoulli [0.99]	CyL expert opinion, 2008
27M	Probability of tracing a vehicle/people movement	Bernoulli [0.70] and Bernoulli [0.40]	CyL expert opinion, 2008
28M	Duration (days) of general movement restriction	30	CyL expert opinion, 2008
29S	Number of farms visited by a person during one trip	3	CyL expert opinion, 2008
30S	Number of farms visited by a vehicle during one trip	4	CyL expert opinion, 2008
31M	Radius (Km) applied for control and surveillance zones	3km/10km	CyL expert opinion, 2008
32M	Number of days for tracing	60	CyL expert opinion, 2008
33M	Duration (days) of control and surveillance zones	30/40	CyL expert opinion, 2008

Table 2. Mean value and mean relative error value (%), see equation (1), of the outputs generated by performing the internal validity (IV), the data validity (DV) and when perturbing randomly by +/-10% the parameters of the Be-FAST model (P10).

ID	Output	Mean Value	IV	DV	P10
O_1	Number of infected farms	3	3.41	11.41	6.49
O_2	Epidemy length (days)	56	1.66	2.37	3.56
O_3	Infection due to local spread (%)	68.8	1.50	4.41	3.85
O_4	Infection due to animal movements (%)	7.7	3.68	19.58	11.40
O_5	Infection due to people (%)	6.7	5.47	27.09	11.57
O_6	Infection due to Integrator group (%)	16.8	3.84	11.24	7.73
O_7	Detection due to zoning (%)	35	2.53	18.16	10.84
O_8	Detection due to clinical signs (%)	40	1.49	4.67	2.59
O_9	Detection due to tracing (%)	25	2.59	19.91	5.11
O_{10}	Number of farms affected by zoning	123	2.19	6.61	7.33
O_{11}	Number of traced farms	82	3.95	16.70	5.57
O_{12}	Mean R0 value (MR0)	1.75	2.62	17.9	3.07
O_{13}	Mean Risk value (MRI)	1.76	5.16	16	5.90
Mean			**3.08**	**13.54**	**6.54**

Table 3. Error values (%) of the outputs generated in the Be-FAST model by deactivating the transmission by animal movements (AM), by local spread (LS), by contact with people (PE) and by contact with vehicles (V); the measures of tracing (TC), zoning (ZO) and restriction of movements (RM); and the within-farm transmission process (SIR) .

ID	AM	LS	PE	V	TC	ZO	RM	SIR
O_1	7.17	51.04	7.17	9.94	20.66	15.89	10.51	100.47
O_2	0.70	13.51	0.70	2.34	15.31	19.42	1.97	18.20
O_3	4.42	*	4.42	6.61	4.12	8.75	8.90	12.43
O_4	*	10.22	*	22.84	1.61	39.08	30.74	21.99
O_5	33.74	14.89	33.74	41.66	51.87	34.66	22.75	22.02
O_6	4.69	10.62	4.69	*	3.02	4.09	13.31	32.09
O_7	1.47	91.15	1.47	1.75	24.61	*	2.09	19.64
O_8	5.44	97.43	5.44	4.42	21.38	28.95	9.13	49.63
O_9	6.75	29.63	6.75	9.65	*	18.60	17.75	52.85
O_{10}	0.29	48.55	0.29	2.77	*	26.07	28.16	108.98
O_{11}	7.62	6.03	7.62	1.11	18.35	*	20.07	25.91
O_{12}	10.17	79.38	10.17	16.33	33.33	31.02	16.38	159.72
O_{13}	11.17	79.61	11.17	15.36	32.63	30.56	16.42	158.88
Mean	**7.87**	**43.38**	**9.67**	**11.46**	**20.37**	**22.56**	**15.08**	**58.23**

* = Not applicable.

Table 4. Comparison of the methods and results of three independent models developed in Belgium, Netherlands and Germany with those of the Be-FAST model considering or not the within farm spread (SIR) component.

