

Effects of Model Resolution and Statistical Postprocessing on Shelter Temperature and Wind Forecasts

M. D. MÜLLER

Institute of Meteorology, Climatology and Remote Sensing, University of Basel, Basel, Switzerland

(Manuscript received 9 August 2010, in final form 31 January 2011)

ABSTRACT

Shelter temperature and wind forecasts from numerical weather prediction models are subject to large systematic errors. Kalman filtering and model output statistics (MOS) are commonly used postprocessing methods, but how effective are they in comparison with steadily increasing resolution of the forecast model? Observations from over 1100 stations in central Europe are used to compare the different postprocessing methods and the influence of model resolution in complex and simple terrain, respectively. A 1-yr period with hourly, or at least 3-hourly, data is used to achieve statistically meaningful results. Furthermore, the importance of real-time observations as MOS predictors and the effects of daily training of the MOS equations are studied.

1. Introduction

The quality of numerical weather prediction has steadily increased over the last decades. Nevertheless, temperature and wind forecasts in particular are still subject to large systematic errors. These errors are not solely due to imperfect initial conditions and model deficiencies, but also are due to errors of representativeness. The latter errors are caused by the fact that temperature and wind are computed for the area of a grid cell and not the particular location where a meteorological station is located. To overcome this problem, postprocessing methods like mean bias removal, Kalman filtering, or model output statistics (MOS) are used. The purpose of this paper is to compare different postprocessing schemes on a large statistical basis and to put them into the perspective of steadily increasing resolution of numerical weather prediction models. The errors due to representativeness are expected to decrease with increasing resolution so that postprocessing might become less important.

To the knowledge of the author, there are very few studies comparing different postprocessing methods with a significant number of stations. Recently, a detailed study was carried out by Cheng and Steenburgh (2007), who focused on 145 stations using forecasts of the Eta

model. In this work, the next-generation forecast model is used and it is set into a perspective of steadily increasing model resolution. Among the different postprocessing methods, Kalman filtering in particular is becoming more popular, as it does not require the long time series needed for the development of MOS equations. Thus, an increasing number of publications are dedicated to presenting slightly different approaches to Kalman filtering surface temperature. Some of these studies (Galanis and Anadranistakis 2002; Libonati et al. 2008; Anadranistakis et al. 2004) achieve really good results. However, only a small number of stations or selected time periods are used, which makes comparisons of the methods rather difficult.

In this study, over 1100 stations from central Europe are considered. The complexity of the terrain ranges from flat plains in the Netherlands to the highest peaks of the Alps. The statistical basis for data analysis is rather large. A time series of one year, with most stations reporting hourly data, is used. The study focuses on shelter temperature as it is available from every station, relatively reliably measured, and often subject to a large systematic error in the forecast. Furthermore, we explore how successfully the same techniques can be applied to 10-m wind speed.

2. Methods and data

a. Model forecast data

The Nonhydrostatic Mesoscale Model (NMM) (Janjic et al. 2001; Janjic 2003) was run for 2.5 years, computing

Corresponding author address: M. D. Müller, University of Basel, Institute of Meteorology, Climatology and Remote Sensing, Klingelbergstr. 27, Basel, Switzerland.
E-mail: mathias.mueller@unibas.ch

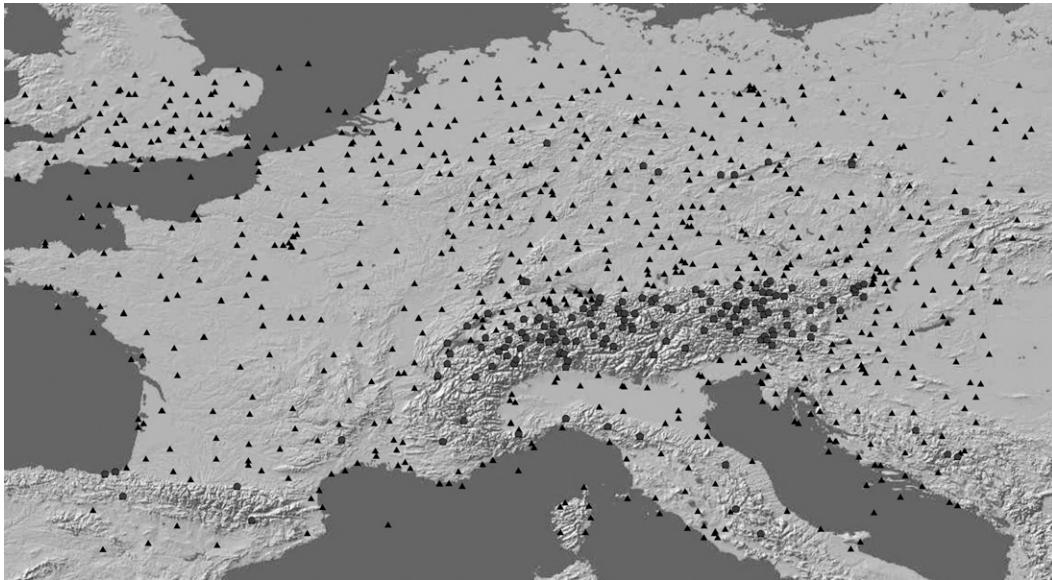


FIG. 1. Map of the 3-km NMM forecast domain. Observational sites in complex terrain are shown as circles; others are shown as triangles.

daily 72- and 144-h forecasts at 3- and 12-km horizontal resolution, respectively. The 12-km model domain stretches from southern Greenland to Iraq, thus covering all of Europe. The much smaller extent of the 3-km domain is nested into the 12-km domain and is shown in Fig. 1. The smaller domain, which is covered by both models, will be used for the analysis in this work. Note that the domain contains the alpine mountain range and thus stations within extremely complex terrain. Initial and boundary conditions of the 12-km domain were derived from the Global Forecast System (GFS) model having a resolution of 0.5° . The raw forecasts of the GFS model are also used in the analysis to underline the benefits of higher-resolution modeling.

