Tellus (2006), 58A, 558–564
Printed in Singapore. All rights reserved

TELLUS

Probability forecasts for weather-dependent agricultural operations using generalized estimating equations

By NINA K. DETLEFSEN*, *Department of Agroecology, Danish Institute of Agricultural Sciences, Research Centre Foulum, DK-8830 Tjele, Denmark*

(Manuscript received 23 November 2005; in final form 24 May 2006)

ABSTRACT

In agricultural production many operations depend on the weather. In this paper, a model is investigated that calculates the probability for the execution of a given operation which depends on several meteorological parameters. The model is based on a 48-hr numerical weather forecast with hourly resolution. The probability forecasts are compared to the numeric forecasts for the operation based on the numeric weather forecast.

The model is a logistic regression model with generalized estimating equations. The Brier skill score, sharpness and reliability diagrams and relative operating characteristic curves are used to evaluate the model.

The model setup described is dynamic in the sense that on a given date, parameters are estimated based on history and these parameter estimates are used for calculating the probability forecasts. This means that parameter estimates adapt automatically to seasonal changes in weather and to changes in numerical weather forecasts following developments in the forecast models.

In this paper, we perform model output statistics, which tune the numeric weather forecast to an operation that depends on several meteorological parameters rather than only tuning a single weather parameter. Although some problems occurred, the model developed showed that the numerical forecast for such an operation could be improved.

1. Introduction

In Denmark an internet-based information and decision-support system for plant production, PlanteInfo (Jensen et al., 2000), has been in operation for more than 10 yr. A subscription system enables the information to be personalized (e.g. location). Many of the applications in this system are weather-based and therefore require either weather observations, forecasts or both. The weather data in PlanteInfo are provided by the Danish Meteorological Institute (DMI) who has developed the AgroMeteorological Information System (AMIS) to interpolate meteorological data (Steffensen et al., 2001) to a 10×10 km grid.

Some of the forecasts, the ones analysed in this paper, are calculated by the High Resolution Limited Area Model (HIRLAM; see Sass et al., 2002). The HIRLAM model is rerun every 6 hr and the timespan of the model gives results for a 48-hr period with an hourly resolution. The spatial resolution in the model is about 15 km. The results from the model are then interpolated to the grid points in AMIS. DMI has analysed the uncertainty of the forecasted parameters in Rasmussen et al. (2000). One of the conclusions from this analysis is that HIRLAM overpredicts small amounts of precipitation in general.

Changnon (2004) discuss usage of climate predictions by United States agribusinesses during 1981–2001. They found that an increased usage overtime was attributed to five factors including growing economic pressures in agriculture, improvements in access to predictive information, improved accuracy of predictions, better formats and timeliness of predictions and increasing employment of atmospheric science expertise, either in firms or as advisors. They state that results indicate that major opportunities exist for further enhance future usage of prediction sin United States agribusinesses. One example where meteorological data was used for agronomic purposes is reported by Suleiman. Suleiman and Crago (2004) proposed a linear regression procedure to estimate evapotranspiration from grassland using a dimensionless temperature, which is a function of surface temperature, air temperature and net radiation.

Other examples of model output statistics can be found in the literature. Precipitation is often the subject of investigation. Sohn et al. (2005) described an application of model output

*Correspondence.
e-mail: Nina.Detlefsen@agrsci.dk
DOI: 10.1111/j.1600-0870.2006.00196.x

statistics where they compare four statistical models to improve the prediction of heavy rain in South Korea. The four models was a linear regression model, logistic regression model, a neural network and a decision tree model. Their results showed that the logistic regression model gave the best results. Sokol and Rezacova (2000) compared two statistical methods, one based on model output statistic and one based on successive learning technique, to increase the accuracy of the local categorial (yes/no) precipitation forecast. For each approach several statistical models for the relationship between the numerical weather prognoses and the categorial precipitation were studied. They found that both approaches improved the numerical weather prognoses. Steffensen (2002) developed a logistic Kalman filter for cases of heavy precipitation. This analysis shows that the HIRLAM model underestimates these cases. The model is good at forecasting cases of no heavy precipitation. Applequist et al. (2002) has conducted an analysis of different methodologies for creating probabilistic precipitation forecasts. They used the Brier skill score as en evaluation method and their analysis of several methods showed that logistic regression gave the best results.

