Leaf wetness prediction 2018

Validation and improvements

0 Introduction

0.1 Company

meteoblue is a Swiss specialist company producing high precision weather data for the entire world, using observation data, high-resolution Numerical Weather Predictions (NWP) and specialised data output methods adapted to the needs of different user groups.
meteoblue produces weather data since 2007 and produces the largest daily available data volume of any private EU weather service. The available weather archives cover 5 years in maximum detail and reaching back for more than 20 years for verification purposes. Quality verification results are shown on https://content.meteoblue.com/en/verified-quality/verification.

0.2 Distribution

meteoblue offers products, services and project resources to clients worldwide.
For representation in certain countries or market segments, meteoblue works with selected distributors, who represent, sell and service meteoblue products, services and /or project resources.

1 Summary

The goal of this study was to improve meteoblue leaf wetness prediction. The main outcomes are two kinds of improvements:

1. Product improvement (Better or more information for customers)
2. Quality improvements (Reducing forecast errors)

The necessity of the product improvement is an outcome of the study as well: When comparing the leaf wetness prediction with measurements, it turned out, that the appearance of dew and the intensity of evaporation are very different from site to site. This is logic, as they depend on microclimatic conditions and especially the land use or vegetation at a local scale. Therefore, the customer should have the option to customize the leaf wetness forecast for his specific use case (e.g. crop, surrounding land use, microclimate).

1.1 Product improvements

The product improvement of leaf wetness prediction is to introduce four new additional variables. Beside the binary leaf wetness index (0 or 1), there will be a linear leaf wetness probability and three linear indicators for precipitation, dew and evaporation (between 0 and 1). Thus, the leaf wetness prediction will contain 5 variables:

1. Leaf wetness index (binary)
2. Leaf wetness probability (%)
3. Precipitation intensity (%)
4. Dew intensity (%)
5. Evaporation intensity (%)

All five variables are consistent with each other: The leaf wetness probability of the actual hour depends on, the value of the last hour plus precipitation intensity and dew intensity of the actual hour and reduced by the actual evaporation intensity. When the leaf wetness probability is higher than 50 %, the leaf wetness index is equal to one.
Beside the information whether the leaf is wet, the customer has additional information about how likely leaf wetness will appear and based on which weather component it appears. Sophisticated users can customize their own leaf wetness by giving localized weights to each of the components (precipitation intensity, evaporation intensity, dew intensity).

1.2 Quality improvements

Quality improvements refer to improvement of the binary leaf wetness prediction as described in the validation section of this report (Chapter 3). Using the new approach, the day ahead forecasting skill of leaf wetness prediction was more then doubled. The probability to detect a leaf wetness event was improved from 34% to 54%, while the false alarm rate was reduced from 44% to 37%, based on single model data. Due to the post processing of precipitation by multi-model and nowcasting techniques within the operational meteoblue forecasting system, these values can be assumed to be about 10% better. For further improvement the forecast has to be tuned on the specific microclimatic conditions of a site as described in the section above.

2 Study outlines

2.1 Importance of leaf wetness prediction

Many plant diseases are triggered by wet and warm weather conditions, because they are favourable for fungi and bacteria. This primary infection is usually, when a minimum temperature is reached, and the leaves are wet for several hours. Thus, the primary infection of these plant diseases is highly dependent on leaf wetness. The knowledge about leaf wetness risk is crucial for the timing of disease control actions like the spraying of fungicides. The peronospora, for example is one of the economically most important plant diseases and triggered, when leaves are wet for more than four hours and the temperature is above 12° Celsius.

Spraying activities are usually most effective if they are applied just before the primary infection. If this was missed out, it has to be applied just a few hours after the primary infection. Therefore, it is very helpful to have an accurate forecast of the leaf wetness.

2.2 What is leaf wetness

Leaf become wet due to precipitation or dew and get dry again by evaporation. This means the main variables, that influence leaf wetness are:

- Precipitation
- Humidity (Dew)
- Evaporation

There are meteorological sensor measuring the leaf wetness index as a binary variable (see Figure 2.1).

