

## PREDICTION OF DISEASE RISK USING SITE-SPECIFIC ESTIMATES OF WEATHER VARIABLES

K.S. KIM, R.M. BERESFORD and W.R. HENSHALL

*HortResearch, Private Bag 92169, Auckland 1142, New Zealand*

*Corresponding author: kkim@hortresearch.co.nz*

### ABSTRACT

To improve the implementation of weather-based disease risk models, a spatial interpolation method was investigated to provide weather estimates for specific sites. Two sites in the HortResearch horticultural weather station network, one in Marlborough and one in Hawke's Bay, were selected as validation sites. Interpolated weather data were estimated for these sites from November to March in 2003-04 and 2004-05, using actual weather data from nearby stations that were selected as natural neighbours using the geometrical technique, Voronoi tessellation. Wetness duration was also estimated using interpolated weather data as inputs to an empirical wetness model. Air temperature estimates were comparable to actual measurements but wetness duration was overestimated. When interpolated and actual data were used as inputs to the grape botrytis model, Bacchus, predicted risks were comparable to each other for short periods rather than the whole growing season. This suggests that risk of botrytis bunch rot could be predicted reliably at a specific site using the spatial interpolation method.

**Keywords:** botrytis bunch rot, weather data, spatial interpolation, natural neighbour.

### INTRODUCTION

Tools to predict disease risk have the potential to improve disease management on farms, vineyards and orchards. Many models have been developed to provide objective prediction of disease risk using hourly weather data. Weather data can be obtained from measurements, but it is difficult to install and manage weather stations at every site because of equipment and labour costs.

Weather data can also be estimated using existing weather station networks located near to the site of interest. The simplest way to do this is to use the nearest station, but geospatial variability of weather limits the usefulness of this approach. A more reliable method would be to use spatial interpolation, which calculates a weighted average of weather measurements from nearby weather stations (Snell et al. 2000). Methods for spatial interpolation that have been used include inverse distance weighed averages, splines and krigging (Meyers 1994).

The set of nearby stations for spatial interpolation can be selected with an algorithm that uses, for example, the stations within a specified distance from the site, or a specific number of stations. However, the optimum distance or number of stations is likely to differ for different locations because the configuration of the weather station network is not uniform. One way to avoid setting a fixed distance or number is to find the natural neighbours of the site based on geometry (Sibson 1980).

When weather data are estimated using spatial interpolation, errors will arise even with an optimal set of nearby stations and it is important to validate any spatial interpolation method that is used. The natural neighbour method of spatial interpolation has been used in ground water modelling (Tsai et al. 2005), but not to obtain weather data for implementation of agricultural models. The objectives of this study were to determine

the accuracy of interpolated weather estimates using the natural neighbour method in comparison with actual weather measurements. The influence of such estimates on a model to predict risk of botrytis bunch rot on grapes was also assessed.

### MATERIALS AND METHODS

Two sites in the HortResearch horticultural weather station network, Renwick (41°30'S 173°51'W) in Marlborough and Tomoana (39°37'S 176°53'W) in Hawke's Bay, were selected as validation sites to evaluate accuracy of interpolated weather data. Air temperature, vapour pressure and wind speed were estimated from weather measurements at nearby stations selected using a natural neighbour (NN) method (Sibson 1980; Watson 1999). The natural neighbour algorithm takes into account distance and bearing of weather stations using the geometrical technique known as Voronoi tessellation (Sibson 1980).

In this study, four weather stations were selected as neighbour stations for each of the Renwick and Tomoana sites. Three of the neighbour stations for Renwick were in vineyards on the Wairau plain, with the nearest 2.2 km and the furthest 5.6 km from Renwick. The other neighbour station was in the Nelson region, which was 86 km distant from Renwick. Three neighbour stations for Tomoana were in orchards on the Heretaunga plain, with the nearest 4.1 km and the furthest 7.8 km from Tomoana. The fourth neighbour station was in Gisborne, which was 136 km distant from Tomoana.

Weather data were collected at neighbour weather stations as well as validation sites from November to March in 2003-4 and 2004-5. Interpolated weather data,  $V$ , were obtained using measurements at the neighbour stations,  $M$ , as follows:

$$v = \sum \left[ M_i \left( \frac{d_i}{\sum d_i} \right) \right] / \sum \left( \frac{d_i}{\sum d_i} \right)$$

where  $d_i$  is the distant between the validation site and neighbour weather station,  $i$ .

An empirical model based on a fuzzy logic system, the FL model, was used to estimate wetness duration (Kim et al. 2004). Cloud parameterisation was added to the FL model to estimate incoming long wave radiation (Collins et al. 2004) in this study. Estimates of air temperature, vapour pressure and wind speed using spatial interpolation were used as input to the FL model to obtain interpolated datasets of wetness duration. These processes were carried out using software written in C++.