	Item	Be-Fast	Be-Fast (without SIR)	Saatkamp et al. 1996	Jalvingh et al. 1999	Karsten et al., 2005a and b.
Methods	Unit of analysis	Pig and farm (within and between- farms spread)	Farm (between-farm spread)	Farm (between- farm spread)	Farm (between-farm spread)	Farm (between-farm spread)
	Technique	Monte Carlo	Monte Carlo	Markov chain	Monte Carlo	Monte Carlo
	Simulations (runs)	1000	1000	1	100	100
	Environment	MATLAB	MATLAB	Quattro-Pro	InterSpread	C++
	Country of study	Spain (province of Segovia)	Spain (Segovia province)	Belgium	The Netherlands	Germany (fictitious province)
	Ways of transmission used	Local spread, pig movement, vehicle and person contacts	(=)	(=)	(=)	(=)
	Control measures applied	Zoning and restriction of movements, stamping-out and tracing	(=)	Two scenarios: I → (=); II → (=) + pre-emptive slaughter	(=) + pre-emptive slaughter	(=) + pre-emptive slaughter
Results	O_1	3 [1, 17]	6 [1, 27]	I →389 II→39	[1] 465 [268, 2087]	[2] 5
	O_2	56 [26, 177]	66 [27, 183]	I →>365 II→112	[1] 306 [254, >365]	[2] 71
	O_3	68.8% [0,100]	77.4% [0, 100]	*	[1] 77.4% [44.1, 100]	*
	O_4	7.7% [0, 100]	6.0% [0, 100]	*	[1] 0% [0, 4]	*
	O_5	6.7% [0,100]	5.2%[0, 100]	*	[1] 6.7% [2.6, 25.6]	*
	O_6	16.8% [0,100]	11.4%[0,100]	*	[1] 6.9 [3.0, 18.7]	*
	O_7	35 [0, 100]	41.8% [0,100]	*	*	*
	O_8	40 [0, 100]	20.3% [0, 100]	*	*	*
	O_9	25 [0, 100]	37.9% [0, 100]	*	*	*
	O_{10}	123 [2, 394]	172 [0, 2464]	I →2,117 II→1,545	*	[2] 695
	O_{11}	82 [0, 1391]	154 [16, 457]	*	*	*
	O_{12}	1.75 [0, 9]	4.5 [0, 50]	*	[1] 1.26 [0.925, 1.625]	*
	O_{13}	1.76 [0, 9]	4.6 [0, 24]	*	*	*

[1] median and 90% IP; [2] mean; (=) = same that in the Be-Fast model;* = not available;
O_1=Number of infected farms; O_2=Epidemy length (days); O_3=Infection due to local spread (%); O_4=Infection due to animal movements (%);O_5=Infection due to people (%);O_6=Infection due to Integrator group (%);O_7=Detection due to zoning (%); O_8=Detection due to clinical sign detection (%); O_9=Detection due to tracing (%);O_{10}=Number of farms affected by zoning; O_{11}=Number of traced farms; O_{12}=Mean R0 value (MR0); O_{13}=Mean Risk value (MRI).