b. Surface observations

Observational data were obtained from the U.S. National Centers for Environmental Prediction (NCEP) using the NCEP Automated Data Processing (ADP) Global Surface Observational Weather Data (dataset ds461.0). For the region of interest, this dataset contains 1150 official weather stations, shown in Fig. 1. All stations report temperature on an hourly or at least three-hourly basis. Around 800 stations report wind speed. Depending on the standard deviation of the model topography in a 3×3 filter, the stations are split into a group of complex as well as a group of simple terrain, as the effects of the higher model resolution are more visible in complex terrain. The group of stations in complex terrain is much smaller (180 stations) and thus not really visible in the statistics of all stations.

c. Description of the Kalman filter

To point out the specifics of the Kalman filter used in this study, a brief description is given here. Further details can be found in Kalman (1960), Brockwell and Davis (1987), and Homleid (1995). The Kalman filter is used to iteratively predict the expected systematic model errors \mathbf{x}_t at time t using previous values of observable errors \mathbf{y}_t . We define the error as the difference between observed and modeled temperature. According to the Kalman filter theory, the evolution of \mathbf{x}_t and \mathbf{y}_t is given by

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{w}_t \quad \text{and} \quad (1)$$

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t, \quad (2)$$

with the coefficient matrices \mathbf{F}_t and \mathbf{H}_t and the random vectors \mathbf{w}_t and \mathbf{v}_t that all need to be defined. The prediction equations

$$\tilde{\mathbf{x}}_t = \mathbf{F}_t \mathbf{x}_{t-1} \quad \text{and} \quad (3)$$

$$\tilde{\mathbf{P}}_t = \mathbf{F}_t \mathbf{P}_{t-1} \mathbf{F}_t^T + \mathbf{W}_t, \quad (4)$$

are used to give an optimal estimate of $\tilde{\mathbf{x}}_t$ and $\tilde{\mathbf{P}}_t$ at time t with the help of the previous values of \mathbf{x}_{t-1} and \mathbf{P}_{t-1} . The latter is a covariance matrix needed in the updating Eqs. (5)–(7), which can be computed as soon as the new observations for \mathbf{y}_t become available. Here \mathbf{W}_t is the covariance matrix of \mathbf{w}_t :

$$\mathbf{x}_t = \tilde{\mathbf{x}}_t + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\tilde{\mathbf{x}}_t), \tag{5}$$

$$\mathbf{K}_t = \tilde{\mathbf{P}}_t\mathbf{H}_t^T(\mathbf{H}_t\tilde{\mathbf{P}}_t\mathbf{H}_t^T + \mathbf{V}_t)^{-1}, \text{ and} \tag{6}$$

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{P}}_t. \tag{7}$$

The term \mathbf{K}_t is the so-called Kalman gain and determines how quickly the filter adapts to changing conditions. At the start of the filter, initial values for \mathbf{x}_t and \mathbf{P}_t have to be specified, but they will very easily adapt to the real values in just a few iterations. So far, the values of \mathbf{F}_t and \mathbf{H}_t as well as \mathbf{w}_t and \mathbf{v}_t were assumed to be known. As we cannot really determine the evolution of \mathbf{x} and \mathbf{y} , we have to assume an identity matrix for \mathbf{F}_t and \mathbf{H}_t , which significantly simplifies the original Eqs. (1)–(7). The system covariance matrix \mathbf{W}_t and observation covariance matrix \mathbf{V}_t can be reduced to diagonal form, if we assume that correlations of system and observation errors of different forecast times are negligible. To estimate \mathbf{V}_t , Libonati et al. (2008) used the mean square error of a linear regression of observations and model forecast and a constant tuned value, following Homleid (1995) for \mathbf{W}_t . As we did not favor to use time invariant values, we implemented the procedure of Galanis and Anadranistakis (2002) to compute the scalar values based on data of the last 7 days:

$$w_t = \frac{1}{6} \sum_{i=0}^6 \left\{ (x_{t-i} - x_{t-i-1}) - \left[\frac{\sum_{i=0}^6 (x_{t-i} - x_{t-i-1})}{7} \right] \right\}^2 \tag{8}$$

and

$$v_t = \frac{1}{6} \sum_{i=0}^6 \left\{ (y_{t-i} - x_{t-i}) - \left[\frac{\sum_{i=0}^6 (y_{t-i} - x_{t-i})}{7} \right] \right\}^2. \tag{9}$$

d. Description of the MOS

The MOS technique (Glahn and Lowry 1972) develops a statistical relationship between observed and forecasted weather elements and applies these relationships to raw model output. A multiple linear regression is used to express the predictand \tilde{z} as a linear combination of predictors m_i :

$$\tilde{z} = a + b_1m_1 + b_2m_2 + \dots + b_nm_n, \tag{10}$$

where a and b_i are the regression constant and coefficients, respectively.

MOS equations were automatically derived for each individual station. To assure temporal consistency of the

MOS prediction, the statistical relation was based on all seasons of the year and all forecast hours of the day. Different MOS equations for different seasons lead to an inconsistency when the MOS equations are switched. This problem can be reduced by overlapping the time periods during the training phase but it never completely disappears. Furthermore, the partitioning into seasons significantly reduces the statistical sample size, which increases the length of the time series needed to train the MOS. Without temporal splitting, a sample size of over 6000 data points per year for a 24-h forecast horizon can be achieved, which leads to a solid statistical relation. Surface variables and upper-air model output up to a height of 500 hPa were used as input for the multiple linear regression, yielding about 100 possible predictors. Because of the large number of stations and predictors, the selection of the best predictors was done by an automated process, using the stepwise approach. It has to be noted that regression equations have to be derived for the 3- and 12-km model runs separately, as differences in the model output can be significant for some stations.

