Application of a non-linear temperature forecast post-processing technique for the optimization of powdery mildew protection on strawberry

Emanuele Eccel¹, Stefano Fratton², Luca Ghielmi¹, Andrea Tizianel, Dani Shtienberg², Ilaria Pertot²

Abstract: Strawberry powdery mildew, caused by the fungus Podosphaera aphanis, is a dangerous disease in warm and dry climates as well as in greenhouses or plasticulture. Plant protection against P. aphanis is mainly based on chemical fungicides. More than ten chemical treatments for each growing cycle are often applied in tunnel strawberry production in northern Italy. SafeBerry is a decision support system, which optimises, and often reduces, the use of chemicals against this disease. The system is based on a correct fungicide application based on the disease risk level in each tunnel and on the specific action mechanism of the fungicides. The level of risk is based on crop and environmental parameters. The temperature assessment and its forecast represent two key points in the system. The decision-making procedure uses day-time temperatures forecasted for the three following days. They were calculated by post-processing of the operational weather model output (Model Output Statistics, MOS). MOS was carried out for three sites with a “machine learning”, multivariate, non-linear technique (“Random Forest”), which uses many meteorological predictors. With this system in 2007 we obtained a strong reduction in the number of treatments (up to 60%).

Keywords: pesticide, fungicide, integrated pest management, MOS, temperature downscaling

INTRODUCTION
Powdery mildew (PM) of strawberry, caused by Podosphaera aphanis, is a dangerous disease in the Mediterranean climate or greenhouses and polyethylene tunnels. The pathogen can attack all the aerial part of the plant, flowers and fruits included, and cause high economic losses especially on the most susceptible cultivars (Amsalem et al., 2006). Plant protection is based on the application of chemical fungicides. In northern Italy (Trentino region) strawberries are mainly produced in high polyethylene tunnels on peat in suspended pots. Under tunnel conditions up to 10-12 fungicide treatments are often applied by growers in order to protect plants. A reduction of pesticides can be obtained with an optimization of their use, which consists in applying them when they can reach the highest efficacy and in the selection of the most suitable fungicide according to its mechanism of action against the disease (Pertot et al., 2008). Decision support systems may help in optimizing treatments in agriculture (Madden and Ellis, 1988). We developed a decision support system (SafeBerry) based on plant, pathogen, efficacy and mechanism of action of fungicides and both on past (measured) and future (forecasted) temperature. In order to make available a calibrated temperature forecast for the sites involved in this work, we applied a “Model Output Statistics” (MOS) to the meteorological model output. Indeed, local topography is ill-represented by general circulation models, due to the coarse grid used (scores of km). Hence, in an Alpine territory, orography strongly affects the goodness of

¹ Corresponding author e-mail: emanuele.eccel@iasma.it
² Fondazione Edmund Mach – Centro Ricerca e Innovazione, S. Michele all’Adige (TN) Italy.
³ Department of Plant Pathology, The Volcani Center, Bet Dagan, Israel.
the representation of true elevation of sites, with errors up to 1000 m at some model grid points, resulting in errors as large as 4 – 6 °C in temperature forecasts. The latter figures representing mean biases of estimation. This effect is particularly evident in the case of deep valleys such as Adige Valley, which is one of the largest in the Alps in terms of both length and depth.

Several approaches have been proposed to downscale temperature predictions from the raw (direct) model output (DMO) to the calibrated values. Not only do such algorithms allow to cope with the different elevation of sites with respect to the model orography, but they also improve the forecast by post-processing DMO with many predictors, other than the predicted temperature itself. The problem can be simply tackled by use of univariate methods as, for example, the application of site-specific offsets (fixed or seasonal) or Kalman filter techniques (Homleid, 1995; Galanis and Anadranistakis, 2002; Anadranistakis et al., 2004; Cane et al., 2004). Although univariate methods have been well tested, multivariate methods have the potential for modelling the influence of both the properties of the site and prognostics provided by meteorological models. Among these techniques, non-linear, "machine learning" algorithms, proved suitable for this task (Schizas et al., 1991; Abdel-Aal and Ellhadidy, 1994; Robinson and Mort, 1997; Arca et al., 1998; Verdes et al., 2000, Basili et al., 2006). Experiences in Trentino have shown that non-linear, multivariate techniques, namely "Random Forest" algorithm, do improve temperature forecast with respect to univariate methods (Eccel et al., 2007). The main objective of this work was to test the goodness of the coupling of a 3-day temperature forecast method to the application of SafeBerry.