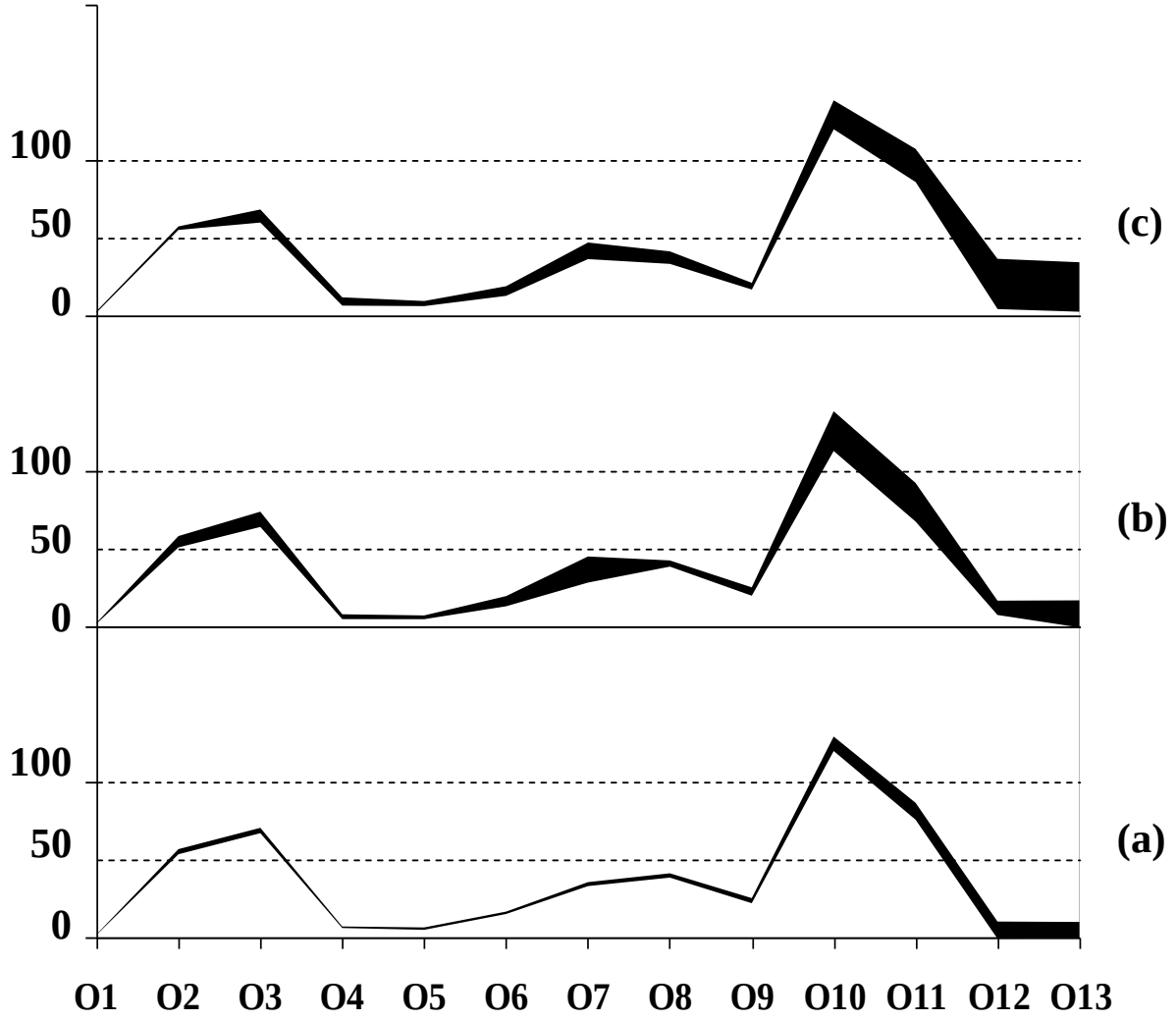

Figure 1. Range (min - max) for each output O_k obtained during (a) the internal validity, (b) the random perturbation of 10% of all parameters and (c) the data validity experiments.

