Land surface initialization strategy for a global reforecast dataset

M. Boisserie,* B. Decharme, L. Descamps and P. Arbogast
CNRM/GAME (Méteo-France/CNRS), Toulouse, France

*Correspondence to: M. Boisserie, 42 Avenue Gaspard Coriolis, 31057 Toulouse, France. E-mail: marie.boisserie@meteo.fr

A 32-year global ensemble reforecast dataset has recently been developed at Météo-France that is approximatively consistent with the operational global ensemble forecast system PEARP. Unlike ECMWF or NCEP, Météo-France does not possess a reanalysis of its own operational forecast system. Therefore, the initial atmospheric state and boundary conditions of the reforecasts are from the ECMWF ERA-Interim reanalysis. This article presents a study of the sensitivity of the reforecasts to the method of land-surface initialization. To this end, two sets of short-range hindcasts using different land-surface initialization approaches are compared. The first set is initialized from interpolated ERA-Interim land-surface fields based on a transfer function. The second set is initialized from offline simulations of the Météo-France land-surface model (SURFEX) driven by the 3-hourly near-surface atmospheric fields of the ERA-Interim reanalysis. Each set is run from 1800 UTC initial conditions and up to +108 h.

Because better results overall are found using offline SURFEX simulations, this latter approach was chosen to perform an ensemble reforecast dataset. Then, this ensemble reforecast database will be used to build a climatology of the operational ensemble prediction system of Météo-France, which will, in turn, help to better estimate systematic forecast errors and, more importantly, improve the forecasting of rare extreme weather events.

Key Words: reforecast; ensemble forecasting; land surface initialization

Received 12 June 2014; Revised 1 September 2015; Accepted 29 September 2015; Published online in Wiley Online Library 2 February 2016

1. Introduction

An ensemble reforecast dataset is a long set of hindcasts initialized from dates in the past using the same operational ensemble forecasting system. The goal of producing such a dataset is to build a model climatology of the operational forecast system, which then can be used in numerous applications. Several articles have demonstrated the need for such a dataset in weather forecasting (Hamill et al., 2004; Whitaker et al., 2006; Hamill and Whitaker, 2006; Hamill and Juras, 2006). Systematic errors have become challenging to correct by increasing the model resolution, improving the physical parametrizations, or improving the production of the initial conditions using a better data assimilation system. The use of an ensemble reforecast dataset will allow us to reduce those errors by learning from the past using post-processing methods. Several statistical post-processing methods have been proposed in the literature: rank histogram recalibration (Hamill and Colucci, 1997), logistic regression (Hamill et al., 2004), non-homogeneous Gaussian regression (Gneiting et al., 2005), Bayesian model averaging (Raftery et al., 2005), and best member dressing (Wang and Bishop, 2005). More importantly, it will allow us to better detect extreme weather events through the computation of indices, such as the Extreme Forecast Index (EFI; Lalaurette, 2003) and the Shift Of Tails (SOT; Zsoter, 2006). Those two indices, which both measure the abnormality of a weather forecast, have recently been developed at the European Centre for Medium-Range Weather Forecasts (ECMWF; Zsoter, 2006). Improving early detection of extreme weather events is of crucial importance for national meteorological services. Providing reliable forecasts of intense weather events three or four days ahead is still challenging today. In most cases, such events have low predictability. Although they have considerably improved in the last decades, numerical weather prediction (NWP) models can still have difficulties to predict, beyond 48 h, the intensity and the trajectory of the strongest storms and high precipitation events. Therefore, the computation of the EFI and SOT indices using the ensemble reforecast dataset will greatly help stakeholders to take preventive protective actions to reduce losses of lives and property caused by those devastating events.