In general, MOS uses not just model data as predictors but also the most current observations, usually from the day before. As the NMM computes very reliable forecasts of 2-m temperature, MOS equations without recent observations are also developed and will be shown for comparison in the result section.

e. Statistical analysis

To better understand the effects of different post-processing approaches, statistics resolving the stations but integrating over time, statistics resolving time but integrating over stations, and statistics on an event basis are carried out. For all analyses, the temperature or wind speed of the closest model grid point is taken without performing horizontal interpolation to the exact location of a station. All errors are computed on the hourly or 3-hourly raw data. No spatial or temporal averaging is done prior to the computation of errors. Thus, if $x_{t,i}$ represents a modeled temperature at time t and station i , and $y_{t,i}$ is the corresponding observation, the RMS error $\overline{E}_{\text{rms}}$ and absolute error $\overline{E}_{\text{abs}}$ are computed as

$$\overline{E}_{\text{rms}} = \sqrt{\frac{1}{n} \sum (y_{t,i} - x_{t,i})^2} \text{ and} \tag{11}$$

$$\overline{E}_{\text{abs}} = \frac{1}{n} \sum |y_{t,i} - x_{t,i}|, \tag{12}$$

where n is the number of considered data pairs and the summation is done over t or i , respectively.

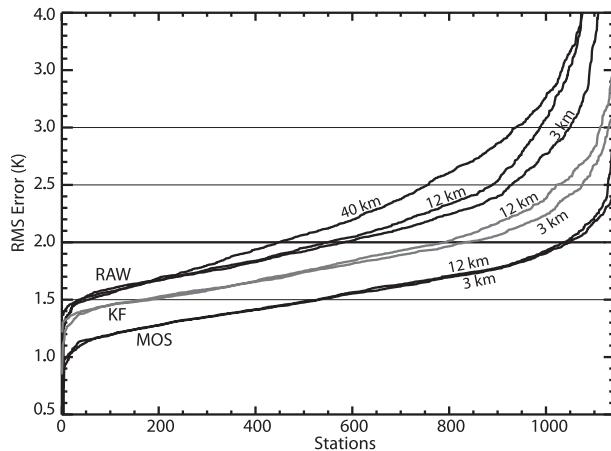


FIG. 2. RMS errors of the 2-m temperature forecasts computed at 40-, 12-, and 3-km resolution. Shown are the raw model forecasts of the closest grid point, as well as the Kalman-filtered (KF) and MOS forecasts for the first 24 h. All errors are plotted in increasing order.

3. Results

a. Station-based yearly statistics

1) TEMPERATURE

To see the total impact of postprocessing methods, the overall temperature error for the whole year is shown in Fig. 2. The RMS, as well as the absolute error (Fig. 3), is plotted for every station, respectively. For better readability, the stations are ordered corresponding to their errors in each diagram and also in each group. Hence, the 500th station of the 3-km MOS can be another station in the 12-km MOS. It can be clearly seen that MOS achieves the smallest errors throughout all stations, followed by the Kalman filter, and finally the raw forecasts. This ranking is seen equally in the absolute as well as in the RMS error. An absolute error smaller than 1.5 K is achieved at over 1000 stations with MOS, at about 750 stations with Kalman filtering, and at only 400–550 stations with the raw forecasts at different resolutions. Interestingly, the benefit of a higher model resolution is almost invisible for the MOS forecast, increased with Kalman filtering, and largest for the raw forecasts. The postprocessing methods slowly approach a minimal RMS error of around 1 K for the best predicted stations. At this lower end, MOS can still reduce the RMS of the raw forecast by about 0.5 K, whereas little can be gained with Kalman filtering. In fact, the MOS curve of the RMS error is always about 0.5 K less than the Kalman-filtered forecast. Note that this does not imply that at any station MOS is 0.5 K better than Kalman filtering, as the stations do not correspond in this diagram.

To visualize the improvement at each station, Fig. 4 shows the 3- and 12-km raw and 3-km MOS forecast at

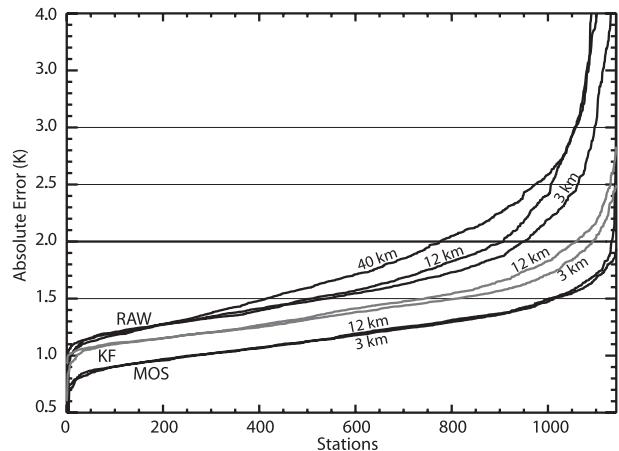


FIG. 3. As in Fig. 2, but for absolute error.

corresponding stations. Hence, the 500th station is the same station in MOS and raw forecasts. As can be expected, the postprocessing methods are most effective for stations with large errors and become less effective at stations already having a small error. Furthermore, the MOS forecast is better than the raw model output at every station. However, the raw forecasts at 3-km resolution do not show such a consistent improvement over the 12-km raw forecasts.

