

# Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe

Bart van den Hurk · Francisco Doblas-Reyes ·  
Gianpaolo Balsamo · Randal D. Koster ·  
Sonia I. Seneviratne · Helio Camargo Jr

Received: 15 July 2010 / Accepted: 23 November 2010 / Published online: 8 December 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** The Second Global Land Atmosphere Coupling Experiment (GLACE2) is designed to explore the improvement of forecast skill of summertime temperature and precipitation up to 8 weeks ahead by using realistic soil moisture initialization. For the European continent, we show in this study that for temperature the skill does indeed increase up to 6 weeks, but areas with (statistically significant) lower skill also exist at longer lead times. The skill improvement is smaller than shown earlier for the US, partly because of a lower potential predictability of the European climate at seasonal time scales. Selection of extreme soil moisture conditions or a subset of models with similar initial soil moisture conditions does improve the forecast skill, and sporadic positive effects are also demonstrated for precipitation. Using realistic initial soil moisture data increases the interannual variability of temperature compared to the control simulations in the South-Central European area at

longer lead times. This leads to better temperature forecasts in a remote area in Western Europe. However, the covered range of forecast dates (1986–1995) is too short to isolate a clear physical mechanism for this remote correlation.

**Keywords** Soil moisture initialization · Seasonal forecasting · Potential predictability · Europe

## 1 Introduction

The contribution of land surface conditions to the predictability of meteorological features is of interest to a wide community. A major portion of predictability at monthly to seasonal time scales is attributed to anomalies in the sea surface temperature (SST), in particular those related to El Niño events (Kirtman and Pirani 2009). However, Koster et al. (2004) identified a number of key regions where anomalous soil moisture conditions may systematically affect precipitation variability in the boreal summer season, based on a model experiment involving multiple Global Circulation Models (GCMs). In combination with a realistic initialization of soil moisture and a long enough memory in the soil water reservoir, increased predictability may be feasible in these regions (Koster et al. 2010b). Dirmeyer et al. (2009) explored a systematic soil moisture–precipitation interaction using a range of observations and (offline) land models for all seasons, roughly confirming the existence of areas where adequate soil moisture information could lead to improved forecasts at the monthly to seasonal time scale. In general these areas are found in transitional zones between dry and wet climates, where the coupling between soil moisture and evapotranspiration is expected to be strong and large enough to affect climate (Koster et al. 2004). Several observational and modelling-based studies

---

B. van den Hurk (✉) · H. Camargo Jr  
Royal Netherlands Meteorological Institute,  
KNMI, PO Box 201, 3730 AE De Bilt, Netherlands  
e-mail: hurkvd@knmi.nl

F. Doblas-Reyes  
Catalan Institute of Climate Sciences (IC3),  
Doctor Trueta 203, 08005 Barcelona, Spain

F. Doblas-Reyes · G. Balsamo  
European Centre for Medium-range Weather Forecasts  
(ECMWF), Shinfield Park, RG2 9AX Reading, UK

R. D. Koster  
Global Modeling and Assimilation Office, NASA/Goddard  
Space Flight Center, Greenbelt, MD, USA

S. I. Seneviratne  
Institute for Atmospheric and Climate Science,  
ETH Zurich, Zurich, Switzerland

approximately agree on the location of these regions (Seneviratne et al. 2010).

Douville (2010) showed that soil moisture conditions in late spring played an important role in successfully modelling contrasting summers concerning precipitation and temperature in the Eurasian continent using a single GCM. A more systematic evaluation of the contribution of soil moisture to the forecast skill with up to two-month lead-time was presented by Koster et al. (2010a) in the context of the second Global Land–Atmosphere Coupling Experiment (GLACE2). This experiment consists of an extensive series of subseasonal ensemble forecasts with multiple models (see below for details). Concentrating on the North American area, the study showed that using realistic soil moisture initial conditions contributes to temperature forecast skill at subseasonal (2 months) lead-times. For precipitation, prediction skill was gained only when a subset of starting dates was selected based on the size of the initial soil moisture anomaly: more extreme soil conditions are found to have a stronger effect on the atmosphere than moderate or small anomalies. These results are consistent with those of Huang et al. (1996), who used observation-driven soil moisture anomaly estimates and statistical techniques to demonstrate the possible contribution of soil moisture anomalies to temperature prediction at multi-month time scales in the continental US.

This study evaluates GLACE2 results over Europe, another area where adequate observations permit a sound evaluation of skill. The metric analysed (proportion of explained variance of 2-week averaged standardized model outputs) is similar to the one presented by Koster et al. (2010a). The results are compared to the potential predictability, defined here as the ability of a collection of models to reproduce temperature or precipitation anomalies generated by any one model in this collection, which is treated as a pseudo-observation. This measure sets an upper limit on the skill improvement that can be expected from a multi-model experiment, bypassing the effect of systematic model biases with respect to observations. The potential predictability of temperature and precipitation in Europe differs significantly from that in the US, due to different characteristics of the variability and remote influences on the local climate (Rodwell and Doblas-Reyes 2006). We present first a brief outline of the general set-up of GLACE2 and the applied post-processing. This is followed by the main results.

## 2 Set-up of GLACE2, observations and diagnostics

### 2.1 The GLACE2 experiment

The multi-model experiment GLACE2 was designed to isolate the contribution of realistic soil moisture

initialization to forecast skill of temperature and precipitation at lead-times of up to 60 days (Koster et al. 2010a). Each participating GCM produced two sets of 60-day, 10-member ensemble forecasts for 100 starting dates: the 1st and 15th day of the months between April and August of the years 1986–1995. The different ensemble members were generated using a range of different techniques by the different participants, depending on their technical constraints or preferred methods of ensemble generation; see Koster et al. (2010b) for details.

In the first set of forecasts (series 1), initial land surface states were extracted from a continuous offline land surface model simulation forced with observed precipitation, radiation, temperature, humidity and wind speed, as provided by the second Global Soil Wetness Project (GSWP2, Dirmeyer et al. 2006). This approach was followed because available in situ soil moisture information is not spatially comprehensive enough in itself to be useful for model initialization. Although the soil moisture fields generated by the offline models may substantially deviate from (highly localized) direct in situ observations (Guo et al. 2006), they generally do represent the effects of major anomalies in the hydrological conditions (precipitation, evaporation) that are captured by the offline forcing data. In addition, the modelled soil moisture products have the advantage of being consistent with the representation of soil moisture in the GCMs participating in GLACE2. In Series 2, initial land conditions were randomized, either by shuffling the GSWP2 fields (for a given day-of-year) in time, or by generating initial conditions for the day-of-year using a free climate run. In all experiments sea surface temperatures were prescribed during the 60-day forecasts. For this, an SST dataset was provided that was an estimate of the observed state on the start date of the forecast with a gradual relaxation to climatology as time proceeds. This set-up mimics the operational application of seasonal forecasting where future SSTs are derived from (uncertain) ocean model simulations (see Koster et al. (2010a, b) for details).

### 2.2 Data processing

Output from the ten participating models (see Table 1) were interpolated to  $2.5^\circ$  longitude  $\times$   $2^\circ$  latitude gridboxes and averaged to 15-day values, starting at the forecast start date. All ensemble members were averaged with equal weights and only the ensemble means are considered throughout this study. For each model and each day-of-year used as a forecast start date, a mean climatology was computed from the 10 years of integrations, as well as a standard deviation  $\sigma_x$ . Results were normalized by recasting individual outputs  $x(t)$  in terms of the standard normal deviate  $Z$ :

**Table 1** GLACE2 participating models used for this study

Acronym	Model resolution	Remark
CCCMA	$2.8^\circ \times 2.8^\circ$	
COLA	$1.9^\circ \times 1.9^\circ$	Version 3.2
COLA_CAM	$1.4^\circ \times 1.4^\circ$	NCAR CAM 3.5
ECHAM	$1.9^\circ \times 1.9^\circ$	Version 5; initial soil moisture series 1 derived from different land surface model simulations
ECMWF	$1.1^\circ \times 1.1^\circ$ <sup>oa</sup>	Integrated Forecasting System (IFS), ocean–atmosphere coupled
FSU	$1.9^\circ \times 1.9^\circ$	Soil initialization from data assimilation suite
KNMI	$1.1^\circ \times 1.1^\circ$ <sup>oa</sup>	As ECMWF, with prescribed sea surface temperatures
NCAR	$2.8^\circ \times 2.8^\circ$	CAM 3.0
NCEP	$0.9^\circ \times 0.9^\circ$	GFS/Noah
NSIPP	$2.5^\circ \times 2^\circ$	GMAO forecasting system

For references see Koster et al. (2010a)

<sup>a</sup> Models using a Gaussian grid; indicated resolution is approximate

$$Z(t) = \frac{x(t) - \bar{x}}{\sigma_x} \quad (1)$$

We will refer to these normalized anomalies when discussing the temperature and precipitation results below. To avoid the effects of differences in atmospheric initialization or methods used to create the model ensemble members, results for the first 15-day period of each forecast are not analysed; we analyze instead the averages over days 16–30, days 31–45, and days 46–60 of each forecast. Similarly, for each day-of-year used as a forecast start date, the 15-day averages from the observations were expressed as standard normal deviates by calculating the mean and standard deviation of these averages over the 10 years in the sample.

