Operational setup of the soil-perturbed, time-lagged Ensemble Prediction System at the Institute of Meteorology and Water Management – National Research Institute

Grzegorz Duniec, Witold Interewicz, Andrzej Mazur, Andrzej Wyszogrodzki
Institute of Meteorology and Water Management – National Research Institute, Podleśna 61, 01-673 Warsaw, e-mail: grzegorz.duniec@imgw.pl, witold.interewicz@imgw.pl, andrzej.mazur@imgw.pl, andrzej.wyszogrodzki@imgw.pl

Abstract. The usage of Ensemble Prediction System (EPS)-based weather forecasts is nowadays becoming very popular and widespread, because ensemble means better represent weather-related risks than a single (deterministic) forecast. Perturbations of the lower boundary state (i.e., layers of soil and the boundary between soil and the lower atmosphere) applied to the governing system are also believed to play an important role at any resolution. As a part of the research project of the Consortium for Small-scale Modelling (COSMO) at the Institute of Meteorology and Water Management – National Research Institute (IMWM-NRI), a simple and efficient method was proposed to produce a reasonable number of valid ensemble members, taking into consideration predefined soil-related model parameters. Tests, case studies and long-term evaluations confirmed that small perturbations of a selected parameter(s) were sufficient to induce significant changes in the forecast of the state of the atmosphere and to provide qualitative selection of a valid member of the ensemble members. Another important factor that added a significant increment to ensemble spread was the time-lagged approach. All these aspects resulted in the preparation of a well-defined ensemble based on the perturbation of soil-related parameters, and introduced in the COSMO model operational setup at the IMWM-NRI. This system is intended for the use in forecasters’ routine work.

Keywords: ensemble forecast, soil parameterization, operational setup, time-lagged ensemble

Submitted 5 February 2016, revised 31 January 2017, accepted 5 May 2017

1. Introduction

Over the last two decades, the use of Ensemble Prediction Systems (EPS) has become a popular technique for weather forecasting and climate studies in various operational centres (Sivillo et al. 1997; Krishnamurti et al. 2000; Buizza et al. 2005). Initially, an ensemble forecasting was designated to accompany deterministic predictions and to present a wide spectrum of possible scenarios. As model(s) resolution increases, the stochastic nature of many weather phenomena becomes more evident. So far, uncertainties of surface conditions have been rarely considered in operational EPS implementations, however several recent surveys show sensitivity of weather forecasts to perturbations of the land surface state within the range of their uncertainty (Sutton et al. 2006; Quintanar, Mahmood 2012; Tennant, Beare 2014; Bonanno, Loglisci 2017). Sutton et al. demonstrated large sensitivity of the spread of ensemble surface temperature and the timing and location of convective precipitation – in certain warm-season situations – to changes in soil moisture. Quintanar and Mahmood also analysed the impact of soil moisture on the spread of an ensemble forecast of near-surface atmospheric conditions and precipitation during three different summertime synoptic events. In these experiments, soil moisture changes not only the spatial and temporal distribution of precipitation but also affects the overlying near-surface circulation. Tennant and Beare investigated perturbations of sea-surface temperature and soil moisture content, showing positive benefits in decreasing spread deficit of surface temperature and improving probabilistic verification scores. More recently Bonanno and Loglisci introduced lower boundary soil moisture perturbations in a high-resolution convection-permitting EPS. The results show a significant spread increase in the main prognostic variables near the soil as well as further up in the atmosphere. In order to implement techniques for the perturbation of the soil state in the operational mode, the sensitivity to different soil moisture and temperature initializations should be further independently verified for other modeling systems.

In order to obtain ensemble forecasts, we need to prepare members of EPS (i.e., a set of independent deterministic forecasts) initialized with different atmospheric states, or account for changes in model parameterizations. Members can be obtained via, for example, the use of the different parameterizations of certain physical processes, changes in numerical schemes (e.g. Tiedtke or Kuo convection schemes, turbulence schemes, mixing length parameterization: Houtekamer et al.1996; Sattler, Feddersen
and long-range (Chen et al. 2013) forecasts. This simple and more popular for short-, medium- (Lu et al. 2007) order to improve ensemble statistics, we have adopted the significant factor in the preparation of a valid ensemble. In order to improve ensemble statistics, we have adopted the concept of time-lagged IC/BCs, which is becoming more and more popular for short-, medium- (Lu et al. 2007) and long-range (Chen et al. 2013) forecasts. This simple approach was chosen as a first step in the process of building the EPS system at IMWM-NRI, based on the current operational settings which are driven by the single deterministic ICON (Zängl et al. 2015) global model forecasts. In the future, information from several different global models or the global EPS may be utilized to enhance the present approach and to build more advanced probabilistic systems at the IMWM-NRI.

