

PAPER • OPEN ACCESS

## Analysis of the effects of uncertainties on agrometeorological models

To cite this article: A. Garinei *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **275** 012012

View the [article online](#) for updates and enhancements.

# Analysis of the effects of uncertainties on agrometeorological models

**A. Garinei<sup>1</sup>, A. Marini<sup>2</sup>, E. Piccioni<sup>3</sup>, M. Marconi<sup>1</sup>, L. Biondi<sup>1</sup>, A. Bertolini<sup>4</sup> and G. Ferrari<sup>4</sup>**

<sup>1</sup> Guglielmo Marconi University, Department of Sustainability Engineering, via Plinio 44, 00193, Roma, Italy

<sup>2</sup> INFN Section of Perugia, Via A.Pascoli, 06123, Perugia, Italy

<sup>3</sup> Idea-re S.r.l., Via Cornelia, 498, 00166, Roma, Italy

<sup>4</sup> Radarmeteo S.r.l., via IV Novembre 119, Due Carrare, 35020, Padova, Italy

E-mail: [a.garinei@unimarconi.it](mailto:a.garinei@unimarconi.it)

**Abstract.** We analyze how the uncertainties in meteorological data affect the outcome of an agrometeorological model. In particular we focus on a model simulating the development of primary infections of downy mildew on grapevine. This model takes as inputs hourly measurements of temperature, relative humidity and rainfall and it simulates all the phases that eventually lead to the downy mildew primary infections. In order to assess the robustness of the model against the uncertainties of the input data we set up a Monte Carlo simulation. This procedure enhances the potentiality of the model since allows to extract, as an additional valuable piece of information, the probability of the infection event. The results we obtain are highly significant and suggest that the effects of the uncertainties are actually important and should be carefully evaluated.

*Keywords:* Uncertainties in Meteorology, Agrometeorological Models, DSS.

## 1. Introduction

Smart techniques applied to the agricultural sector are becoming increasingly spread and important. In this scenario, agrometeorological models are formidable tools in farmers' hands which can provide a helpful guidance in taking timely decisions and consequently boost the crop productivity. Naively, the outcomes of these models are heavily dependent on the input data, both in term of the amount of data available and of their quality. In many cases, it is reasonable to expect that even relatively small changes in the input data could drastically change the outcome of the model. So, considering that each measurement is affected by a certain amount of uncertainty, a valuable question that we can ask is whether and to what extent we can trust the results of a model. Despite being crucial, this aspect is usually neglected in real implementation of these models and in their actual use in the agricultural sector. Currently many of the agrometeorological models are based on some internal deterministic dynamics and, once fixed the input data, they yield a certain and unique output; there are no shades in the results. If one took into account the uncertainties of the inputs and determined how they propagate, it would be possible to enhance the amount of information that can be extracted from these models: namely the latter could tell not just if a certain event will happen but rather what is the odd that it will happen. This could make such models more useful and also make easier



the integration of the results with other sources of information. For example, in the context of decision support systems (DSS), the knowledge of the probability that a certain disease may be present in a field could be integrated by the farmer experience and help in taking the decision to start the suitable phytosanitary treatment. Note furthermore, that this kind of investigations could be extremely interesting and have relevant applications in the agricultural insurance field: for instance it could provide the guidelines on the good practices giving access to the benefits offered by mutual aid societies when natural events or disasters harm the crop productivity.

In this paper, we analyze from a quantitative point of view how the uncertainties in the input data affect the outcome of an agrometeorological model. In particular, we focus on a model simulating the development of primary infections of downy mildew on grapevine, proposed by Rossi et al. [1, 2]. Downy mildew is considered the most devastating disease affecting grapevines and it is caused by a heterothallic oomycete, the *plasmopara viticola* [3]. During the wintertime, this oomycete lives in the soil in the form of oospores, which in spring germinate and spawn macrosporangia. Under wet conditions, macrosporangia release zoospores, which can be dispersed and eventually reach grapevine's leaves through rain splashes causing primary infections on plants. The infection becomes manifest after few days when yellow lesions appear on leaves. During favorable weather, these lesions can sporulate and also cause new secondary infections.

The agrometeorological model we consider in our study takes as inputs hourly measurements of temperature, relative humidity and rainfall and it simulates all the phases that eventually lead to the downy mildew primary infections. We start from a set of real meteorological data, collected by a weather station and we set up a Monte Carlo simulation in order to assess the robustness of the model against the uncertainties of the input data. The simulation is carried out by adding statistical noise to the real data with amplitudes determined according to the level of uncertainty affecting the measured data. This is given by the measurement error, which is determined by the combination of the instrumental error and the siting error, induced by the environment surrounding the weather stations. Both these two sources of error can be estimated on the basis of the compliance classification made by the World Meteorological Organization (WMO) [4, 5, 6]. Running the simulation several times we were able to determine the impact of the uncertainties on the model outcome for all the possible compliance classes.