### 2.3 Validation data sets

Observations over Europe were taken from the E-OBS data base (Haylock et al. 2008), in which carefully quality-checked station observations are gridded to  $0.25^\circ$  resolution. The observations were interpolated and time-averaged to the same grid and time axis as the model data. Care was taken to average the observations over the same calendar days as used for averaging the model output, implying slightly different intervals for different lead times. Over North America we used the data sets used by Koster et al. (2010a, b).

### 2.4 Predictability measures: forecast skill and potential predictability

The collection of models, ensemble members and start dates implies that a total of 6,000 forecasts are used to construct results for the June–July–August (JJA) season. As in Koster et al. (2010a, b), the diagnostic of interest here is  $R^2$  (square of correlation coefficient) between

observations and the ensemble mean model results, effectively a measure of the explained fraction of variance. For this metric, the normalized ensemble mean model outputs for each of the ten participating models were plotted against the corresponding observations, resulting in a scatter plot with 600 different points.  $R^2$  values from this scatter plot were separately calculated for series 1 and 2, and the contribution of realistic land initialization to skill is measured as the skill difference:

$$\text{sign}(R(1))R^2(1) - \text{sign}(R(2))R^2(2) \quad (2)$$

where the sign of  $R(1)$  and  $R(2)$  is considered to avoid rewarding large negative correlations over small positive ones. For testing the hypothesis that the skill of series 1 exceeds that in series 2 in a statistically significant way, a 1,000-member bootstrapping procedure was applied in which, for each member of the procedure, the 60 observational values were shuffled. The significance level is indicated by the fraction of redrawn sets of data for which the correlation  $R$  between the observations and series 1 simulations  $R(1)$  exceeds  $R(2)$ .

To estimate the maximum possible value of land-derived skill that could be obtained from the multi-model experiment, we derived a measure of the “potential predictability”— $R^2$  calculated as above, but instead of using the observations as the reference “truth”, we used the ensemble mean results from an individual model. This calculation was repeated using each of the ten models in turn as the reference truth, and the ten resulting score values were averaged after transforming them to a normal distribution using Fisher’s Z-score statistic  $0.5 \ln(1+r)/(1-r)$ . Note that this metric is different from the average potential predictability calculated using the individual ensemble members as truth for every model separately. The procedure was applied to both the series 1 and series 2 simulations.

## 2.5 Data subsets for extreme soil moisture initializations

For various analyses presented below, a subset of forecasts was constructed based on initial soil moisture content. Extreme wet or dry soil moisture values were identified at each grid point from the 60 initial soil moisture conditions there (one for each start date providing data during the JJA period) by subtracting the mean seasonal cycle from the 60 values and then ranking the 60 anomalies. The extreme 20%-values refer to the 12 wettest and 12 driest start dates in the sample, and the 10%-values are the 6 wettest and 6 driest start dates. The fields used for this selection are a representative set of GSWP2-derived initial soil moisture fields, namely the fields constructed for the models ECMWF and KNMI, generated using the HTESSEL land surface model (Balsamo et al. 2009) which carries a 4-layer soil scheme. The total water content in the top three layers (top 1 m of soil) was taken as the grid point value. These soil moisture fields do represent the effects of anomalous hydrological forcings and are statistically very similar to the fields used by a majority of GLACE2 participants (see below).

The anomalies were calculated and ranked at each individual grid point to produce a subset of start dates specific to that grid point, ignoring the possible spatial coherence of the anomalies. In one analysis below, however, this coherence was retained by examining how the subset of start dates generated at one location affects the skill score generated in a predefined remote target domain.

## 3 Results

### 3.1 Potential predictability

The term “potential predictability” is often interpreted as an intrinsic property of a geophysical system, expressing the degree to which chaos would limit forecast skill assuming a perfect model configuration. The predictability inherent in nature is not measurable; the best we can do is quantify the effects of chaos within a given model or set of models, for purposes of understanding better the models’ behaviour. Here, we estimate predictability from the ability of the multi-model simulations to predict the behaviour of a single participating model (Sect. 2.4).

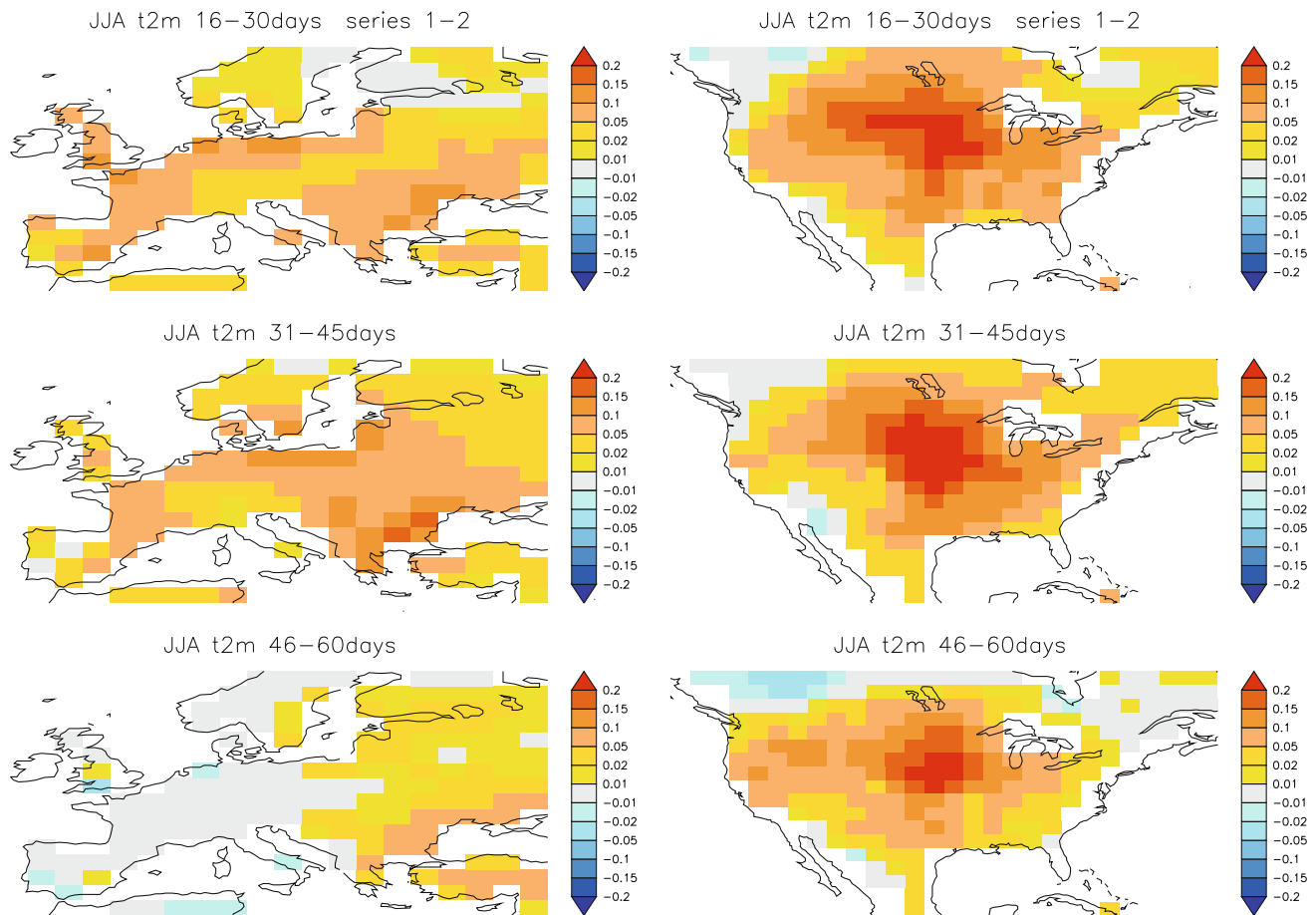
The potential predictability of two-weekly mean near surface temperature in JJA in Europe is generally higher for series 1 than for series 2, particularly for shorter lead times (see Fig. 1). A similar result (but at much lower levels of  $R^2$ ) is obtained for precipitation, although the spatial patterns of the potential predictability differ from the values for 2 m temperature. This implies that using

similar initial soil moisture values (derived from common external data) for this collection of models increases the reproducibility of the temporal variations generated by the individual models, resulting in a smaller inter-model spread in series 1 than in series 2. Again, these estimates of predictability reflect the modelling systems used; systematic model biases may be producing, for example, predictability levels larger than those present in nature, resulting in overconfidence of the predictions (Huang and Van den Dool 1993; Hagedorn et al. 2005).

From Fig. 1, it is also evident that the soil moisture related potential predictability increase is generally much higher in the US than in Europe. This is true for all lead times.