2) WIND SPEED

Similar to temperature, the RMS error of the 10-m wind speed is shown in Fig. 5. The overall result looks remarkably similar to temperature. However, for wind speed the improvement of MOS compared to Kalman filtering is larger than for temperature. Even though not

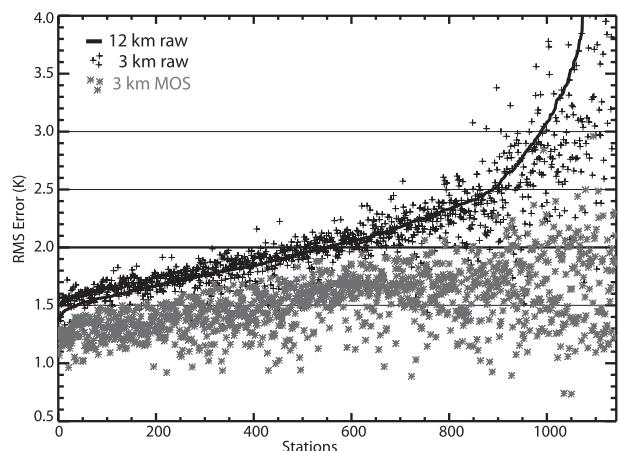


FIG. 4. RMS errors of the 2-m temperature forecasts computed at 12- and 3-km resolution. Shown are the raw model forecasts of the closest grid point, as well as the 3-km MOS forecasts at corresponding stations for the first 24 h.

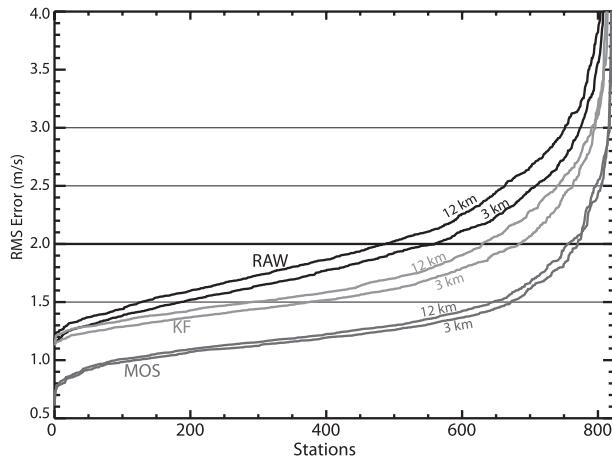


FIG. 5. RMS errors of the 10-m wind speed forecasts computed at 12- and 3-km resolution. Shown are the raw model forecasts of the closest grid point, as well as the KF and MOS forecasts for the first 24 h. All errors are plotted in increasing order.

shown, it is interesting to note that both postprocessing methods managed to remove the bias of the forecast. Again, the benefit of the higher model resolution is decreased by postprocessing but is generally larger for wind speed than for temperature.

b. Temporal variation of forecast errors

1) TEMPERATURE

From the previous analysis, one might conclude that MOS is always better than Kalman filtering, but it is worth looking at the statistics on a daily basis. Hence, all RMS and absolute errors occurring in the 24 h of one day at all stations are summarized in Figs. 6 and 7, respectively. For clarity of presentation, the focus is on the 3-km resolution. The 40-km raw forecast is also shown as a reference. It is interesting to see that the raw forecasts are still worse than the postprocessed forecasts, but the difference between Kalman filter and MOS becomes less apparent. In fact, the Kalman-filtered forecast is often almost as good as MOS, with the exception of a few situations where it is much worse, which explains

the large differences between MOS and Kalman, visible in Figs. 2 and 3. However, we have to keep in mind that these results are aggregated over all stations and 24 h, thus we cannot conclude about the number of correct forecasts. Interestingly, the postprocessed as well as the raw forecast errors do not have a seasonal trend. This is a sign for the high quality of the forecast model but also a consequence of the not-very-continental regime in central Europe. Hence, training the MOS equations for specific seasons did not result in improved forecasts.

Whenever the weather conditions change, the Kalman filter has to adapt to the change, which leads to significant errors in the Kalman-filtered forecast. In these situations, the Kalman filter has significant shortcomings, which was demonstrated, for example, by Cheng and Steenburgh (2007). Note that because of the relatively small study area, shown in Fig. 1, more than half of the stations used in the analysis can be affected by a relatively small frontal system within one day. Thus, the mean error over all stations presented here will show such an event. However, the findings focusing on such special weather situations might overestimate the discrimination between MOS and Kalman filter, which will be shown in the section dealing with event-based statistics.

2) WIND SPEED

Figure 8 shows the time series of wind speed RMS errors for the time range of available wind observations. For clarity, only the 3-km resolution is shown but only little difference to the 12-km resolution was found. In fact the MOS forecasts are almost identical. Noteworthy, the advantage of MOS over Kalman filtering is more apparent than for temperature. Furthermore, the forecast errors do not show a seasonal trend.

c. Event-based statistics

1) TEMPERATURE

For a user of model forecasts it might be more helpful to see how many times a forecast achieves a certain level of quality. In Fig. 9 the percentage of all individual

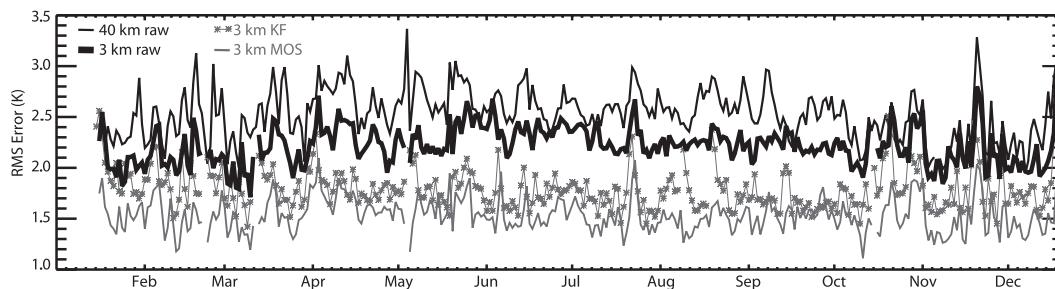


FIG. 6. Time series of 2-m temperature RMS errors based on all stations for the 3-km-resolution raw, KF, and MOS forecasts. As a reference the 40-km raw forecast is shown as well.