Another parameter subject of research is temperature. Anadranistakis et al. (2002) and Galanis and Anadranistakis (2002) developed models for correction of operationally forecasted near surface temperatures in Greece. Their models use the Kalman filter techniques. In Anadranistakis et al. (2004) the authors has, in addition, improved the forecast for relative humidity.

Other research has concentrated on making a probabilistic forecast using quantile regression methods. In Bremnes (2004) the author investigates the possibilities for making daily precipitation forecasts using this method and reports good results.

The planning of many agricultural operations could be improved by a good weather forecast with an hourly resolution. One example is the spraying of crops. Depending on crop, pesticide/fungicide and equipment, this operation requires certain weather conditions to be fulfilled. The operation typically lasts only a limited number of hours giving the farmer a choice of when to perform the operation in near future. Therefore, a reasonably precise local weather forecast will help the planning of this type of operation. The aim of this paper is to investigate the possibilities for using the numerical weather forecasts to calculate the probability for an operation taking place that requires several meteorological parameters to be fulfilled such as the spraying of crops, as described above. These probability forecasts will be compared with the numerical forecasts for the operation to see if the numerical forecast is improved.

2. Data description

We choose to perform the analysis for the gridcell where the weather station in Foulum is located. Data used in this analysis from 2002 and 2003 consist of:

(1) Numerical 48-hr period weather forecasts in the AMIS grid where Foulum is located with hourly resolution. These forecasts are based on the HIRLAM model, which is run every 6 hr.
(2) Hourly observations from Foulum weather station.

The parameters used in both the numerical weather forecast and the observed data are temperature, precipitation and wind speed.

Let t denote an hour. Then for hour t the weather is observed (by the three weather parameters above) at Foulum weather station. Let $\mathbf{o}_t$ denote these observations. For the gridcell where Foulum is located and for each hour t, we have forecasts of different age for each of the three weather parameters. Since the HIRLAM model is run every 6 hr with a time span of 48 hr there will be up to eight different forecasts for the same hour t.

New probability forecasts are only needed when new information is received, that is every time a new numerical weather forecast is received from the DMI. Figure 1 illustrates the information available when the analysis is performed.

For a probability forecast t hours ahead, the amount of available information will depend on the size of t. If $t \in [1; 6]$, there will be eight different forecasts available that include all three weather parameters, whereas if $t \in [43; 48]$ then there will only be one forecast of the three weather parameters available.

When more than one forecast is available, only the most recent forecast is taken into account. One could, on one hand, argue that the older forecasts are outdated and that the information is already included via the HIRLAM model, from which the newer forecast originate, and on the other hand, that they contain valuable information regarding predictability. We believe that correlation between hours is more important for predictability and have chosen to model this using generalized estimating equations (GEE). In this paper, we therefore report results from the analysis where we take the first 6 hr of each forecast, which means that in the data set containing forecasts we have

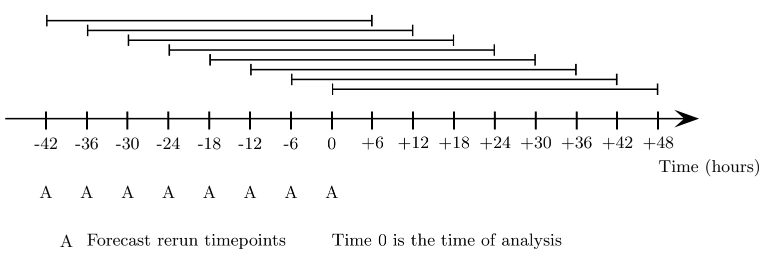

Fig. 1. Available information at time 0 and the next 48 hr.

exactly one observation for each hour. This is a prerequisite for the statistical model framework we are using, which will be described later.