### 3.2 Forecast skill at different lead times

Figure 2 shows that, in analogy to the North America results shown by Koster et al. (2010a), initial soil moisture in Europe affects temperature forecast skill more than it affects precipitation forecast skill, and that forecast improvement (difference in  $R^2$  between series 1 and 2) reduces with lead time. For reference, the results for the longest lead time for North America have been reprocessed using the same set of models and plotted with the same colour scale. Grid points with significant differences between series 1 and 2 ( $p = 98\%$ ) are shaded. For precipitation no meaningful skill improvement could be detected using the GSWP2 soil moisture data, but temperature is positively affected up to 1 month ahead in all areas except the land area around the Eastern Mediterranean Sea. At longer lead times, the contribution of realistic land initialization to temperature forecast skill decreases, and for the 46–60 day period, the realistic initialization appears to lead to a decrease in skill around the Baltic Sea. The potential predictability from the use of realistic initial soil moisture in this area is fairly low (Fig. 1). A change of the initial soil moisture can therefore affect the forecast skill in multiple directions, certainly at longer lead times. Reasons for this may be effects of snow treatment around the initialization in April/May, systematic model drifts (e.g. due to persistent low intensity model precipitation), natural variability, or a wide variety of initial soil moisture fields used by the individual models, in spite of the fact that most models used offline GSWP2 simulations to generate series 1 soil moisture fields. Cross-correlations between pairs of model-specific Western European time series of soil moisture anomalies averaged over the first 15 days of the simulations yielded values as high as 0.98 between KNMI and ECMWF (which used identical land models and identical fields at time zero) to as low as 0.26 between NCAR and NSIPP. While a disagreement between 15-day soil moisture averages does not necessarily imply a



**Fig. 1** Difference in potential predictability (expressed as  $R^2$  averaged over all ten models serving as reference; see Sect. 2.4) between series 1 and series 2 of two-weekly 2 m temperature in JJA for (left column) Europe and (right column) the US domain. Results are shown

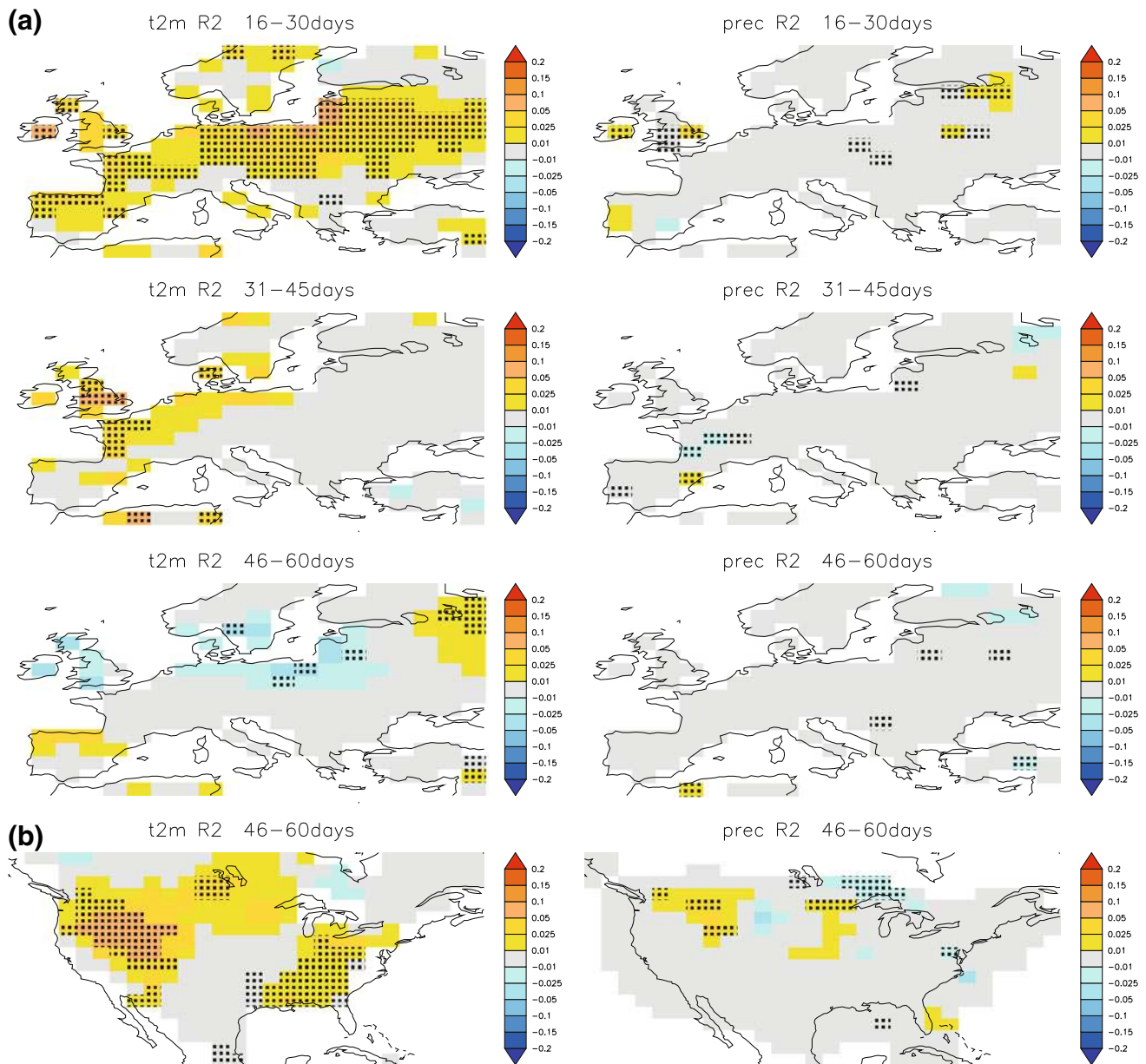
at lead times ranging between 16–30 days (top), 31–45 days (second row) and 46–60 days (bottom row). Positive numbers indicate a gain in potential predictability owing to using realistic soil moisture values

disagreement in initial conditions (given differences in model structure and given variations in rainfall during the first 15 days) it is interesting that the five models with the highest mutual correspondence of day 1–15 soil moistures (ECMWF, KNMI, FSU, NCEP and COLA) do give rise to higher skill scores for temperature, particularly at longer lead times (Fig. 3), with hardly any grid points showing decreased skill.

In general, the positive results for series 1 forecasts are less convincing over Europe than over North America, both for temperature and precipitation. This is consistent with the lower values of potential predictability calculated for Europe (Fig. 1). Figure 4 shows the fraction of the potential predictability of temperature actually gained by applying the realistic soil moisture initializations, using results from all participating models. The mean potential predictability of series 1 and 2 is taken as reference. Results for precipitation are much noisier and are therefore not shown. Also shown is the statistical significance

( $p$  value) of the difference between series 1 and 2. Note that discrepancies between actual and potential predictability may have several causes, including an overestimation of potential predictability in the models (insufficient spread between the models), imperfect soil moisture data or initialization procedures, imperfect observations, and systematic model errors leading to imperfect predictions.

Reasonable fractions of potential predictability (>20%) are attained at short lead times in a major part of the European continent. This fraction drops with lead time, but less so in the Western half of Europe. Over the Iberian peninsula the fraction tends to increase, but this is at least partly an artefact of normalizing a low skill increase by a low potential predictability in that region. Within the limitations of the methodology followed here, Fig. 4 suggests that soil moisture initialization as implemented in the GLACE2 simulations does close the gap between actual and potential predictability at short lead times to some extent, and that more can be gained from other sources of