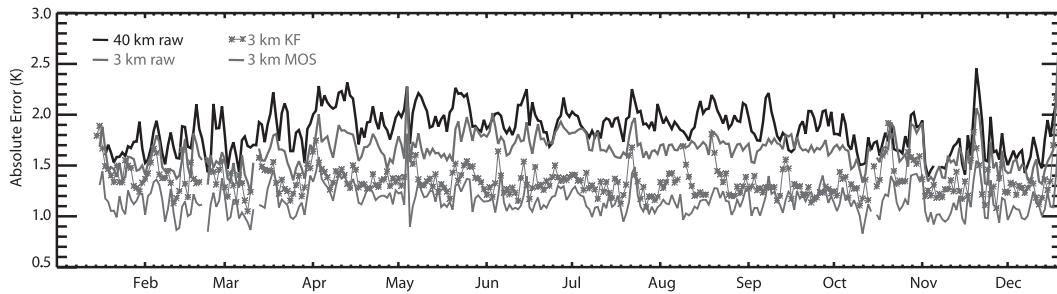


FIG. 7. As in Fig. 6, but with absolute error.

forecasts with an RMS error smaller than 1.5 and 2 K are shown. It can be seen that throughout the first 24 forecast hours, 80% of the MOS forecasts achieve an RMS error smaller than 2 K, which is significantly better than a raw forecast and also better than a Kalman-filtered forecast. Note that the forecast hours in Fig. 9 correspond with an offset of one hour to the local time, as model forecasts start at 0000 UTC. Thus, the raw model forecast shows a diurnal course with a minimum accuracy in the morning and early afternoon. The Kalman filter can close this gap but not significantly improve the raw forecasts during nighttime. Differences between the 3- and 12-km MOS are again negligible.

The cases documented in literature where the Kalman filter achieves best results are related to stable anticyclonic conditions. Hence, rather large areas and thus many stations are affected simultaneously. Furthermore, anticyclonic conditions last several days and the signal should be visible for a few consecutive days. Hence, given the rather small size of the study area, there should be time periods with a significantly larger number of events where the Kalman filter clearly outperforms MOS. To see how relevant such cases are, Fig. 10 shows the number of individual forecasts on each day where the Kalman filter was better than MOS and also the number of cases where MOS was better than the Kalman filter. In order for an event to be considered, the difference between Kalman filtering and MOS has to be larger than 0.5 K. It can be seen that MOS is

consistently better than the Kalman filter, except for four days where the Kalman filter is similar to MOS. However, there are no periods of consecutive days where the Kalman filter is better than MOS. In relation to MOS, the Kalman filter is slightly worse during wintertime. Considering the large areas, and thus the large number of stations that are affected by cyclones or high pressure systems, the daily variations fluctuate little and seem to be only weakly influenced by synoptic conditions. For the majority of locations, it thus seems very difficult to identify in advance if a Kalman-filtered forecast or MOS will be better. A lot of local experience will be needed, as if there existed a simple pattern, MOS would have implicitly used it.

The synoptic conditions heavily influence the quality of the forecast, but cannot clearly be used to identify the ideal postprocessing method. Furthermore, to maximize the number of good forecasts, MOS is the preferred choice.

2) WIND

In analogy to Fig. 9, the percentage of all individual wind forecasts with an RMS error smaller than 1 and 1.5 m s^{-1} is shown in Fig. 11. Interestingly, there is no daily course of the forecast skill, neither in the raw nor in the postprocessed results. In comparison to temperature, there is again a larger difference between MOS and the Kalman filter. The differences between the high and low resolutions are about 5% for the raw and Kalman-filtered results and negligible for the MOS forecast.

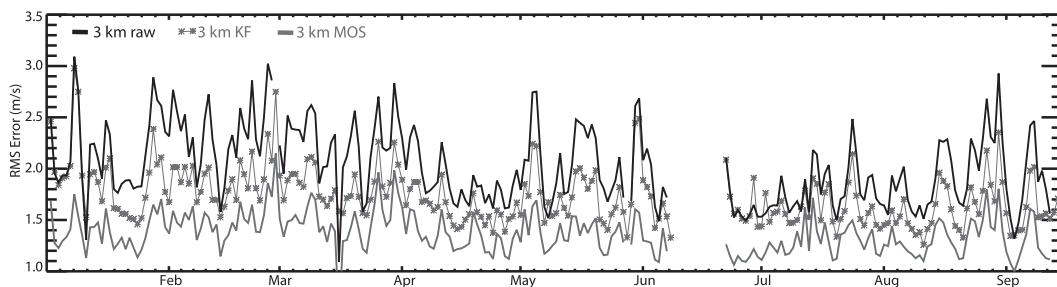


FIG. 8. Time series of 10-m wind speed RMS errors based on all stations for the 3-km-resolution raw, KF, and MOS forecasts.

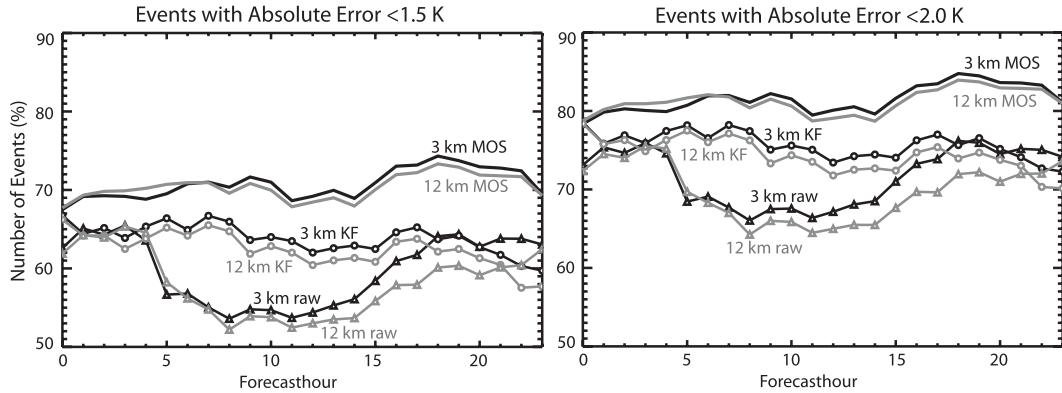


FIG. 9. Percentage of hourly forecast with an error smaller than (left) 1.5 and (right) 2.0 K. Data of all 1050 stations from the year 2009 are used.