We also perform an analogous analysis in the case in which the input data that feed the agrometeorological model are not measured directly from a physical weather station located in the field, but rather they are determined by means of geostatistical procedures. Indeed, weather forecast providers have highly sophisticated services that allows to retrieve local meteorological data even when there are no stations in the fields, using data collected by surrounding stations and interpolating these data through geostatistical methods. In such a way, one can reconstruct punctual meteorological data for every location in the territory covered by the network of weather stations. In this case the uncertainties of the data are generally higher since, besides the measurement errors, one has to consistently take into account also the statistical error introduced by the interpolation procedure.

## 2. Materials and methods

We consider as a case study an agrometeorological model for the development of primary infection of *plasmopara viticola* (downy mildew) on grapevine [1, 2]. The model uses hourly measurements of temperature, relative humidity and rainfall as inputs to simulate in a mechanistic way the dynamics that leads to downy mildew primary infections. The model's dynamics is quite involved but for our scope a full understanding of how it works is not required; we refer the interested reader to the original articles for a detailed description.

The aim of our study is to assess the robustness of the model against the uncertainties of input data. We consider two different scenarios: the first in which the weather data come from

**Table 1.** Maximum instrumental errors for each of the three classes according to the WMO classification. The errors for temperatures and that for rainfalls below 5mm are absolute while the ones for relative humidity and rainfalls above 5mm are relative (expressed as percentage).

	Temperature	Relative Humidity	Rainfall	
	(°C)	(%)	if < 5 mm (mm)	if > 5 mm (%)
Fully compliant (class 1)	0.2	3	0.3	5
Compliant (class 2)	1	10	0.5	10
Non-compliant (class 3)	2	20	1	20

direct measurements taken by stations installed in the field we want to monitor; the second one where the weather data are obtained through geostatistical interpolations of data coming from surrounding stations.

### 2.1. Data from real stations

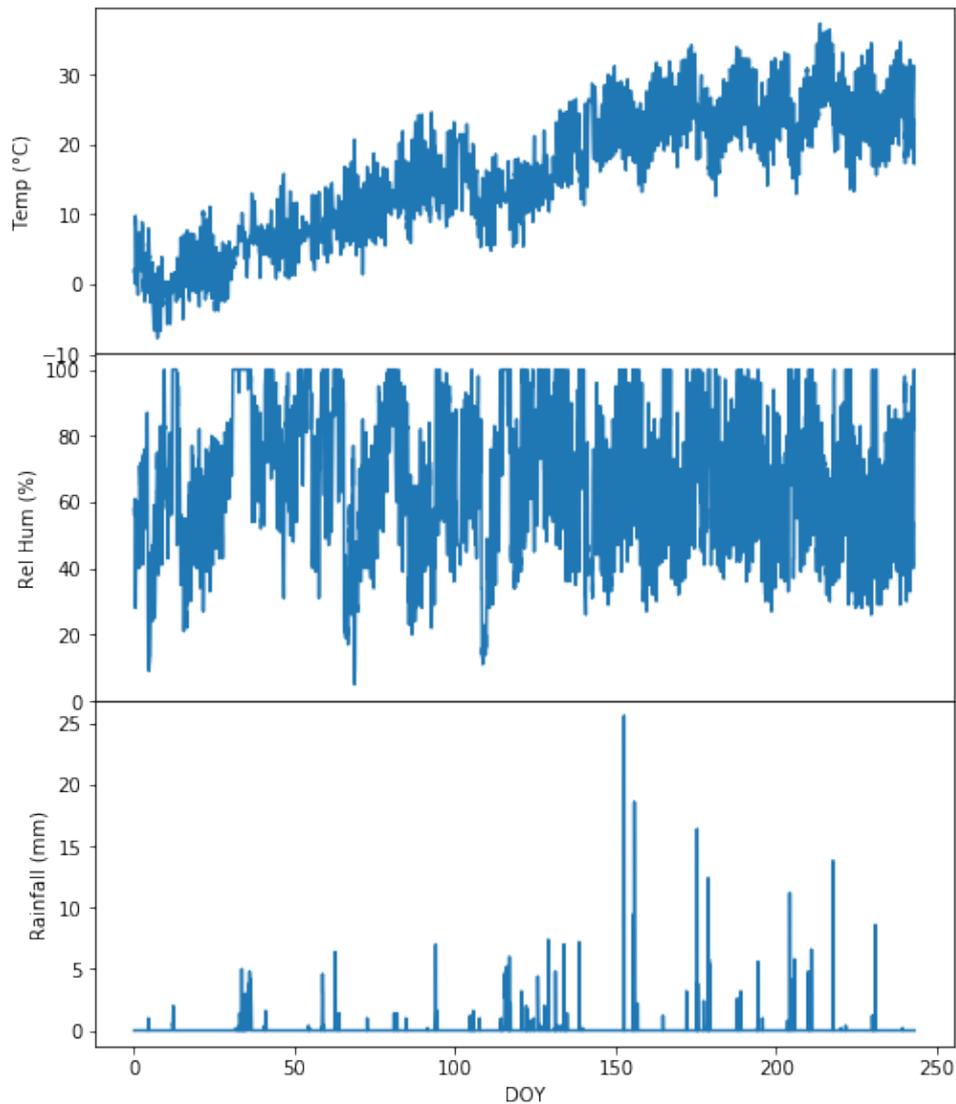
We considered six datasets of meteorological measurements collected from weather stations in Veneto, a region of Italy particularly devoted to viticulture and renowned in the world for its wines, like Prosecco and Amarone. More precisely, we used data from the following six locations:

- Malo (Vicenza);
- Tribano (Padova);
- Illasi (Verona);
- Marano di Valpolicella (Verona);
- Brendola (Vicenza);
- Conegliano Veneto (Treviso).