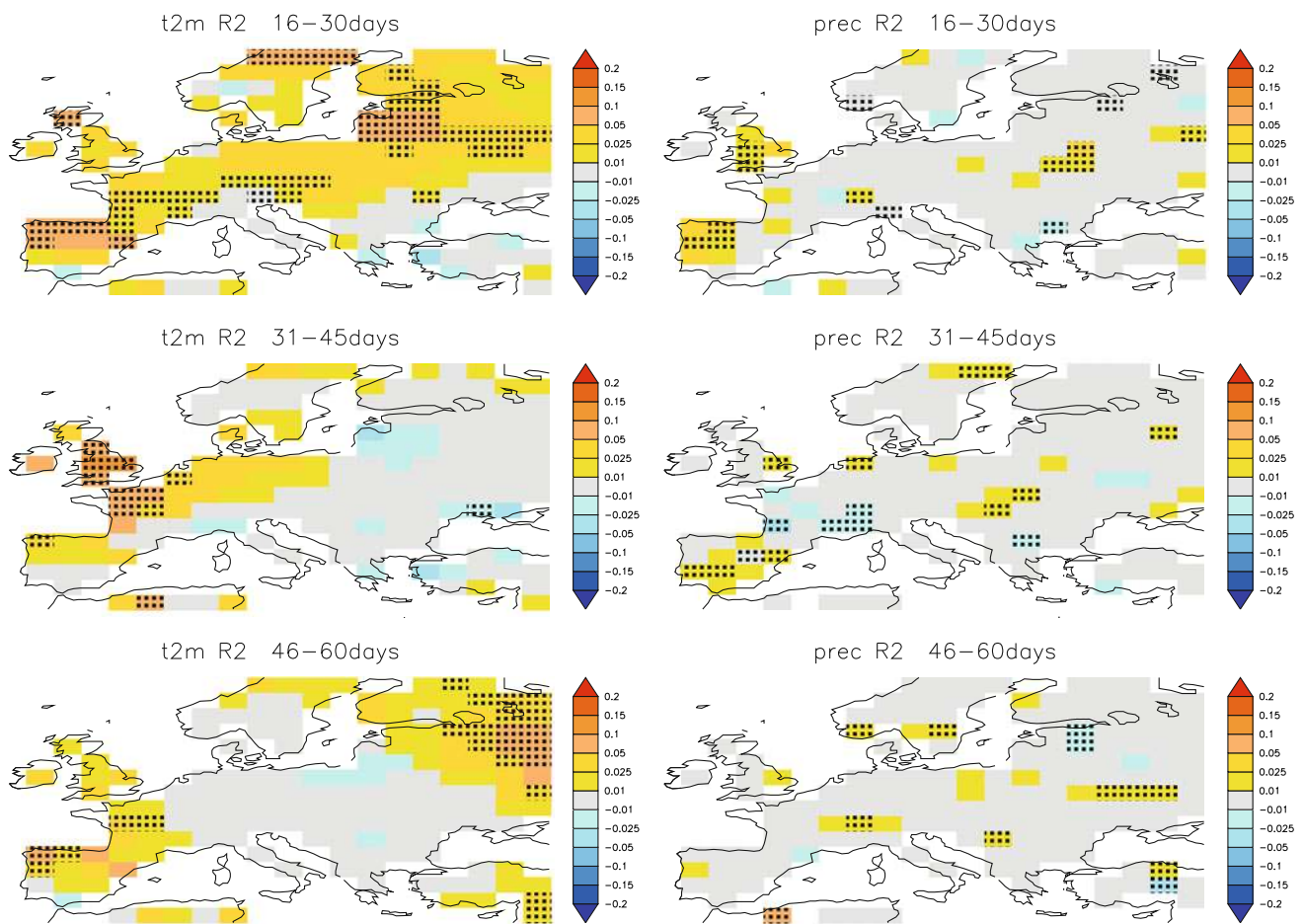
**Fig. 2** Gain in forecast skill by using realistic soil moisture initialization [ $R^2(1) - R^2(2)$ ] for (left) temperature and (right) precipitation for three different lead times: 16–30 days (top), 31–45 days (second row) and 46–60 days (bottom two rows). Results

for the US are similar to the results published earlier by Koster et al. (2010a). Grid points for which the difference between series 1 and 2 are significant at 98% confidence are shaded

skill, such as better model representations, higher resolutions, and improved datasets for the initialization and validation of the model variables.

As before, one can infer from Fig. 4 that the situation in North America is more promising than in Europe. One potential reason for the difference is the shorter autocorrelation time scale of soil moisture in Europe. Figure 5 shows the multi-model mean correlation between the average soil moisture anomaly for days 1–15 and the average anomaly for days 46–60. Only forecasts ending in

the JJA season are considered in the calculation. Correlations are calculated separately for series 1 and series 2 forecasts; the difference in correlation between these two series is fairly small. Both for North America and for Europe, many areas with relatively high utilized fractions of potential predictability in series 1 (Fig. 4) at 46–60 days lead time coincide with areas with high temporal correlation across the 2-month forecast interval: Southern Europe, the US West coast and South-West of the Great Lakes region. This is consistent with findings of Weisheimer



**Fig. 3** As Fig. 2, but using a selection of five models with a high mutual cross-correlation of 1–15 day anomalous soil moisture time series in Western Europe ( $10^{\circ}\text{W}$ – $25^{\circ}\text{E}$ ,  $35^{\circ}\text{N}$ – $55^{\circ}\text{N}$ )

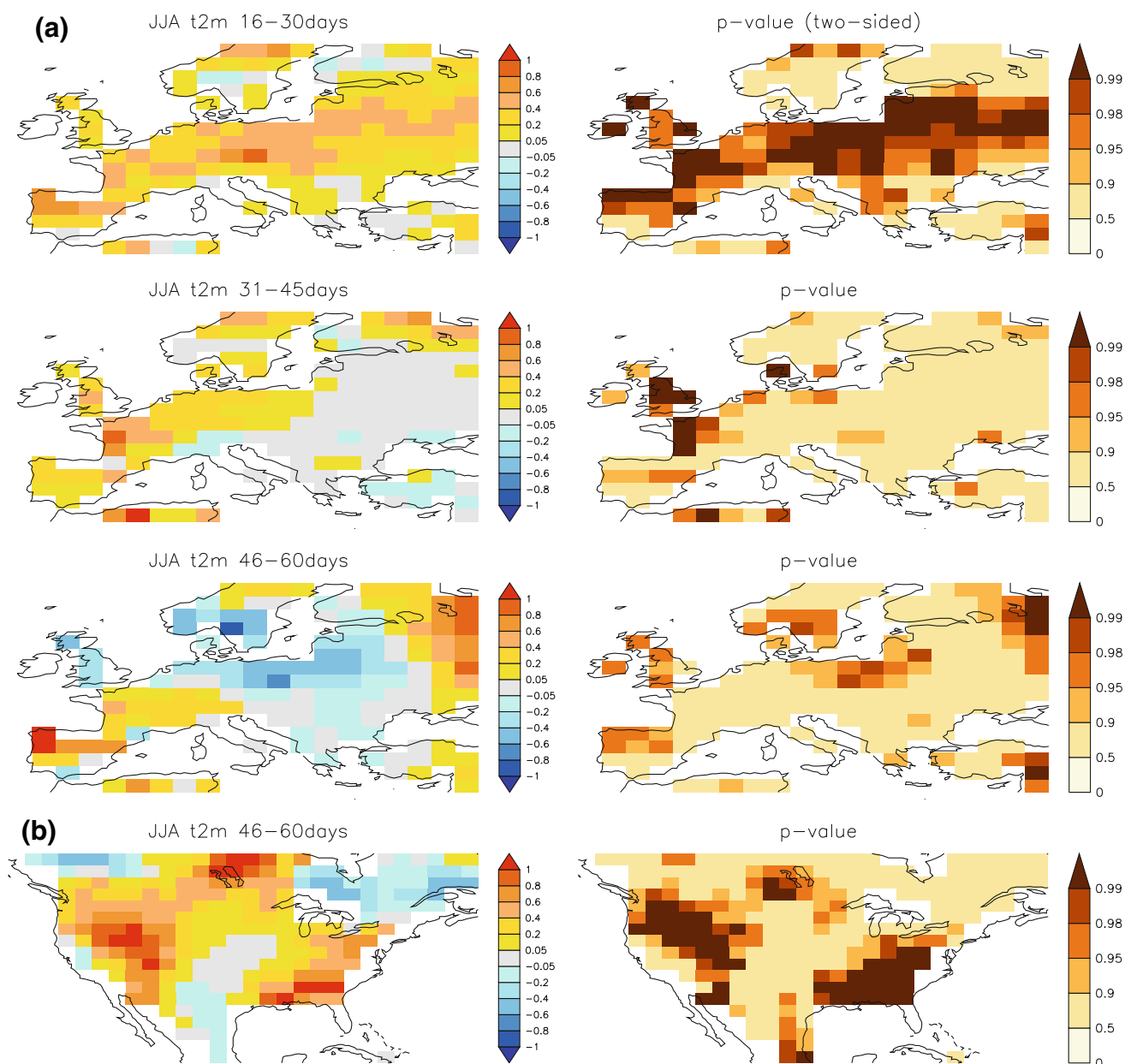
(private comm), who demonstrated that soil moisture persistence is an important factor for explaining the skill of seasonal forecasts for the anomalous 2003 European summer, for which anomalously low soil conditions in spring gave rise to improved predictability of the summer temperature anomaly. However, high soil moisture autocorrelation is not the only factor determining the positive skill of series 1 forecasts: some areas with low soil moisture autocorrelation (e.g. South-East US) also have a relatively high skill.

### 3.3 Forecast skill for extreme initial soil moisture conditions

A slightly more optimistic picture emerges when a selection of dates is used for the temperature forecast skill calculation, based on the size of the initial soil moisture anomaly (Fig. 6; see Sect. 2.5). This analysis does not explicitly account for spatial correlation between soil moisture values at different grid points; different selections of start dates may apply to adjacent grid points.