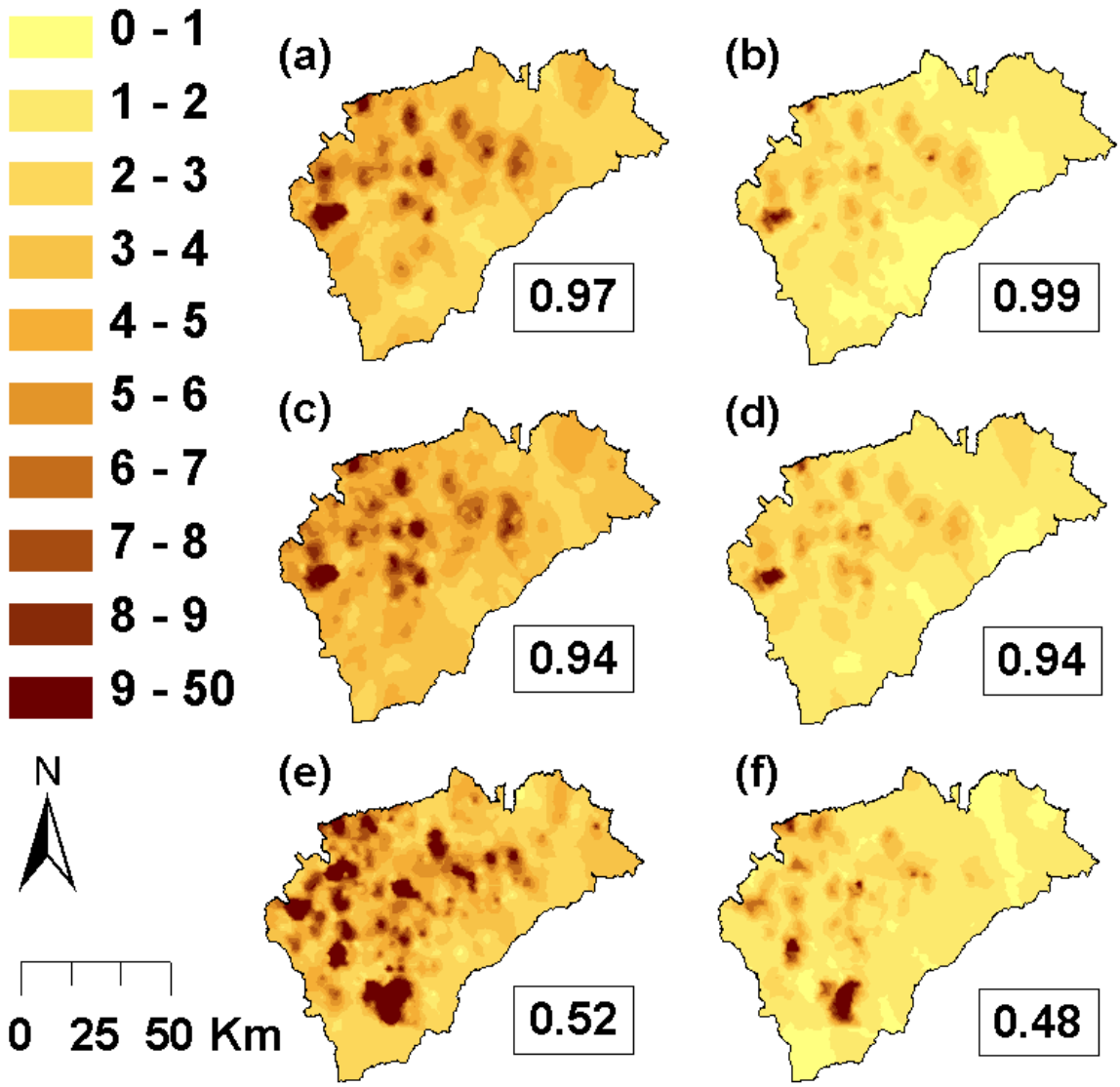

Figure 2. Spatial distribution of the mean error for the RI (Left) and R0 (Right) values obtained during the experiments of (a,b) the internal validity, (c,d) the 10% random perturbation of input parameters and (e,f) the data validity, respectively. The Pearson correlation coefficient (R^2) between the reference R0 and RI values and the values for each experiment is represented in the boxes.