**MATERIALS AND METHODS**

**The decision support system**

The decision-making procedure in SafeBerry is comprised of three major steps: the determination of the risk of disease outbreak, the suitability of temperature to the disease and the recommendation of action. The potential risk index for the disease is the result of the integration of a basic risk index, which includes the risk factors that do not vary in the season, and the daily risk index, which includes the risk factors that vary on daily base. The risk factors were identified and quantified by interviewing growers and technicians and by searching literature. Each risk factor was rated in levels and risk points were empirically attributed to each level. The risk factors included in the basic risk index are: cultivar susceptibility, presence of disease at less than 50 or 10 m at planting day, incidence of PM in the nursery where plants were grown, type of sprayer, overhead irrigation (sprinkler-cooling), tunnel height, density of plants per meter, mulching with black plastic. The risk factors included in the daily risk index are: phenological stage of the plants, disease incidence in the tunnel, time of disease onset, presence of disease in less than 10 m, time since last treatment, presence of runners.

Even in cases where the level of potential risk is high, its realization depends on the existence of a disease-conducive environment. We used a model that describes the relationship between day-time temperature (DTT; average between 5:00 and 17:00) and the rate of PM development (daily disease increase) and it is based on data collected in experiments in Trentino (northern Italy) from 1999 to 2003. Suitability of the weather to PM development is divided into three categories: low suitability (LS; DTT ≤ 18°C or DTT > 26°C) when temperature is expected to limit PM development; medium suitability (MS; 18°C < DTT ≤ 20°C or 25°C < DTT ≤ 26°C) when temperature is suitable for PM development; and high suitability (HS; 20°C < DTT ≤ 25°C) when temperature is highly conducive for PM development. The daily disease increase is based on DTT in the previous six days (measured) and in the next three days (forecasted).

In the last step of SafeBerry, the potential risk and the likelihood of its realization (suitability of the temperature to the disease) are integrated by taking into account the phenological stage of the crop and finally a recommendation for action is given. The possible recommendations for action are: “Do not spray today,” and “Spray as soon as possible.” In the latter case, a list of preferable fungicides is recommended according their mechanism of action, risk of developing resistance to the pathogen and restriction on time of application before harvest.

**The investigation area**

The pilot area for this investigation is located in Trentino region, northern Italy. Three locations were chosen in the neighbourhood of meteorological stations within important strawberry growing areas; two are located in the lowest part of Valsugana valley (Pergine Valsugana, 460 m a.s.l. and Borgo, 420 m a.s.l.) while a third (Baselga di Pinè) is in a highland, at 950 m a.s.l.

SafeBerry was evaluated in five experiments carried out in 2007 at the same three locations (Baselga, Pergine1, Pergine2, Borgo1 and Borgo2). The
cultivar was Elsanta, which is the most widely used by strawberry growers in this area and is highly susceptible to PM. Strawberry plants were planted in peat, in suspended pots in rows in walk-in plastic tunnels and treated according to SafeBerry recommendation. Untreated control was included. Treatments were arranged in a randomized complete block design with four replicates. Each plot contained 24 potted plants (six plants/pot × four pots). The fungicides were applied using a gun sprayer. To avoid any drift to adjacent plots, untreated plots were covered with polyethylene film during spraying. Plants in untreated plots were inspected visually starting at planting and continuing at three to five day intervals until disease onset, after which point disease was assessed in all plots on a weekly basis. 40 randomly chosen leaves were inspected per replicate. For each replicate, disease incidence on leaves (percentage of infected leaves) was determined. Incidence was arcsin-transformed before ANOVA to obtain homogeneity of variance (Levene's test). Statistical analyses were performed using the Statistica software 6.0 (Statsoft, Tulsa, OK, USA).

Downscaling temperature prediction
A MOS approach was used to downscale temperature forecast, which is available daily from the meteorological model. In the MOS approach, relationships are obtained by using the model outputs as predictors (e.g., temperature forecasts at different grid points at a given lead time), and measured quantities as predictands – in this case, temperature at the chosen sites at the same times. This approach also takes into account offsets that are intrinsic to the model itself, by applying non-linear functions to correct systematic, grid-point-specific biases.