(An additional analysis below will deal with this issue.) The patterns shown in Fig. 6 have roughly the same spatial structure as those shown in Fig. 2 (but with more noise due to the smaller sample size), but overall, the skill levels have increased. Figure 6 confirms the notion that initial soil moisture is not equally informative across the entire range: extreme wet or dry conditions have a greater ability to affect near surface temperature. In analogy to the North America results of Koster et al. (2010a), the positive impact is most pronounced at short forecast lead times, while at longer lead times areas with positive and negative skill remain. For precipitation, the results show an overall increase of the field significance (grid points with positive skill appearing more frequently than those with negative skill), but the results are very noisy and, hence, are not shown.

A supplemental analysis was performed in which the start dates were subsetting into two bins: those for which the initial soil moisture was lower than the climatological mean, and those for which it was higher. Temperature forecast skill levels were then computed for each subset to



**Fig. 4** Left panels gain in forecast skill  $[R^2(1) - R^2(2)]$  of JJA temperature as fraction of the potential predictability averaged for series 1 and 2 for temperature at lead times as in Fig. 2. Right panels  $p$  value of the difference between series 1 and 2 (two-sided)

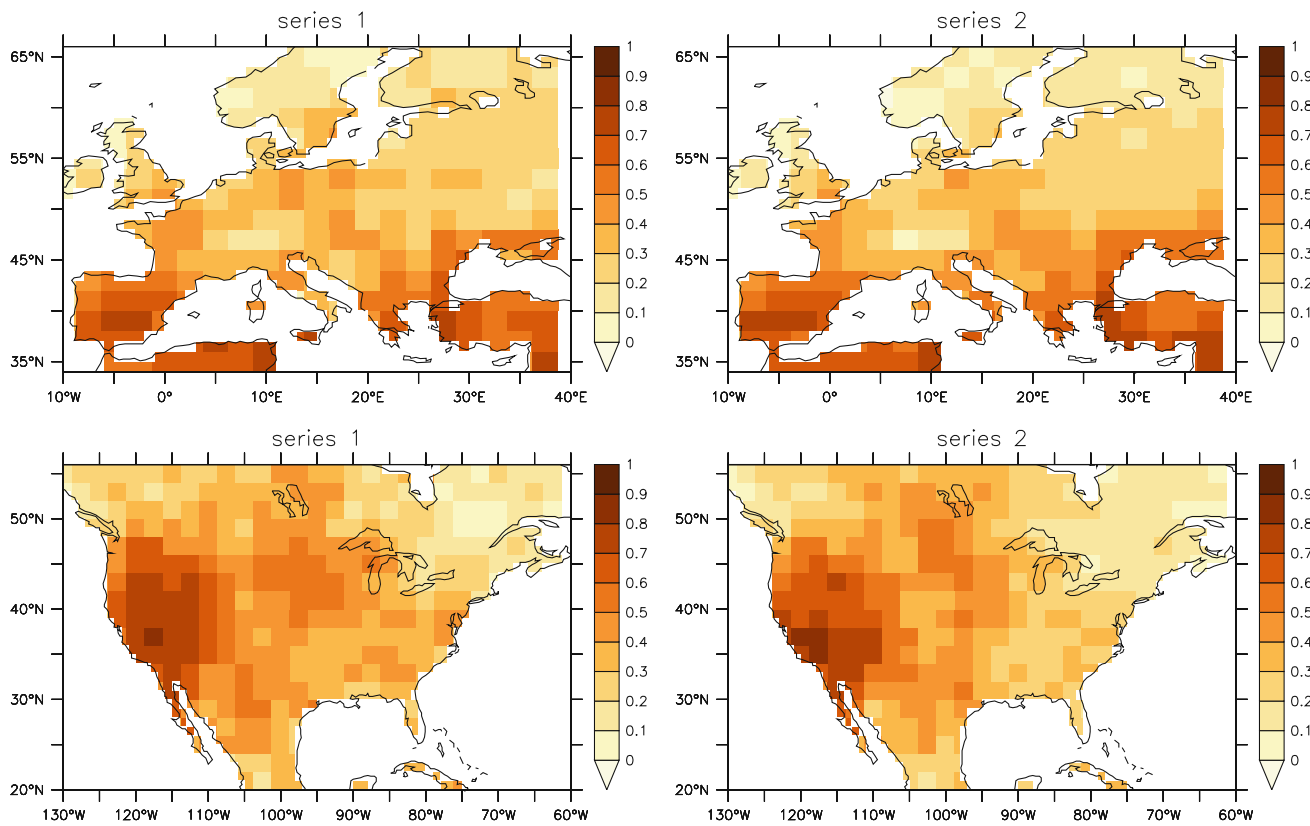
determine if drier conditions might lead to less (or more) skill than wetter conditions, in analogy to the analysis of Koster et al. (2010b). However, for the European area, no clear patterns of asymmetry were evident, and results are not shown.

The seasonal evolution of the effects of soil moisture initialization on forecast skill is shown in Fig. 7. Here the average score difference  $\overline{R^2(1) - R^2(2)}$  in an area roughly covering the Iberian peninsula through Poland ( $10^\circ\text{W}$ – $25^\circ\text{E}$ ,  $35^\circ\text{N}$ – $55^\circ\text{N}$ ) is shown for different lead times and initial soil moisture selections. For temperature the selection of extreme quintile or decile soil moisture content has

a strongly favourable effect on the forecast scores in all months for days 16–30. For the longer lead times, the strongest impact is during the late summer season (particularly August). Note that the apparent negative skill in September for the decile calculation likely reflects the very small sample size available during this particular month. For all forecasts after day 30, all subsettings produce very little temperature forecast skill for May, June, and July.

For precipitation, the noise level at individual grid points is too high to detect a clear signal, especially at longer lead times. Further investigation, however, reveals (for the all data case) that when the individual precipitation





**Fig. 5** Temporal correlation between average soil moisture anomalies in the first forecast interval (days 1–15) and the last interval (days 46–60), calculated from simulations of all ensemble members of all

models ending in the JJA season ( $N = 6,000$ ). Results for series 1 (left) and series 2 (right) are shown separately

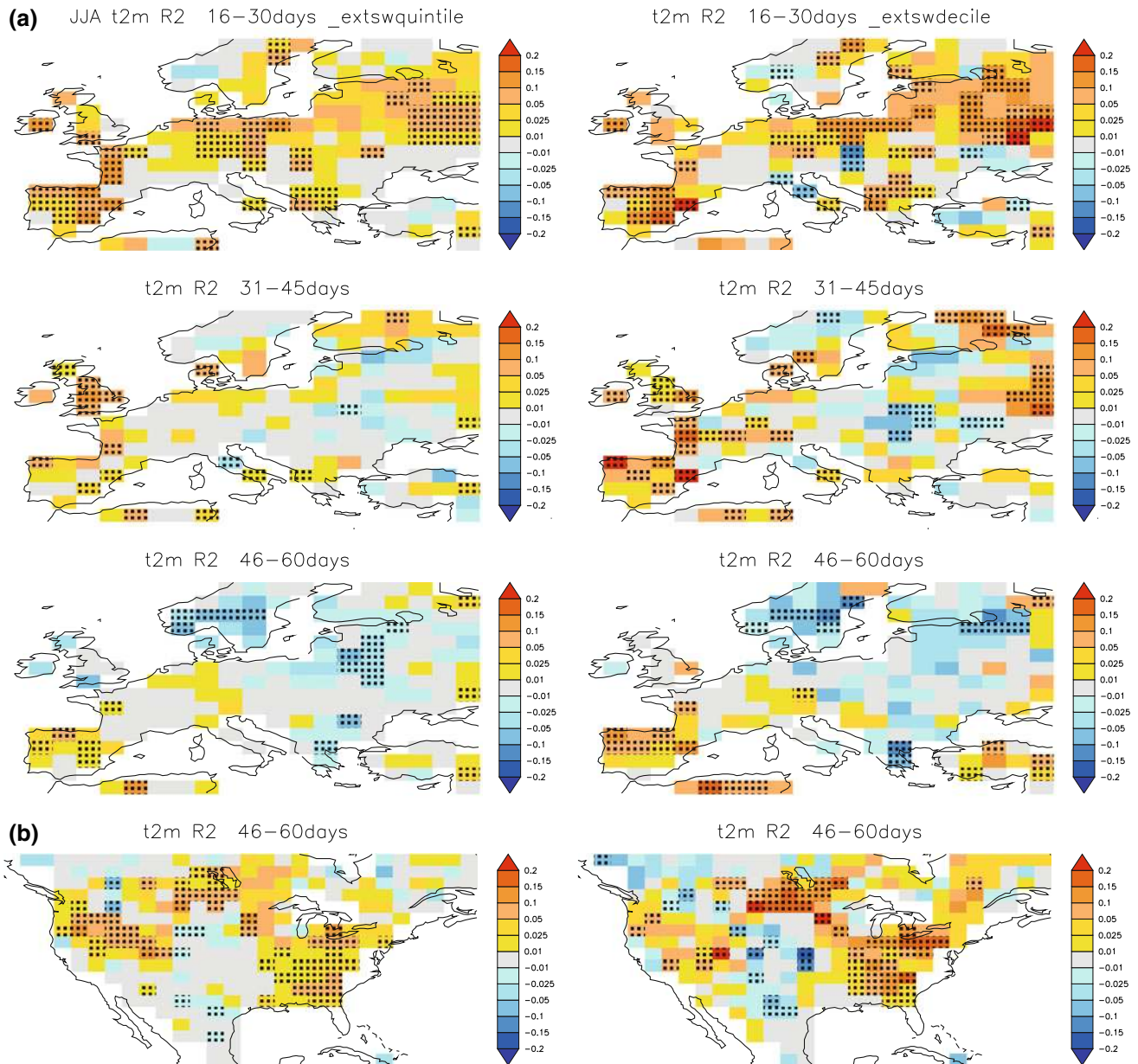
forecasts are spatially averaged to coarser resolution prior to computing the skill levels, realistic land initialization provides a larger positive impact. Figure 8 shows the seasonal evolution of the skill obtained for large spatial averages, i.e. for the individual temperature and precipitation forecasts averaged over the same Western European domain ( $R^2(\bar{x})(1) - R^2(\bar{x})(2)$ , with  $\bar{x}$  indicating the time series of spatially averaged temperature or precipitation). The results for temperature are similar to those in Fig. 7. For precipitation, larger improvements are seen at both short and long lead times. The shorter spatial correlation length scale of precipitation contributes to the difficulty of detecting the effects of initial soil moisture conditions on forecast skill at grid point spatial scales.