The explaining variables or covariates are based on the weather forecasts, whereas the dependent variable is a function of the observations.

3. Description of the agricultural operation

In this paper, we use spraying conditions as the operation to be analysed. Spraying conditions differ for different fungicides/pesticides. We analyse a case which requires the following meteorological conditions:

(1) Precipitation less than or equal to 0.2 mm.
(2) Wind speed less than or equal to 3 m s^{-1}.
(3) Temperature greater than or equal to 12 °C.

From the numerical weather forecast a numerical forecast for the operation, z_t, can be calculated. $z_t = 1$ if the weather forecast fulfills the conditions for the operation and otherwise $z_t = 0$. Equally, $y_t = 1$ if the observed weather for hour t, $\mathbf{o}_t$, fulfills the conditions above and otherwise $y_t = 0$. In the rest of the paper 'numerical forecast' refers to z_t and 'observed operation' refers to y_t.

The numerical forecasts and the observed operation can be compared. A contingency table, which is given in Table 1, can subsequently be computed. Hourly data from Foulum weather station (gridcell = Foulum) for the years 2002 and 2003 have been used.

From Table 1 it can be calculated that only in about 11% of observed hours were the weather conditions suitable for spraying. In about 61% of these hours there is a correct numerical forecast for the observed operation that is the hit rate is about 61%. On the other hand, we can see that of all the hours where the conditions are not suitable, the numerical forecast have forecasted 97% of them correctly and only 3% were false alarms.

4. The model framework

In this section, the statistical model framework used to develop models that produce probability forecast for the case described in Section 3 is investigated. The method of GEE is used. When specifying a GEE model, in addition to specifying the systematic part it is necessary to impose a correlation structure for the observations as part of the random part of the model. This method is often used when the data set contains observations from the same unit. We consider days as the same unit. A general description of GEE can be found in Dobson (2002).

Table 1. Contingency table for numerical forecasts and observations for the case. Data from 2002 and 2003 is used

Forecast (z_t)	Observed operation (y_t)		
	0	1	Total
0	**3362**	**339**	3701
1	**251**	**602**	853
Total	3613	941	4554

First it is noted that y_t is Binomial/Bernoulli distributed, $y_t \sim B(1, \pi_t)$ where π_t is the probability that $y_t = 1$. In statistical literature, y_t is often referred to as the response.

The systematic part of the statistical model is of the following form:

$$\operatorname{logit}(\pi_t) = \log\left(\frac{\pi_t}{1-\pi_t}\right) = \mu + \beta^T \mathbf{x}_t, \tag{1}$$

where $\mathbf{x}_t$ is a vector of covariates for the given model which is based on the weather forecast, μ is a model parameter usually called the intercept and $\boldsymbol{\beta}$ is a vector of model parameters for the covariates. Later it will be discussed which covariates were chosen. π_t is referred to as the 'probability forecast' since this is the forecasted probability that $y_t = 1$.

The responses, y_t's, are not expected to be independent, since precipitation in one hour increases the likelihood of precipitation in the following hour. The responses for two hours in the same day are assumed to be correlated, whereas the responses for 2 hr on different days are independent. We therefore consider observations from the same day as observations taken within the same unit. The working correlation matrix $R(\alpha)$ is chosen to have an autoregressive correlation structure with parameter α. The assumption behind this correlation structure is that observations close in time are closer related than observations further apart. The correlation decays exponentially with time. The analysis is limited to hours 6 a.m. to 8 p.m. which are the interesting ones from an agronomic perspective, in order to reduce the size of the working correlation matrix. $R(\alpha)$ then becomes a 15×15 matrix with the following structure:

$$R(\alpha) = \begin{bmatrix} 1 & \alpha & \cdots & \alpha^{13} & \alpha^{14} \\ \alpha & 1 & \cdots & \alpha^{12} & \alpha^{13} \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ \alpha^{14} & \alpha^{13} & \cdots & \alpha & 1 \end{bmatrix}.$$

Note here that α is raised to the power of the superscripts.