The only two operational centres to have produced a global ensemble reforecast dataset are the National Centers for Environmental Prediction (NCEP; Hamill et al., 2013) and the ECMWF (Hagedorn et al., 2012). Both of those centres have the capability to have a reanalysis of their own NWP model. Typically, operational centres archive their real-time analyses and...
forecasts for only a few years, i.e. not long enough to generate a reforecast. Further, the assimilation and modelling system may have changed over that period. Hence, for a centre to initialize a longer reforecast, reanalyses from another centre must be utilised. MeteoSwiss is one of the first centres to have done so, producing a 30-year reforecast dataset of the limited-area model ensemble prediction system (COSMO-LEPS) initialized from ECMWF reanalyses (ERA-40; Fundel et al., 2010). However, for land-surface fields, the initialization from a reanalysis developed using a different NWP model is not straightforward due to many differences of land-surface parametrizations (e.g. the number of soil layers) and the physiographic databases that represent soil and vegetation characteristics between two land-surface schemes. Moreover, many studies have shown that near-surface atmospheric forecasts are very sensitive to the initial land-surface state, in particular soil moisture characteristics because of its long memory and its role in governing the surface heat fluxes (Shukla and Mintz, 1982; Rind 1982; Yeh et al., 1984; Sud and Fenessy, 2000; Hong and Kalnay, 2000, etc). An appropriate land-surface initialization is then crucial for near-surface forecasts. Therefore, in this study, two different land-surface initialization approaches are evaluated for the development of an ensemble reforecast at Météo-France. The first approach consists of interpolating the interim ECMWF reanalysis (ERA-Interim; Dee et al., 2011) land-surface fields based on a transfer function. The second approach consists of using offline simulations of the Météo-France land-surface model SURFEX (Surface Externalisée; Masson, 2013) driven by the 6-hourly near-surface atmospheric fields from ERA-Interim.

The objective of this article is to determine the most appropriate land-surface initialization approach for producing a long set of ensemble reforecasts; this ensemble reforecast dataset will document the climatology of the Météo-France operational NWP model. In section 2, the experimental design is presented. The impacts of two land-surface initialization approaches on retrospective weather forecasts are compared in section 3.

2. Experimental design

The land-surface initialization from an analysis that is produced by another model is not straightforward because of the many differences of land-surface parametrizations and physiographic databases between the two land-surface schemes, which must be taken into account. For instance, the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) scheme of ERA-Interim uses four soil layers with fixed thicknesses; from top to bottom, they are 7, 21, 72 and 189 cm (each layer has its own water content). In contrast, the land-surface scheme of ARPEGE (Action de Recherche Petite Echelle Grande Echelle; the operational global NWP model of Météo-France) uses only two layers, the top layer with a fixed size of 1 cm and the second layer overlaps the first one and has a variable size. The water saturation fraction depends on the soil type (e.g. coarse, medium, medium-fine, fine, very fine, organic, reserved), and the soil types are also very different between the two land-surface schemes.

Therefore, using the raw land-surface datasets from ERA-Interim (i.e. without interpolation) as initial conditions would likely lead to strong biases in the near-surface atmospheric forecasts rising from an incorrect land-surface initialization. This

Figure 1. Soil moisture content at 1800 UTC from (a) ERA-Interim and (b) its difference from SURFEX simulations over the three-month study period (1 June to 31 August 2009). (c, d) and (e, f) are as (a, b) but for surface and deep soil temperatures, respectively.
is due to the role that soil moisture plays in regulating land-surface heat fluxes.

2.1. Initial conditions from ERA-Interim

The first approach consists of interpolating ERA-Interim land-surface fields so as to preserve as much as possible the surface fluxes (sensible heat, latent heat and momentum fluxes). The interpolation technique of the ERA-Interim land-surface scheme TESSEL is based on the Soil Wetness Index (SWI) which is a good indicator for soil water availability. Soil water availability then regulates the partition of latent and heat fluxes, which, in turn, influence energy and water exchanges or interactions between the atmosphere and the land-surface. This index is calculated for each soil layer, \( i \), as follows:

\[
SWI_i = \frac{\omega_{\text{liq},i} - \omega_{\text{wilt}}}{\omega_{\text{cap}} - \omega_{\text{wilt}}},
\]