d. The role of recent observations

Traditionally MOS includes the most recent observations as predictors. However, the quality of today’s high-resolution NWP models has reached a level where the use of recent observations might become unnecessary, making operational implementations of MOS much easier. This can be seen in Fig. 12, where the 12-km MOS as well as the 3-km MOS with and without recent observations is shown. In fact, the benefit of including recent observations in the MOS forecast is not visible anymore. Furthermore, the differences between the 3- and 12-km MOS can be neglected, especially if put in perspective to the 3-km raw forecast shown in gray, which has a much larger error. It has to be noted that the benefit of recent observations naturally decreases with increasing forecast lead time and thus should have the most effect on the first day analyzed here. The importance of observations in frequently updated nowcasting applications was not studied, but the results presented here are for the first 24 h of the forecast, thus also including the nowcasting period.

Obviously, Kalman filtering requires a steady flow of recent observations, which makes its application for operational use more complicated and less reliable because of observational gaps and errors.

e. Adaptive short-term MOS

The most significant disadvantage of MOS is the long time series of modeled and observed data needed for deriving the MOS equations. Mesoscale models typically are subject to a relatively fast development and implementation cycle, particularly in the physics routines. But changing model physics can have large impact on the model forecast; thus, previously trained MOS equations are not optimal anymore for the changed model. The introduction of updates in an operational forecast model might thus not increase the forecast skill, if unchanged MOS postprocessing is used. The costs to historically rerun an entire year with an updated model are often too high; hence the old MOS equations have to be used for the updated model. If MOS equations are trained on a short time period, the regressions are unstable and likely to produce outliers in conditions slightly different than used for training. However, as the weather generally has some persistence, it might be sufficient to develop MOS equations based on a smaller number of days directly preceding the forecast. This is done in analogy to the Kalman filter. Hence, every day the MOS equations are newly derived based on the last 30, 60, or 90 days and then only applied to the next forecast run. Obviously, this procedure is more complex to implement and

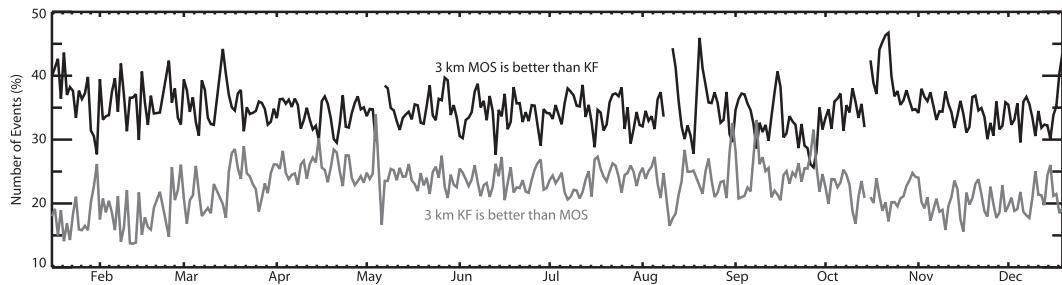


FIG. 10. Percentage of hourly forecast in which MOS has a smaller error than the Kalman filter and in which the Kalman filter has a smaller error than MOS. Data of all 1050 stations from the year 2009 are used.

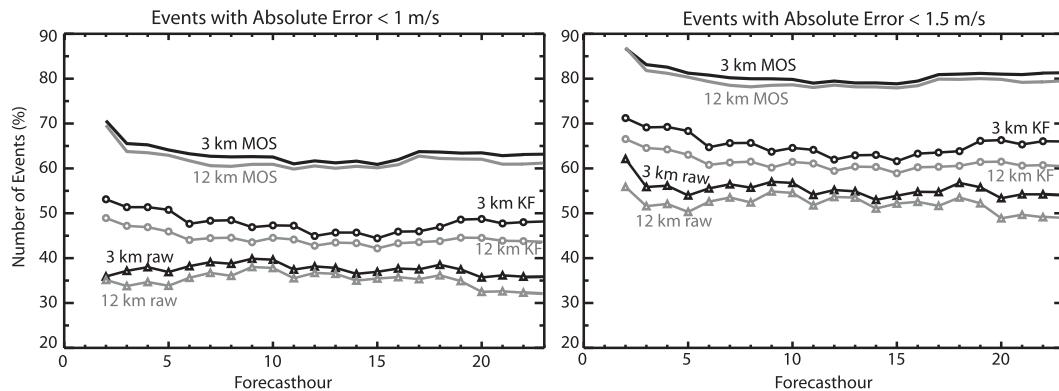


FIG. 11. Percentage of hourly forecast with an error smaller than (left) 1.0 and (right) 1.5 m s^{-1} . Data of all stations with wind observations (800) are used.

much more time consuming in operations than a standard MOS, as lots of data have to be kept in fast accessible archive storage. This procedure was simulated for 180 days and the results are presented in Fig. 13. The time windows used for training are set to 30 and 90 days. For comparison, the standard MOS as well as the raw forecast is shown. It can be clearly seen that the 90-day training period is superior to the shorter 30-day period, but both are less accurate than the standard MOS. A second important result is the presence of outliers, which we define as forecasts being worse than the raw forecast. Clearly, the number of days with outliers is reduced when the training period is increased from 30 to 90 days, but some still remain. It also has to be kept in mind that every point in Fig. 13 represents 1150 stations evaluated on every hour of that day. Hence it is a mean error that hides the real magnitude of the error at some stations. In summary, and unfortunately, this approach does not give a useful solution to shorten the training period needed for a MOS.