The estimation procedure for GEE requires that there are no missing data for the days included in the analysis. But the days included do not have to be consecutive.

The models are expected to run in a dynamic sense. That is, at a given date the model parameters μ, $\boldsymbol{\beta}$ and α are estimated on the most recent data and then these parameter estimates are used to calculate or predict a probability forecast (according to 1) for future hours. In this way parameter estimates will change continually over the year to accommodate the variation in weather throughout the year.

The fitted probabilities for the past data on which the parameters were estimated are referred to as 'fitted probabilities' and the probability forecasts ahead in time which have not been used to estimate parameters are referred to as 'future probability forecasts'.

One issue to decide is the amount of history to include when estimating the model parameters. Experiments showed that if more historic data were included, this did not necessarily improve the quality of the future probability forecasts as it slowed down the adaptation to changes in weather. If the amount of historic data is too small, it is more likely that the procedure does not converge and therefore no probability forecasts will be available at all.

All calculations were performed using the statistical software SAS®, SAS Institute Inc., Cary, NC, USA.

5. Methods for evaluating the model

The Brier skill score (BSS), relative operating characteristic (ROC) curves and sharpness and reliability diagrams are chosen for evaluating the model. A short description of the methods used is given below. A good overview over forecast verification is given in Jolliffe and Stephenson (2003), where also the methods used in this paper are described in detail.

The BSS is used to evaluate the quality of the probability forecast. The form used here is based on the Brier scores (BS):

$$\mathrm{BS}_{\mathrm{model}} = \frac{1}{n}\sum_t (\pi_t - y_t)^2 \qquad \mathrm{BS}_{\mathrm{ref}} = \frac{1}{n}\sum_t (z_t - y_t)^2,$$

where n is the number of hours included in the analysis, π_t is the probability forecast, z_t is the numerical forecast and y_t is the observed operation. $\mathrm{BS}_{\mathrm{model}}$ and $\mathrm{BS}_{\mathrm{ref}}$ are the Brier scores for the probability forecast and the numeric forecast, respectively. The numeric forecast is used as a reference since this is the best forecast for the case not including the model in this paper. Then:

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}_{\mathrm{model}}}{\mathrm{BS}_{\mathrm{ref}}}.$$

For each day, the BSS for the fitted probabilities (past BSS) and future probability forecasts (future BSS) are calculated. The same number of hours is used in both BSSs.

Another method of evaluating the model is to look at sharpness and reliability diagrams. For a given model, a data set containing the future probability forecast for each hour and the observed operation is created. Then the data set is partitioned into categories according to probabilities (0–0.05, 0.05–0.1, ..., 0.95–1) and for each category the frequency of hours for which $y_t = 1$ is found. Reliability requires that this frequency is about the same value as the probabilities in the interval used to create the category. Plotting the frequencies against the midpoints of the probability intervals gives a reliability diagram. If the number of members of each category is calculated and these are plotted against the midpoints of the categories, a sharpness diagram is obtained.

Finally, the ROC curves is used to evaluate the results. For each possible decision probability threshold, a forecast based on whether the predicted probability forecast is above the decision probability threshold or not is calculated. This forecast and the observed operation are used to calculate a frequency table, which gives the hit rate and the false alarm rate. A ROC diagram is obtained by plotting the hit rate against the false alarm rate for each possible decision probability threshold.

6. The case study for Foulum

This section reports the results of the model for the case described in Section 3 for Foulum. Two years of data (2002 and 2003) were extracted from Foulum weather station. When we ran the model, it turned out that during the winter season the model was unable to estimate parameters since there were none or only very few instances where $y_t = 1$. For this reason we concentrated on the growing season, which is also the interesting one from a farming perspective.