where \( \omega_{\text{liq},i} \) is the soil water content of layer \( i \) (m\(^3\) m\(^{-3}\)), \( \omega_{\text{wilt}} \) the soil-water content at the wilting point (defined as the minimal soil moisture the plant requires not to wilt) and \( \omega_{\text{cap}} \) the soil water content at field capacity (the maximum value that the land model can sustain without precipitation for more than a few hours). Both two variables \( \omega_{\text{cap}} \) and \( \omega_{\text{wilt}} \) depend on the soil type. In order to preserve the land-surface fluxes, we assume that the SWI calculated for the TESSEL land-surface model is equal to that of SURFEX (i.e. ARPEGE land-surface model in offline mode; more details in section 2.2):

\[
SWI_{\text{TESSEL},i} = SWI_{\text{SURFEX},i}.
\]

Then, from Eqs (1) and (2), the interpolated soil water content from ERA-Interim to SURFEX, \( \omega_{\text{liq},i} \), is determined for each SURFEX layer \( i \) as follows:

\[
\omega_{\text{liq},i} = SWI_{\text{TESSEL},i}(\omega_{\text{cap}} - \omega_{\text{wilt}}) + \omega_{\text{wilt}}.
\]
The soil ice content, $\omega_{\text{ice},i}$, is calculated for each SURFEX layer $i$ as follows:

$$\omega_{\text{ice},i} = f(T) \omega_{\text{liq},i},$$

where the function $f(T)$, when $T_{f2} < T < T_{f1}$, is given by:

$$f(T) = 0.5 \left[ 1 - \sin \left\{ \frac{\pi (T-0.5T_{f1}-0.5T_{f2})}{T_{f1}-T_{f2}} \right\} \right].$$

(3)

If $T < T_{f2}$ then $f(T) = 1$, and if $T > T_{f1}$ then $f(T) = 0$. $T$ is the temperature of each layer and $(T_{f1}, T_{f2})$ are the temperatures within which the freezing process follows the sinusoidal function given by Eq. (3); $T_{f1} = 270.15$ and $T_{f2} = 276.15$ for the first soil layer and $T_{f1} = 267.15$ and $T_{f2} = 280.15$ for the second soil layer. As mentioned earlier, the vertical discretization of SURFEX (two soil layers) and ERA-Interim (four soil layers) is different. In SURFEX, the liquid and ice soil moisture content of its top layer is calculated using the ERA-Interim soil moisture content of its own top layer, while for the deep layer the average of the four soil layers of ERA-Interim is used.

2.2. Initial conditions from offline SURFEX simulations

The second approach consists of producing offline simulations of the land-surface model of Météo-France, SURFEX, driven by the 3-hourly near-surface atmospheric variables from ERA-Interim. SURFEX is a surface modelling method which contains three independent models: the Town Energy Balance (TEB) model for urban areas, the FLAKE model for lakes/inland water, and the Interaction Sol–Biosphère–Atmosphère (ISBA; Noilhan and Planton, 1989) for land-surface processes. ISBA is the operational land-surface model of ARPEGE and is used in coupled mode with the atmospheric part of ARPEGE. It computes the energy and water exchange between the continuum soil–vegetation–snow and the atmosphere above, and the surface hydrology (surface runoff and deep gravitational drainage). The standard version of ISBA uses a limited number of variables to represent the soil state and the soil–plant–atmosphere exchanges. A single surface temperature characterizes a grid cell composed of vegetated surfaces and bare soil, and is related to a single soil temperature. Regarding the water budget in SURFEX, the
Figure 4. Location of SYNOP stations. The grey rectangles represent the six study areas. The number of stations per day used to compute the scores at daytime (night-time) are about 100 (90) over North America, 400 (370) over Europe, 250 (220) over Asia, 160 (140) over Africa+Arabia, 45 (44) over Australia+New-Zealand and over 120 (70) South America.

