Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe

Chloé Prodhomme[1] , Francisco Doblas-Reyes[1,2,3], Omar Bellprat[1,3], Emanuel Dutra[4]

[1] Catalan Institute of Climate Sciences (IC3), Barcelona, Spain

[2] Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain

[3] Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS), Barcelona, Spain

[4] European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK

Submitted to Climate Dynamics

Corresponding author: Chloé Prodhomme, chloe.prodhomme@ic3.cat

25th October 2015

Abstract

Land surfaces and soil conditions are key sources of climate predictability at the seasonal time scale. In order to estimate how the initialization of the land surface affects the predictability at seasonal time scale, we run two sets of seasonal hindcasts with the general circulation model EC-Earth2.3. The initialization of those hindcasts is done either with climatological or realistic land initialization in May using the ERA-Land re-analysis. Results show significant improvements in the initialized run occurring up to the last forecast month. The prediction of near-surface summer temperatures and precipitation at the global scale and over Europe are improved, as well as the warm extremes prediction. As an illustration, we show that the 2010 Russian heat wave is only predicted when soil moisture is initialized. No significant improvement is found for the retrospective prediction of the 2003 European heat wave, suggesting this event to be mainly large-scale driven. Thus, we confirm that late-spring soil moisture conditions can be decisive in triggering high-impact events in the following summer in Europe. Accordingly, accurate land-surface initial conditions are essential for seasonal predictions.

1. Introduction

In the context of global warming and the associated increasing number of extreme events, such as heat waves, droughts and floods, the predictability at seasonal time scale of extreme temperature and precipitation events appears to be crucial for climate services, adaptation and risk management (Challinor et al. 2005; García-Morales and Dubus 2007; Thomson et al. 2006). The feasibility of seasonal prediction largely rests on the existence of slow, and predictable, variations in the ocean surface temperature, sea ice, soil moisture and snow cover, and how the atmosphere interacts and is affected by these boundary conditions (Shukla and Kinter 2006). Ocean anomalies associated with El Niño - Southern Oscillation (ENSO) and other ocean phenomena, soil moisture, snow, and ice cover should be taken into account when initializing the predictions (Balmaseda et al. 2008; Balmaseda and Anderson 2009). Unfortunately, less information is available about the state of the climate components other than the atmosphere (Balmaseda et al. 2007; Saha et al. 2010). Due to this source of initial-condition uncertainty, but also other limitations as model inadequacy, and lack of appropriate computational resources, the ability to make predictions on time scales longer than two weeks is still limited (Palmer et al. 2005; Palmer et al. 2005b; Lee et al. 2011).

However, in the past years due to increase of the resolution (Fosser et al. 2014; MacLachlan et al. 2014), the development of better initialization products (Guemas et al. 2014; Balmaseda et al. 2009; Balsamo et al. 2015; Dee et al. 2009), and the improvement of model physics (Hourdin et al. 2013; Frenkel et al. 2012) the skill of climate predictions at seasonal and longer time scales has improved (Doblas-Reyes et al. 2013).

Despite the global improvement of seasonal prediction, our ability to forecast temperature and precipitation in some regions such as Europe remains relatively low. On one side, the pronounced warming trend since the 1980s is well captured by most of the seasonal retrospective forecasts over Europe, which provides significant skill for two meter temperatures (t2m, hereinafter) in this region (Doblas-Reyes et al. 2006). On the other side, the skill of predicting the variability around the warming trends is much lower (Weisheimer et al. 2011). This is mainly because the climate variability over Europe is controlled by a variety of mechanisms, such as, the North Atlantic Oscillation (NAO, Rogers 1997; Rodwell et al. 1999), the anomalous frequency of a set of weather regimes (Reinhold and Pierrehumbert 1982; Cassou et al. 2005; Wang et al. 2011), complex teleconnections with the Arctic (Cohen et al. 2014) and with the tropics (Kutiel and Benaroch 2001; Shaman and Tziperman 2011; Behera et al. 2012), and the coupling between the atmosphere and the land surface (Fischer et al. 2007; Orsolini, Yvan and Kvamstø 2009; Wang et al. 2011). All these processes are not properly represented in coupled models, which could explain the poor skill over Europe (Seneviratne et al. 2010; Kim et al. 2012; Scaife et al. 2011). In an early study, Schär et al. (1999) had shown the existence of a soil-precipitation feedback over Europe. Later on, soil has been shown to influence precipitations, temperature and extreme temperature over Europe; (Fischer et al. 2007; Douville 2010; Seneviratne et al. 2006, 2010, 2013; Quesada et al. 2012; Bellprat et al. 2013).

For instance, Seneviratne et al. (2010) described the soil moisture-temperature coupling feedback loop in which, when an anticyclonic anomaly is present over Europe the soil moisture content will either amplify or moderate the surface temperature response. If the soil is moist (energy limited regime) the available surface energy will preferentially dissipate into latent heat fluxes and dampen surface heating. Conversely, when the soil is dry (soil moisture limited regime) more energy is available for sensible heating, inducing an increase of near-surface air temperature (Seneviratne et al. 2010; Hirschi et al. 2011).

As soil moisture partly controls the occurrence of warm events over Europe, a correct initialization of soil moisture content might be essential to correctly forecast summer extreme temperatures. This problem was studied by the global land-atmosphere coupling experiment (GLACE) intercomparison project (http://gmao.gsfc.nasa.gov/research/GLACE). The first phase (GLACE-1) focused on predictability that arises from soil moisture anomalies and determined the geographical regions where soil moisture exerts a significant influence on surface air temperature and precipitation (hot spots) of land-atmosphere coupling (Koster et al. 2004). The second phase (GLACE-2) focused on forecast quality, and assessed the impact of accurate soil-moisture initialization on actual skill using a multimodel approach (Koster et al. 2011). The multimodel mean in GLACE-2 indicates a significant soil-moisture contribution to surface temperature forecast skill in summer with forecast times of up to two months over North and South America (Koster et al. 2010, 2011). While Europe was not then found as a main region of improvement when soil moisture is initialized the GLACE project, numerous other studies have found an impact of soil moisture initialization in Europe (Douville 2010; van den Hurk et al. 2010; Materia et al. 2014).

In the present study, the predictability associated with soil moisture at seasonal time scales is revisited with a focus on Europe. The originality of the present study resides in three different aspects. First, the experiments described in this manuscript cover a long period of 30 years instead of the 10 used in GLACE-2. Second, the forecast time has been extended up to 4 months, which is longer than most of the GLACE-2 experiments (Koster et al. 2004, 2010). Finally, the initialization of the soil moisture has been performed using the new ERA-Land reanalysis (Balsamo et al. 2015), which is expected to provide a good and consistent estimate of soil-moisture initial conditions.

The paper is structured as follows. In section 2, the EC-Earth2.3 forecast system, the experimental set up and the forecast quality assessment methods, as well as the definition used for “mid-extreme” events, are described. Section 3 illustrates the impact of the soil initialization on temperature and precipitation skill at the global scale and for extremes over Europe. Section 4 describes in detail the role of soil moisture for the two case studies of summer 2003 and summer 2010. Finally, section 5 offers a summary and the future prospects of the work.

2. Model and data description

2.1 The EC-Earth2.3 forecast system

The seasonal hindcast experiments are conducted using the EC-Earth2.3 forecast system (Hazeleger et al. 2011). EC-Earth2.3 consists of three model components, the Integrated Forecasting System (IFS) cycle 31r1 for the atmosphere, NEMO2 for the ocean and LIM2 for the sea ice. The model resolution chosen for the atmosphere is a spectral triangular truncation at a wavenumber 159 and for the computation of physical processes reduced Gaussian grid N80, which corresponds to a mesh resolution of around 120 km in the mid-latitudes, with 62 layers in the vertical. EC-Earth uses the H-TESSEL (TESSEL for Tiled ECMWF Scheme for Surface Exchanges over Land) scheme for the land surface (van den Hurk et al. 2000), which includes an improved representation of hydrology over the TESSEL scheme, in agreement with more recent IFS cycles (Balsamo et al. 2009). The model has four active soil layers extending to a depth of 2.89 meters, without considering capillary rise of groundwater or horizontal exchange of soil water. The oceanic component is NEMO (Madec 2008) using the ORCA1 horizontal resolution (which is 1º although with a highly irregular, tripolar grid) and 42 vertical levels. The LIM2 sea-ice model is coupled to the ocean (Fichefet and Maqueda 1997). All model components are coupled through the Ocean Atmosphere Sea Ice Soil version 3 (OASIS3; Valcke 2006) coupler.

2.2 Experimental set up

To assess the impact of a realistic land-surface initialization on sub-seasonal and seasonal forecasts two seasonal hindcast experiments have been performed. A 10-member, four-month long hindcast experiment has been performed over the period 1981-2010 with start dates the first of May of each year. The ocean, sea-ice and atmospheric components are initialized with ORAS4 (Balmaseda et al. 2013), IC3 sea-ice analysis (Guemas et al. 2014) and ERA-Interim

(Dee et al. 2011), respectively. In the INIT experiment the land surface is initialized with the soil moisture and temperature and snow from ERA-Land (Balsamo et al. 2015), which provides consistent land surface conditions to the forecast system since both share the same land-surface model version. The ensemble is constructed by using atmospheric singular vectors and the five ocean analyses available from ORAS4. The CLIM experiment initializes the land surface using the climatology of ERA-Land for the corresponding start date, this being the only difference between INIT and CLIM. With this set up, the impact of the land-surface initialization can be isolated from all the other factors that influence the forecast quality in climate forecasting.

2.3 Forecast quality assessment

The objective of the present study is to assess how the land-surface initialization affects different aspects of the forecast quality of summer precipitation and temperature, with a specific focus over Europe.

For 500-hPa geopotential height, t2m, precipitation and sea level pressure data from the ERA-Interim reanalysis have been used as reference (Dee et al. 2011). For precipitation, the 0 to 12 hour forecasts have been used. For soil moisture the ERA-Land reanalysis product is used (Balsamo et al. 2015).

The skill has been estimated using the correlation of the ensemble mean and the mean anomaly spatial correlation coefficient (MACC, hereafter). We use the Student distribution with N degrees of freedom to estimate the significance level of correlation, N being the effective number of independent data calculated following the method of von Storch and Zwiers (2001). The significance of the difference between two correlations is estimated using the methodology of Steiger (1980), which takes into account the dependence from sharing the same observations in both correlation coefficients. . In addition, the two methods to assess the significance of correlation and the significance of the difference of two correlations takes into account the independent number of data, which is necessary given the serial correlation typical of the time series considered. As there is no standard method to assess the significance of the MACC and difference between two MACC, we estimated their significance with a bootstrap of 100 random drawings, following the methodology of Masson and Mimack (1992). The drawings are done over the members (random selection of the members with repetition) and over the space (bootstrap by square blocks over the considered region). The block size is estimated by estimating the independent number of data on the longitude and latitude dimensions.