2. Selection of the valid soil state perturbations

As a part of COSMO Towards Ensembles at the Km-scale In Our Countries (COTEKINO) priority project, an effective method was proposed at IMWM-NRI for producing a reasonable number of valid ensemble members, taking into consideration predefined soil-related model parameters. The first phase of this project was based on tentative analysis of the influence on reference results from various model settings (i.e., numerical schemes, physical parameterizations, etc.) combined with rough changes of soil model parameters: e.g., the surface-area index of evaporating fraction or the bottom of the last active hydrological soil layer. This analysis provided an answer to the “importance” of soil parameters and the possibility of neglecting “less significant” ones (see Duniec, Mazur 2014; Mazur, Duniec 2014a). In the second phase, further sensitivity tests were performed. These allowed for the detailed selection of various system configurations and for the assessment of the efficiency of methods of perturbation of the most important soil parameters (Mazur, Duniec 2014b, 2015). Two perturbation methods were tested: (i) by setting a single random parameter value, globally – on the whole domain (easier to perform, by changing namelists instead of modifying the model source code), (ii) an alternative approach – by distributing the random values of the soil parameters for every point in the entire model domain (requires modification of the model source code).

Eleven case studies were chosen with different start dates of the forecast. These covered four seasons and a diversity of synoptic situations. After the completion of numerous model runs, the following conclusions were drawn: (i) surface area index of evaporating fraction ($c_{\text{soil}}$; see Doms et al. 2007) had a noteworthy impact on values of relative humidity, air and dew point temperatures at 2 m AGL., as well as on the wind speed and wind direction at 10 m AGL.; (ii) other soil-related parameters had rather an insignificant impact on forecast meteorological variables when compared with reference ones. Representative ensemble members were prepared by distribution of the random values of $c_{\text{soil}}$ from point to point over the...
entire model domain. This new algorithm was incorporated into the model source code. All of the results are described in terms of spatial distribution of forecast “spread” – i.e., the standard deviation against mean value, as seen in Fig. 1 for a c_soil based ensemble.

3. Ensemble prediction system design

At IMWM-NRI, the COSMO model runs in a deterministic mode using initial (IC) and boundary (BC) conditions from the ICON global model, as shown in Fig. 2. The non-hydrostatic ICON model runs at DWD using an icosahedral-hexagonal grid, with the highest spatial resolution over Europe equal to 13 km and a time span of at least 78 hours, four times per day. The ICON generates a set of IC/BCs for COSMO mesoscale model runs with a basic spatial resolution of 7 km and the domain covering the central part of Europe. This generates 78-hour forecasts. The COSMO model with 7 km horizontal resolution (COSMO-7 km) applies nudging-based data assimilation to correct global model forecasts, ingesting the most recent set of meteorological data acquired from the GTS/WMO network. Forecast results from the COSMO-7 km are further used as IC/BCs for a nested instance of a COSMO model with a higher resolution of 2.8 km and 36-hour forecasts. A set of the deterministic COSMO-2.8 km forecasts define the basis for the operational configuration of an ensemble forecasting system. More details of the configuration of deterministic models that run operationally at IMWM-NRI are presented in the Table 1.

In the recently developed EPS configuration, twenty ensemble members were selected, based on the COSMO-2.8 km convection permitting (CP) forecasts. Every member of an ensemble applies perturbed lower boundary conditions, composed of the random noise of a specified amplitude added to parameters of the soil-model physical parameterization. The proper collection of adequate initial and boundary conditions (IC/BC) is another crucial issue. In the basic EPS configuration, the whole set of IC/BC was nested from a single 78-hour run of a deterministic COSMO model with spatial resolution of 7 km, as shown in Fig. 2. In order to increase the spread of model parameters generated by forecasts, we have further adopted the concept of time-lagged IC/BCs (see e.g. Lu et al. 2007; Chen et al. 2013), in which the set of deterministic CP forecasts is subdivided into groups starting at the consecutive time windows: 00, 06, 12, and 18 UTC.