### 3.4 Spatial patterns of initial soil moisture and forecast skill improvement

Similar to the notion that soil moisture is not equally informative across the entire range of its distribution, the potential contribution of soil moisture to forecast skill (both local and remote) also varies spatially. This is

illustrated by Fig. 9, which shows the results of a special calculation. In effect, forecast skill for spatial averages in the outlined red box (again, the value of  $R^2(\bar{x})(1) - R^2(\bar{x})(2)$ ) is computed for different subsets of start dates. To generate these subsets, each grid cell in Europe is considered in turn. For a given grid cell, the extreme quintiles are established as above, and the corresponding subset of dates for that one cell are used to compute the skill level for the outlined red box; this skill level is then plotted at the location of the given grid cell. The process is repeated at the next grid cell, with the skill for the outlined box plotted at that cell, and so on. Note that in contrast to the results presented in Figs. 6, 7, 8, the spatial structure of the soil moisture analyses is retained here.

For temperature and—to some extent—precipitation forecasts for days 16–30, a great majority of the grid cells provide subsets of start dates for which land initialization contributes positively to skill in the Western European area. However, for precipitation, no grid cell provides a useful soil moisture subsetting for lead times longer than 4 weeks. For temperature, the 4–6 week and 6–8 week forecast in Western Europe is improved when the extreme

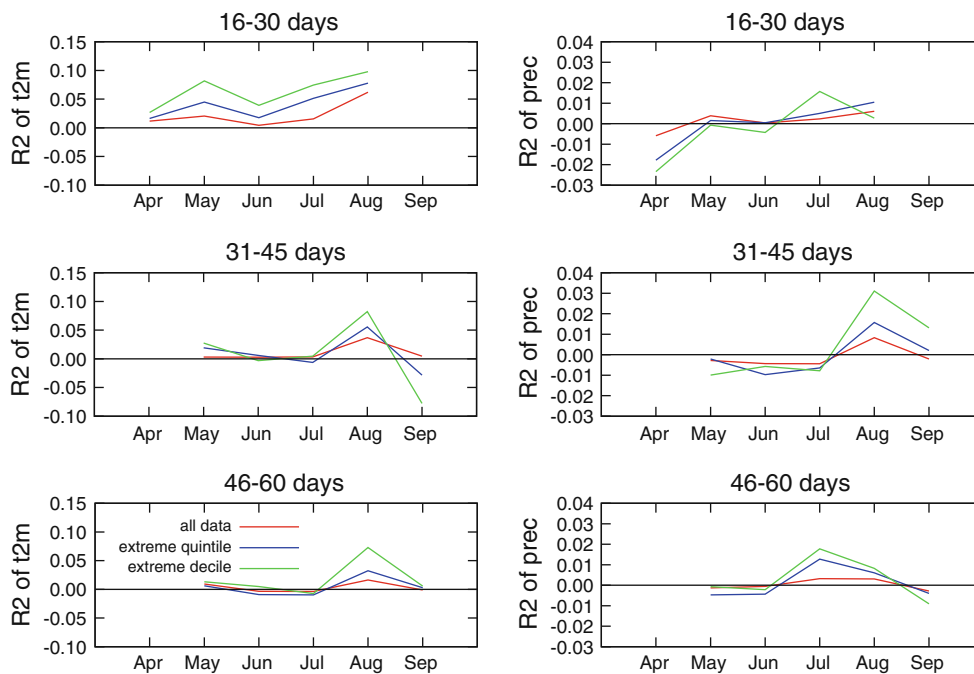


**Fig. 6** Gain in forecast skill [ $R^2(1) - R^2(2)$ ] of JJA temperature for forecasts where initial soil moisture is within the extreme quintile (*left*) or extreme decile (*right*) range at lead times and significance levels as in Fig. 2

soil moisture time slots are determined from Balkan and central-eastern European grid cells. Interestingly, this area is outside the domain in which the skill is improved, suggesting a potential physical or statistical connection between the two areas.

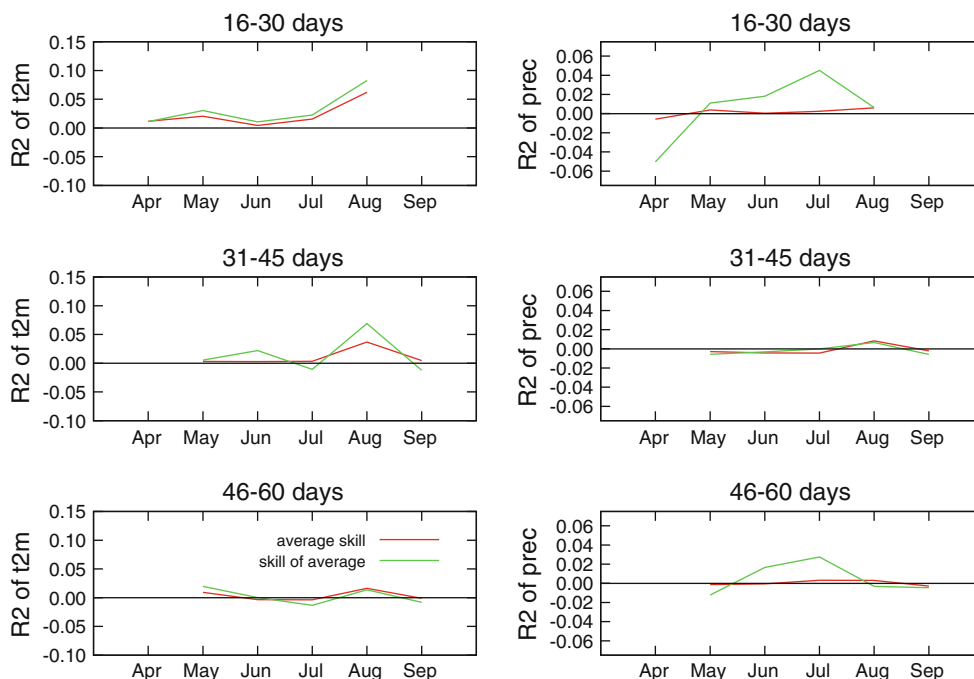
The south-central European area is roughly co-located with the area where the GLACE2 multi-model ensemble shows a large interannual variability in temperature (Fig. 10). The interannual variability of temperature (defined as the standard deviation of ensemble mean JJA forecasts over the ten simulation years, averaged across all models) is higher in series 1 than in series 2, and it shows a

marked pattern with a local minimum in West-central Europe and maxima to the east and west of this area. The difference between the temperature variabilities of series 1 and 2 gradually decreases with lead time, as the models approach their own equilibrium climate values. In the South-central European area, the temperature variability remains relatively high at longer lead times, which might be related to the impact of the subsetting there on west European skill, as shown in Fig. 9. However, the 10-year time range of the experiment does not allow the isolation of a clear physical mechanism behind this potential remote connection.