The time window covered by the datasets spans from the January 1st to August 31th, 2017, in accordance to what required by the model [1, 2]. Figure 1 shows the behavior of the measured data as a function of time (day of the year).

Starting from the measured data we set up a Monte Carlo procedure aimed to the derivation of a statistical evaluation of the uncertainties affecting the outcome of the model. For each measurement, we generated a random variable having the actual measured value as mean and the measurement error as standard deviation. In particular we used Gaussian distributions for all the three input variables, namely the temperature, the relative humidity and the rainfall. Since, naively, relative humidity is allowed to be between 0 and 100 and rainfall to be strictly non-negative we also imposed these constraints on the generated variables. We considered two contributions to the measurement error associated to the three meteorological quantities used as inputs in the model: the instrumental and the siting errors. The instrumental error is the one associated to the accuracy of the measurement devices installed in the station. According to the WMO classification we considered three classes: fully compliant, compliant and non-compliant instruments. The maximum instrumental error associated to each class follows the WMO specifications [4, 5] and it is reported in Table 1.

The measurement error is not only determined by the quality of the instruments installed but also by the location of the station. The environmental conditions can indeed influence the outcome of the measurements. The WMO defined a set of relevant criteria to assess the goodness of the sites and, according to these criteria, it provided a fivefold classification, where class 1 corresponds to an optimal location and class 5 to the worst environmental condition for



**Figure 1.** Temporal behavior of the physical quantities used as inputs in the model as measured from the station located in Malo. Horizontal axis: the day of the year (DOY); vertical axis: temperature, relative humidity and rainfall.

the weather station [6]. Table 2 shows the siting induced errors for temperature and rainfall measurements for each of the five classes. The errors for relative humidity measurements are not given so we will assume that they are not significantly influenced by the environmental conditions.

In the Monte Carlo simulations, we will consider different combinations of instrumental and siting classes, fixing the uncertainties to the maximum value of the error for the considered class, according to Tables 1 and 2. The instrumental and siting errors are then combined in quadrature yielding the overall uncertainty associated to the measurement.

**Table 2.** Maximum error for each site class according to the WMO classification. The error for the temperature is absolute while the one for rainfall is relative (%).

	Temperature (°C)	Rainfall (%)
Class 1	–	–
Class 2	–	5
Class 3	1	15
Class 4	2	25
Class 5	5	100