The model is only run once everyday because GEE cannot handle missing data for a unit, which in our case is days.

We have chosen to select covariates manually instead of using a more automated procedure, because we believe it is important to build a model that can be explained to users and that makes sense to people with a knowledge of the application. A few simple arguments will limit the number of covariates and help the most important ones to be chosen.

For this case it is expected that the numerical precipitation forecast for hour $t - 1$ and $t + 1$ will influence the responses. Given that the precipitation forecast for hour t has some value less than or equal to 0.2 mm, then if the precipitation forecast for the 2 hr either side of t are small, it is more likely that the response will be 1 than if the precipitation forecast for the two proximate hours are large. The temperature and wind speed forecast for the hours $t - 1$ and $t + 1$ are not expected to influence the response as much as the precipitation forecast. These are therefore left out. The result is that the covariates include the most recent forecast for all parameters in addition to the forecast for precipitation for hour $t - 1$ and $t + 1$.

The final model included seven covariates and an intercept. The covariates were precipitation, wind speed, temperature, precipitation times wind speed, wind speed times temperature, precipitation for hour $(t - 1)$ and precipitation for hour $(t + 1)$.

The amount of history to be included to estimate parameters for the model also has to be determined. It was found that a history of 40 d of hourly values was reasonable. This gave a good fit to historic data and the best future probability forecasts. As already noted, the 40 d need not be consecutive, but it was necessary for each day not to have missing values from 6 a.m. to 8 p.m. in order to use the estimation procedures for GEE.

The procedure were that for each day at 6 a.m. we estimated parameters for the model based on 40 d of history. Then we used these parameter estimates to calculate future probability

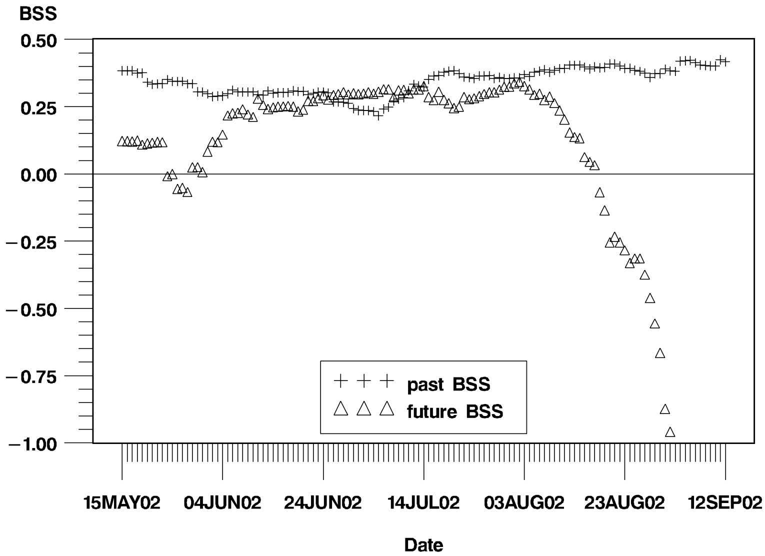

Fig. 2. BSS for 2002. Past BSS are calculated on basis of data in the training set (40 d) and future BSS is calculated on basis of predictions 40 d ahead (15 hr each day).