f. Comparison in complex topography

So far, little difference in the quality of the 3- and 12-km forecasts was noticeable. This suggests that the increase in resolution from 12 to 3 km has very little

effect on the temperature forecast and little on the wind forecast. This seems in fact to be true in flat terrain. However, if a time series of only the 180 stations in complex terrain are considered (Fig. 14), differences become much more apparent. Figure 14 appears a bit confusing, but we can easily identify two groups. The first group consists of the two curves with highest error and contains the raw forecasts at 12 and 3 km. Among the raw forecasts, the higher resolution is clearly superior. The second group shows the postprocessed forecasts at 3- and 12-km resolution. Almost every day, MOS is better than the Kalman-filtered forecast. Note that the 12-km MOS is not shown, because it is again nearly identical to the 3-km MOS. In the Kalman-filtered forecast, some days can be identified where the 3-km resolution beats the 12-km model, but differences are smaller than for the raw forecast. As postprocessing methods can remove systematic errors, like those caused by height differences, nonlinear processes are playing the key role. In fact, in complex terrain the 12- and 3-km forecasts often differ in cloud cover and precipitation, which results in different temperature forecasts. Müller et al. (2010) studied cold air dynamics and fog/low stratus formation in complex terrain and noticed the need for very high resolution to

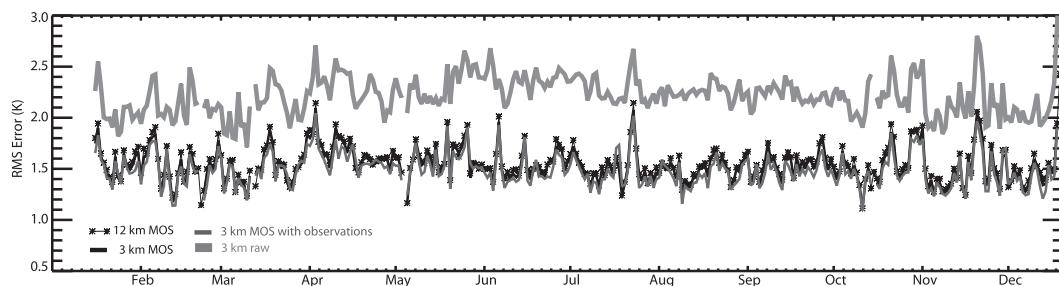


FIG. 12. Time series of 2-m temperature RMS errors based on all stations. The 3-km raw forecast, MOS forecasts at 12- and 3-km resolution using no current observations, and a 3-km MOS forecast using observations from the previous day are shown.

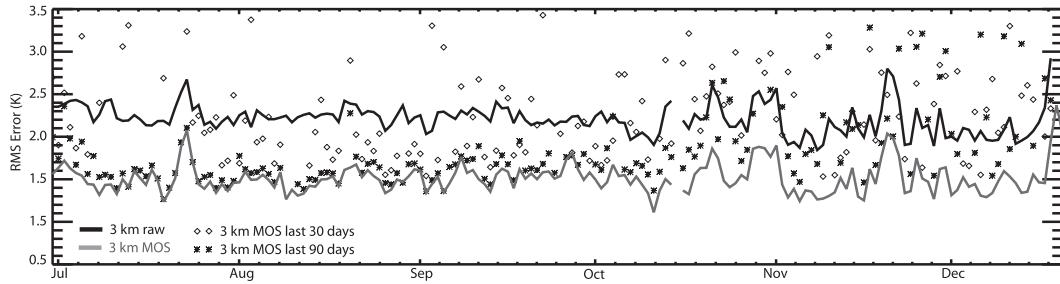


FIG. 13. Time series of 2-m temperature RMS errors based on all stations. The 3-km raw forecast, the standard MOS forecast, and MOS forecasts trained over the most recent 30 and 90 days are shown.

resolve cold air flows, cold air pooling, and cloud formation. As MOS considers many variables, it can implicitly know about systematically wrong cloud cover and adjust temperature accordingly, which is not possible for the Kalman filter that relies on a single variable. Hence, the Kalman filter is much more dependent on the raw temperature forecast than MOS, and the effect of model resolution is more visible.

Generally, the differences caused by the choice of the postprocessing method are larger than those caused by the resolution of the forecast model. The MOS forecast quality seems not to improve by the increase in resolution from 12 to 3 km. This is very interesting as it seems that the true predictive skill of the temperature forecast does not increase beyond a resolution of 12 km. This can also be seen in Fig. 2, where over half the stations have about the same error at 3 and 12 km in the raw forecast. Clearly, the higher resolution improves the raw forecast in complex terrain, but postprocessing can eliminate systematic errors. The resolution increase beyond about 10 km is less visible for Kalman-filtered forecasts and almost unnoticeable for MOS forecasts.

It can be expected that increasing the resolution to about 1 km would have a significant impact on the raw forecast but almost no impact on the MOS forecast. Note that raw forecasts in complex terrain like the Alps have a significant bias, as even the 3-km resolution is

unable to resolve the individual valleys. At 1 km, the valleys are fairly well resolved, which would result in a much better representation of local wind systems and station elevation in the model. Nevertheless, even with perfect representation of height and valley wind systems, the postprocessed forecasts are expected to be better, as it is already the case in flat terrain.

4. Discussion

Overall, MOS outperforms the Kalman filter. This is clearly seen in the RMS error as well as in event-based statistics. To the knowledge of the author, Libonati et al. (2008) achieved the most impressive results with Kalman filtering, where the RMS error of stations in Portugal was around 1.1 K, which is actually very difficult to achieve even for MOS. When we applied their approach, the average RMS error of the 1150 stations was significantly higher around 1.8 K. The reason might be that many more stations located in all kinds of terrain were used. Furthermore, all stations of this study are located further north and thus are affected by more frontal systems than Portugal, which introduces fast weather changes that are more difficult to forecast.