As we need to assess the contribution of the trend to the skill, we have compared the correlation and the MACC calculated on “raw” and detrended data. The detrended values are the residual of the regression on the global mean two meter temperature (GMT, hereafter) of the concerned variables; the observations are regressed on the observed GMT and both experiments are regressed on their simulated GMT (van Oldenborgh et al. 2013).

All the verification, as well as part of the plotting, have been done using the version 2.1.1 of the R-based s2dverification package (http://cran.r-project.org/web/packages/s2dverification/index.html).

Contrary to the common evaluation in seasonal forecasting, where seasonal means of the variables are analyzed, the skill of daily extremes is also evaluated in this paper. To estimate the daily extremes, we follow the same methodology as Pepler et al. (submitted), which was inspired by Hamilton et al. (2012) and the CECILIA EU project definitions (http://www.cecilia-eu.org/index.htm). The extremes have been calculated using Tx and Tn, the daily maximum and minimum temperature, respectively, estimated from the 6 hourly t2m.

The first set of extremes are the monthly 90th and 10th percentile of Tn and the 90th percentile of Tx, named hereafter q10 and q90 of Tn, and q90 of Tx, respectively. For the second set of extremes, the climatological 90th and 10th percentile of Tx and Tn are estimated using data from all years between 1981 and 2010. This is done separately for the ERA-Interim data and for the hindcasts. The frequency of days and nights in a month over and under the corresponding climatological percentile are then estimated. To summarize, the present study will focus on six of these variables:

- q10 of Tn for each month and the percentage of nights in a month under the climatological value of the q10 of Tn, also called number of cold nights.
- q90 of Tn for each month and the associated number of nights in a month over the climatological value of the q90 of Tn, also called number of warm nights.
- q90 of Tx for each month and the associated number of days in a month over the climatological value of the q90 of Tn, also called number of warm days.

The two first variables, q10 of Tn and the number of cold nights, correspond to cold extremes while the other four variables are related to warm extremes.

3. Results

3.1 Impact of land-surface initialization during boreal summer

Figure 1 illustrates the skill of the EC-Earth2.3 system for predicting land t2m and precipitation using the correlation between the ensemble-mean prediction and the observational reference.

The results of the CLIM experiment are used as a benchmark. As in most state-of-the-art forecast systems, EC-Earth2.3 shows high skill for t2m over land almost everywhere except over some areas where the observational reference might not be trustworthy (Doblas-Reyes et al. 2013). Statistically significant correlations appear mainly in tropical regions. In contrast, the predictions exhibit lower skill for precipitation, except over a few regions such as those neighboring the Pacific basin and sub-Saharan Africa. An important part of the skill in both temperature and precipitation is linked to ENSO (Landman and Beraki 2012; Phelps et al. 2004; Doblas-Reyes et al. 2013) whose teleconnections over land is well reproduced by the model in most of the relevant areas (Fig. S1).

The use of a realistic initialization of soil variables (snow, soil moisture and soil temperature) such as the one used in the INIT experiment compared to the one used in CLIM has generally a positive impact on the skill of seasonal mean t2m (Fig. 1c). Nevertheless, only very few of the positive changes are statistically significant at the 95% confidence level (black dots), which is the likely result of the small differences and the reduced sample size of the experiment, an aspect that is limited by the observational data available to reliably initialize the hindcasts. The impact of land-surface initialization on the precipitation skill is patchy, although with a tendency to show positive differences in correlation. There is no area with a significant decrease of correlation, whereas a few areas show an important increase of skill (Fig. 1d). The patterns of improvement cannot be simply described by a modification of the ENSO teleconnections over land in INIT compared to CLIM (Fig. S1), because they are very similar in both experiments, and an alternative explanation is needed.

It has to be borne in mind that our study considers longer forecast time scales than the GLACE-2 experiment. For instance, no improvement in seasonal skill over the Great Plains emerges in Fig. 1 compared to previous studies. However, consistently with previous studies (Koster et al. 2004, 2010, 2011; van den Hurk et al. 2010), there is an important improvement of skill in June (second forecast month) over the United States, which disappears in July and August (Figs. S2, S3, S4).

In order to quantify precisely the impact on skill seen on Fig. 1, Fig. 2 shows scatter plots of the difference of correlation between INIT and CLIM against the correlation in CLIM for both precipitation and t2m in different regions. Figure 2 shows the improvement due to the soil initialization for temperature prediction: 65.3% of the land points have a positive impact (Fig. 2d). Nevertheless, the correlation difference between INIT and CLIM is significant only in very few cases (red dots), with no significant negative difference (dark blue dots). In general, in all regions, more improvements (positive differences, points where the skill is significant in INIT but not in CLIM and statistically significant negative correlation decrease) than degradations (negative differences, points where the skill is significant in CLIM but not in INIT and statistically significant negative correlation decrease) are found. This comprehensive analysis shows that the land-surface initialization has on average a positive impact on the temperature skill when large regions are considered.

Conversely, for precipitation no clear improvements are visible on Fig. 2h: on one side 53.9% of the grid points have an increased skill in INIT. On the other side, more points are located in the bottom-right quadrant than in the top-right quadrant, which suggests that more degradations than improvements occur in the areas where CLIM has skill.

Figure 2 shows that, for both temperature and precipitation, the lower the skill in CLIM is, the stronger the improvements in INIT are. In addition, the degradation in INIT tends to occur when the skill is already positive in CLIM. This suggests that the land-surface initialization brings skill to regions where the forecast system has no skill, but it can also negatively perturb the system in regions of high skill, suggesting that the large-scale signal can be perturbed by soil moisture initialization. This can be partly explained by the biases in the soil initialization products (Balsamo et al. 2015) and by the initial shock and drift of the soil variables in the forecasts (Dirmeyer 2005; Materia et al. 2014). Furthermore, model inadequacies in the representation of the land and/or land atmosphere coupling might explain the decrease of skill in INIT. Error compensations may take place in CLIM, in other words, CLIM may have skill in some region for the wrong reasons. In this case, a better representation of the soil state might in some region lead to a decrease of skill. An illustration of possible error compensations can be seen over North-Western South America, where the relation between ENSO and t2m is reversed compared to the observed one (Fig. S1) while still CLIM has a high t2m skill in this region (Fig. 1 and 3).

Another factor that can explain the difference in skill between CLIM and INIT is their representation of the recent temperature trends. In fact, recent trends can explain a large part of the seasonal forecast temperature skill (Doblas-Reyes et al. 2006). Figure 3 is similar to Fig. 1, but this time the correlation has been computed using the residuals of the regression of the temperature and precipitation fields on the GMT. The comparison of Figs. 1a and 3a illustrate the important contribution of the trend in the skill of temperature. An important part of the t2m skill is related to the trend, especially over Europe where most of the skill in CLIM is related to the trend (Doblas-Reyes et al. 2006, 2013). Conversely, Figs. 1 and 3 suggest that there is almost no impact on the skill of precipitation from the temperature trend.

For both precipitation and t2m, the impact of the land-surface initialization remains very similar when the effect of the global-mean temperature is removed (Figs. 1b, 3b). This result, and the inspection of the regression coefficients, suggests that the land-surface initialization affects only marginally the representation of the temperature trend, consistently with the results of Jaeger and Seneviratne (2010). The comparison between Figs. 1c and 3c (see also Fig. S5) gives a hint that the skill improvement in INIT compared to CLIM is slightly stronger when the trend is removed.

As most seasonal forecast systems, EC-Earth2.3 shows widespread skill in seasonal-mean t2m and relatively low skill for precipitation forecasts. An important part of the skill for forecasting t2m is linked to the warming trend. The soil moisture initialization leads to a general improvement of t2m skill and to a lesser extent of precipitation skill, occurring mainly in regions where the skill is low in the CLIM experiment. This improvement remains robust when the global-mean trend effect is removed. The rest of the paper focuses on Europe, a region where soil moisture has been shown to have a strong impact, an aspect that is also evidenced in our experiments (Figs. 1, 2; Jaeger and Seneviratne 2010; Hirschi et al. 2011; Quesada et al. 2012; Douville 2010).

3.2 Summer skill over Europe

The previous section showed that Europe is one of the regions with the largest impact of the land-surface initialization. However, all the results described concentrate on seasonal averages of temperature and precipitation. Instead, various studies have demonstrated that soil moisture plays an important role in the occurrence of extreme warm events (Jaeger and Seneviratne 2010; Hirschi et al 2011; Hamilton et al. 2012). The prediction of extreme events is highly relevant to society (Wang et al. 2009). Hence, any skill improvements on this aspect might have a larger impact than the more traditional result of the increase in seasonal mean skill. This section focuses on the predictability of “seasonal extremes” or “daily extremes” as defined in Hamilton et al. (2012), Eade et al. (2012) and Pepler et al. (submitted). The extreme variables considered, which were selected because they are the most relevant in summer, are classified in two categories (see Section 2.3):

- The warm extremes: q90 of Tx, number of warm days, q90 of Tn and number of warm nights
- The cold extremes: q10 of Tn and number of cold nights

Figure 4 shows the correlation of the ensemble-mean predictions of CLIM for the JJA (one-month lead time) seasonal mean for the different extreme variables. The correlation for the individual months is provided in the supplementary material (Figs. S6, S7 and S8, for June, July and August, respectively).

Consistent with previous studies, the pattern of extreme temperature skill tends to be similar to that of the mean temperature (Figs. 3a and 4a-f; Hamilton et al. 2012; Eade et al. 2012). The skill is also similar for all the variables inside the two groups of extreme variables (Figs. 4 a-d, e-f). The similarity is found also when undetrended anomalies are considered (Figs. 1a, S9a-f; Pepler et al., submitted). However, as for the mean temperature, the skill is lower for all extreme variables when the correlation is calculated on detrended anomalies. However, some regional differences appear. The skill of the CLIM experiment for the warm extreme variables in the Mediterranean region is slightly higher than for the seasonal-average t2m skill, while the skill of the cold extreme variables tends to be higher than the mean t2m skill in eastern and northern Europe (Figs. 4 a-f, 3a).