In the final EPS configuration run operationally at IMWM-NRI, a set of twenty ensemble members is grouped into four bundles, each containing five elements as shown in Fig. 3. The uppermost group – members no. 01 to 05 – is finalized and ready for post-processing 18 hours before the nominal EPS start-up time in the current time window; the second group – members no. 06 to 10 –

Table 1. Details of the configuration of deterministic models

<table>
<thead>
<tr>
<th>Model</th>
<th>Resolution min. [km]</th>
<th>Grid size N×M×L</th>
<th>Forecast length [hours]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICON (DWD)</td>
<td>13</td>
<td>2949120 triangles</td>
<td>78</td>
</tr>
<tr>
<td>COSMOv5.01</td>
<td>7</td>
<td>415×460×40</td>
<td>78</td>
</tr>
<tr>
<td>COSMOv5.01</td>
<td>2.8</td>
<td>380×405×50</td>
<td>36</td>
</tr>
</tbody>
</table>
are gathered 12 hours before the current time window; the third group – members no. 11 to 15 – are prepared 6 hours before the current time window; and finally the lowermost group – members no. 16 to 20 – starts and is finalized in the current time window, completing the whole EPS forecast.

The available forecast results (i.e., various meteorological fields) are ready for post-processing in order to provide selected statistics (ensemble mean, ensemble spread, probability of exceedance etc.) in graphical form, suitable for forecasting applications. This will be presented in the following section.

4. Results of the EPS forecasts

The exemplary results of the 36-hour ensemble forecast are shown in Figs. 4-8, for the initial forecast time 00 UTC, of 5 January 2016. The synoptic situation for this period started after the week-long easterly inflow of continental cold and dry air masses. During 5 January 2013 Poland was affected by an increased high pressure system. The western part of Europe was under the influence of deep low pressure system with two fronts: occlusion (located over the UK, Netherlands and Germany) and an undulated warm front (settled over the UK, Germany and the southern part of Europe up to the Black Sea), which, however, disappear over the following hours. At the beginning, a cold front related to a shallow low pressure system settled over Russia, moved in a western direction and disappeared over the next few hours. The southern part of Europe was under the influence of frontal systems related to shallow low pressure settled over Italy and Greece. The eastern part of Europe was under the influence of a stable high pressure system. A Polar Sea air mass moved into Poland from a north-westerly direction, bringing rain and snowfall (mainly in the southern part of Poland) and producing areas with fog. Temperature fluctuated from –7°C (in the southern part of Poland) to +8°C (in central and western parts of the country). Clouds covered the sky completely in most of the country, wind was moderate – up to 10 m/s, blowing from the west and north-west. During the follow-
Operational setup of the soil-perturbed, time-lagged Ensemble Prediction System...

Fig. 4. Synoptic situation for 5 January, 12 UTC, and 6 January, 12 UTC, 2016; source: http://meteo.imgw.ad/pogoda/analizasyn.php, modified (data access 06.01.2016)

Fig. 5. Ensemble mean (left) and ensemble spread (right) of total cloud cover (octants) at start (top), at 18th hour (middle) and at 36th hour (bottom) of forecast

Fig. 6. Ensemble mean (left) and ensemble spread (right) of wind speed at 10 m above ground level (m/s) at start (top), at 18th hour (middle) and at 36th hour (bottom) of forecast
ing days (between the 6 and 15 of January), the synoptic situation in Poland was polarized with the incursion of arctic air masses.

The new low pressure systems developed over the Scandinavian and Baltic areas, and the Polar front settled over southern Europe – ranging from the Azores to the Black Sea – along which several low-pressure systems travelled eastwards (Fig. 4). The deep low pressure with its centre over the UK moved towards the Baltic Sea with warm frontal precipitation – both snow and freezing rain events. The weather uncertainty over Poland was dependent upon the speed of the eastward front movement and the decay of the low-pressure system over the warmer surface of the Baltic Sea.

The EPS skill provides an estimate of the most probable and alternative realisations of the event and its extremes. The first two charts in Figs. 5 and 6 demonstrate the ensemble mean and ensemble spread of selected meteorological fields (i.e., total cloud cover and wind speed at 10 m above ground level). While ensemble spread has the general tendency for underestimation, it helps to objectively estimate the uncertainty of generated forecasts. The low values of the spread mean that forecasts converge toward a similar solution, leading to stronger anomalies. If the spread is large, the atmosphere is in a state where the small changes in the initial conditions may have a large effect on the forecast quality.

In Figure 7 an example of the probability of the exceedance of a selected threshold in the meteorological field is presented. This option is especially useful for calculating the probability of a predefined amount of precipitation (here: 1.5 mm / 3 hours).