Figure 5. Root mean square error (solid lines), the bias (dashed lines) and standard deviation (dotted lines) for the 2 m temperature (°C). The results from the set of the operational ARPEGE hindcasts are represented in black, FCT in red, and FCT in blue. These statistics are computed over six different areas (North America, Europe, Asia, Africa+Arabia, Australia+New Zealand and South America) defined in Figure 4.
soil can be represented by either two or three layers. In this study, the number of soil layers is set to two in order to be in accordance with the version of the land-surface scheme used in ARPEGE.

SURFEX can be used offline, forced by atmospheric analyses, or online, coupled with an atmospheric model. In this study, SURFEX is used offline driven by the 3-hourly near-surface atmospheric variables from ERA-Interim from 1 January 1980 to 31 December 2012. The near-surface atmospheric variables needed to drive SURFEX are: precipitation, incoming solar radiation, air temperature, air specific humidity, wind speed, and the incoming long-wave radiation. Because precipitation is the most important forcing variable, the ERA-Interim 3-hourly precipitation was corrected so as to match the global observed monthly climatology produced by the Global Precipitation Climatology Centre (GPCC; Rudolf et al., 2005). More details about this correction technique can be found in Decharme and Douville (2006).

2.3. Two sets of hindcasts

Two sets of short-range hindcasts are produced in this study. One set is initialized from the interpolated ERA-Interim land-surface dataset (hereafter $\text{FCT}_{\text{ERA}}^{\text{I}}$) based on the transfer function described in section 2.1. The other set is initialized from the offline SURFEX simulations (hereafter $\text{FCT}_{\text{SFX}}$) described in section 2.2. In both sets, the initial atmospheric and sea-surface temperature (SST) states are provided by ERA-Interim. ERA-Interim uses a succession of different SST data products (Dee et al., 2011, give more details).

The hindcasts of each set are run using the ARPEGE model. It is a spectral model that uses a variable horizontal resolution with an enhanced spatial resolution over France (Courtier et al., 1991). The SST data used in ARPEGE are the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) from the UK Met Office (Donlon et al., 2011). Hindcasts are run from 1800 UTC initial conditions and up to +108 h at a T538L65 spatial resolution (spectral triangular truncation of T538 with a stretching coefficient of 2.4 and 65 vertical levels).

The land-surface initialization approach that is most similar to the operational ARPEGE initial conditions will be chosen for use in subsequent longer-term reforecast computations. Therefore, the two sets are evaluated against a reference set, which correspond to a set of ARPEGE hindcasts initialized from the operational ARPEGE analyses. This reference set thus differs from the two other sets of hindcasts in both its atmospheric and land-surface initial conditions. Finally, the evaluation is carried out across the globe and over the boreal summer of 2009 (1 June to 31 August).

3. Results

3.1. Comparison of the land-surface initial states

In this section, the land-surface initial states of the two sets of hindcasts are compared. Recall that $\text{FCT}_{\text{ERA}}^{\text{I}}$ uses the ERA-Interim reanalyses and $\text{FCT}_{\text{SFX}}$ uses SURFEX simulations for land-surface initial conditions. The land-surface prognostic variables analyzed are soil moisture and soil temperature (surface and deep layer).

First, Figure 1(a) displays the ERA-Interim soil moisture content (integrated over the soil column) at 1800 UTC after interpolation, averaged over the three-month study period. The
The second column of Figure 1 displays the difference between the ERA-Interim and the SURFEX simulations. It is found that SURFEX is much drier than the interpolated ERA-Interim over most of the globe. Because of the long memory of soil moisture and its role in controlling surface heat fluxes, a drier initial state of the soil will have an impact on the forecasts of the near-surface atmospheric state. Because the study period is a summer period, not much soil ice is found except over the Poles.

For the deep layer, the temperature difference is smaller, but SURFEX remains cooler than ERA-Interim over most of the globe, except Greenland (Figure 1(f)).

### 3.2. Comparison against ARPEGE hindcasts

Here, we evaluate how close the spatial distribution of each set of hindcasts is to the reference set. The temporal average difference between each set of hindcasts and the reference set is calculated at 24 and 72 h forecast lead times for the following near-surface meteorological parameters: 2 m temperature, 2 m relative humidity, sea-level pressure, and 6 h precipitation amount (Figures 2 and 3).