The correlation changes in INIT with respect to CLIM are very similar for all the extreme warm variables (Fig. 4g-j). Substantial improvements are found over the Mediterranean region, central Europe and Scandinavia for extreme warm variables (Figs. 4g-j), which are areas of low skill in CLIM (Fig. 4a-d). For the extreme cold variables, the soil moisture initialization leads to a weak improvement over the Mediterranean region and Western Europe and a strong degradation in northeastern Europe (Fig. 4k-l) that might be linked to the different behaviour of the snow melting in the two experiments. These patterns obey to a strong intraseasonal evolution of the skill improvement (Figs. S6-S8), with the skill decrease in northeastern Europe occurring mainly in June and the skill increase in western Europe in July, especially for the warm extremes.

To better understand the intraseasonal evolution of the impact of the soil initialization on the skill, the MACC calculated over Europe (20ºW70ºE-25ºN75ºN) for CLIM and INIT and the difference between the MACC in both experiments are displayed in Fig. 5. In May (first month of the forecast), Fig. 5a shows that the skill for predicting the mean and the cold extremes is high (up to 0.7), while the skill for the warm extremes is substantially lower (around 0.25). For the CLIM experiment, the skill of all the variables decreases along the forecast time and reaches almost zero in July (Fig. 5a). In INIT, as in CLIM, the skill sharply decreases between May and June, but remains almost constant at ~0.1 for all variables, which is statistically significant at 95% but not high enough to be considered useful in term of seasonal forecasting (Fig. 5a). Hence, the positive impact of the soil initialization over Europe is more obvious a few weeks after the forecasts have been initialized and is found for all the variables considered. This can be better observed in Fig. 5b, which displays the difference of MACC between the two experiments. There is almost no difference for the variables in May, while in June the cold extremes and the mean t2m exhibit a negative impact of land-surface initialization and a positive or neutral impact for the warm extremes. The decrease in skill of the cold extremes in June is related to the important decrease of skill in central Europe (Figs. 6, S5), which occurs for all variables but is stronger for the cold extremes. As in previous cases, the degradation of skill due to the land initialization happens for regions and periods where the skill is high in CLIM (Figs. 4e-f, 5, S3). Conversely, in July and August, when the skill is low in CLIM (Fig. 5a), we observe an important improvement in INIT for all variables, especially for the warm extremes (Fig. 5b).

In spite of the positive impact of the land initialization over Europe, different regions experience a different impact. Figure 6 shows the correlation difference for the mean t2m and the extreme variables averaged in some of the regions defined in Christensen and Christensen (2007). In two regions of low skill in the CLIM experiment (Scandinavia and eastern Europe; Fig. 4), the land-surface initialization has a positive impact for all variables and during the complete forecast length (Fig. 6d, f, g). In the Alps and Mediterranean area, despite a degradation of skill during one forecast month, the skill is generally higher in INIT than in CLIM. In the three other regions considered, France, central Europe and the Iberian Peninsula, the results are less clear with improvement for some variables occurring simultaneously to degradation of other variables. No statistically significant differences can be found, except for the number of warm days in eastern Europe and for the number of warm nights in Scandinavia.

In summary, the impact of the land-surface initialization is generally positive on predictions of both the mean t2m and extreme temperature variables and is slightly stronger for the warm than for the cold extremes. The improvements last the whole forecast length. However, the results vary from one region to another, and might be associated with the correct prediction of a few events. An analysis of the impact on two of the most relevant events recorded recently over Europe might help interpreting these results.

4. Predictions of the European summers of 2003 and 2010

Dry soils seem to have played a key role in the development of the 2003 and 2010 heat waves over Western Europe and Russia (Weisheimer et al. 2011; Quesada et al 2012; Fischer 2014). The CLIM and INIT experiments allow investigating the soil contribution to these events and to understand their role in determining the seasonal forecast skill.

Figures 7 and 8 illustrate the summer 2003 and 2010 events from observational estimates and their representation in both INIT and CLIM. The left column shows the observed anomalies for five variables: t2m, precipitation, 500-hPa geopotential height (z500, hereafter), sea level pressure (SLP, hereafter) and vertically integrated soil moisture. Dots are used to mark the areas where the anomalies are higher than the climatological upper quintile for t2m, z500 and SLP and are lower than the climatological lower quintile for precipitation and soil moisture. The CLIM and INIT results are displayed in the central and right columns, respectively. Instead of displaying ensemble-mean anomalies, which usually are seriously damped when compared to the reference, the forecast odds are computed from the ensemble. The odds are the ratio between the probability for the anomalies to be in the upper quintile, the interquintile range or the lower quintile, and the climatological probability of these three categories (respectively 20%, 60% and 20%). Each point is attributed to the category corresponding to the highest odds ratio. If the point is attributed to the interquintile range or if there is no category assigned (the categories with two highest odds ratio have an equal value) the point is drawn in white. If the point is attributed to the lower/upper quintile category, the corresponding odds ratio is plotted with the left/right color scale. The odds ratio is a useful way of representing the signal in a probabilistic way because it gives an estimate of how anomalous the probability of the event is (i.e. the number of times it can occur above its climatological frequency) independently of the baseline. These figures allow visualizing how the hindcasts predict the extreme quintile categories for each point.

For the 2003 heat wave, Fig. 7 confirms the occurrence of the warm and dry event over Western Europe in 2003. A blocked regime is visible in the geopotential height, with negative anomalies over north-eastern Europe and positive anomalies over the North Atlantic and Western Europe (Fig. 7g; Garcia-Herrera et al. 2010). The blocking regime is also clearly visible on SLP, except over Western Europe where, consistently with García-Herrera et al. (2010) and Fischer et al. (2007b), the heat low mechanism takes place.

Both INIT and CLIM are able to forecast with high probability this warm and dry anomaly over Western Europe (Fig. 7 b, c, e, f). A successful prediction of the 2003 heat wave has previously been achieved with retrospective forecasts presented in Weisheimer et al. (2011), where the authors highlighted the crucial role of the land surface for the correct prediction of this event. An initial dry anomaly in spring has further been discussed to have been pre-requisite for the development of the 2003 heat wave (Fischer et al. 2007b; Ferranti and Viterbo, 2006). The fact that both experiments are able to forecast the 2003 heat wave is hence surprising and suggests that the exceptional high temperatures in 2003 may be largely a consequence of a strong dynamical forcing. This is supported further by the fact that, in spite of starting from climatological initial conditions, the CLIM experiment develops a high probability of extremely low soil moisture over the Mediterranean and Western Europe. This result is consistent with the studies of Feudale and Shukla (2011a, 2011b), which suggest oceanic conditions to be a major driver of the heat wave. However, the soil moisture, precipitation and temperature are forecasted with higher probabilities in INIT than in CLIM (Fig. 7 b, c, e). Moreover, the spatial pattern of the observed anomalies is better reproduced in INIT than in CLIM. For instance, the dipole structure of temperature and precipitation between north-eastern and western Europe, a characteristic of a blocking regime, is, in contrast to CLIM, reproduced more realistically in INIT. These differences between the two experiments suggest that soil moisture plays a role in maintaining the blocking regime over Europe and for the occurrence or maintenance of the baroclinic anomalies of the heat low mechanism over western Europe, consistently with the studies of Fischer et al. (2007b) and Miralles et al. (2014).

In the case of the 2010 heat wave Fig. 8a, d, g and m show the occurrence of the warm and dry event over Russia in 2010 associated with a dry soil moisture anomaly and an anticyclone over Russia. This warm and dry anomaly associated to high sea level pressure is substantially higher (or lower for soil moisture and precipitation) than the climatological higher quintile for all concerned variables, consistently with Dole et al. (2011). Unlike 2003, for the summer 2010 event no heat low mechanism takes place associated with the anticyclone and warm and dry anomalies over Russia, although the z500 anomaly is shifted with respect to the SLP anomaly. Figure 8 shows that CLIM is not able to predict with probabilities substantially different from the climatological ones the extreme characteristics of the 2010 Russian heat wave for none of the considered variables, except for the soil moisture anomaly. Conversely, in INIT, high probabilities for warm and dry anomalies are found in Eastern Europe (Fig. 8c, f). Figure 8i shows that INIT predicts relatively well the z500 anomalies, indicating that soil moisture initialization might have a feedback on the atmospheric circulation. Nevertheless as for 2003, the SLP pattern of anomalies is not reproduced correctly in INIT nor in CLIM (Fig. 8k, l). The anomalies of temperature, precipitation are misplaced compared to the observational reference (Fig. 8c, f).

In order to better understand how the soil initial conditions can affect the predictability of the event, figure 9 shows for both 2003 and 2010, the soil moisture anomalies, with respect to the daily climatology calculated over 1981-2010, for May 1st. Previous studies have suggested that the 2003 spring was possibly drier than usual (Fischer et al. 2007), however more recent analyses have shown that it was actually likely close to climatology. For instance, in a catchment in Northeastern Switzerland with measurements of whole surface water balance (including soil moisture and evapotranspiration), a recent study (Seneviratne et al. 2012) has shown that soil moisture was not particularly low prior to June 2003. This result was confirmed more broadly for a large part of Central Europe in another study (Whan et al. 2015) based on a newly derived soil moisture dataset (Orth and Seneviratne 2015). However, Fig. 9 shows that according the ERA-Land product, the soil moisture over western Europe in 2003 exhibit a large dry anomaly over the whole western Europe at the beginning of May. During the course of May, soil moisture dries over western Europe and recovers at the end of the month and then decreases during the whole summer. This behaviour is similar to the one described in Whan et al. (2015) based on the soil dataset of Orth and Seneviratne (2012). The two products have the same evolution during summer 2003 but ERA-Land does have larger soil anomalies than the other product in both June and May.

In CLIM, logically, the soil initial condition is very close to 0, while in INIT the simulation starts to a dryer state, however probably due to the interpolation errors (from T511 to T106) and the drift of the first time steps, the first day is less dry than the observed state. The progressive drying of soil during summer is well reproduced by the two simulations (due to the ensemble averaging the evolution is smoother in the simulation than in the reanalysis). Independently of the initial condition of soil, the successful forecast of temperature and precipitation leads to the correct evolution of the soil during the summer. However, it is not clear from the present experiment to know if in July or August, after the strong drying of June (Fig. 9b and Whan et al. 2015), the soil conditions are important, additional experiments with more start dates would be needed.

In 2010, INIT starts with a dry anomalies equivalent to the observed one, while again CLIM starts from 0 (Fig. 9d). Conversely to 2003, the model is unable to forecast the evolution of the soil moisture during the forecast, while the ERA-Land shows a drastic drying during summer, INIT and CLIM keep the same anomaly. So while in INIT, the dry conditions will allow the heat wave to develop the neutral condition in INIT will inhibit its development.