The Bacchus model was used to predict the risk of botrytis bunch rot on grapes. This model (B. Rengasamy & R. Edwards, unpubl. data) calculates a risk index ( $I$ ) for each hour with surface wetness:

$$I = 84.37 - 7.238T + 0.1856T^2,$$

where  $T$  is the mean air temperature during a wet hour. The  $I$  values were summed for the duration of each wet period using the implementation of Bacchus in the Metwatch™ program (HortPlus Ltd, Cambridge, New Zealand). Sets of risk indices were obtained using actual and interpolated weather data as inputs. They were compared for a whole growing season (November to March) as well as for the periods relevant to grape development, from November to January (flowering to bunch closure) and from February to March (veraison to harvest).

Root mean square error (RMSE) was determined for estimates of air temperature. Mean error (ME) and mean absolute error (MAE) were obtained to evaluate accuracy in estimation of wetness duration using interpolated weather data. Because the Bacchus risk from the MetWatch™ output was characterised by both the timing of wet periods and the degree of risk, it needed to be represented in two dimensions. Thus, the two-dimensional Kolmogorov-Smirnov statistic (Peacock 1983; Fasano & Franceschini 1987; Press et al. 1997) was used to test the null hypothesis that the Bacchus risk calculated from actual weather data agreed for each seasonal period with that calculated from the interpolated weather data.

## RESULTS

**Estimation of air temperature and wetness duration**

Interpolated air temperature using the NN method was reasonably accurate at both Renwick and Tomoana sites (Table 1), with values of  $R^2$  and slopes of the regression lines close to 1. The RMSE was  $< 1^\circ\text{C}$  at both sites, although Tomoana had a slightly greater RMSE than Renwick during both 2003-4 and 2004-5 seasons.

Interpolated wetness duration using the FL model was slightly overestimated, with an ME of 0.7 h/day at Renwick and 1.7 h/day at Tomoana. The MAE in estimation of wetness duration was smaller at Renwick than Tomoana during both 2003-4 and 2004-5 seasons.

**TABLE 1: Root mean square error (RMSE), regression coefficient ( $R^2$ ) and slope of regression line (b) for estimates of air temperature ( $^\circ\text{C}$ ) and mean error (ME), standard error of the mean difference (SEM) and mean absolute error (MAE) for wetness duration (h/day) at Renwick and Tomoana obtained by spatial interpolation.  $N_H$  and  $N_D$  represent the number of hours and days, respectively.**

	Air temperature ( $^\circ\text{C}$ )				Wetness duration (h/day)		
	$N_H$	RMSE	$R^2$	b	$N_D$	ME <sup>1</sup> (SEM)	MAE <sup>2</sup>
Renwick							
2003-4	3648	0.5	0.99	0.97	152	0.7 (0.3)	2.5
2004-5	3624	0.5	0.99	0.98	151	0.8 (0.3)	2.8
Both	7272	0.5	0.99	0.97	303	0.7 (0.2)	2.7
Tomoana							
2003-4	3648	0.7	0.99	1.00	152	1.1 (0.3)	3.4
2004-5	3624	1.1	0.98	1.06	151	2.3 (0.3)	3.6
Both	7272	0.9	0.98	1.03	303	1.7 (0.2)	3.5

<sup>1</sup>Mean error defined as  $\Sigma[\Sigma(\text{Estimate-Measurement})]/N_D$ .

<sup>2</sup>Mean absolute error given by  $\Sigma[\Sigma|\text{Estimate-Measurement}|]/N_D$ .

**Prediction of botrytis bunch rot risk**

At the Renwick site there was no significant difference ( $P>0.05$ ) in the predicted risk of botrytis bunch rot using interpolated or actual weather data as input into the Bacchus model for all or part of either season (Table 2).

In contrast, at Tomoana the Bacchus risk calculated using interpolated data was significantly different (Table 2) from that using actual data for the whole season in 2003-4 and 2004-5, and for parts of the season in 2004-5. The statistical confidence with which the null hypothesis should be rejected was greater for a whole season than for the shorter periods from November to January and from February to March (Table 2). Thus, there tended to be greater similarity between the actual and interpolated data when shorter parts of the season were used to calculate Bacchus risk than when data for the whole season were used.