It is found that, at 24 h forecast lead time, the use of SURFEX land-surface initialization leads to smaller temperature differences with ARPEGE hindcasts (Figure 2(a,b)). For instance, over parts of Asia and Northern America, $FCT_{ERA-I}$ is 15°C cooler than ARPEGE hindcasts, while $FCT_{SURFEX}$ is only around 2°C cooler. For the 2 m relative humidity and sea-level pressure, $FCT_{SURFEX}$ is also closer to the reference set than $FCT_{ERA-I}$ (Figure 2(c–f)).

For instance, over Northern America, $FCT_{ERA-I}$ shows sea-level pressure differences of about 200 hPa; this is reduced to less than 50 hPa in $FCT_{SURFEX}$. Not much impact of the land-surface initialization is found on the 6 h precipitation amount. Finally, similar results are found at 72 h forecast lead time (Figure 3).

Therefore, using SURFEX simulations for the land-surface initialization helps to produce near-surface hindcasts that are closer to the operational ARPEGE hindcasts than using the interpolated ERA-Interim land-surface fields based on a transfer function.

### 3.3. Forecast skill

In this section, the skill of each set of hindcasts is calculated using the land SYNOP station network. Each skill is then compared to that of the reference set. Classical forecast skill verification measurements (scores) computed over different areas across the globe are used: the root mean square error (RMSE), the mean error (bias) and the error standard deviation (SDE). Figure 4 shows the SYNOP station locations and the different areas where the scores are computed in grey rectangles. The same meteorological variables as in section 3.2 are analyzed in this section. We will first compare $FCT_{SURFEX}$ with $FCT_{ERA-I}$. Then, those two sets of hindcast will be compared to the reference set.

First of all, in all the results, one can note a diurnal cycle which may be due to the larger number of available SYNOP observation data in the daytime than in the night-time.

For the 2 m temperature, at all forecast lead times, $FCT_{SURFEX}$ shows better skill than $FCT_{ERA-I}$ over most of the areas, except for South America (Figure 5). For instance, over North America, $FCT_{SURFEX}$ shows sea-level pressure differences of about 200 hPa; this is reduced to less than 50 hPa in $FCT_{SURFEX}$. Not much impact of the land-surface initialization is found on the 6 h precipitation amount. Finally, similar results are found at 72 h forecast lead time (Figure 3).

Therefore, using SURFEX simulations for the land-surface initialization helps to produce near-surface hindcasts that are closer to the operational ARPEGE hindcasts than using the interpolated ERA-Interim land-surface fields based on a transfer function.

Figure 7. As Figure 5, but for the sea surface pressure (hPa).
Table 1. The first row indicates how many times (over the forecast lead times and the four meteorological variables) each set shows a smaller RMSE than the other. The second row indicates how many times each set shows a closer RMSE to that of ARPEGE hindcasts than the other.

|                | FCTERA-1 | FCTSFX |
|----------------|----------|--------|
| Smaller RMSE   | 54       | 252    |
| Closer to ARPEGE RMSE | 57     | 249    |

Table 2. As Table 1, but for the bias.

|                | FCTERA-1 | FCTSFX |
|----------------|----------|--------|
| Smaller bias   | 77       | 229    |
| Closer to ARPEGE bias | 55   | 251    |

between the two sets is found, except over Asia. Finally, it is interesting to note that these results are consistent with the earlier results of Figure 1; drier soils found in SURFEX, used as initial conditions in FCTSFX, induce higher near-surface temperatures forecasts.

For the 2 m relative humidity, FCTSFX shows better skill than FCTERA-1 most of the time (Figure 6). The RMSE of FCTSFX is clearly smaller than that of FCTERA-1 at all forecast ranges and over all the areas, except Australia/New Zealand where not much RMSE difference is found. Over the areas in the Northern Hemisphere (North America, Europe, Asia), the RMSE reduction between FCTERA-1 and FCTSFX is of about 20% due to both a bias and a SDE reduction. Over the areas in the Southern Hemisphere (Africa/Arabia, South America), the RMSE reduction is mostly due to a bias reduction only. The bias reduction is due to smaller values of 2 m relative humidity forecasts for FCTSFX than FCTERA-1, which is again consistent with the results of Figure 1, in particular for soil moisture; a drier soil found in SURFEX simulations than in ERA-Interim in the initial conditions will lead to drier near-surface atmospheric forecasts.