**Fig. 7** Time series of score differences [ $R^2(1) - R^2(2)$ ] averaged over a large part of Western Europe ( $10^\circ\text{W}-25^\circ\text{E}$ ,  $35^\circ\text{N}-55^\circ\text{N}$ ) for (left) temperature and (right) precipitation at different lead times (top

panel 16–30 days, middle panel 31–45 days, bottom panel 46–60 days) and different initial soil moisture selections (red lines all data, blue lines extreme quintiles, green lines extreme deciles)

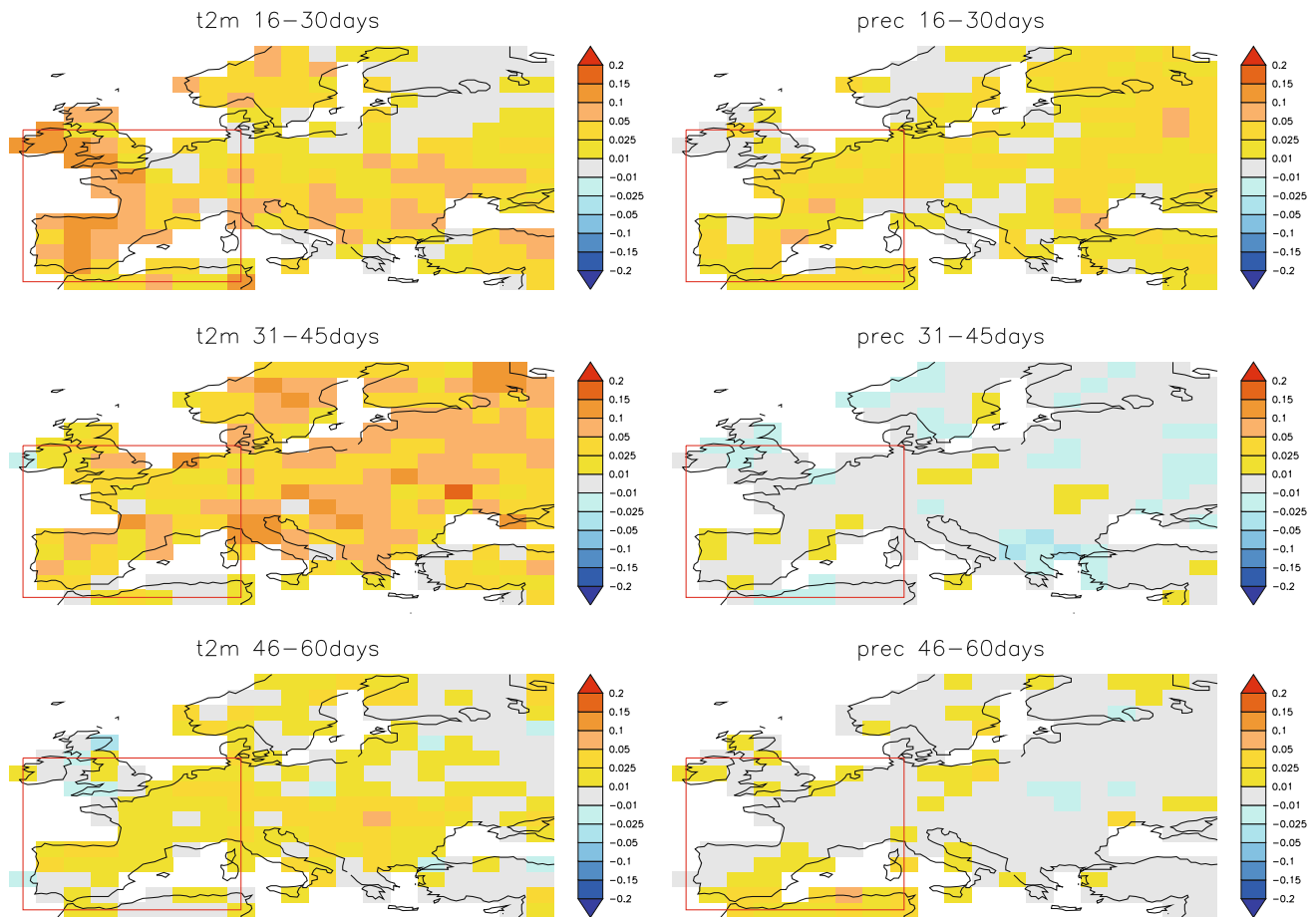


**Fig. 8** As Fig. 7, comparing, the spatially averaged skill (“average skill”, same values as “all data” in Fig. 7), and the skill of the spatially averaged temperature (left) or precipitation (right)

**4 Discussion and conclusions**

Results from the second Global Land Atmosphere Coupling Experiment (GLACE2) for Europe show that realistic soil

moisture initialization in the spring and summer seasons does lead to improved forecast scores for temperature across the entire area at short lead times (16–30 days). At longer lead times the areas with improved scores decrease, and even



**Fig. 9** Locations where selection of extreme initial soil moisture has a strong effect on the  $R^2$ -difference for (*left panels*) temperature and (*right panels*) precipitation averaged over the area indicated by the red box. The difference in JJA  $R^2$  averaged over the indicated area is

plotted at locations where soil moisture values in the extreme quintiles of the distribution were used to make a selection of time slots

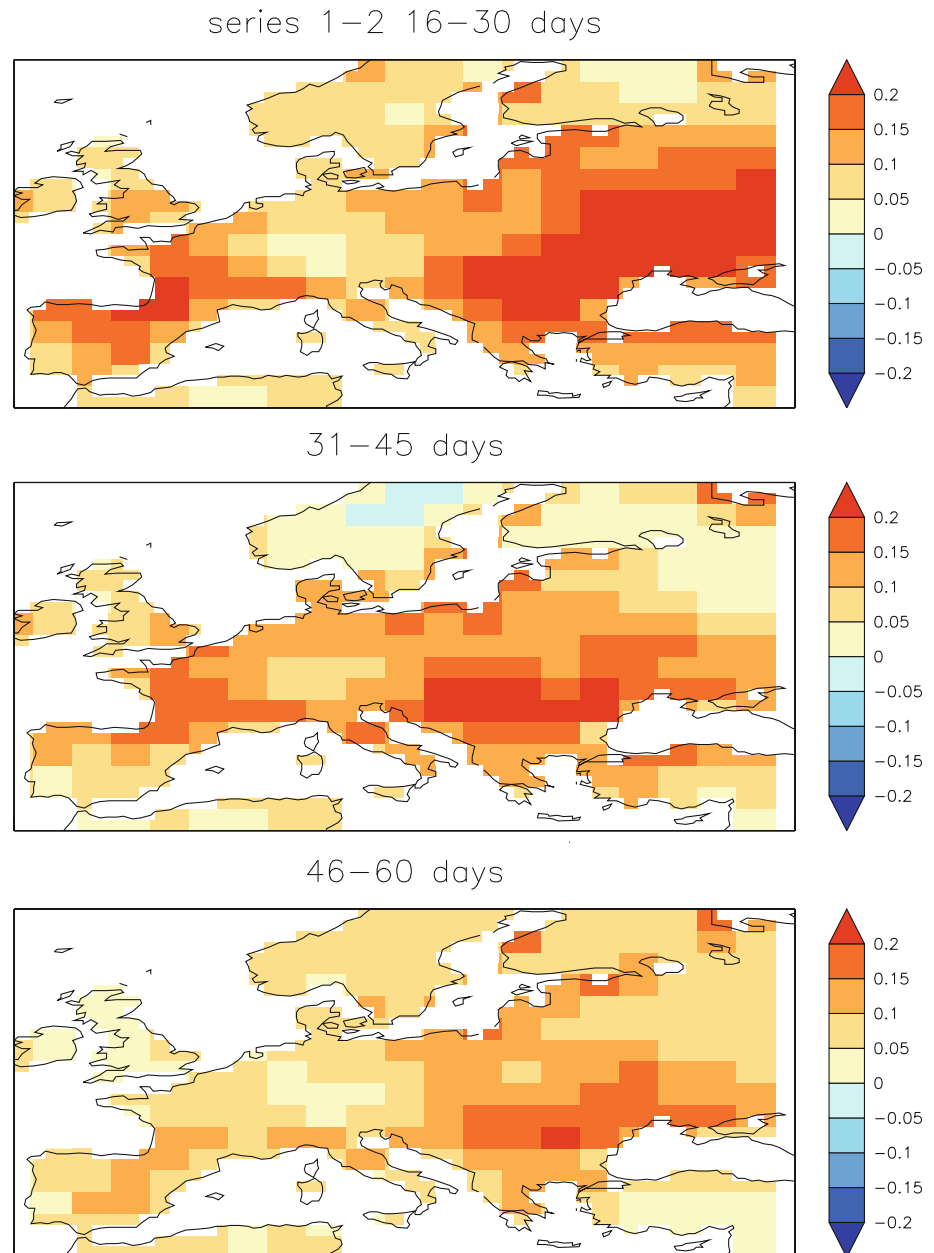
some negative scores emerge at long lead times. The relatively low potential predictability in Europe may be related to the relatively large influence of remote (Atlantic) air masses on temperature and precipitation anomalies. Larger predictability and skill levels are seen in North America, perhaps due to the more continental (less maritime) nature of the climate there (especially in the central US), allowing soil moisture processes there to be more effective. In addition, the northern half of Europe is on average situated at higher latitudes with lower radiation levels (and thus lower evaporation and/or evaporation variability), and it contains fewer areas that might have soil moisture deficits.

As expected, the precipitation forecasts do not improve. Precipitation in most parts of Europe is dominated by atmospheric advection of moisture from the Atlantic (e.g. Van der Ent 2010), and local adjustments of soil moisture conditions may on average have a small impact on precipitation.