To summarize, it seems that the 2003 warm event was predictable even without the correct initialization of the land surface, consistently with the studies of Feudale and Shukla (2011a, 2011b). The atmospheric and ocean conditions are enough to generate the dry soil moisture anomalies (Fig. 7h and 9b). This last feature shows that the atmospheric circulation was predictable by the model even without the correct soil-moisture initial condition. It hence suggests that the anticyclonic circulation over Europe was driven by the large scale conversely to what has been suggested by previous studies (Garcia-Herrera et al. 2010). However, it appears that the soil moisture is also an important factor for the occurrence of the 2003 heat wave. First, the precipitation and temperature are better predicted when the soil moisture is initialized (Fig. 7b-c, e-f). Moreover, the results of both experiments show that the soil moisture has a feedback on the atmospheric circulation. Finally, the soil moisture seems important to simulate the cold and moist anomalies in Eastern Europe, which are occurring with the heat wave over Western Europe. Conversely, the 2010 heat wave is predictable only when the soil moisture is adequately initialized suggesting that the dry soil moisture anomalies at the beginning of spring might have been crucial for the development of the heat wave over Russia.

5. Summary and conclusions

While European climate is hardly predicted by coupled models, many studies have shown the essential role played by soil moisture in this region (Schär et al. 1999, 2004; Fischer et al. 2007; Douville 2010; Seneviratne et al. 2006, 2010, 2013; Quesada et al. 2012). In the framework of the GLACE project, the role of soil-moisture initialization has been assessed at sub-seasonal time scales (Koster et al. 2010, 2011; van den Hurk et al. 2010). Nevertheless, fewer studies evaluated how soil-moisture initialization can affects skill, especially over Europe, at longer time scales (Douville 2010).

The present study aims to assess the added value of land-surface initialization for seasonal forecasts. Two sensitivity experiments, consisting in 30 years of 10-member ensemble hindcasts of 4 month length have been run. Both sensitivity experiments have been carried out with the EC-Earth2.3 forecast system initialized in the same way for ocean, atmosphere and sea ice. The difference between both sensitivity experiments resides in the initialization of the soil. While the soil temperature, moisture and snow are initialized with ERA-Land in INIT, these variables are initialized with a climatology of ERA-Land in CLIM.

The comparison of those two experiments for summer (June-to-August average, one-month lead time) shows that land-surface initialization has a positive impact on temperature skill and also, to a lesser extent, on precipitation, which is consistent with previous studies (Koster et al. 2004, 2010; Douville et al. 2010; Materia et al. 2014). This improvement is robust whether the warming trend is considered or not. At regional scale, particularly over Europe, the skill improves in a similar way for t2m and a set of associated extreme variables. The improvement occurs up to the last forecast month, which contrasts with the results described in van den Hurk et al. (2010), who found that the improvement goes up to six weeks over Europe. As they found in their study, land initialization can degrade the skill during the second month of the forecast, while important improvements occur at longer forecast times.

Land initialization is also crucial for the prediction of the 2010 heat wave over Russia. The prediction of the 2010 event is successful only when the soil moisture is initialized, showing that the dry conditions preceding the heat wave were decisive in the occurrence of the event. Conversely, the 2003 European heat wave is predicted by both experiments, with either a climatological or a realistic land-surface initialization, suggesting that the event was driven by the large-scale atmospheric circulation. The slightly better skill of the INIT experiment for this event still suggests a positive feedback of dry soil on temperature, consistently with Weisheimer et al. (2011).

This study shows an improvement of temperature skill when land is initialized, Nevertheless, while initializing realistically soil moisture improves skill in regions and for variables of low skill, for regions and variables with high skill, the land-surface initialization could lead to a skill degradation, consistently with the findings of Materia et al. (2014). This means that the land-surface initialization can increase the skill in regions where the original forecast system has no skill but, at the same time, can negatively perturb the large-scale signal or the local conditions in regions of positive skill. A better knowledge of the interaction between the large-scale circulation and the local land-atmosphere coupling as well as an evaluation of the role of the soil-moisture drift on the temperature anomalies simulated is needed to understand the skill degradation. This assessment will require an inspection of the daily evolution of different variables, such as temperature, soil moisture, precipitation and fluxes. The comparison of different soil initialization products and different initialization techniques, such as for example the one known as anomaly initialization, could also help better understanding the processes involved. It is also important determining to what measure the findings of the current study are model dependent. The authors have plans to perform the analysis in a multi-model framework.

An interesting result of the study is the ability to predict the 2003 European heat wave even without realistic land-surface initialization from May, suggesting that there is a role for the large-scale circulation. Deeper analysis are needed to confirm the robustness of this result, first large discrepancies seems to exist between different dataset suggesting that different soil product should be tested for the initialization. Moreover, a large drying which occurs at the beginning of June, new simulation would be needed to know the influence of this drying on the heat wave. The INIT and CLIM simulation will be extended in this purpose. With the help of those extended simulations, the authors will analyze the possible remote forcing of the blocking events over Europe in 2003 and the land-atmosphere feedbacks, which took place that summer.

Acknowledgements

The research leading to these results has received funding from the EU Seventh Framework Programme FP7 (2007-2013) under grant agreements 308378 (SPECS), 282378 (DENFREE) and 607085 (EUCLEIA), and from the Spanish Ministerio de Economía y Competitividad (MINECO) under the project CGL2013-41055-R. We acknowledge the s2dverification R-based package (http://cran.r-project.org/web/packages/s2dverification/index.html). We also thank

ECMWF for providing the ERA-Land initial conditions and computing resources through the SPICCF Special Project.

References

Balmaseda M, Anderson D, Vidard A (2007) Impact of Argo on analyses of the global ocean. Geophys Res Lett 34:L16605. doi: 10.1029/2007GL030452

Balmaseda M, Vidard A, Anderson DLT (2008) The ECMWF Ocean Analysis System: ORA-S3. Mon Weather Rev 136:3018–3034. doi: 10.1175/2008MWR2433.1

Balmaseda M, Anderson D (2009) Impact of initialization strategies and observations on seasonal forecast skill. Geophys Res Lett 36:L01701. doi: 10.1029/2008GL035561

Balmaseda M, Alves O, Arribas A, et al. (2009) Ocean initialization for seasonal forecasts. Oceanography 22:154–159.

Balmaseda M, Mogensen K, Weaver AT (2013) Evaluation of the ECMWF ocean reanalysis system ORAS4. Q J R Meteorol Soc 139:1132–1161. doi: 10.1002/qj.2063

Balsamo G, Beljaars A, Scipal K, et al. (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. doi: 10.1175/2008JHM1068.1

Balsamo, G., et al (2015) ERA-Interim/Land: a global land surface reanalysis data set, Hydrol. Earth Syst. Sci., 19, 389-407, doi:10.5194/hess-19-389-2015 2015.

Behera S, Ratnam J V., Masumoto Y, Yamagata T (2012) Origin of extreme summers in Europe: the Indo-Pacific connection. Clim Dyn 41:663–676. doi: 10.1007/s00382-012-1524-8

Bellprat O, Kotlarski S, Lüthi D, Schär C (2013) Physical constraints for temperature biases in climate models. Geophys Res Lett 40:4042–4047. doi: 10.1002/grl.50737

Cassou C, Terray L, Phillips A (2005) Tropical Atlantic influence on European heat waves. J Clim 2805–2811.

Cohen J, Screen J a., Furtado JC, et al. (2014) Recent Arctic amplification and extreme mid-latitude weather. Nat Geosci 7:627–637. doi: 10.1038/ngeo2234

Challinor AJ, Slingo JM, Wheeler TR, Doblas-Reyes FJ (2005) Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles. Tellus A 57:498–512. doi: 10.1111/j.1600-0870.2005.00126.x

Christensen JH, Christensen OB (2007) A summary of the {PRUDENCE} model projections of changes in {E}uropean climate by the end of this century. Clim Change 81:7–30.

Dee DP, Berrisford P, Poli P, Fuentes M (2009) ERA-Interim for climate monitoring. ECMWF Newsletter 119: 5–6.

Dee DP, Uppala SM, Simmons a. J, et al. (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

Dirmeyer P (2005) The land surface contribution to the potential predictability of boreal summer season climate. J Hydrometeorol 6:618–632. doi: 10.1175/JHM444.1

Doblas-Reyes FJ, Hagedorn R, Palmer TN, Morcrette J-J (2006) Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts. Geophys Res Lett 33:L07708. doi: 10.1029/2005GL025061

Doblas-Reyes FJ, García-Serrano J, Lienert F, et al (2013) Seasonal climate predictability and forecasting: Status and prospects. Wiley Interdiscip Rev Clim Chang 4:245–268. doi: 10.1002/wcc.217

Dole R, Hoerling M, Perlwitz J, et al. (2011) Was there a basis for anticipating the 2010 Russian heat wave? Geophys Res Lett 38:1–5. doi: 10.1029/2010GL046582

Douville H (2010) Relative contribution of soil moisture and snow mass to seasonal climate predictability: A pilot study. Clim Dyn 34:797–818. doi: 10.1007/s00382-008-0508-1

Eade R, Hamilton E, Smith DM, et al. (2012) Forecasting the number of extreme daily events out to a decade ahead. J Geophys Res Atmos 117:D21110. doi: 10.1029/2012JD018015

Feudale L, Shukla J (2011) Influence of sea surface temperature on the European heat wave of 2003 summer. Part I: an observational study. Clim Dyn 36:1691–1703. doi: 10.1007/s00382-010-0788-0

Feudale L, Shukla J (2011) Influence of sea surface temperature on the European heat wave of 2003 summer. Part II: a modeling study. Clim Dyn 36:1705–1715. doi: 10.1007/s00382-010-0789-z

Fichefet T, Maqueda Morales MA (1997) Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. J Geophys Res 102:12609. doi: 10.1029/97JC00480

Fischer EM, Seneviratne SI, Lüthi D, Schär C (2007) Contribution of land-atmosphere coupling to recent European summer heat waves. Geophys Res Lett 34:L06707. doi: 10.1029/2006GL029068

Fischer EM, Seneviratne SI, Vidale PL, et al (2007b) Soil Moisture–Atmosphere Interactions during the 2003 European Summer Heat Wave. J Clim 20:5081–5099. doi: 10.1175/JCLI4288.1

Fischer EM (2014) Climate science: Autopsy of two mega-heatwaves. Nat Geosci 7:332–333. doi: 10.1038/ngeo2148

Fosser G, Khodayar S, Berg P (2014) Benefit of convection permitting climate model simulations in the representation of convective precipitation. Clim Dyn 44:45–60. doi: 10.1007/s00382-014-2242-1

García-Herrera R, Díaz J, Trigo RM, et al. (2010) A Review of the European Summer Heat Wave of 2003. Crit Rev Environ Sci Technol 40:267–306. doi: 10.1080/10643380802238137