For sea-level pressure and 6 h precipitation amount, differences between the FCTERA-1 and FCTSFX are not as significant as for the two previous variables (Figures 7 and 8). Over most of the study areas, no impact is found. Nevertheless, for sea-level pressure, the bias of FCTSFX is smaller than that of ARPEGE hindcasts over most of the study areas, except over Europe and Africa/Arabia. In Figure 8, the discontinuity found in the results over Australia/New Zealand is due to a lack of SYNOP data at 0000 UTC.

Finally, all Figures 5–8 show another positive impact of using SURFEX simulations for land-surface initial conditions, that is, to bring the skill of FCTSFX closer to that of ARPEGE hindcasts. Tables 1 and 2 summarize the latter results (Figures 5–8) by counting how many times each set of hindcasts shows the better skill and how many times each set is closer to the skill of ARPEGE hindcasts. It is clear that the skill of FCTSFX is better most of the time and the closer to the skill of ARPEGE hindcasts.

4. Conclusion

This article compares two land-surface initialization approaches for the development of an ensemble reforecast dataset at Météo-France. Because no reanalysis of our operational NWP model is available, the ensemble reforecast is initialized from the
ERA-Interim reanalysis produced by ECMWF. Nevertheless, the initialization of the land-surface using the ERA-Interim reanalysis is not straightforward. Two approaches for land-surface initialization are then analyzed:

(i) the first one interpolates ERA-Interim fields based on a transfer function that preserves the surface fluxes as much as possible, and
(ii) the second one uses offline SURFEX (the land-surface model of ARPEGE) simulations driven by the 3-hourly near-surface atmospheric fields from ERA-Interim.

First, in general the results have shown a non-negligible sensitivity of near-surface forecasts to the land-surface initialization approach, which is in accordance with several past studies. Then, over the study period, the hindcasts have most of the time shown better skill using SURFEX simulations than the interpolated ERA-Interim land-surface initial conditions. Moreover, the use of SURFEX simulations allows production of near-surface hindcasts that are much closer to ARPEGE hindcasts. A positive impact using SURFEX simulations has also been found for the boreal winter season of 2009/2010 (not shown), but this is not as significant as for the boreal summer season reported.

Therefore, we have have decided to develop a Météo-France ensemble reforecast dataset using a blended initialization approach in which ERA-Interim provides the initial atmospheric state and boundary conditions and the SURFEX simulations provide the initial land-surface state. The ensemble reforecast dataset is a long set of retrospective weather forecasts to build a climatology of the operational ensemble prediction system PEARP (Prévision d’Ensemble ARPege). Thus, we have produced hindcasts over 32 years (from 1981 to 2012), run once every four days at 1800 UTC up to 4.5 days. For each date, a ten-member ensemble hindcast is run, each member using a different physical package. All physical packages are validated separately by experts on the physics of the ARPEGE model to verify using statistical tests and minimum crps estimation. Mon. Weather Rev. 133: 1098–1118.

Hagedorn R, Buizza R, Hamill T, Leutbecher M, Palmer TN. 2012. Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts. Q. J. R. Meteorol. Soc. 138: 1814–1827.

Hamill TM, Bates GT, Whitaker JS, Murray DR, Fiorino M, Galarneau TJ, Jr, Zhu Y, Lapenta W. 2013. NOAA’s second-generation global medium-range ensemble reforecast data set. Bull. Am. Meteorol. Soc. 94: 1533–1565.

Hamill TM, Whitaker JS. 2006. Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. Mon. Weather Rev. 134: 3209–3229.

Hamill TM, Whitaker JS, Wei X. 2004. Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. Mon. Weather Rev. 132: 1434–1447.