**TABLE 2: Test statistic (D) and P-value for the two-dimensional Kolmogorov-Smirnov test on predicted risks of botrytis bunch rot using measured and interpolated weather.**

	Renwick				Tomoana			
	2003-4		2004-5		2003-4		2004-5	
	D	P	D	P	D	P	D	P
Nov-Jan	0.16	0.59	0.14	0.77	0.18	0.24	0.24	0.05
Feb-Mar	0.12	0.95	0.25	0.18	0.23	0.22	0.36	0.01
Nov-Mar	0.11	0.74	0.19	0.10	0.20	0.03	0.27	0.00

## DISCUSSION

These results suggest that the risk of botrytis bunch rot can be predicted reasonably accurately when weather estimates using the NN method are used as inputs to the weather-driven disease warning system. It appears that accuracy of botrytis bunch rot predictions using interpolated data may be greater over shorter periods. For spray decisions, the risk of botrytis bunch rot would be assessed over short periods, e.g. less than one month.

Interpolated air temperature was relatively accurate (RMSE of  $< 1^{\circ}\text{C}$ ) for both Renwick and Tomoana sites. The greater accuracy of interpolated weather data at Renwick than Tomoana was likely due to the shorter distance to the neighbour stations in Marlborough than Hawke's Bay. For example, the three nearest natural neighbours of Renwick were closer than those of Tomoana. In addition, the furthest stations of both sites had little impact on estimation due to small weight values.

Estimation of air temperature using the NN method had advantages over simply using weather data measured at the nearest site. For example, the NN method reduced the RMSE by 12% at Renwick and 26% at Tomoana compared with actual weather measurements at the nearest station to these (data not shown). The NN method also minimised the problem of data missing due to sensor malfunction at the nearest site by estimating weather data from other neighbour stations.

The overestimation error of wetness duration by the FL model appeared to result in the deviation between the Bacchus risks calculated from interpolated and actual wetness data. For example, the predicted risks using interpolated and actual data were significantly different at Tomoana during the 2004-5 season. This coincided with relatively high overestimation of wetness duration, e.g. 2.3 h/day, at Tomoana during the season. It is thought that such overestimation errors occurred because the FL model was developed for painted sensors whereas unpainted sensors were used to measure wetness duration at validation sites. Sentelhas et al. (2004) reported that painted sensors were more sensitive to wetness occurrence, e.g. small droplets on the sensor surface, than unpainted ones. Thus, it is likely that a part of overestimation errors resulted from hours in which wetness occurrence was predicted by the FL model, but was not identified by the unpainted sensors.

The present study illustrates the potential of estimating weather parameters for an individual site that will lead to reliable prediction of disease risks. The interpolation method is purely based on distance between a site of interest and nearby weather stations without considering terrain. Therefore, improvements in accuracy might result from inclusion of geographical information, such as topology and proximity to sea, in the interpolation procedure.

### ACKNOWLEDGEMENTS

We thank Rob Agnew and Peter Wood for maintaining the weather stations used for this study, the Foundation for Research, Science and Technology for funding this study, and B. Rengasamy and R. Edwards who developed the Bacchus model.

### REFERENCES

- Collins WD, Rasch PJ, Boville BA, Hack JJ, McCaa JR, Williamson DL, Kiehl JT, Briegleb B, Bitz C, Lin S-J, Zhang M, Dai Y 2004. Description of the NCAR Community Atmospheric Model (CAM3.0). Technical Report NCAR/TN-364+STR. National Center for Atmospheric Research, Boulder, Colorado, USA. 101 pp.
- Fasano G, Franceschini A 1987. A multidimensional version of the Kolmogorov-Smirnov test. *Monthly Notices of the Royal Astronomical Society* 225: 155–170.
- Meyers DE 1994. Spatial interpolation: An overview. *Geoderma* 62: 17-28.
- Kim KS, Taylor SE, Gleason ML 2004. Development and validation of a leaf wetness duration model using a fuzzy logic system. *Agricultural and Forest Meteorology* 127: 53-64.
- Peacock JA 1983. Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society* 202: 615–627.
- Press W, Teukolski S, Vetterling W, Flannery B 1997. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK. 645 pp.
- Sentelhas PC, Monteiro JEB, Gillespie TJ 2004. Electronic leaf wetness duration sensor: why it should be painted. *International Journal of Biometeorology* 48: 202-205.
- Sibson R 1980. A vector identity for the dirichlet tessellation. *Mathematical Proceedings of the Cambridge Philosophical Society* 87: 151-155.
- Snell SE, Gopal S, Kaufmann RK 2000. Spatial interpolation of surface air temperature using artificial neural networks: evaluating their use for downscaling GCMs. *Journal of Climate* 13: 886-895.
- Tsai FTC, Sun N, Yeh WWG 2005. Geophysical parameterization and parameter structure identification using natural neighbors in groundwater inverse problems. *Journal of Hydrology* 308: 269-283.
- Watson D 1999. The natural neighbor series manuals and source codes. *Computer and Geosciences* 25: 463-466.