García- Morales M, Dubus L (2007) Forecasting precipitation for hydroelectric power management: how to exploit GCM's seasonal ensemble forecasts. Int J Climatol 1705:1691–1705. doi: 10.1002/joc

Guemas V, Doblas-Reyes FJ, Mogensen K, et al. (2014) Ensemble of sea ice initial conditions for interannual climate predictions. Clim Dyn. doi: 10.1007/s00382-014-2095-7

Hamilton E, Eade R, Graham RJ, et al. (2012) Forecasting the number of extreme daily events on seasonal timescales. J Geophys Res 117:D03114. doi: 10.1029/2011JD016541

Hazeleger W, Wang X, Severijns C, et al. (2011) EC-Earth 8.2: description and validation of a new seamless earth system prediction model. Clim Dyn 39:2611–2629. doi: 10.1007/s00382-011-1228-5

Hirschi M, Seneviratne SI, Alexandrov V, et al. (2011) Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. Nat Geosci 4:17–21. doi: 10.1038/ngeo1032

Hourdin F, Grandpeix J-Y, Rio C, et al. (2012) LMDZ5B: the atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection. Clim Dyn 40:2193–2222. doi: 10.1007/s00382-012-1343-y

Jaeger EB, Seneviratne SI (2010) Impact of soil moisture–atmosphere coupling on European climate extremes and trends in a regional climate model. Clim Dyn 36:1919–1939. doi: 10.1007/s00382-010-0780-8

Kim H-M, Webster PJ, Curry J a. (2012) Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. Clim Dyn 39:2957–2973. doi: 10.1007/s00382-012-1364-6

Kutiel H, Benaroch Y (2002) North Sea-Caspian Pattern (NCP)–an upper level atmospheric teleconnection affecting the Eastern Mediterranean: Identification and definition. Theor Appl Climatol 28:17–28. doi: 10.1007/s704-002-8205-x

Koster RD, Suarez MJ, Liu P, et al (2004) Realistic Initialization of Land Surface States: Impacts on Subseasonal Forecast Skill. J Hydrometeorol 5:1049–1063.

Koster RD, Mahanama SPP, Yamada TJ, et al (2010) Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. Geophys Res Lett 37:L02402. doi: 10.1029/2009GL041677

Koster RD, Mahanama SPP, Yamada TJ, et al (2011) The Second Phase of the Global Land–Atmosphere Coupling Experiment: Soil Moisture Contributions to Subseasonal Forecast Skill. J Hydrometeorol 12:805–822. doi: 10.1175/2011JHM1365.1

Landman W a., Beraki A (2012) Multi-model forecast skill for mid-summer rainfall over southern Africa. Int J Climatol 32:303–314. doi: 10.1002/joc.2273

Lee S-S, Lee J-Y, Ha K-J, et al. (2010) Deficiencies and possibilities for long-lead coupled climate prediction of the Western North Pacific-East Asian summer monsoon. Clim Dyn 36:1173–1188. doi: 10.1007/s00382-010-0832-0

MacLachlan C, Arribas A, Peterson KA, et al. (2014) Global Seasonal Forecast System version 5 (GloSea5): a high resolution seasonal forecast system. Q J R Meteorol Soc. doi: 10.1002/qj.2396

Madec G (2008) NEMO ocean engine. Note du Pole de modelisation, Institut Pierre-Simon Laplace (IPSL) No 27. ISSN No 1288-1619

Mason S, Mimmack G (1992) The use of bootstrap confidence intervals for the correlation coefficient in climatology. Theor Appl Climatol 45:229–233.

Materia S, Borrelli A, Bellucci A, et al (2014) Impact of Atmosphere and Land Surface Initial Conditions on Seasonal Forecasts of Global Surface Temperature. J Clim 27:9253–9271. doi: 10.1175/JCLI-D-14-00163.1

Miralles DG, Teuling AJ, Heerwaarden CC Van (2014) Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat Geosci 7:345–349. doi: 10.1038/ngeo2141

Orsolini YJ, Kvamstø NG (2009) Role of Eurasian snow cover in wintertime circulation: Decadal simulations forced with satellite observations. J Geophys Res 114:D19108. doi: 10.1029/2009JD012253

Rene Orth, Seneviratne SI (2015) Introduction of a simple-model-based land surface dataset for Europe. Environ Res Lett 10:044012. doi: 10.1088/1748-9326/10/4/044012

Palmer TN, Shutts GJ, Hagedorn R, et al. (2005) Representing Model Uncertainty in Weather and Climate Prediction. Annu Rev Earth Planet Sci 33:163–193. doi: 10.1146/annurev.earth.33.092203.122552

Palmer TN, Doblas-Reyes FJ, Hagedorn R, Weisheimer A (2005b) Probabilistic prediction of climate using multi-model ensembles: from basics to applications. Philos Trans R Soc Lond B Biol Sci 360:1991–8. doi: 10.1098/rstb.2005.1750

Pepler AS, Diaz L, Karthik Kashinath K, et al, The ability of a multi-model seasonal forecasting ensemble to forecast the seasonal distribution of daily extremes. Weather and Climate Extremes (submitted)

Phelps M, Kumar A, O'Brien J (2004) Potential predictability in the NCEP CPC dynamical seasonal forecast system. J Clim 17:3775–3785.

Quesada B, Vautard R, Yiou P, et al. (2012) Asymmetric European summer heat predictability from wet and dry southern winters and springs. Nat Clim Chang 2:736–741. doi: 10.1038/nclimate1536

Reinhold B, Pierrehumbert R (1982) Dynamics of weather regimes: Quasi-stationary waves and blocking. Mon. Weather Rev.

Rodwell M, Rowell D, Folland C (1999) Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. Nature 398:25–28.

Rogers J (1997) North Atlantic storm track variability and its association to the North Atlantic Oscillation and climate variability of northern Europe. J Clim 10:1635–1647.

Saha S, Moorthi S, Pan H-L, et al. (2010) The NCEP Climate Forecast System Reanalysis. Bull Am Meteorol Soc 91:1015–1057. doi: 10.1175/2010BAMS3001.1

Scaife A a., Copsey D, Gordon C, et al. (2011) Improved Atlantic winter blocking in a climate model. Geophys Res Lett 38:L23703. doi: 10.1029/2011GL049573

Schär C, Lüthi D, Beyerle U, Heise E (1999) The soil-precipitation feedback: A process study with a regional climate model. J Clim 12:722–741.

Seneviratne SI, Lüthi D, Litschi M, Schär C (2006) Land-atmosphere coupling and climate change in Europe. Nature 443:205–9. doi: 10.1038/nature05095

Seneviratne SI, Corti T, Davin EL, et al (2010) Investigating soil moisture-climate interactions in a changing climate: A review. Earth-Science Rev 99:125–161. doi: 10.1016/j.earscirev.2010.02.004

Seneviratne S, Wilhelm M, Stanelle T, et al. (2013) Impact of soil moisture- climate feedbacks on CMIP5 projections: First results from the GLACE- CMIP5 experiment. Geophys Res Lett 40:5212–5217. doi: 10.1002/grl.50956

Seneviratne SI, Lehner I, Gurtz J, et al. (2012) Swiss prealpine Rietholzbach research catchment and lysimeter: 32 year time series and 2003 drought event. Water Resour Res 48:1–20. doi: 10.1029/2011WR011749

Shaman J, Tziperman E (2011) An Atmospheric Teleconnection Linking ENSO and Southwestern European Precipitation. J Clim 24:124–139. doi: 10.1175/2010JCLI3590.1

Shukla J, Kinter III JL (2006) Predictability of seasonal climate variations: A pedagogical review. Predictability in Weather and Climate, 306-341.

Steiger (1980) Tests for comparing elements of a correlation matrix. doi: 10.1037/0033-2909.87.2.245

Valcke S (2006) OASIS3 user guide (prism_2-5). PRISM support initiative report 3, pp 64

van den Hurk BJJM, et al (2000) Offline validation of the ERA40 surface scheme. European Centre for Medium-Range Weather Forecasts Technical Memorandum 295. Available from http://nwmstest.ecmwf.int/publications/library/do/references/show?id=83838

van den Hurk BJJM, Doblas-Reyes F, Balsamo G, et al (2010) Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. Clim Dyn 38:349–362. doi: 10.1007/s00382-010-0956-2

van Oldenborgh GJ, Doblas Reyes FJ, Drijfhout SS, Hawkins E (2013) Reliability of regional climate model trends. Environ Res Lett 8:014055. doi: 10.1088/1748-9326/8/1/014055

von Storch H, Zwiers FW (2001) Statistical Analysis in Climate Research. Cambridge University Press, 484 pp.

Wang A, Bohn TJ, Mahanama SP, et al. (2009) Multimodel Ensemble Reconstruction of Drought over the Continental United States. J Clim 22:2694–2712. doi: 10.1175/2008JCLI2586.1

Wang E, Xu J, Jiang Q, Austin J (2008) Assessing the spatial impact of climate on wheat productivity and the potential value of climate forecasts at a regional level. Theor Appl Climatol 95:311–330. doi: 10.1007/s00704-008-0009-5

Wang G, Dolman a. J, Alessandri a. (2011) A summer climate regime over Europe modulated by the North Atlantic Oscillation. Hydrol Earth Syst Sci 15:57–64. doi: 10.5194/hess-15-57-2011

Whan K, Zscheischler J, Orth R, et al. (2015) Impact Of Soil Moisture On Extreme Maximum Temperatures In Europe. Weather Clim Extrem 1–11. doi: 10.1016/j.wace.2015.05.001

Weisheimer A, Doblas-Reyes FJ, Jung T, Palmer TN (2011) On the predictability of the extreme summer 2003 over Europe. Geophys Res Lett 38. doi: 10.1029/2010GL046455

Figure captions

Fig. 1: **a** Correlation of the ensemble mean t2m averaged in JJA (one-month lead time) in the CLIM experiment. The dots mark the areas where the correlation is significant at the 95% confidence level. **b** Same as **a**, but for precipitation. **c** Difference of correlation of the ensemble mean between the INIT and CLIM experiments for the t2m in JJA. The dots mark the areas where the difference of correlation is significant at the 95% confidence level. **d** Same as **c** but for precipitation.

Fig. 2: **a** Scatter plot of the difference of correlation of the JJA t2m (one-month lead time) between INIT and CLIM against the correlation of the JJA t2m in CLIM over the the Northern Hemisphere land grid points. The numbers in the corners correspond to the percentage of grid points in the respective quadrants. The grey dots correspond to the values in the grid points where neither the correlation in CLIM, INIT, nor the difference of correlation between INIT and CLIM is significant. The black dots represent the points where the correlation is significant at 95% confidence level in both CLIM and INIT, the orange dots the points where the correlation is significant at 95% confidence level in INIT but not in CLIM, the light blue dots to the points where the correlation is significant at 95% confidence level in CLIM but not in INIT and the red (dark blue) dots to the points where the correlation difference is significantly positive (negative) at 95% confidence level. **b** Same as **a**, but in the tropics. **c** Same as **a**, but in the Southern Hemisphere (without Antarctica). **d** Same as **a**, but over the whole globe. The **e**, **f**, **g** and **h** panels show the equivalent results for precipitation.