Hamill TM, Bates GT, Whitaker JS, Murray DR, Fiorino M, Galarneau TJ, Jr, Zhu Y, Lapenta W. 2013. NOAA’s second-generation global medium-range ensemble reforecast data set. Bull. Am. Meteorol. Soc. 94: 1533–1565.

Koster RD, Suarez MJ. 2000. Soil moisture memory in climate models. J. Hydrometeorol. 2: 558–570.

Lalourette F. 2003. Early detection of abnormal weather conditions using a probabilistic extreme forecast index. Q. J. R. Meteorol. Soc. 129: 3037–3057.

Masson V. 2013. The SURFEXv7.2 land and ocean surface platform for coupled offline simulation of earth surface variables and fluxes. Geosci. Model. Dev. 6: 929–960.

Noilhan J, Planton S. 1989. A simple parameterisation of land surface processes for meteorological model. Mon. Weather Rev. 117: 536–549.

Raftery A, Gneiting ET, Balabdaoui F, Polakowski M. 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather Rev. 133: 1155–1174.

Rind D. 1982. The influence of ground moisture conditions in North America on summer climate as modeled in the GISS GCM. Mon. Weather Rev. 110: 1487–1494.

Rudolf B, Beck C, Grieser J, Schneider U. 2005. Global Precipitation Analysis Products, Global Precipitation Climatology Centre (GPCC). Internet publication, DWD: Offenbach, Germany.

Shukla J, Mintz Y. 1982. Influence of land-surface evapotranspiration on the Earth’s climate. Science 215: 1498–1500.

Sud YC, Fennelly M. 1984. Influence of evaporation in semi-arid regions on the July circulation: A numerical study. J. Climatol. 4: 383–398.

Wang X, Bishop CH. 2005. Improvement of ensemble reliability with a new dressing kernel. Q. J. R. Meteorol. Soc. 131: 965–986.

Whitaker JS, Wei X, Vitart F. 2006. Improving week-2 forecasts with multimodel reforecast ensembles. Mon. Weather Rev. 134: 2279–2284.

Yeh TC, Wetherald RT, Manabe S. 1984. The effect of soil moisture on the short-term climate and hydrology change – A numerical experiment. Mon. Weather Rev. 112: 474–490.

Zoeter E. 2006. Recent developments in extreme weather forecasting. ECMWF News. 107: 8–17.

We would like to thank Hervé Douville for the scientific discussions and insights he has provided and Aurore Voldoire for her technical support.

References

Courtier P, Freydidier C, Geleyn JF, Rabier F, Rochas M. 1991. ‘The Arpege model at Météo-France’. In Proceedings of Workshop on Numerical Methods in Atmospheric Models, 193–231. ECMWF: Reading, UK.

Decharme B, Douville H. 2006. Uncertainties in the GSWP-2 precipitation forcing and their impacts on regional and global hydrological simulation. Clim. Dyn. 27: 695–713.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Holm EV, Isaksen L, Källberg P, Kohler MK, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137: 553–597, doi: 10.1002/qj.828.

Donlon JC, Martin M, Stark JD, Roberts-Jones J, Fiedler E, Wimmer W. 2011. The operational sea surface temperature and sea ice analysis (OSTIA). Remote Sens. Environ. 116: 155–173.

Fendel F, Waber A, Liniger M, Frei C, Appenzeller C. 2010. Calibrated precipitation forecasts for a limited-area ensemble forecast system using reforecasts. Mon. Weather Rev. 138: 176–189.

Gneiting T, Raftery AE, Westveld AH, Goldman T. 2005. Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. Mon. Weather Rev. 133: 1098–1118.

Hagedorn R, Buizza R, Hamill T, Leutbecher M, Palmer TN. 2012. Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts. Q. J. R. Meteorol. Soc. 138: 1814–1827.

Hagedorn R, Buizza R, Hamill T, Leutbecher M, Palmer TN. 2012. Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts. Q. J. R. Meteorol. Soc. 138: 1814–1827.