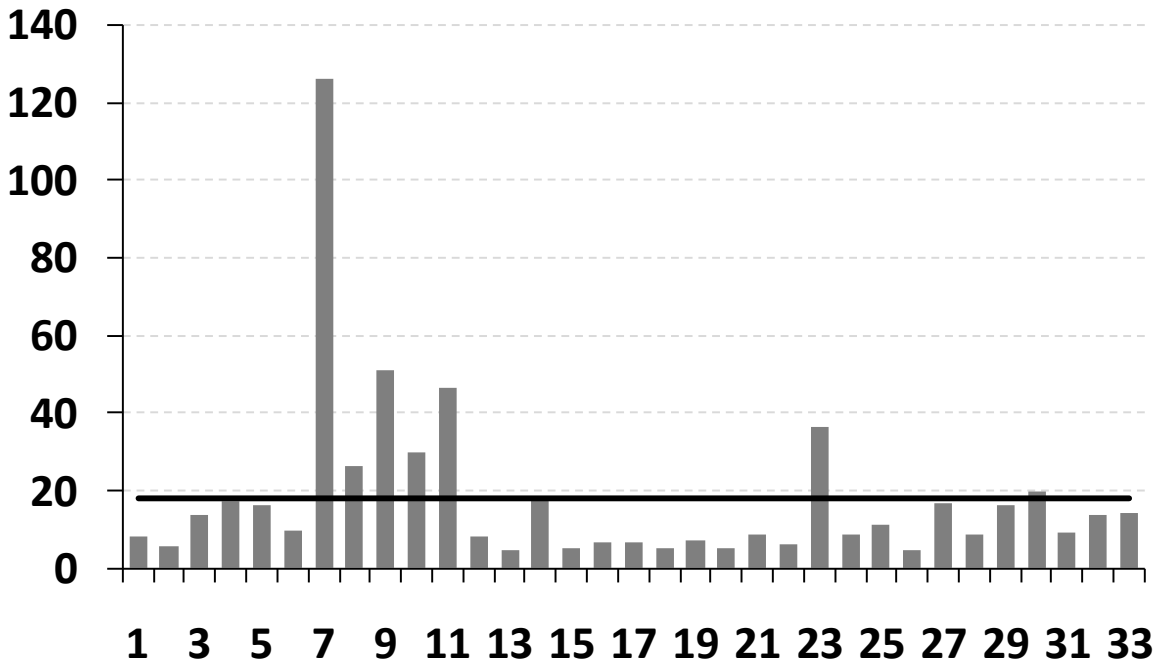

Figure 3. Mean error value of the outputs obtained by perturbing by +/-80% of their initial value each input parameters of the Be-FAST model. The mean value considering the 33 experiments is represented by a black line.