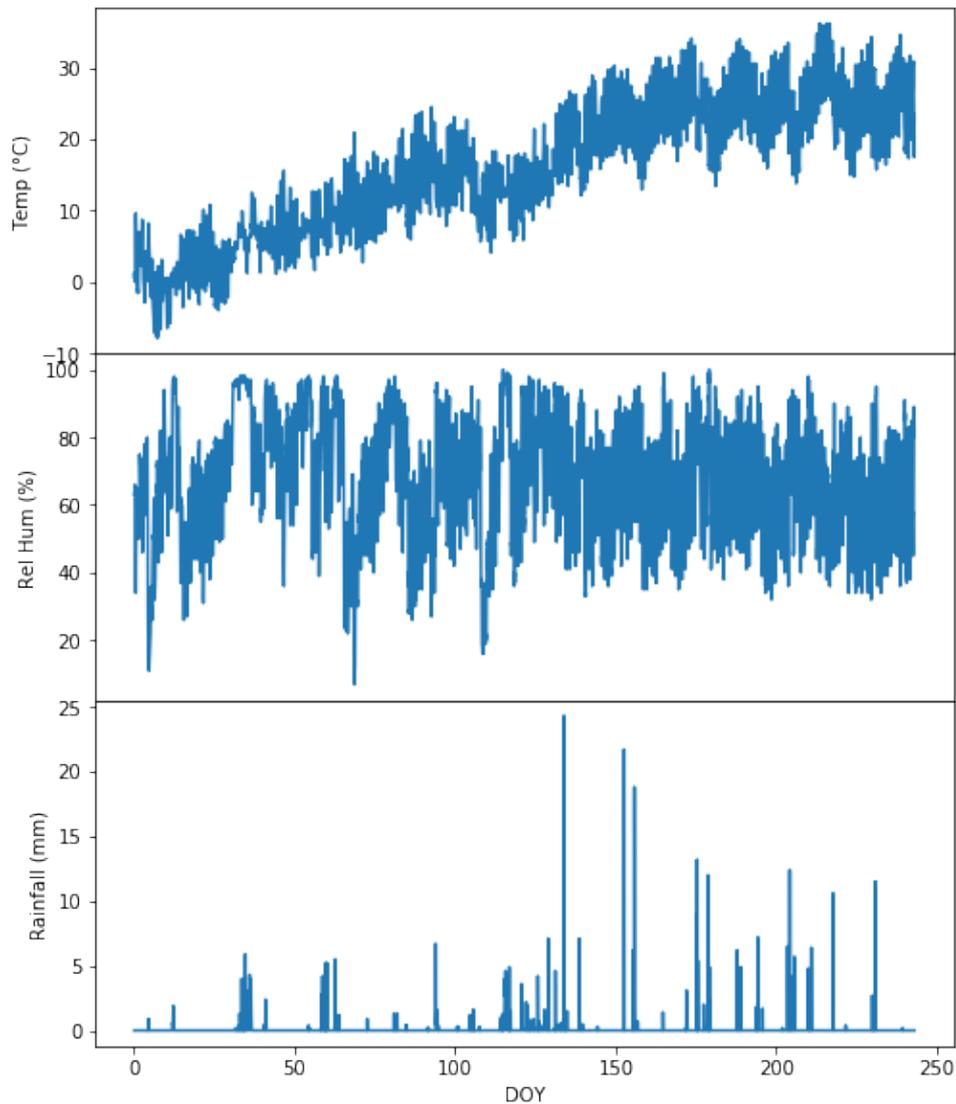
## 2.2. Data from virtual stations

Besides the analysis explained in the previous section, we consider also the case in which the data used as inputs of the agrometeorological model are not measured directly by devices located in the field, but rather are determined by means of geostatistical interpolations [7, 8]. This is a particularly relevant case, since it is quite common that there are no weather stations installed on or nearby the field one wants to monitor. In such situations, one can still retrieve the needed data using the services provided by companies operating in the weather forecast sector. Indeed, usually these companies have access to the meteorological data from a wide network of stations spread on the territory and thus they can use geostatistical interpolations to derive an estimate of the variables at any locations. Since in this case there is not a real physical station installed in the field, we say that the data comes from a “virtual station”.

We chose to apply the model in the same time period and in the same six positions considered in the previous section, using this time the “virtual” measurements instead of the real ones recorded by the physical stations. Of course, in each case we pretended that there was not a weather station in that place, *i.e.* in the interpolation procedure carried on to derive the “virtual” measurements we discarded the data from the real station located in that place. Specifically, the data are obtained by means of an interpolation technique particularly suited for this kind of applications, called kriging method [9, 10, 11].

Figure 2, analogously to what displayed in Figure 1 for the real station, shows the temporal behavior of the input variable of the models as “measured” by the virtual station.

The procedure we use in order to estimate the uncertainties induced on the model outcome by the data is the same as for the previous case of data collected from real station. What changes in this case is the error associated to the input data, since we now have to take into account also the contribution coming from the geostatistical interpolation itself. The statistical error associated to the interpolated data has been estimated applying the leave-one-out cross-validation procedure to the measurements coming from the set of  $N$  stations involved in the interpolation [12, 13]. In practice, for each of these stations one estimates the meteorological data applying the kriging method, using the data from the remaining  $N - 1$  stations. Then for every physical quantity, one defines an associated “error field” whose value in each station location is given by the absolute value of the difference of the real measurement and the estimated one. Finally, using again the geostatistical interpolation applied to these error fields one can estimate the values of uncertainties at the location of the virtual station. Notice that this procedure yields uncertainties which are specific for each location and, differently from the ones associated to the real measurements, variable in time, since the cross validation is carried out independently for each hourly measurement. The statistical error due to the geostatistical procedure is then combined with the measurement errors associated to the real measured data

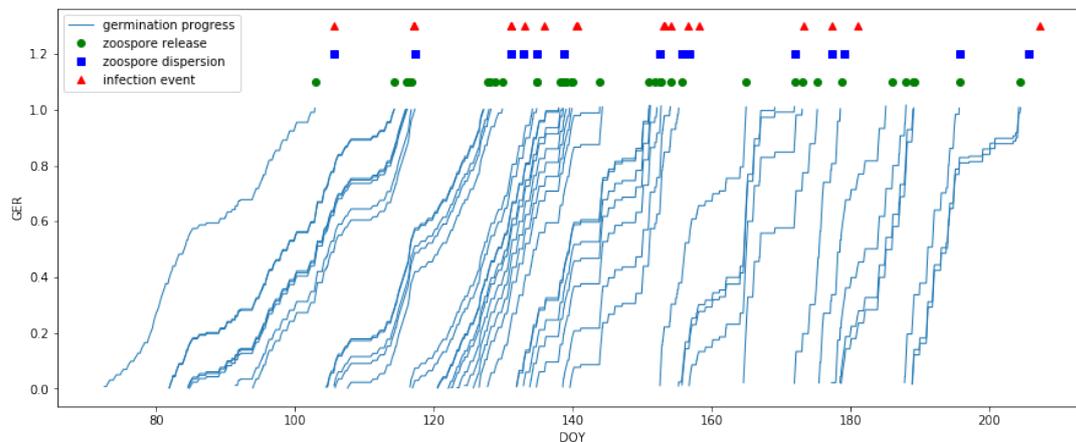


**Figure 2.** Temporal behavior of the physical quantities used as inputs in the model as obtained from the virtual station in Malo: on the horizontal axis, the day of the year (DOY) is reported, and on the vertical axis temperature, relative humidity and rainfall.