For temperature forecast processing ECMWF's operational meteorological model “T511 L60” was used (60 vertical levels, 0.5° horizontal resolution). The nine grid points surrounding the target area were chosen. The operational run at 00 UTC was chosen in the 6-hourly outputs (at daily times 00, 06, 12 and 18) till +72 hours from issue. The absence of a suitable calibration period for the present operational release of ECMWF's model (“T799 L91” - 91 vertical levels, 0.25° horizontal resolution) lead us to develop the MOS with the previous version, even considering that previous grid points are available also in the latest version. In a previous experience in the same area, using the same model outputs as predictors, different approaches were tested for temperature prediction, limited to morning minimum at 06 UTC (Eccel et al., 2007). The investigation had pointed out that non-linear methods, as artificial neural networks or Random Forest, performed better than simple bias corrections or linear models for output post-processing. The latter experience suggested to apply Random Forest to perform the MOS of GCM output. The freeware R (package “RandomForest”; Liaw and Wiener, 2005) was used. Some features of RF algorithm are given; for more details, refer to the abovementioned work and to Breiman (2001). RF model is a non-linear, multivariate regressive method that can be comprised in the “machine learning” category of algorithms, to which “artificial neural networks” also belong. It is only moderately prone to overfitting (Breiman, 2001) and, under this aspect, it is preferable to neural networks. RF model yields an ensemble processing, pooling several variables for the calculation of the predictands. In detail, having identified an optimal set of predictors (by means of a selection of an even high number of potential variables), RF performs the MOS by running a series of decision trees. A regression tree (Breiman et al., 1984) consists of a set of nodes that branch out from a root node. Each node contains a question with several possible answers, each leading either to another node or a “leaf” (a terminal node with an associated prediction). At each branch in every tree the values of prediction variables are considered and consequently the direction of the “decision flux” is determined, based on the fulfilment (or not) of given logical conditions on predictors, and according to the conformation taken by each individual tree during the “training” stage. At every node, the vector of predictors drives the choice through different branches to the final result (a “leaf”); such operation is carried out on each tree in the same way. In general, the prediction stemming by individual trees is strongly inaccurate; on the contrary, the value obtained averaging results from all trees (1000 in this application) yields a stable prediction, with a limited variance (Geman et al., 1992). In this application a RF model was created and calibrated for every lead time (00, 06, 12, 18).

Obtaining day-time values
Having obtained the predictions for the four synoptic lead times, hourly values are calculated via interpolation after a refined version of the “TM” model (Cesaraccio et al., 2001). The latter needs fixing expected times of occurrences of maximum and minimum temperatures; these parameters were set with a preliminary analysis. Curves are traced with four analytical functions, each one valid in a
given time range. Being $H_{\min}$ and $H_{\max}$ times when minimum and maximum temperatures are expected, respectively, the functions are defined in the following ranges:

1. from 00 UTC to $H_{\min}$ – parabolic
2. from $H_{\min}$ to $H_{\max}$ – sinusoidal
3. from $H_{\max}$ to 18 – sinusoidal
4. from 18 UTC to 24 UTC - parabolic.

Details on the equations of the interpolating curves are given in the Appendix.

The routine calculates hourly values until time +72 h, by an iterative application of the equations valid for the first 24 hours. Then, hourly values are averaged between 05 and 17 of every day to get day-time values, which are used by SafeBerry.

Comparison of temperatures suitable to the disease in the next three days (measured vs. forecasted)

Temperatures suitable for the disease in the following three days were calculated and rated according to three classes: HS, MS and LS, both on measured and forecasted temperatures, and the two series compared. Differences were expressed as ± 1-2 classes of distance.

RESULTS

Temperature MOS

Implementation of RF required the set-up of the maximum number of predictors for each tree; in general, the total number of predictors is high, while the algorithm is optimised with a limited number of predictors at once, chosen at random from the pool. Usually, the improvement becomes negligible with more than 10 - 20 predictors. The latter were selected, in each trial, from a larger set. The routines available in the package RandomForest were used to select the relative importance of variables and hence the set of predictors. Beyond the output of meteorological model, in each of the grid points, the following variables were taken into account:

- temperatures at 00, 06, 12, and 18 UTC of the previous day;
- forecast errors of the previous three days, for every grid point;
- night length.

The potentially useful variables, for every lead time (00, 06, 12 e 18 UTC), are:

- ECMWF's output (24), for each of 9 grid points;
- temperature measured at every station in the following day;
- forecast errors for every grid point for each of the previous days;
- night length (only for prediction of temperature at 06 UTC).

As a whole, predictors had to be chosen among 245 variables, over a time span of +3 days from every weather forecast issue, and identically replicating in the following three days for every synoptic lead time. The most impacting variables proved to be (beyond the trivial DMO of temperature at 2 m):

- temperatures predicted for the previous day at the same times, dew temperatures, night length (for forecast at 06 UTC), temperature at different atmospheric levels (geopotential heights), and humidity at 850 hPa. The presence of the temperature forecast at 2 m for more than one grid point among the predictors, shows their independent contributions.