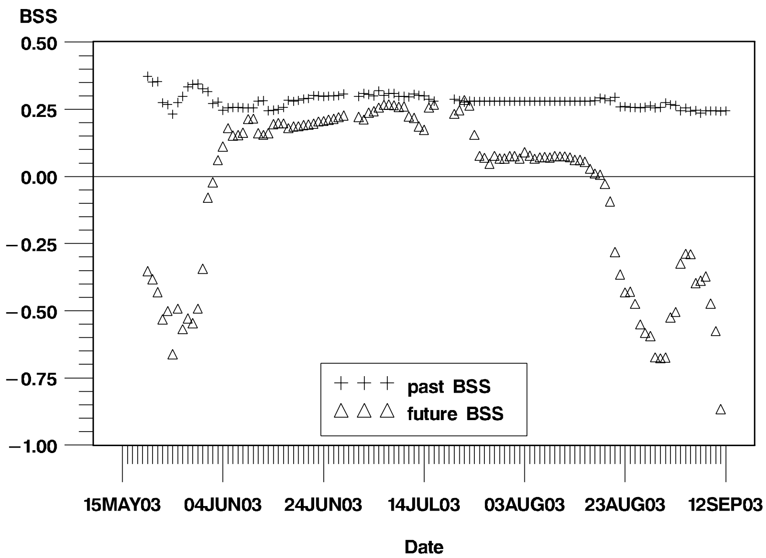

Fig. 3. BSS for 2003. Past BSS are calculated on basis of data in the training set (40 d) and future BSS is calculated on basis of predictions 40 d ahead (15 h each day).

forecast for the 15 hr between 6 a.m. and 8 p.m. on the day of analysis.

The BSS diagrams for 2002 and 2003 are given in Figs. 2 and 3. The horizontal axes in the figures only cover from mid-May to mid-September because outside this time interval there are too few hours where $y_t = 1$ for the estimation procedure to converge.

In Figs. 2 and 3 both the past BSS (calculated from a 40 d history) and the future BSS (calculated from a 40-d future probability forecast) for each day are shown. It can be seen that the past BSS is rather constant around 0.3, whereas the future BSS lies around 0.2 from start-June to mid-August. The latter indicates a reasonably good future probability forecast. In the beginning and at the end of the season there are some difficulties indicated by very negative future BSS. The reason is that at both ends of the growing season the weather is changing rather rapidly. Then, when parameters are estimated based on a 40 d history and these parameters are used to predict 40 d, the weather has changed significantly.

Figure 4 presents the reliability histogram for the future probability forecasts. Data for 2002 and 2003 are used for days where estimation procedures converged. It can be seen that reliability deteriorates slightly decreasing for the high probability categories, but otherwise the histogram shows good reliability.

Figure 5 shows a high reliability for the small probability categories, but a decreasing reliability for the high probability categories. From a farming perspective we would have preferred that there was a higher reliability in the high probability categories.

Figure 6 shows the ROC curve for the future probability forecasts. The area under the curve can be seen as an indicator of how

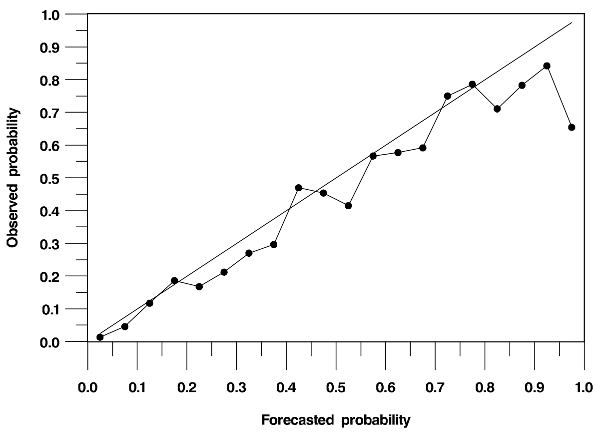

Fig. 4. Reliability histogram for future probability forecasts in 2002–2003.