Fig. 3: As in Fig. 1, but with the correlation computed using the residual of the regression of the temperature and precipitation anomalies on the global mean temperature.

Fig. 4: Correlation of the ensemble mean in JJA (one-month lead time) in CLIM, for **a** q90 of Tx, **b** number of warm days, **c** q90 of Tn, **d** number of warm nights, **e** the q10 of Tn and **f** number of cold nights. The dots mark the areas where the correlation is statistically significant with a 95% confidence level. Difference of correlation between INIT and in CLIM in JJA, for **g** q90 of Tx, **h** number of warm days, **i** q90 of Tn, **j** number of warm nights, **k** the q10 of Tn and **l** number of cold nights. The dots mark the areas where the difference of correlation is significant at 95% confidence level and the correlation has been computed using the detrended anomalies.

Fig. 5: a Mean spatial anomaly correlation coefficient (MACC) calculated for the ensemble-mean hindcasts of CLIM (plain line) and INIT (dotted line) over the land in Europe (10ºW40ºE-

35ºN75ºN) for the monthly mean t2m (black), the q90 of Tx (red), the q90 of Tn (pink), the q10 of Tn (purple), the number warm days (orange), the number of warm nights (green) and the number of cold days (light blue). The MACC is calculated on detrended anomalies. The solid (open) dots mark the values significant at 95% level in INIT (CLIM), estimated with a bootstrap over of 100 drawings. **b** Same as **a** but for the difference of the MACC between INIT and CLIM. None of the difference of MACC is significant at 95% level, estimated with a bootstrap of 100 drawings.

Fig. 6: a Difference of correlation between the INIT and CLIM experiments for the temperature variables averaged in the Iberian Peninsula region (10ºW3ºE-36ºN44ºN). **b** Same as **a**, but for France (5ºW5ºE-44ºN50ºN), **c** central-Europe (2ºW16ºE-48ºN55ºN), **d** Scandinavia (5ºE30ºE-55ºN70ºN), **e** the Alps (5ºE15ºE-44ºN48ºN), **f** the Mediterranean area (3ºE25ºE-36ºN44ºN) and **g** Eastern Europe (6ºE30ºE-44ºN55ºN).

Fig. 7: a Observed anomalies of t2m for 2003 JJA (one-month lead time) mean (K). The dots indicate the area where the anomaly is in the upper quintile (estimated over 1981-2010). **b** Odds in CLIM for t2m. The odds are the ratio between the probability for the anomalies to be in the upper quintile, the interquintile range or the lower quintile and with the climatological probability of these three categories (20%, 60% and 20%, respectively). Each point is attributed to the category corresponding to the highest odds ratio. If the point is attributed to the interquintile range or if there is no category assigned (the categories with two highest odds ratio have an equal value) the point is drawn in white. If the point is attributed to the lower/upper quintile category, the corresponding odds ratio is plotted with the left/right color scale.**c** Same as **b**, but for INIT. **d** Observed anomalies of precipitation for 2003 JJA mean (mm/day). The dots indicate the area where the anomaly is in the lower quintile for the 1981-2010 period. **e** same as **b,** but for precipitation. **f** Same as **c**, but for precipitation. **g**, **h**, **i** same as **a**, **b**, **c**, but for geopotential height at 500 hPa (m). **j**, **k**, **l** same as **a**, **b**, **c**, but for monthly mean of 6 hourly SLP (hPa). **m**, **n**, **o** same as **d**, **e, f,** but for the vertically integrated volume fraction of water in soil (m^3/m^3).

Fig. 8: Same as Fig. 7, but for JJA (one-month lead time) 2010.

Fig.9: a Standardized anomalies with respect to the daily climatology computed over 1981-2010 of ERA-Land for May 1st 2003. **b** Evolution of the daily anomalies of summer 2003 averaged in the black box of a (5W20E-43N55N) in black for ERA-Land, in blue for the ensemble mean of CLIM and in red for the ensemble mean of INIT. **c** Same as a but for May 1st 2010. **d** Same as

b but for the box drawn on c (25E55E-45N60N) during summer 2010. For all the panel the unit is m^3/m^3.

Figure

Figures

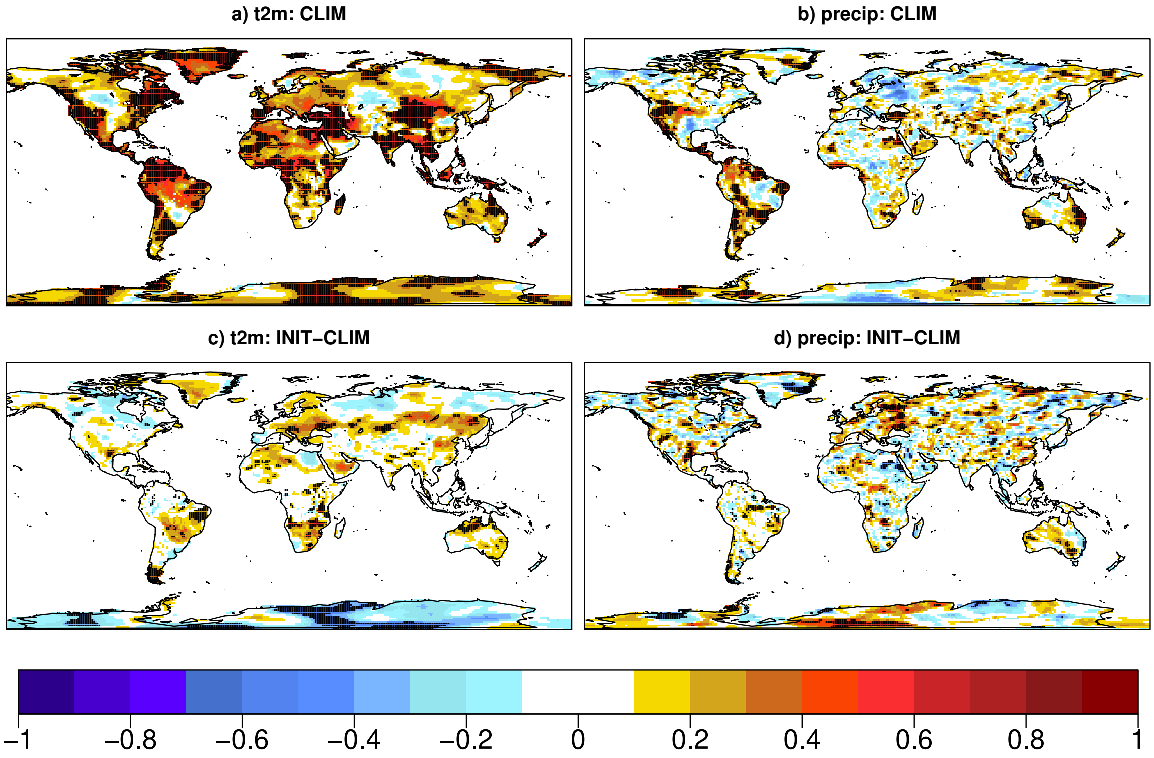

Fig. 1: **a** Correlation of the ensemble mean t2m averaged in JJA (one-month lead time) in the CLIM experiment. The dots mark the areas where the correlation is significant at the 95% confidence level. **b** Same as **a**, but for precipitation. **c** Difference of correlation of the ensemble mean between the INIT and CLIM experiments for the t2m in JJA. The dots mark the areas where the difference of correlation is significant at the 95% confidence level. **d** Same as **c** but for precipitation.