Figure 2.1: Leaf wetness sensor

2.3 Study approach

A feedback to the actual leaf wetness prediction was, that dew and evaporation were underestimated. The goal of this study is to improve the leaf wetness prediction in this matter. Meteorological weather data of the years 2016 & 2017 were provided by the company Geosense for 9 stations, situated in vineyards in different regions in Germany, Switzerland and Austria.
The accuracy of the actual approach was compared to another approach by Geosense and a new optimized approach.
A third approach was developed by trying out many different options, first to estimate each indicator and second weighing the indicator so it fits for most of the available sites.

2.3.1 Leaf wetness models

The leaf wetness prediction model actually used by **meteoblue** (mb_old) is based on the saturation deficit and precipitation. The principle is quite simple: If the saturation deficit is equal or below 0K or the precipitation is higher than 0, than the leaf wetness is 1. This is calculated for each hour.

The leaf wetness prediction model used by **Geosense** is also based on the saturation deficit and precipitation. The principle is: If the saturation deficit is equal or below 1.2K or the precipitation is higher than 0, than the leaf wetness event begins. It lasts until three criteria are fulfilled:

- No precipitation in the last hour
- Relative Humidity below 85 %
- Saturation deficit below 1.2K

After checking the data of one station in Freiburg Germany, it was assumed that with a two-step model approach a higher accuracy could be reached:

1. Calculate intensity
    a. Precipitation intensity
    b. Evaporation intensity
    c. Dew intensity

2. Calculate leaf wetness based on indicators (1)

Beside a higher accuracy the approach with 3 indicators (mb_new) has the advantage, that it can easily be calibrated for a specific location, as for example the appearance of dew is highly dependent on the local vegetation and land use.

Therefore, it is recommended to use these variables (leaf wetness and indicators), not only as a binary variable but as an indicator (e.g. probability of dew in percent) and use them (especially dew) for local calibration.

2.3.2 Validation measures

Leaf wetness is, similar to precipitation, an event-based variable. Furthermore, the leaf wetness is measurements are binary variables and standard error measures like systematic and absolute deviations are not a suitable choice.

There are three different statistical measures, that have been analysed to compare these models:

- Probability of detection (**POD** also: hit rate) of a leaf wetness event for the 24 h forecast horizon
- False alarm rate (**FAR**) of a leaf wetness event for the 24 h forecast horizon
- Heidke Skill Score (**HSS**) of a leaf wetness event for the 24 h forecast horizon. The HSS is a combination of the POD and the FAR.

These measures are suitable, because they can easily be converted into economical reality. A false alarm means a spraying activity with bad timing. The hit rate (POD) shows how many events are predicted correctly, while the others are missed out.

In this report it is always referred to the 24h forecast for precipitation events only. The analyses could be repeated for other thresholds, but earlier studies have shown, that the patterns are similar.

The error measures were calculated for 3-hourly aggregated values.

3 Validation results

For technical reasons, in this approach only single model data was used. Within operational forecast systems of meteoblue best fit multi-model is used for precipitation, which will lead to a higher precision of leaf wetness. Therefore, the precision of the final operational leaf wetness prediction will be much higher, than the results represented here. As an indicator of the difference the precipitation multi-model validation can be cited: HSS raised from 0.39 for a single model to 0.43 for the multi-model and the POD improved from 56% to 70%.

3.1 Summary

As shown in Table 1 the forecast skill could be highly improved for all measures. The mean skill (HSS) was improved from 0.17 to 0.39. While the hit rate (POD) was raised from 34% to 54 %, the false alarm rate (FAR) was reduced from 44% to 37%. When looking at the distribution of each validation measure over different stations (Figure 3.1 - Figure 3.3), it turns out, that the skill could not be improved for all locations. This indicates again the need for local tuning: On one location the validation measures can still be improved by a different weighing of the three component indicators, but this will worsen the forecast at the other sites.