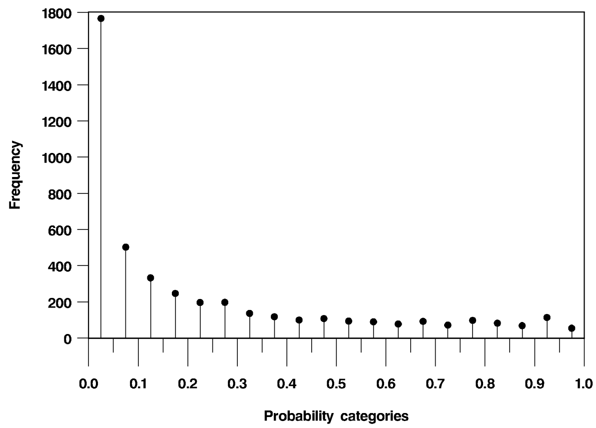

Fig. 5. Sharpness histogram for future probability forecasts in 2002 and 2003.

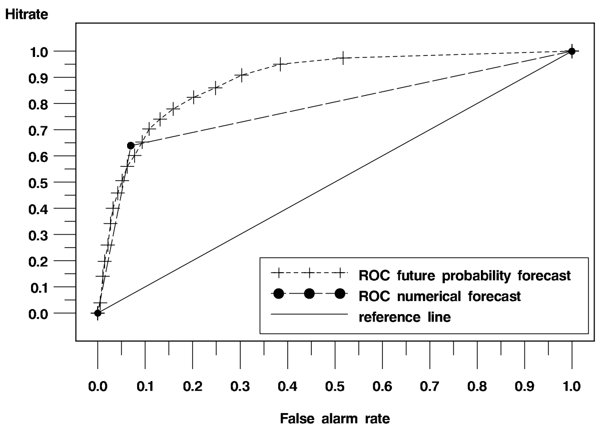

Fig. 6. ROC diagram for future probability forecasts for 2002 and 2003.

much is gained by the model (Jolliffe and Stephenson, 2003), and it can be seen that the future probability forecast improves the numerical forecast slightly. The ROC curve for the numerical forecast is calculated from the days in 2002 and 2003 where the model was able to calculated probability forecasts that is the estimating procedure converged. Therefore, the hit rate and false alarm rate in Fig. 6 differs from the ones calculated in Section 3. For the false alarm rate where the numerical forecast is available it actually has the best hit rate. An advantage of using the model is that the user, depending on the operation, can set a probability level to ensure a high level of hit rates. The future probability forecasts give more detailed information than the numerical forecasts does (only 0 or 1).

7. Summary and discussion

This paper describes a model to forecast spraying weather conditions based on the numerical weather forecast. A dynamic model setup was wanted in the sense that the estimates of the model parameters should be able to continuously adapt to changes in models for the numerical weather forecasts or weather changes in general. One reason for this was that the models calculating the weather forecasts are undergoing continuous development by meteorologists and therefore might change now and then without the user's knowledge.

The results in this paper were based on the first 6 hr of the numerical weather forecasts. The case study was carried out for the gridcell where Foulum is located.

The methods used for evaluating the model are the BSS, the reliability diagram, the sharpness diagram and the ROC curves. All these methods show a good fit with the suggested model. In addition, more detailed information in the form of probabilities is obtained. From the analysis in this paper it can be concluded that it is possible to build models for cases involving several weather parameters and improve the numerical weather forecast.

In this paper, we have made some choices of how to build the model. First, we wanted to investigate dependencies between hours and therefore chose to use the method of GEE. Then we needed a data set that contained exactly one observation and forecast per hour. This was created by taking the first 6 hr of every forecast. It would be interesting to investigate if an analysis based on the numerical weather forecast, for example 6 to 12 hr forecasts, shows similar results. This work is still on the to-do list. Next we chose to manually select covariates based on testing combinations suggested by experts on spraying in different time periods. This is because we think it is important to include our knowledge on spraying weather in the model. If we used an automated procedure the covariates would change overtime. It would, however, be interesting to explore whether a more automated selection of covariates by using for example cross-validation would change the results.

The model in this paper is developed for a single location housing a weather station. A big challenge for future work will be to develop a model that can be applied nationwide, even where weather stations are not available. This is needed in order to make the model operational on our information system PlanteInfo.