The accuracy of forecast was assessed with conventional error statistics. Errors arise from the sum of a meteorological forecast error (in turn, sum of a meteorological model error and a MOS error) and one due to the hourly interpolation. For the single day-time averages over the three-day forecast period, in about one year of simulations errors were assessed in the range of ± 0.1°C (mean error), 1.5°C-1.7 °C (standard error, or RMSE), and 1.1°C-1.3°C (mean absolute error). Errors ranged between –3.3°C (Baselga) and +6.5°C (Pergine). The latter values are rather high; however, the cases with large errors are restricted to limited periods, and they occurred mostly in the first part of the trial.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Mean Apr.–Sept.</th>
<th>Max (+)</th>
<th>Max (-)</th>
<th>25th percent.</th>
<th>75th percent.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baselga di Piné</td>
<td>0.15</td>
<td>0.27</td>
<td>6.3</td>
<td>-3.3</td>
<td>-0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Pergine</td>
<td>0.72</td>
<td>0.63</td>
<td>6.5</td>
<td>-2.4</td>
<td>0.0</td>
<td>1.4</td>
</tr>
<tr>
<td>Borgo</td>
<td>0.84</td>
<td>0.63</td>
<td>5.9</td>
<td>-2.9</td>
<td>0.1</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Tab. 1 – Statistics of errors for the day-time temperature predictions over three days. Values in °C. Mean: mean of errors. Mean Apr.-Sept.: mean limited to the period April – September. Max (+): maximum of positive errors. Max (-): maximum of negative errors. 25th (75th) percent.: values of the 25th (75th) percentile.

(Fig. 1), while in general the inter-quantile range of errors (IQR) is lower than 1.6°C (Pergine), with mean errors lower than 0.9°C (Tab. 1).

**Temperature suitable for powdery mildew**
The highest number of measured days with temperature suitable for the disease during the growing cycle was in Pergine1 (48.9%) followed by Baselga (33.8%), Borgo1 (21.2%), Borgo2 (17.6%) and Pergine2 (14.7%) (Fig. 2). These results indicate that in 2007 between June and July, which is the central part of the growing season in Trentino, temperature was highly conducive to PM. Therefore strawberry growing cycles starting at the end of June- beginning of July had a high component in the risk of infection related to the weather. In Baselga and Pergine1 the suitability of the measured temperature to the disease was initially low, then increased in the last part of the growing cycle; the disease incidence increased very quickly in the period with temperatures highly suitable for the disease. In the last part of the season a low suitability of temperature was related to a steady disease development in the late stage. In fact in Pergine2 and Borgo2 the suitability of temperature to the disease was initially high and then became low for the second part of the growing cycle; the disease
increased quickly, but then it was stable with a slight decrease in the last part of the growing cycle. In Borgo1, in spite of a temperature highly suitable for the disease in 21.2% of the days, PM symptoms did not occur. In this experiment the potential risk index was always low because of the absence of disease in the other tunnels of the area. This case demonstrates that, to initiate the disease, additional factors (i.e. presence of inoculum) are necessary.

The suitability of temperature to the disease based on forecasted DTT compared to measured DTT was satisfactory: in Baselga di Pinè the forecasted DTT was correct in 89.9% of the days; it was overestimated by two and one level respectively in 1.3 and 8.7% of the days (17 April – 31st October 2007) and underestimated by one level in 8.1% of the days. In Pergine Valsugana the forecasted DTT was correct in 74.7% of the days; it was overestimated and underestimated by one level in 4.4 and 14.6% of the days, respectively (Fig. 2). In Borgo the forecasted DTT was correct in 71.1% of the days; it was overestimated and underestimated by one level in 12.0 and 16.9% of the days, respectively (Fig. 3).

**Optimization of treatments with SafeBerry**

No symptoms of the disease were seen in Borgo1 during the growing cycle. In all the other experiments the incidence of the disease on leaves was significantly
reduced in plots managed according to SafeBerry’s recommendation compared to the untreated control (Tab. 2). The number of sprays applied according SafeBerry’s recommendations was reduced in all experiment (by 3 to 6 sprays) compared to the number applied according to the common practice of growers in the area.