used in the interpolation. In principle, each measured data has its own error which is related to the compliance of the station from which it is taken, namely to the instrumental specifications and to location of the station, as discussed in Section 2.1. However, to keep the things simple, we assume that all the stations used to estimate the virtual data are of the same compliance class and we explore how the results of our analysis change by varying the compliance classes.

### 3. Results

As a first step, we just run the agrometeorological model using the “real” data, namely the ones measured by the instruments physically installed in the locations considered in this study. An example of the output of the model, for the station located in Malo, is summarized in Figure 3, showing the temporal behavior of the various phases of disease development.



**Figure 3.** Plot summarizing the progress of the various phases that lead to the development of the disease as a result of the model run over measurement from the real station in Malo. The day of the year is represented on the horizontal axes; the lines indicate the germination progress (GER – ranging from 0 to 1) for each cohort of *plasmopara viticola*; green circles correspond to zoospore release events, blue squares to zoospore dispersion events, red triangle to infection events.

We focus only on the infection events (depicted as red triangle in the plots in Figure 3) whose prediction is indeed the main goal of the model and clearly it is what really matters for agricultural purposes. The main goal of our research is to derive an indicator of the reliability of the outcome of the model in relation to the uncertainties affecting the data used as inputs. We choose to use the probability that a downy mildew primary infection starts at a given day as such indicator. This is obtained using Monte Carlo simulations: we run the model several times using generated data and the probability of infection at a given day is computed as the fraction of simulations which predicted the onset of an infection on that day. In particular we fixed the number of runs,  $N$ , to one thousand, since we observed that this number guaranteed a good stability of the outcomes, *i.e.* further increasing  $N$  did not have significant effect on the results of the simulation.

Note that the model we use takes as inputs hourly data and gives outputs with hourly resolution as well. However, in the computation of the probabilities, we decided to recast the output on a daily grid in order to avoid an excessive complexity of the model and to make our results more stable, reducing the chance that a simple few-hours translation of an event in different runs is interpreted as a discrepancy. This is also motivated by the fact that in practical uses in agriculture it is sufficient to know the response on a daily level.

### 3.1. Data from real stations

As explained before, when the meteorological data necessary in order to run the model come from measurements made by weather station installed in the field, the uncertainties that affect the outcome are due to the instrumental and siting errors. We considered different combinations of instrumental compliance and siting classes, according to the classification reported in the previous section. For ease of notation, we denote a station belonging to the instrumental compliance class  $n_i$  and to the siting class  $n_s$  simply through the ordered pair  $(n_i, n_s)$ . For the sake of brevity we report here only some of the results we obtained: focusing again on the station in Malo, the plots in Figure 4 show the results for the estimated infection probabilities

in the cases of compliance classes (1,1), (1,3), (2,2) and (3,3).

We notice that when the instruments in the station are fully compliant ( $n_i = 1$ ) and optimally located ( $n_s = 1$ ), *i.e.* when the uncertainties associate to the measurements are the least possible, the probabilities associated to the predicted infection events are very clearly polarized in almost all the cases. This means that results provided by the model are quite robust. As one naively expects, the peaks in the probabilities tend to lower and spread out as we take into account lower quality compliance classes, namely as we increase the uncertainties.

It is interesting to note that even if consider a case of a mid-level quality station, for instance in the class ( $n_i = 2$ ,  $n_s = 2$ ), the information that our procedure adds to the simple and dichotomous output of the model becomes in some cases really valuable, as clearly shown in the third plot of Figure 4.

### 3.2. Data from virtual stations

We now repeat the same analysis as before using data coming from interpolations rather than from direct measurements in the field. Firstly, in Figure 5 we report the output of the model run over the interpolated data for the virtual station located in Malo.