When the work presented in this paper was started, it was expected to result in a case specific model that would improve the numerical forecast for a given operation significantly. Although the numerical forecast has been improved, it is difficult to use the results. Firstly, the introduction of another model and the added uncertainty might not be worthwhile. Secondly, the probability forecast obtained from the model is difficult to explain to users. Most users will use a decision probability threshold automatically. Therefore, one has to investigate possible ways of disseminating probability forecasts carefully before operationalizing this kind of model. Still, this work has shown that the numerical weather forecast can be adapted to a specific operation with good results.

8. Acknowledgments

I would like to thank the reviewers for their useful comments which have greatly improved the manuscript.

References

Anadranistakis, M., Lagouvardos, K., Kotroni, V. and Skouras, K. 2002. Combination of Kalman filter and an empirical method for the correction of near-surface temperatur forecast: application over Greece. *Geophys. Res. Lett.* **29**(16), art. no. 1776.

Anadranistakis, M., Lagouvardos, K., Kotroni, V. and Elefteriadis, H. 2004. Correcting temperature and humidity forecasts using kalman filtering: Potential for agricultural protection in northern greece. *Atmos. Res.* **71**, 115–125.

Applequist, S., Gahrs, G. E. and Pfeffer, R. L. 2002. Comparison of methodologies for probabilistic quantitative precipitation forecasting. *Wea. Forecast.* **17**, 783–799.

Bremnes, J. 2004. Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. *Mon. Wea. Rev* **132**(1), 338–347.

Changnon, S. A. 2004. Changing uses of climate predictions in agriculture: implications for prediction research, providers, and users. *Wea. Forecast.* **19**(3), 606–613.

Dobson, A. J. 2002. *An Introduction to Generalized Linear Models*, Chapman & Hall/CRC, Florida.

Galanis, G. and Anadranistakis, M. 2002. A one-dimensional kalman filter for the correction of near surface temperature forecasts. *Meteorol. Appl.* **9**, 437–441.

Jensen, A. L., Boll, P. S., Thysen, I. and Pathak, B. K. 2000. Pl@nteinfo—a webbased system for personalised decision support in crop management. *Comput. Electron. Agric.* **25**, 271–294.

Jolliffe, I. T. and Stephenson, D. B. (Eds.), 2003. *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. Wiley, John Wiley & Sons Ltd., England.

Rasmussen, A., Sørensen, J. H., Nielsen, N. W. and Amstrup, B. 2000. Uncertainty of meteorological parameters from DMI-HIRLAM. *Scientific Report 00-07*, Danish Meteorological Institute.

Sass, B. H., Nielsen, N. W., Jørgensen, J. U., Amstrup, B., Kmit, M. and co-authors. 2002. The operational DMI-HIRLAM system. *2002-version. Technical Report 02-05*, Danish Meteorological Institute.

Sohn, K. T., Lee, J. H., Lee, S. H. and Ryu, C. S. 2005. Statistical prediction of heavy rain in South Korea. *Adv. Atmos. Sci.* **22**(5), 703–710.

Sokol, Z. and Rezacova, D. 2000. Improvement of local categorical precipitation forecasts from an NWP model by various statistical post-processing methods. *Stud. Geophys. Geodaet.* **44**, 38–56.

Steffensen, M. 2002. Logistisk Kalman filter for kraftig nedbør. *Technical Report 02-28*, Danish Meteorological Institute.

Steffensen, M., Vejen, F. A. H., Overgaard, S., Scharling, M. and Jüngling, H. 2001. Evaluation of the AMIS gridded observations and radar derived 24-hour accumulated precipitation by comparison with climate grid - denmark gridded observations. *Technical Report 01-13*, Danish Meteorological Institute.

Suleiman, A. and Crago, R., 2004. Hourly and daytime evapotranspiration from grassland using radiometric surface temperatures. *Agron. J.* **96**, 384–390.