Figure 8 shows an example of so called “spaghetti plots” – the representation of a selected isopleth of a chosen meteorological field (here, the temperature at 2 m above the ground) for every member in an ensemble. Since ensemble members generally diverge in time, these plots show many different model outputs. In the areas where the isopleth pattern is chaotic (i.e., resembling spaghetti) the confidence of the forecast is low. Spaghetti plots provide high confidence of a forecast in regions where members tend to coincide (i.e., contours follow a recognisable pattern through the sequence), and allows for the identifying of the clustering of contours (e.g., bi-modal distributions). This particular chart presented in Fig. 8 shows an example of the amount of uncertainty in the forecast of air temperature. If there is a good agreement and the contours are close to one another (like in the presented case), then the confidence in the forecast can be high. The presented analysis (as the spread and other probabilistic statistics) derived from ensemble forecasts leads to an increase of confidence in the forecast and may provide critical information that increases the awareness of a high impact event in advance.

5. EPS comparison with observations

The EPS results were compared with observations collected from 61 polish SYNOP stations. In Figure 9 presents an analysis of the model generated temperature forecast averaged for the whole day of 5-01-2016, while Fig. 10 presents a similar verification for dew point temperature. In both cases, the EPS mean reproduces the main spatial features of the large-scale temperature distribution. The observed temperature is in a similar range (−20.7°C, −2,5°C) to the computed EPS mean (−20.8°C, −3,8°C). Similarly, the observed dew point temperature is in a similar range (−20,7°C, −2,0°C) to the EPS computed mean (−20,6°C, −1,6°C). In both cases, the model shows a bias when compared to observations with an absolute
Operational setup of the soil-perturbed, time-lagged Ensemble Prediction System...

Error of up to –6.1°C for temperature and –5.7°C for dew point temperature. For the temperature, larger error values (above 3°C) are well correlated with the air masses that came after the warm front crossing Poland during 5 January 2016. Dew point temperature generally shows a small error – below 2°C – except in the north-eastern part of Poland, where we can observe a persistent lack of cloud coverage (Fig. 5) and in south-western areas, which may be correlated to a larger exceedance of precipitation intensity (Fig. 6) that started in the afternoon hours of 05.01.2016 and developed further during the night and the morning hours of the following day.

A larger EPS spread implies more uncertainty in the forecast mean and hence in the ensemble members. The EPS spread is typically underestimated near the surface (Tennant, Beare 2014) where the relations between prognostic parameters are controlled with higher fidelity parameterizations, not accounting enough for observation errors. Thus, in general EPS leads to overconfident forecasts of near-surface weather elements, especially in winter conditions where the natural variability of soil parameters is lower.

In our approach, the perturbation of the evaporating fraction in the soil has a significant impact on the forecast due to the increased stimulation of energy exchange between the surface and atmosphere. The maximum spread for the temperature is 1.1°C, with increased value covering the entire northern and western parts of Poland. For the dew point temperature, we observe a higher spread of the amplitude up to 2.0°C, localized only in the north-western corner of the country.

More advanced analysis of the EPS scores and the correlation between EPS error and spread will be further provided in the context of specific applications. The examples of the probability of the occurrence of dry spells or of rainfall occurrence and distribution are both important in agriculture in assessing the influence of EPS forecasts on irrigation, crop growth or drought occurrence. For synoptic forecasters, the risk of the likelihood of a phenomena and its potential impact will have to be properly estimated to provide the objective basis for a decision to issue a warning.

6. Discussion and summary

Our extensive tests conducted during the COTEKINO Priority Project proved that small perturbations of selected soil parameters were sufficient to induce significant changes in the forecast of the state of the atmosphere and to a provide qualitative selection of a valid member of an ensemble (Duniec, Mazur; 2014; Mazur, Duniec 2014a). Changes of \( c_{\text{soil}} \) had a significant impact on values of air temperature, dew point temperature and relative humidity at 2 m AGL., wind speed/direction at 10m AGL., and
surface specific humidity (ibidem, see also Mazur, Duniec 2015). The use of the concept of time-lagged initial and boundary conditions allowed us to obtain a valid ensemble and use it efficiently in an operational mode. Further work is intended to focus on “tuning” ensemble performance and to provide quantitative quality scores. For this purpose, a random number generator combined with perturbations of initial soil surface temperature and the dependence of the amplitude of perturbation on soil type will be implemented in the COSMO model. While the set of equally weighted time-lagged forecasts improves short-range forecasts, further progress may also be sought by adopting a regression approach to compute a set of weights for different time-lagged ensemble members.