**DISCUSSION**

The method used to forecast temperature gave a satisfactory level of accuracy and resulted in an acceptable level of over/underestimation of the component of risk related to the suitability of temperature to strawberry PM. The central part of summer, at least in 2007, is the period in which temperatures are most suited to the disease, while late spring and autumn are less suited. The decision support system allowed a reduction of 43.2% (3.8 sprays, on the average) in the number of fungicide treatments compared to the common practice in the area and a reduction of the disease if compared to the untreated. This reduction, combined to a satisfactory level of disease control, demonstrated that in Trentino some treatments currently applied to protect strawberry against PM are unnecessary. Further experiments under commercial conditions will be necessary to confirm these experimental results.

**ACKNOWLEDGEMENTS**

This research was supported by SafeCrop Centre, funded by Autonomous Province of Trento. Thanks
APPENDIX

Hourly interpolation of temperature

Two hypotheses are considered in interpolation:

1. \( T_{06} < T_{00} \) (more frequent case)
2. \( T_{06} > T_{00} \)

In this second case, representative of a condition of temperature increase during the second part of the night, temperature comes from an average from two sinusoidal functions, allowing a better approximation of the true value.

1. Determining \( T_{\text{min}} \) and \( T_{\text{max}} \)

In the first part \( T_{\text{min}} \) and \( T_{\text{max}} \) are determined via sinusoidal functions calibrated by forcing the passage through points already calculated by the MOS - \( T_{06} \) and \( T_{12} \), and with the values of \( H_{\text{min}} \) and \( H_{\text{max}} \) calculated from the statistical analysis (mode value) of individual temperature series at each station, generally different month by month. Such values express the time (hour) in which minimum and maximum values are expected to occur, counting from time 00 at the beginning of the simulation of the meteorological model. For instance, 36 indicates 12 UTC of the day following the issue of ECMWF's "run 00" forecast.

Under the hypothesis 1:

\[
T_{\text{min}} = T_{06} - \frac{(T_{12} - T_{06})(0.5 + S_{06})}{(S_{12} - S_{06})}
\]

where:

\[
S_{06} = 0.5 \sin \left( \frac{30 - H_{\text{min}}}{(H_{\text{max}} - H_{\text{min}}) \cdot \pi / 2} \right)
\]

Under the hypothesis 2 an average of two functions is carried out:

\[
T_{\text{max}} = \left[ \frac{T_{06} \cdot (S_{06} - S_{12}) - (T_{06} - T_{12}) \cdot S_{06}}{S_{06} - S_{12}} \right] + \frac{T_{06} - (T_{12} - T_{06}) \cdot (0.5 + S_{06}) \cdot \frac{1}{2}}{(S_{12} - S_{06})}
\]

where:

\[
S_{06} = \sin \left( \frac{30 - H_{\text{min}}}{(H_{\text{max}} - H_{\text{min}}) \cdot \pi / 2} \right)
\]

\[
S_{12} = \sin \left( \frac{36 - H_{\text{min}}}{(H_{\text{max}} - H_{\text{min}}) \cdot \pi / 2} \right)
\]
2. Hourly interpolation

The following equations are applied for each of the three days along which forecast is given. Term $t$ indicates time in hours.

From 00 to $H_{\text{min}}$

$$T(t) = T_0 + \left( \frac{(T_{\text{min}} - T_0)}{(H_{\text{min}} - 0)} \right) (t - 0)^{0.5}$$

From $H_{\text{min}}$ to $H_{\text{max}}$

- Hypothesis 1:

$$T(t) = T_{\text{min}} + \left( \frac{(T_{\text{max}} - T_{\text{min}})}{2} \right) \left[ 1 + \sin \left( \frac{\pi}{2} \frac{t - H_{\text{min}}}{H_{\text{max}} - H_{\text{min}}} \right) \right]$$

- Hypothesis 2:

$$T(t) = T_{\text{min}} + \left( \frac{(T_{\text{max}} - T_{\text{min}})}{2} \right) \left[ 1 + \sin \left( \frac{\pi}{2} \frac{t - H_{\text{min}}}{H_{\text{max}} - H_{\text{min}}} \right) \right]$$

From $H_{\text{max}}$ to 18:

$$T(t) = T_{18} - \left( (T_{\text{max}} - T_{18}) \right) \sin \left( \frac{\pi}{2} \left[ 1 + \frac{(t - H_{\text{max}})}{(18 - H_{\text{max}})} \right] \right)$$

From 18 to 24:

$$T(t) = T_{18} + \left( \frac{(T_{24} - T_{18})}{\sqrt{6}} \right) (t - 18)^{0.5}$$

REFERENCES


http://stat-www.berkeley.edu/users/breiman/Ran- domForest

Madden L.V., Ellis M.A., 1988. How to develop a