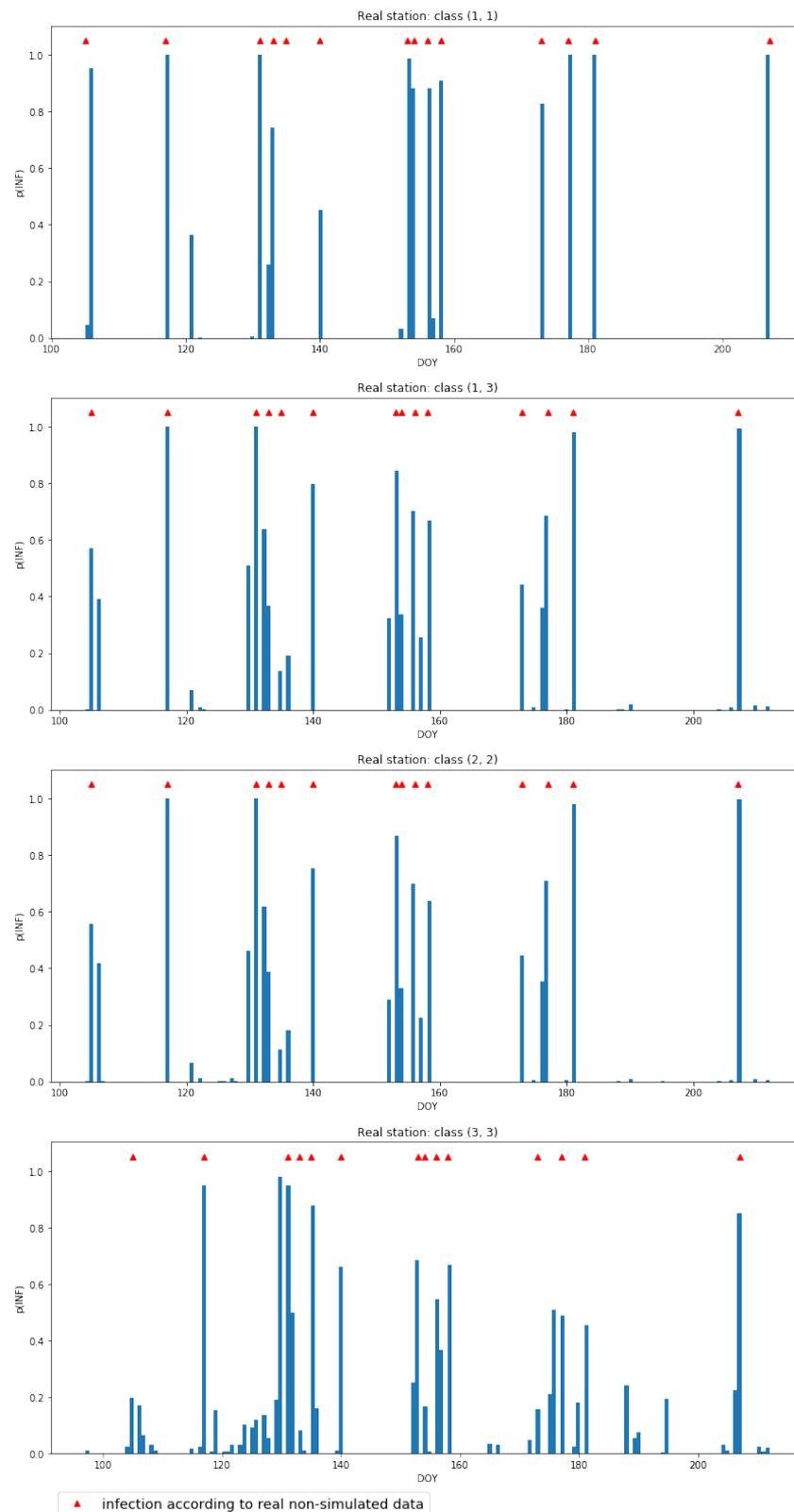
Comparing the results obtained using the real (Figure 3) and virtual (Figure 5) data for the station in Malo we notice that, the two outputs are quite similar, though not exactly coincident. Repeating the same comparison for the other locations considered in our study we found that in some cases (in particular Illasi and Marano) the discrepancies are sharper. This is a hint that the agrometeorological model we are considering is quite sensitive to the precision of the input data and in some cases the interpolated measurements may be not so reliable for the required task.

Then, as we did in our analysis for real stations, we run the Monte Carlo procedure to determine the robustness of the model, considering different combinations of instrumental and siting classes ( $n_i$ ,  $n_s$ ) for the network of stations involved in the interpolation. Figure 6 shows some examples of the results of the Monte Carlo simulation, for compliance classes (1,1), (1,3), (2,2) and (3,3). Note that, differently from what happens for real stations, in this case the results for different compliance classes are almost identical, since for virtual station the dominant contribution to the uncertainties comes from the interpolation error.