The contributions of realistic land initialization to skill in Europe are less pronounced than those shown by Koster

et al. (2010a) for North America. The potential predictability at the time scales considered is lower in Europe than in North America, but in addition, the fraction of the potential predictability captured by the skill calculation is fairly low in Europe, particularly at long lead times, and with a systematic reduction of skill around the Baltic Sea. Although predictability metrics reflect model behaviour rather than intrinsic properties of the real climate, there may be ample room for improvement of the skill, particular through the use of better models, larger ensembles, sampling over a longer period, better initialization methods, and better observations. Koster et al. (2010b) already point at the limited quality of the soil moisture fields used to initialize the series 1 simulations in many areas of the world, largely a reflection of sparse rain gauge density. The verifying temperature and precipitation observations are also not free of errors, which will lead to a systematic gap between skill and potential predictability. Here we also show that the spread in the initial soil moisture content used for series 1 affects the multi-model skill: selecting a

**Fig. 10** Difference in interannual standard deviation of JJA 2 m temperature between series 1 and series 2 at different lead times. Shown is the interannual standard deviation of ensemble mean forecasts, averaged over all models and all JJA time slots



multi-model ensemble characterized by a high similarity in initial soil moisture gives better results.

As demonstrated for North America by Koster et al. (2010a), performing the skill calculations on subsets of the forecast periods as determined by the size of the initial soil moisture anomaly improves the skill scores in many areas of Europe. Soil moisture is not equally informative across the entire wetness range (Koster et al. 2009); selecting extreme soil moisture conditions apparently results in selecting moisture regimes that do affect evaporation and other atmospheric characteristics that in turn determine the surface temperature.

A suggestive result is that temperature forecast skill in Western Europe appears to be related to extreme soil

moisture conditions in South-Central Europe. At longer lead times (46–60 days), computing skill for start dates subsetted on anomalous soil moisture conditions in the remote South-Central Europe region leads to larger skill levels in Western Europe. The South-Central Europe region (a “soil moisture initialization hotspot”) coincides with an area associated with strong soil moisture effects on the surface energy balance in climate simulations (Senviratne et al. 2006) as well as with recent summer heat waves in regional climate simulations (Fischer et al. 2007). This area is also coincident with findings based on GSWP2 simulations and Fluxnet observations regarding the location of regions lying within the soil moisture-limited evapotranspiration regime in Europe (Teuling et al. 2009).

Even with the large number of simulations and models examined here, the noise level in this experiment is rather large. For the highly variable European climate, the 10-year time range covered by the GLACE2 experiment is too short to confirm the existence, for example, of clear atmospheric teleconnections via surface heat low development which can affect the circulation in a large domain. Using a 17-member ensemble climate simulation of 150 years duration, Haarsma et al. (2009) demonstrate an effect of a Mediterranean heat low development in response to excessive soil drying on atmospheric circulation at higher latitudes. This teleconnection could not be confirmed in the multi-model data set explored here, probably due to the limited number of weather situations covered in the experiment. To address such questions, we require an extended version of the GLACE2 experiment, covering a more comprehensive weather history—an experiment utilizing, for example, the multi-decadal forcing dataset of Sheffield et al. (2006) for the soil moisture initialization rather than the 10-year GSWP2 forcing dataset.

**Acknowledgments** Help from Sarith Mahanama (NASA) and Frederic Vitart (ECMWF) for submitting data and executing simulations, as well as from the modelling teams participating in GLACE2, is highly appreciated. Alexander Bakker and Geert Jan van Oldenborgh helped design some of the statistical analyses. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (<http://www.ensembles-eu.org>) and the data providers in the ECA&D project (<http://eca.knmi.nl>).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Balsamo G, Viterbo P, Beljaars A, van den Hurk BJJM, Betts A, Scipal K (2009) A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *J Hydrometeorol* 10:623–643
- Dirmeyer PA, Gao X, Zhao M, Guo Z, Oki T, Hanasaki N (2006) GSWP-2: multimodel analysis and implications for our perception of the land surface. *Bull Am Meteorol Soc* 87:1381–1397
- Dirmeyer PA, Schlosser CA, Brubaker KL (2009) Precipitation, recycling, and land memory: an integrated analysis. *J Hydrometeorol* 10:278–288
- Douville H (2010) Relative contribution of soil moisture and snow mass to seasonal climate predictability: a pilot study. *Clim Dyn* (in press)
- Fischer EM, Seneviratne SI, Luthi D, Schär C (2007) The contribution of land-atmosphere feedbacks to recent European summer heatwaves. *Geophys Res Lett* 34:L06707. doi:10.1029/2006GL029068
- Guo Z, Dirmeyer PA, Hu Z-Z, Gao X, Zhao M (2006) Evaluation of the second global soil wetness project soil moisture simulations. 2. Sensitivity to external meteorological forcing. *J Geophys Res* 111:D22S03. doi:10.1029/2006JD007845
- Haarsma RJ, Selten FM, Van den Hurk BJJM, Hazeleger W, Wang X (2009) Drier mediterranean soils due to greenhouse warming bring easterly winds over summertime Central Europe. *Geophys Res Lett* 36:L04705. doi:10.1029/2008GL036617
- Hagedorn R, Doblas-Reyes FJ, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I. Basic concept. *Tellus A* 57:219–233
- Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded dataset of surface temperature and precipitation. *J Geophys Res* 113:D20119. doi:10.1029/2008JD10201
- Huang J, van den Dool HM (1993) Monthly precipitation-temperature relations and temperature prediction over the United States. *J Clim* 6:1111–1132
- Huang J, van den Dool HM, Georgarakos KP (1996) Analysis of model-calculated soil moisture over the United States (1931–1993) and applications to long-range temperature forecasts. *J Clim* 9:1350–1362
- Kirtman B, Pirani A (2009) The state of the art of seasonal prediction: outcomes and recommendations from the first world climate research program workshop on seasonal prediction. *Bull Am Meteorol Soc* 90:455–458
- Koster RD, Dirmeyer PA, Guo Z, Bonan G, Chan E, Cox P, Gordon CT, Kanae S, Kowalczyk E, Lawrence D, Liu P, Lu C-H, Malyshev S, McAvaney B, Mitchell K, Mocko D, Oki T, Oleson K, Pitman A, Sud YC, Taylor CM, Verseghy D, Vasic R, Xue Y, Yamada T (2004) Regions of strong coupling between soil moisture and precipitation. *Science* 305:1138–1140
- Koster RD, Schubert SD, Suarez MJ (2009) Analyzing the concurrence of meteorological droughts and warm periods, with implications for the determination of evaporative regime. *J Clim* 22:3331–3341
- Koster R, Mahanama S, Yamada T, Balsamo G, Boisserie M, Dirmeyer P, Doblas-Reyes F, Gordon T, Guo Z, Jeong JH, Li Z, Luo L, Malyshev S, Merryfield W, Seneviratne SI, Stanelle T, van den Hurk B, Vitart F, Wood EF (2010a) The contribution of land surface initialization to subseasonal forecast skill: first results from the GLACE-2 project. *Geophys Res Lett* 37:L02402. doi:10.1029/2009GL041677
- Koster RD, Mahanama S, Yamada TJ, Balsamo G, Berg AA, Boisserie M, Dirmeyer PA, Doblas-Reyes FJ, Drewitt G, Gordon CT, Guo Z, Jeong J-H, Lawrence DM, Lee WS, Li Z, Luo L, Malyshev S, Merryfield WJ, Seneviratne SI, Stanelle T, van den Hurk BJJM, Vitart F, Wood EF (2010b) The second phase of the global land-atmosphere coupling experiment: soil moisture contributions to subseasonal forecast skill. *J Hydrometeorol* (submitted)
- Rodwell M, Doblas-Reyes FJ (2006) Predictability and prediction of European monthly to seasonal climate anomalies. *J Clim* 19:6025–6046
- Seneviratne SI, Lüthi D, Litschi M, Schär C (2006) Land-atmosphere coupling and climate change in Europe. *Nature* 443:205–209
- Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, Orlowsky B, Teuling AJ (2010) Investigating soil moisture–climate interactions in a changing climate: a review. *Earth Sci Rev* 99:125–161
- Sheffield J, Goteti G, Wood EF (2006) Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modelling. *J Clim* 19:3088–3111
- Teuling AJ, Hirschi M, Ohmura A, Wild M, Reichstein M, Ciais P, Buchmann N, Ammann C, Montagnani L, Richardson AD, Wohlfahrt G, Seneviratne SI (2009) A regional perspective on trends in continental evaporation. *Geophys Res Lett* 36:L02404. doi:10.1029/2008GL036584
- Van der Ent RJ, Savenije HHG, Schaeffli B, Steele-Dunne SC (2010) The origin and fate of atmospheric moisture over 2 continents. *Water Resour Res*. doi:10.1029/2010WR009127 (in press)