In terms of temperature, the increase in resolution from 12 to 3 km has negligible impact on MOS and a small effect on Kalman-filtered forecast in complex

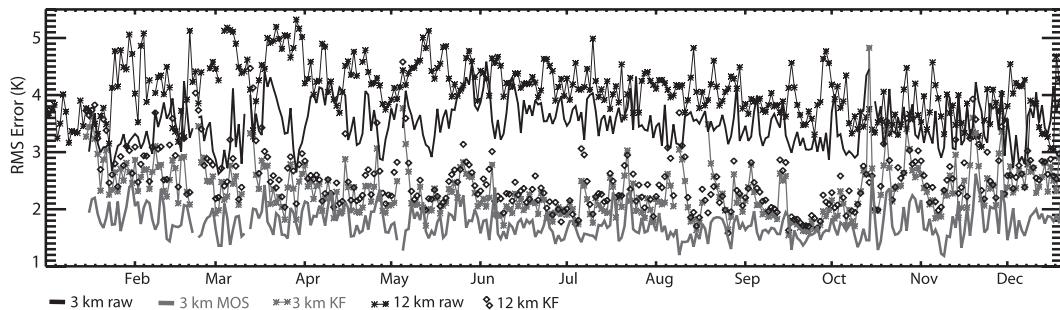


FIG. 14. Time series of 2-m temperature RMS errors based on stations in complex terrain for raw, KF, and MOS forecasts at 3- and 12-km model resolution. Note that the 12-km MOS is not shown because it is almost identical to the 3-km MOS.

terrain. The largest differences are found in the raw model forecast for stations in complex terrain. For wind speed, the increase in resolution is more beneficial than for temperature but is again negligible for MOS forecasts. It is very important to note that postprocessing requires the presence of a meteorological station and can only correct for the conditions at that location. Extrapolating station forecast into the area is quite complicated and error prone, especially in complex terrain. Thus, if forecasts have to be provided for locations other than the observing sites, the higher resolution becomes much more important, as postprocessing methods cannot easily compensate for the lack in resolution. From an operational point of view, considering that the 3-km run requires about 100 times more computing power than the 12-km run, might it be more useful to use the immense computing power of a single high-resolution run for lower-resolution ensemble forecasting or more detailed physics? However, often the more detailed physics are only applicable or useful at a higher resolution. For the development of severe storms and convection in general, the higher resolution carries more physical realism, and can provide higher skill also in other variables not analyzed in this work.

5. Conclusions

An analysis of shelter temperature and wind forecasts at 1150 stations in central Europe, providing hourly or at least 3-hourly data, was carried out for a 1-yr period. Significant differences in the forecast quality of a global model and higher-resolution mesoscale models can be found. For raw model output of temperature and wind speed, the error is reduced with increasing resolution, especially in complex terrain. With Kalman-filtered forecasts, the benefit of a resolution higher than 10 km becomes smaller and almost disappears if MOS is used. It is likely that the current predictive skill of temperature forecasts is around 10 km. A higher resolution will further reduce the model bias, especially in complex terrain, but this can also be achieved with statistical postprocessing if observations are available. Furthermore, a postprocessed temperature forecast at 12-km resolution outperforms a raw forecast at 3 km, regardless of the complexity of the terrain. However, it has to be emphasized that this only holds for locations where observations are available and not everywhere. Statistics integrating over many days demonstrate a clear superiority of MOS. However, if the temporal evolution is analyzed, Kalman filtering achieves comparable results to MOS in many cases for temperature. Almost on any

given day, an MOS temperature forecast is significantly better than Kalman filtering in 35% of all individual forecasts. The events where the Kalman filter beats MOS show a slight seasonal trend from around 18% during winter to approximately 22% during summer. Interestingly, wind forecast are significantly more improved by MOS than by Kalman filtering. With the exception of nowcasting applications, high-resolution NWP models are capable of providing strong MOS predictors that cannot be improved upon by including recent observations. Dynamically training an MOS with always the latest data from the last 30 to 90 days produces many outliers in the forecast. Unfortunately, such an approach cannot be used as a strategy to shorten the time period for training an MOS.

Acknowledgments. We thank the CISL Data Support Section at the National Center for Atmospheric Research for providing the observational data used in this study. The work substantially benefited from the contributions of the anonymous reviewers.

REFERENCES

- Anadranistakis, M., K. Lagouvardos, V. Kotroni, and H. Eleftheriadis, 2004: Correcting temperature and humidity forecasts using Kalman filtering: Potential for agricultural protection in Northern Greece. *Atmos. Res.*, **71**, 115–125.
- Brockwell, P. J., and R. A. Davis, 1987: *Time Series: Theory and Methods*. Springer-Verlag, 519 pp.
- Cheng, W. Y. Y., and W. J. Steenburgh, 2007: Strengths and weaknesses of MOS, running-mean bias removal, and Kalman filter techniques for improving model forecasts over the western United States. *Wea. Forecasting*, **22**, 1304–1318.
- Galanis, G., and M. Anadranistakis, 2002: A one-dimensional Kalman filter for the correction of near surface temperature forecasts. *Meteor. Appl.*, **9**, 437–441.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.
- Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using the Kalman filter. *Wea. Forecasting*, **10**, 689–707.
- Janjic, Z. I., 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.*, **82**, 271–285.
- , J. P. Gerrity, and S. Nickovic, 2001: An alternative approach to nonhydrostatic modeling. *Mon. Wea. Rev.*, **129**, 1164–1178.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35–45.
- Libonati, R., I. Trigo, and C. C. DaCamara, 2008: Correction of 2 m-temperature forecasts using Kalman filtering technique. *Atmos. Res.*, **87**, 183–197.
- Müller, M. D., M. Masbou, and A. Bott, 2010: Three-dimensional fog forecasting in complex terrain. *Quart. J. Roy. Meteor. Soc.*, **136**, 2189–2202.