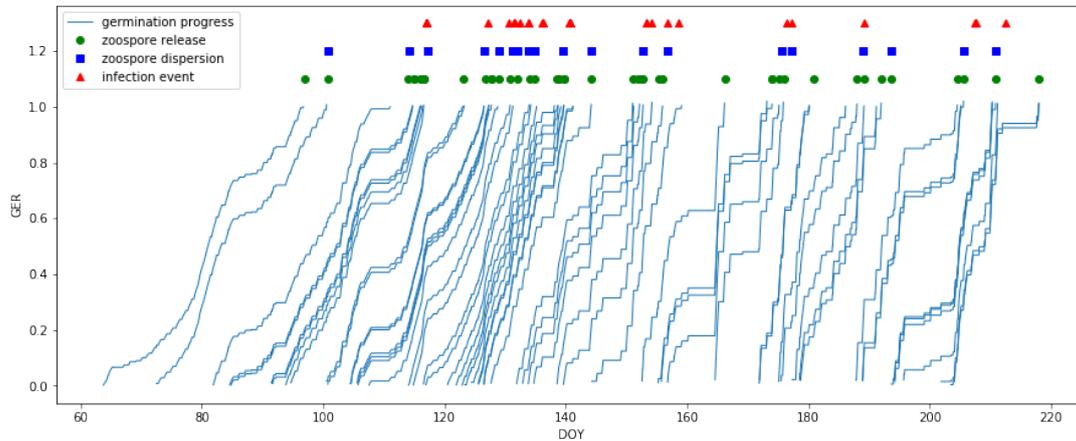
## 4. Conclusions

We developed a procedure to assess the robustness of an agrometeorological model against the uncertainties that affects the input data. These uncertainties cannot be completely removed since they are related to the physical characteristics of the measurement devices installed in the stations (instrumental sensitivities) and to environmental factors of the surrounding area. We considered a specific model as a case study, namely a mechanistic model for the development of primary infections of downy mildew on grapevine. The results that we obtained are highly significant and suggest that the effects of the uncertainties are actually important and should not be neglected.

The procedure we used to estimate the propagation of the uncertainties consisted essentially in a Monte Carlo simulation repeated several times with generated data. At every run, the model gives as outputs the predicted downy mildew infection events. These outputs allowed us to derive an estimation of the probability of the onset of an infection at a given day of the year, simply by counting the fraction of simulations yielding that result. It should be noted that in order to generate the data for the simulations we had to make some strong assumptions on their statistical distributions. We considered in particular normal distributions for all the measured input variables, with mean on the actual measured value (or the interpolated value resulting from the kriging procedure in the case of virtual stations) and variance determined according the estimated error of the measurement. This assumption, although being standard, is



**Figure 4.** Plots showing the probability of the onset of infection at a given day of the year (DOY) for the real station located in Malo in the cases of compliance classes (1,1), (1,3), (2,2) and (3,3). The red triangles in the upper part indicate the days of infection according to the model run with the original (non-simulated) real data.



**Figure 5.** Plot summarizing the progress of the various phases that lead to the development of the disease as a result of the model run over measurement from the virtual station in Malo. The day of the year is represented on the horizontal axes; the lines indicate the germination progress (GER – ranging from 0 to 1) for each cohort of *plasmopara viticola*; green circles correspond to zoospore release events, blue squares to zoospore dispersion events, red triangle to infection events.

quite strong: it implies, for instance, that we are considering measurements that are on average unbiased and no systematic error is present. This is a caveat that one has to take into account when interpreting the results.

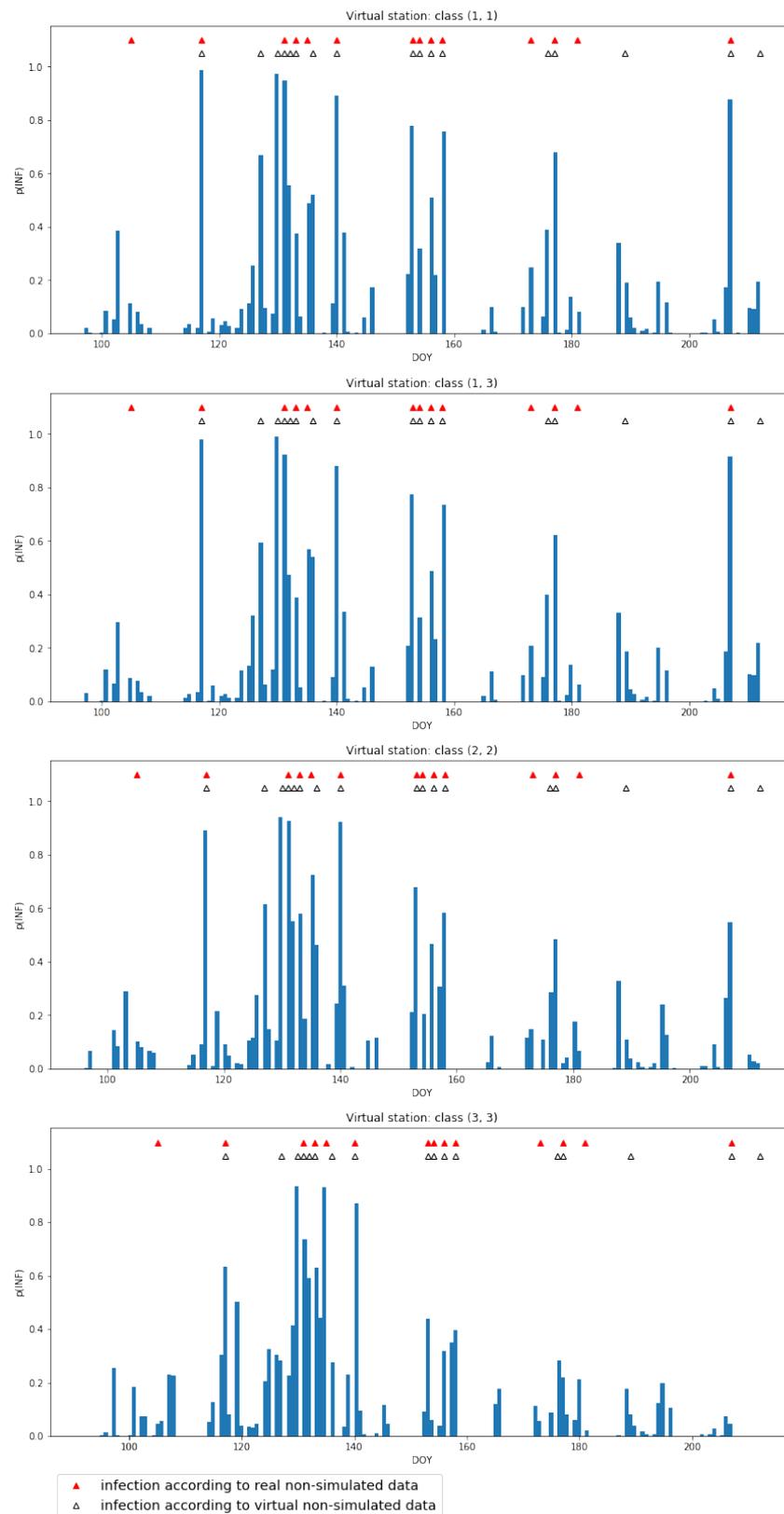
Even if in this paper we considered only a specific agrometeorological model, the procedure we implemented to estimate the propagation of the uncertainties can be applied to any model. Besides the actual results we showed in Section 3, the main and most relevant objective of the paper has been to introduce a new general paradigm of implementation of agrometeorological models and demonstrate that it really increases the information that can be extracted from the models and allows the user to have a clearer interpretability of the outcomes.