Table 1: Overview of Maximum, Minimum and Mean validation measures for 3h values

	old leaf wetness prediction			new leaf wetness prediction		
	HSS	POD	FAR	HSS	POD	FAR
MAX	0.36	0.56	0.81	0.56	0.74	0.63
MEAN	0.17	0.34	0.44	0.39	0.54	0.37
MIN	0.02	0.20	0.18	0.11	0.36	0.16

3.2 Heidke Skill Score (HSS)

Figure 3.1 gives an overview of the distribution of the 3-hourly HSS overall locations in 2017:

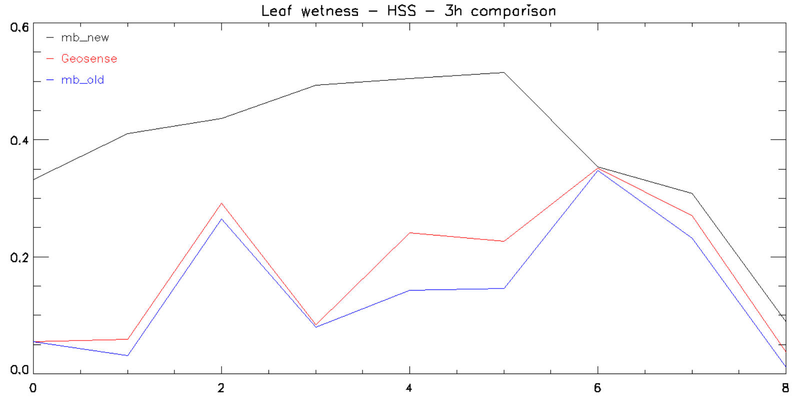

Figure 3.1: Distribution of the HSS calculated for 3-hourly averages overall locations for the three different approaches

3.3 Probability of detection (POD)

Figure 3.2 gives an overview of the distribution of the 3-hourly hit rate (POD) overall locations in 2017:

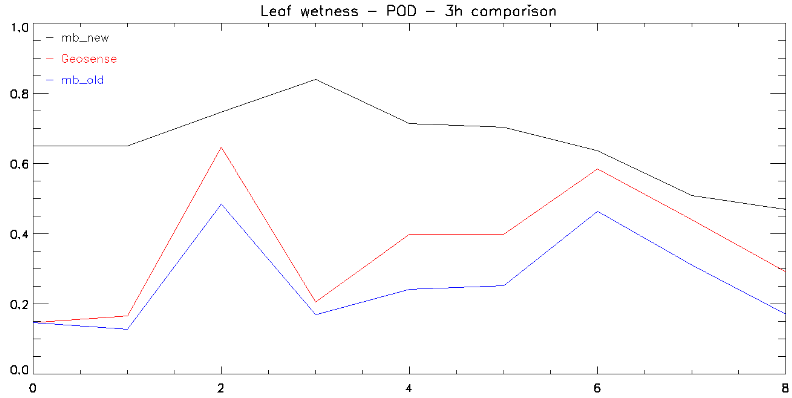

Figure 3.2: Distribution of the POD calculated for 3-hourly values overall locations for the three different approaches

3.4 False Alarm Rate (FAR)

Figure 3.3 gives an overview of the distribution of the 3-hourly false alarm rate (FAR) overall locations in 2017:

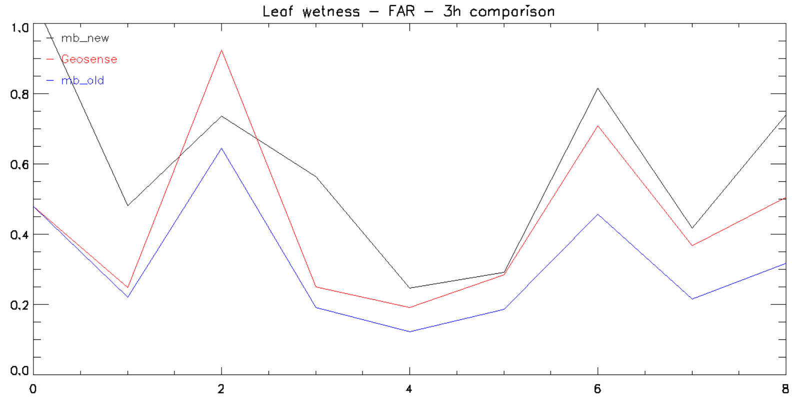

Figure 3.3: Distribution of the FAR calculated for 3-hourly values overall locations for the three different approaches