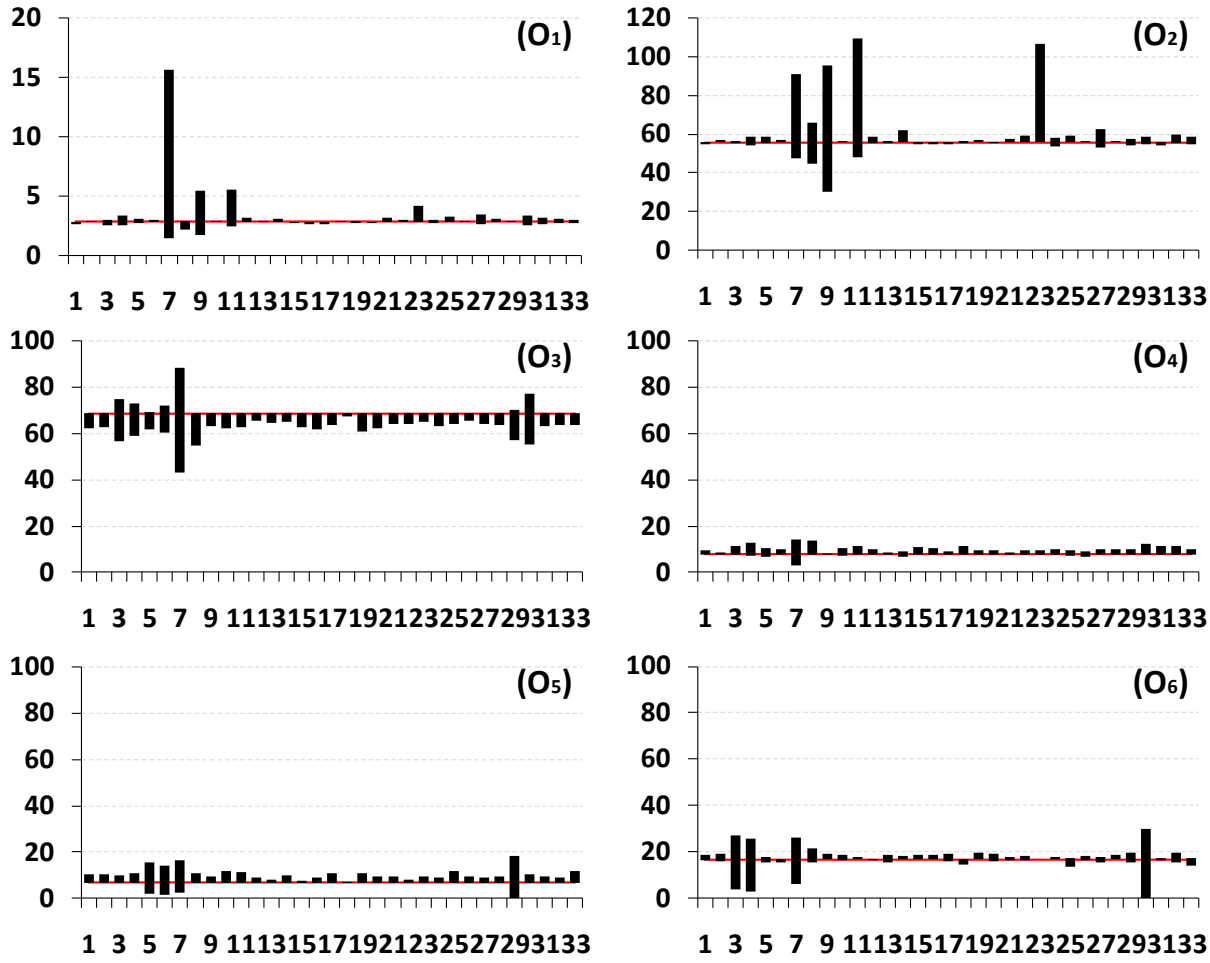

Figure 4. Bar plot representing the error range (min – max) for each outputs (O_1)- (O_6) obtained by perturbing +/-80% each input parameters used in the Be-FAST model.

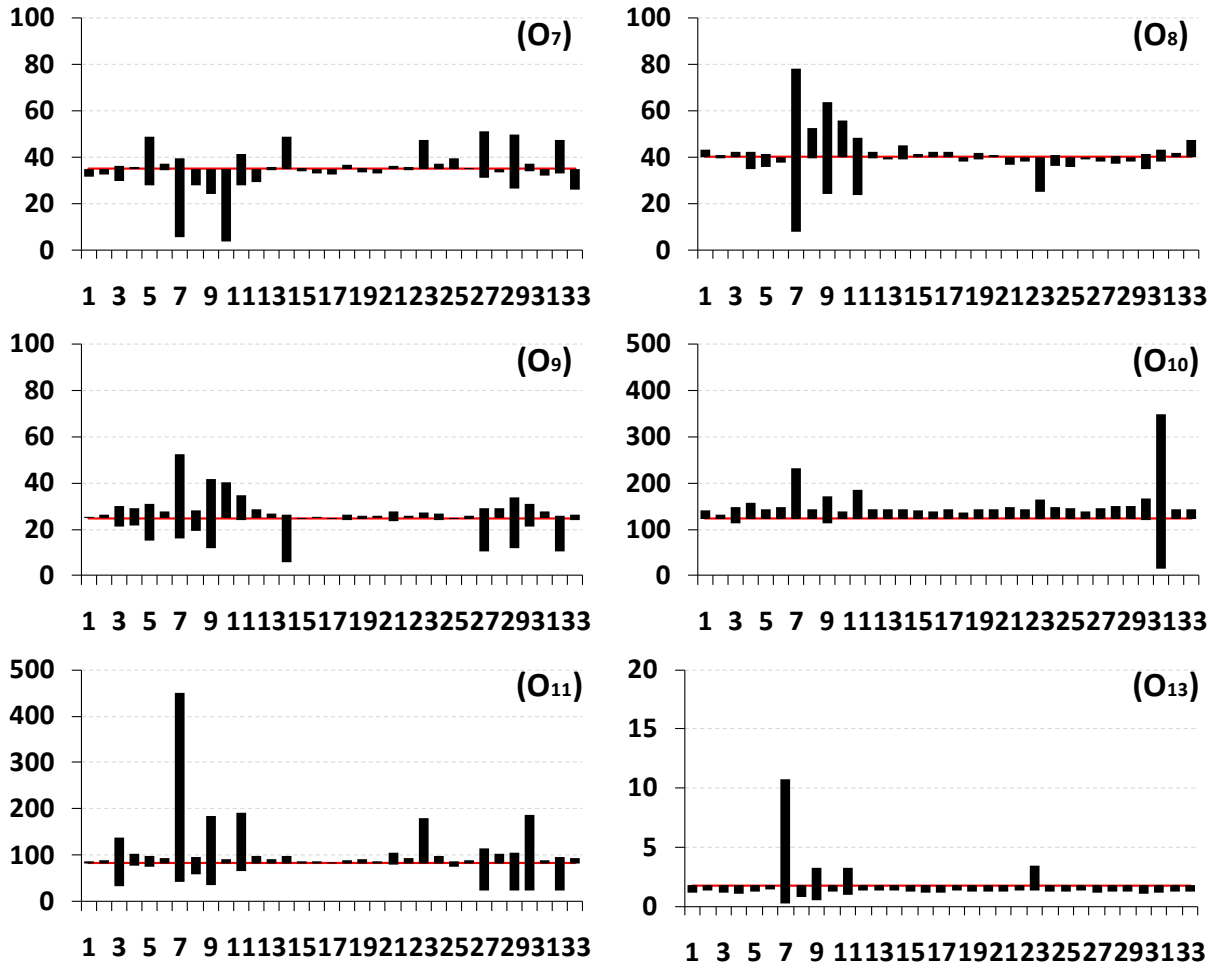

Figure 5. Bar plot representing the error range (min – max) for each outputs (O_7)- (O_{13}) obtained by perturbing +/-80% each input parameters used in the Be-FAST model. The Mean R0 (O_{12}) is not represented as it is very similar to the Mean Risk (O_{13}).