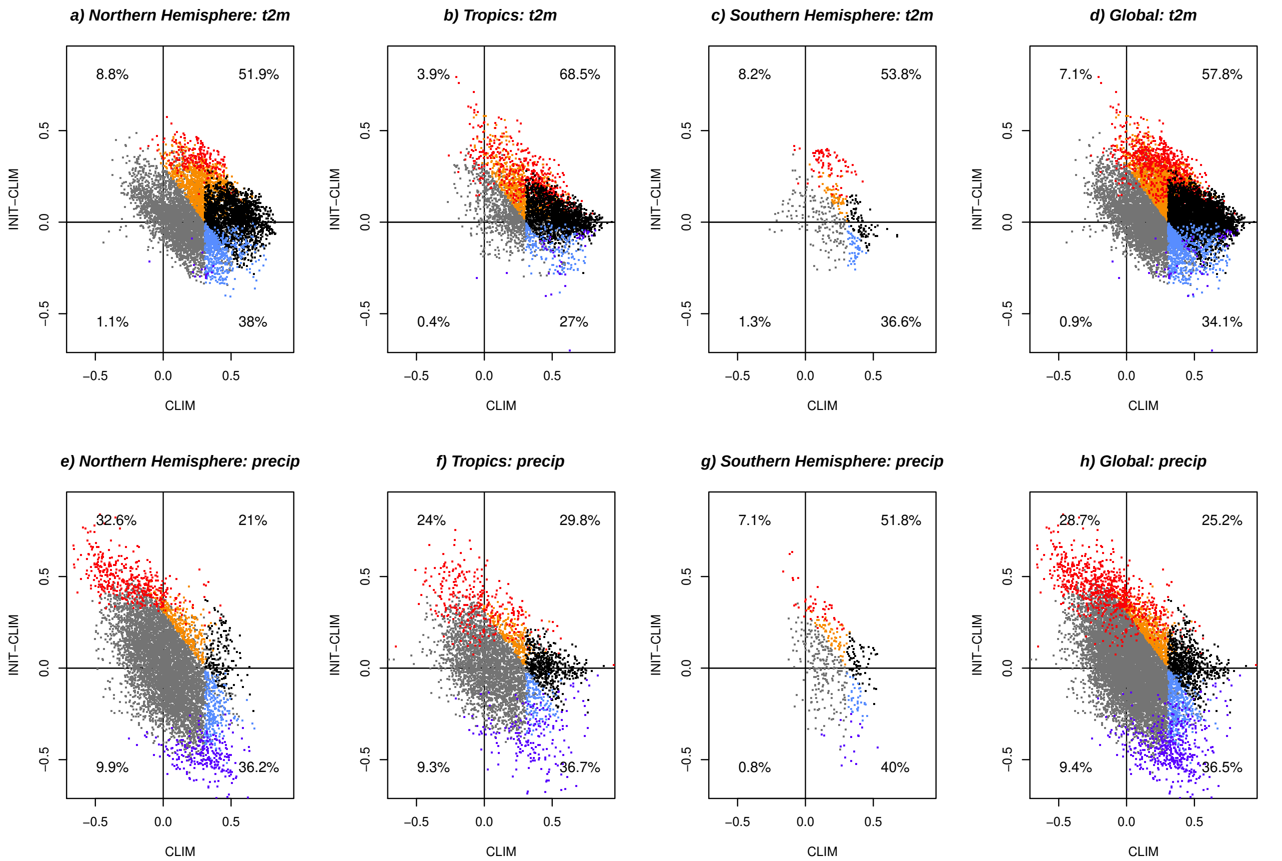

Fig. 2: **a** Scatter plot of the difference of correlation of the JJA t2m (one-month lead time) between INIT and CLIM against the correlation of the JJA t2m in CLIM over the the Northern Hemisphere land grid points. The numbers in the corners correspond to the percentage of grid points in the respective quadrants. The grey dots correspond to the values in the grid points where neither the correlation in CLIM, INIT, nor the difference of correlation between INIT and CLIM is significant. The black dots represent the points where the correlation is significant at 95% confidence level in both CLIM and INIT, the orange dots the points where the correlation is significant at 95% confidence level in INIT but not in CLIM, the light blue dots to the points where the correlation is significant at 95% confidence level in CLIM but not in INIT and the red (dark blue) dots to the points where the correlation difference is significantly positive (negative) at 95% confidence level. **b** Same as **a**, but in the tropics. **c** Same as **a**, but in the Southern Hemisphere (without Antarctica). **d** Same as **a**, but over the whole globe. The **e**, **f**, **g** and **h** panels show the equivalent results for precipitation.

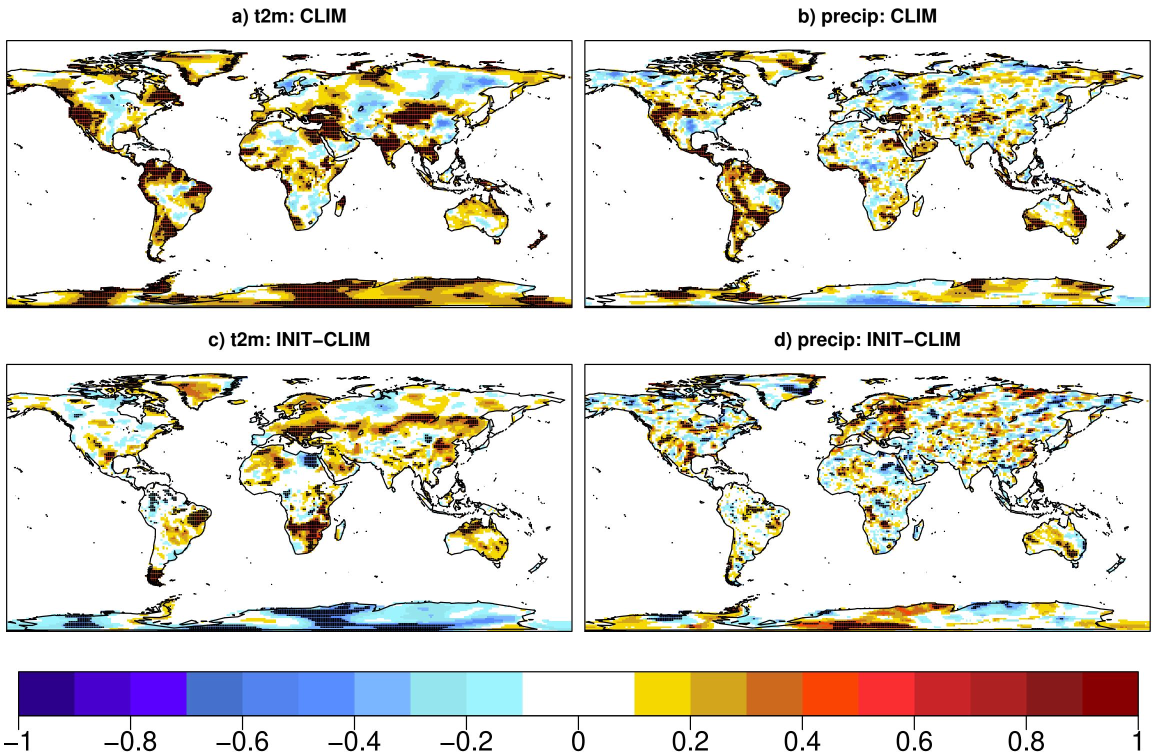

Fig. 3: As in **Fig. 1**, but with the correlation computed using the residual of the regression of the temperature and precipitation anomalies on the global mean temperature.

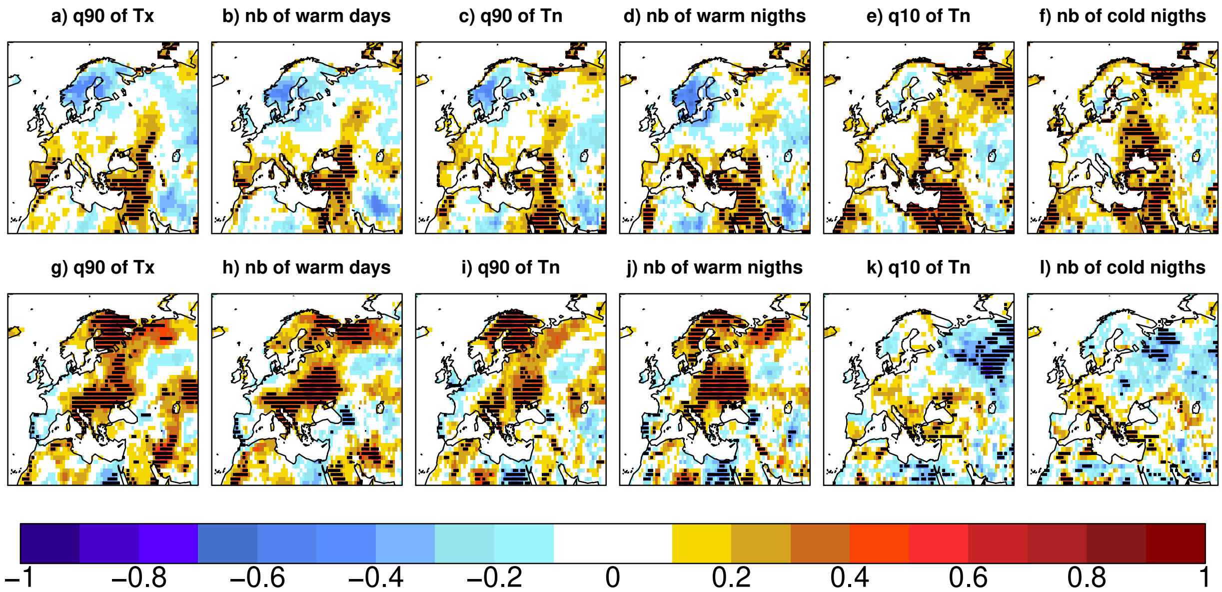

Fig. 4: Correlation of the ensemble mean in JJA (one-month lead time) in CLIM, for **a** q90 of Tx, **b** number of warm days, **c** q90 of Tn, **d** number of warm nights, **e** the q10 of Tn and **f** number of cold nights. The dots mark the areas where the correlation is statistically significant with a 95% confidence level. Difference of correlation between INIT and in CLIM in JJA, for **g** q90 of Tx, **h** number of warm days, **i** q90 of Tn, **j** number of warm nights, **k** the q10 of Tn and **l** number of cold nights. The dots mark the areas where the difference of correlation is significant at 95% confidence level and the correlation has been computed using the detrended anomalies.