We decided to apply the aforementioned procedure in an extremely involved agrometeorological model, which takes as inputs a big amount of data (hourly measurements of temperature, relative humidity and rainfall for an extended time window) and elaborate them in a highly non-linear fashion in order to produce the desired output. Indeed, that of complicated models is exactly the case in which one can expect some chaotic behaviors to emerge when perturbing the input data and it is thus the most interesting instance to consider.

Finally, it is worth noting that the obtained results suggest that one needs to be very cautious when using data from virtual stations since these may be not so reliable in the application to a particular agrometeorological model. Therefore, when one uses such “virtual” data, a careful evaluation of the actual effectiveness of the procedure for the specific case at hand is highly recommended.

### Acknowledgments

We thank *Radarmeteo* company for providing the data used for our elaborations. The study is part of a work carried out in the framework of the *RESA project*, funded by *Regione del Veneto – POR FESR 2014-2020*.



**Figure 6.** Plots showing the probability of the onset of infection at a given day of the year (DOY) for the virtual station located in Malo in the cases of compliance classes (1,1), (1,3), (2,2) and (3,3). The red and black (empty) triangles in the upper part indicate the days of infection according to the model run with the original (non-simulated) real and virtual data, respectively.

## References

- [1] Rossi V, Caffi T, Giosuè S and Bugiani R 2008 *Ecological Modelling* **212** 480–491
- [2] Caffi T, Rossi V, Bugiani R, Spanna F, Flamini L, Cossu A and Nigro C 2009 *Journal of Plant Pathology* 535–548
- [3] Gessler C, Pertot I and Perazzolli M 2011 *Phytopathologia Mediterranea* **50** 3–44
- [4] WMO 2008 *Guide to meteorological instruments and methods of observation* WMO-No. 8 (Geneva: World Meteorological Organisation)
- [5] Fisler J, Kube M, Stocker C and Grüter E 2017 *Quality assessment using METEO-Cert: The MeteoSwiss classification procedure for automatic weather stations* (World Meteorological Organisation)
- [6] WMO 2010 *Commission for instruments and methods of observation – Report of the fifteenth session* WMO-No. 1064 (Helsinki: World Meteorological Organisation)
- [7] Bivand R S, Pebesma E J, Gomez-Rubio V and Pebesma E J 2008 *Applied spatial data analysis with R (Use R vol 10)* (New York: Springer)
- [8] Diggle P and Ribeiro P J 2007 *Model-based Geostatistics* Springer Series in Statistics (New York: Springer)
- [9] Matheron G 1963 *Economic geology* **58** 1246–1266
- [10] Cressie N 1990 *Mathematical geology* **22** 239–252
- [11] Sluiter R 2009 *Interpolation methods for climate data* (KNMI)
- [12] Pebesma E J 2004 *Computers & Geosciences* **30** 683–691
- [13] Pebesma E J and Wesseling C G 1998 *Computers & Geosciences* **24** 17–31