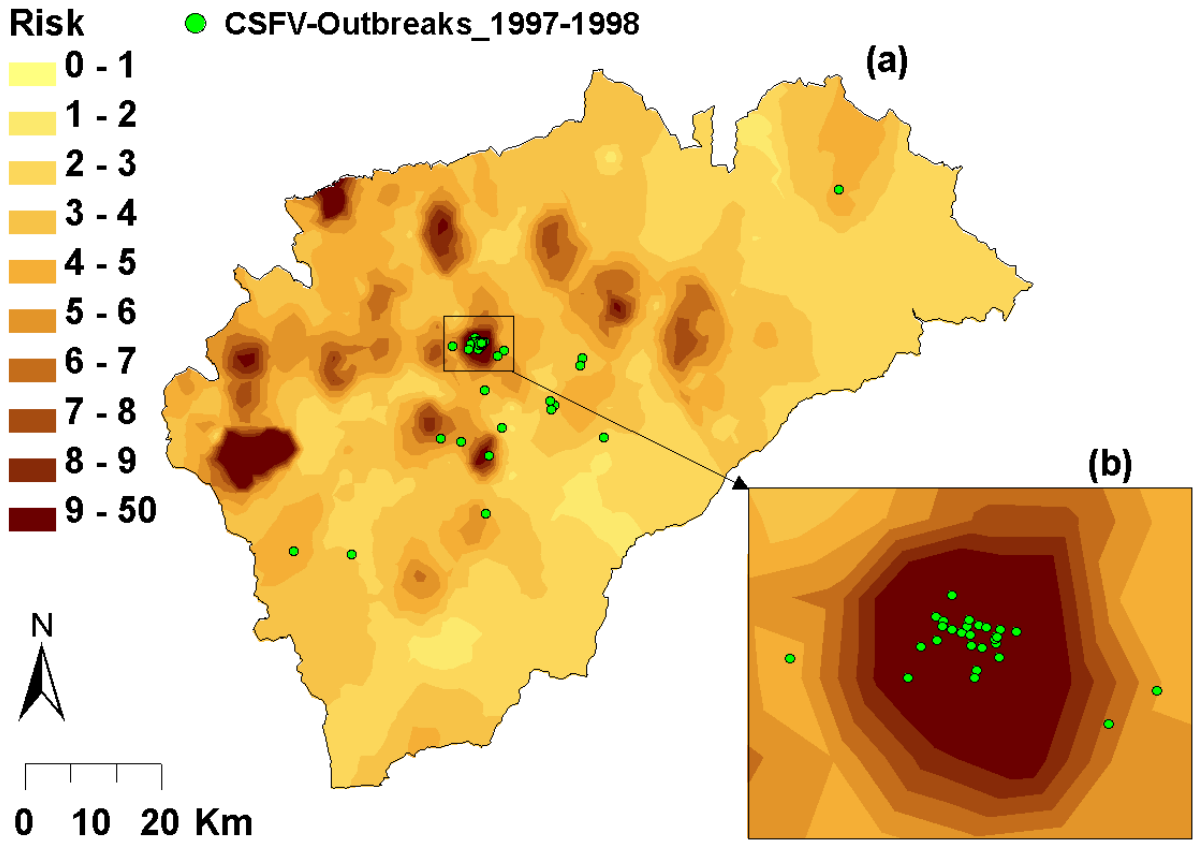

Figure 6. (a) Spatial distribution of the CSFV outbreaks occurring during 1997-1998 in the Spanish region of Segovia (green points) and the estimated RI map obtained in the Be-FAST model (background). (b) The most affected area in the CSF 1997-1998 epidemic, which allocated 69% of the outbreaks, has been highlighted.