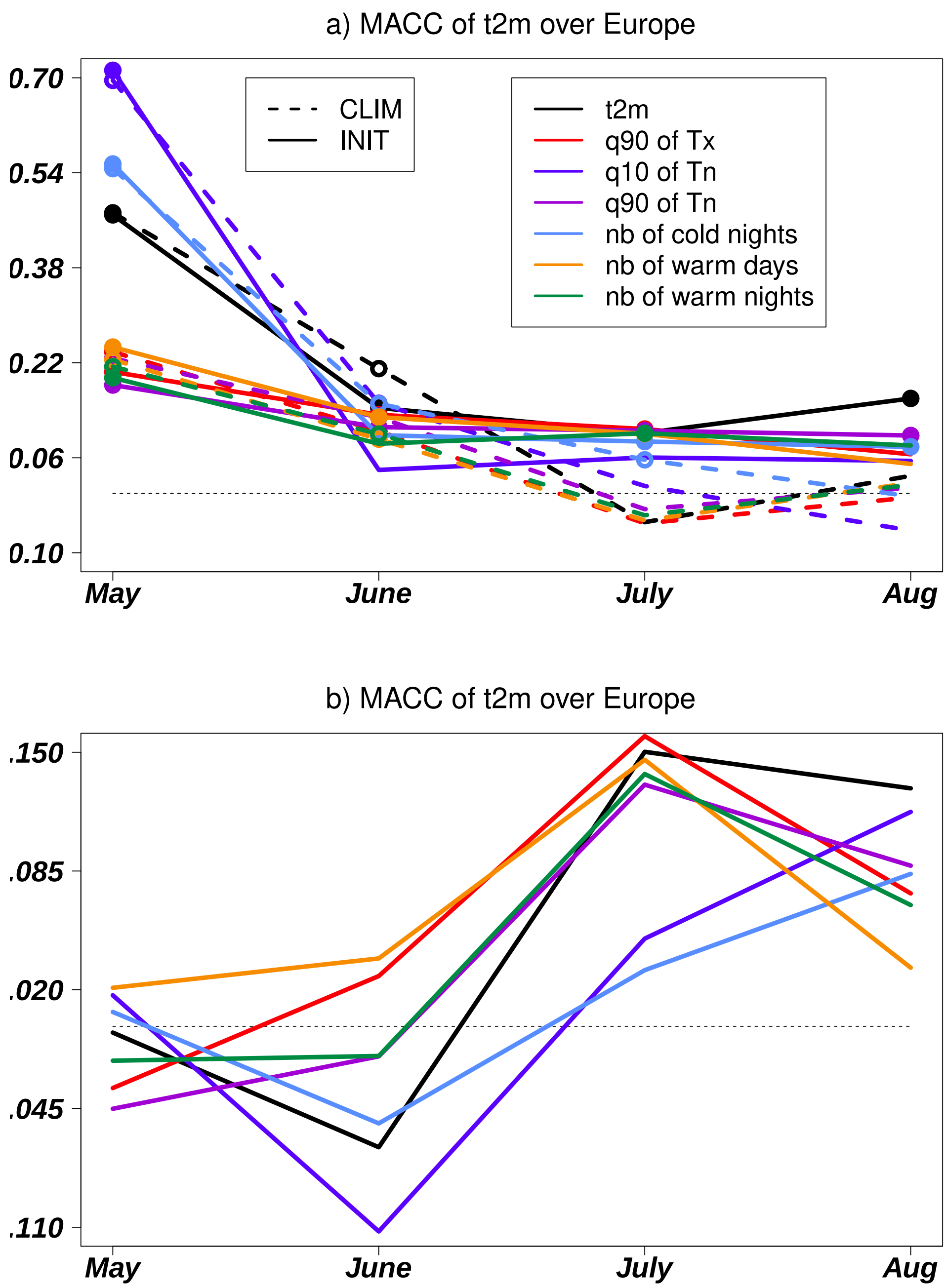

Fig. 5: **a** Mean spatial anomaly correlation coefficient (MACC) calculated for the ensemble-mean hindcasts of CLIM (plain line) and INIT (dotted line) over the land in Europe (10°W40°E-35°N75°N) for the monthly mean t2m (black), the q90 of Tx (red), the q90 of Tn (pink), the q10 of Tn (purple), the number warm days (orange), the number of warm nights (green) and the number of cold days (light blue). The MACC is calculated on detrended anomalies. The solid (open) dots mark the values significant at 95% level in INIT (CLIM), estimated with a bootstrap over of 100 drawings. **b** Same as a but for the difference of the MACC between INIT and CLIM.

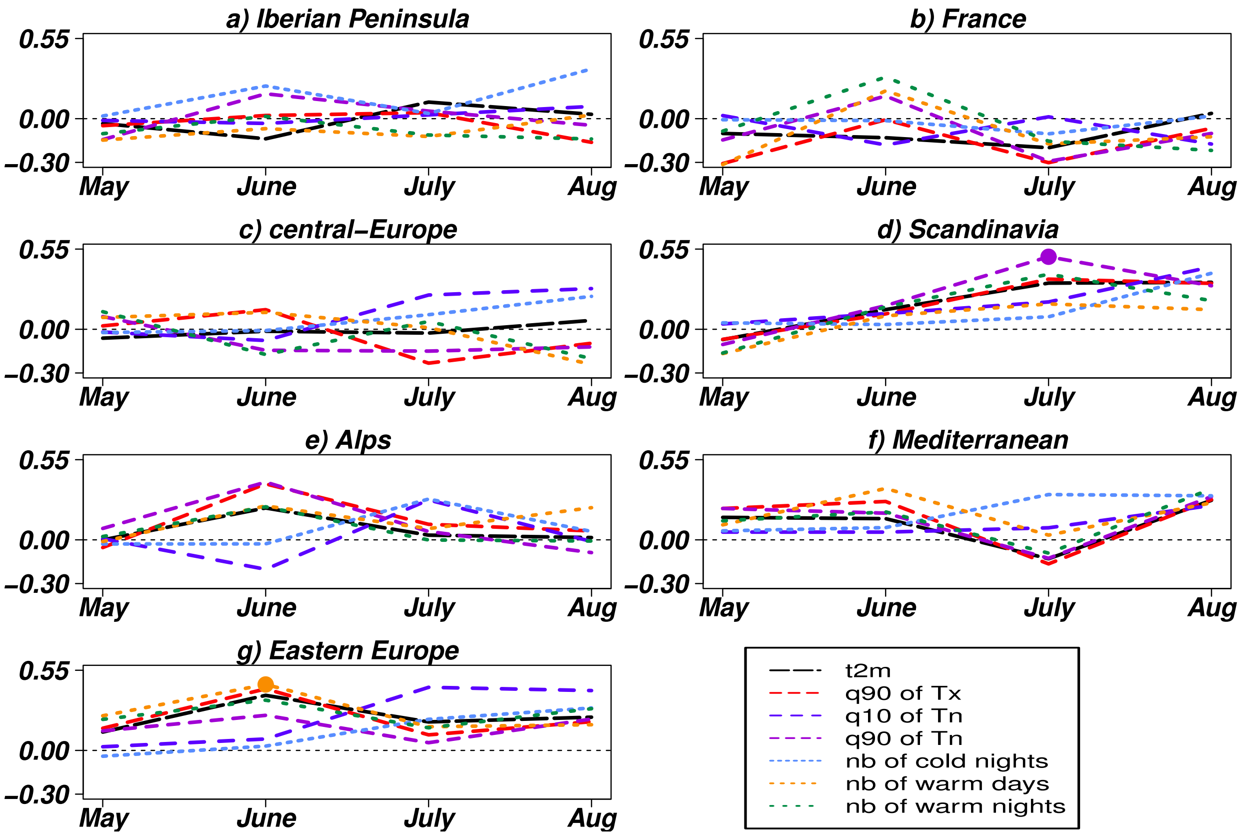

Fig. 6: **a** Difference of correlation between the INIT and CLIM experiments for the temperature variables averaged in the Iberian Peninsula region (10°W3°E-36°N44°N). **b** Same as **a**, but for France (5°W5°E-44°N50°N), **c** central-Europe (2°W16°E-48°N55°N), **d** Scandinavia (5°E30°E-55°N70°N), **e** the Alps (5°E15°E-44°N48°N), **f** the Mediterranean area (3°E25°E-36°N44°N) and **g** Eastern Europe (6°E30°E-44°N55°N).

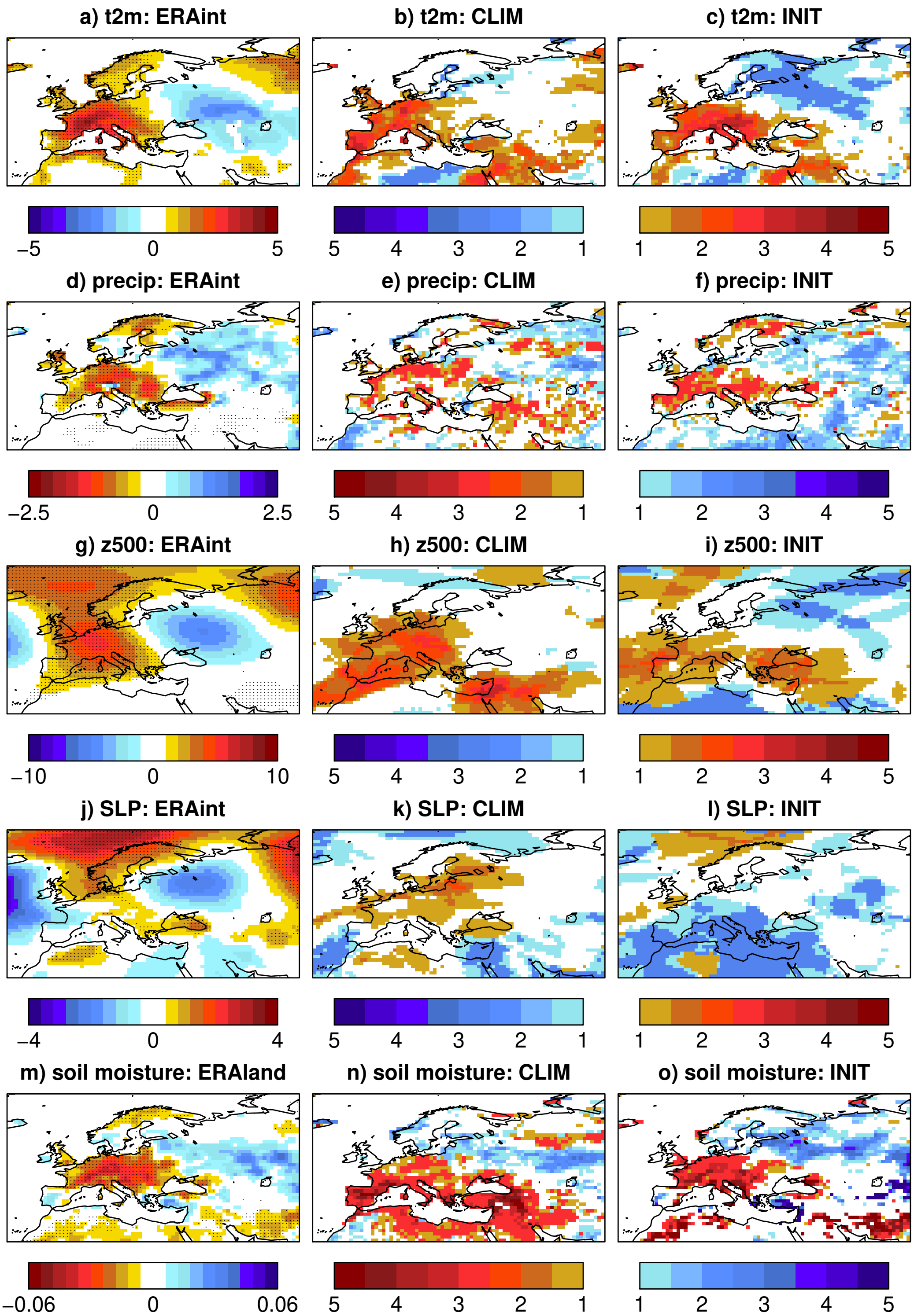

Fig. 7: **a** Observed anomalies of t2m for 2003 JJA (one-month lead time) mean (K). The dots indicate the area where the anomaly is in the upper quintile (estimated over 1981-2010). **b** Odds in CLIM for t2m. The odds are the ratio between the probability for the anomalies to be in the upper quintile, the interquintile range or the lower quintile and with the climatological probability of these three categories (20%, 60% and 20%, respectively). Each point is attributed to the category corresponding to the highest odds ratio. If the point is attributed to the interquintile range or if there is no category assigned (the categories with two highest odds ratio have an equal value) the point is drawn in white. If the point is attributed to the lower/upper quintile category, the corresponding odds ratio is plotted with the left/right color scale. **c** Same as **b**, but for INIT. **d** Observed anomalies of precipitation for 2003 JJA mean (mm/day). The dots indicate the area where the anomaly is in the lower quintile for the 1981-2010 period. **e** same as **b**, but for precipitation. **f** Same as **c**, but for precipitation. **g**, **h**, **i** same as **a**, **b**, **c**, but for geopotential height at 500 hPa (m). **j**, **k**, **l** same as **a**, **b**, **c** but for monthly mean of 6 hourly SLP (hPa). **m**, **n**, **o** same as **d**, **e**, **f** but for the vertically integrated volume fraction of water in soil (m^3/m^3).

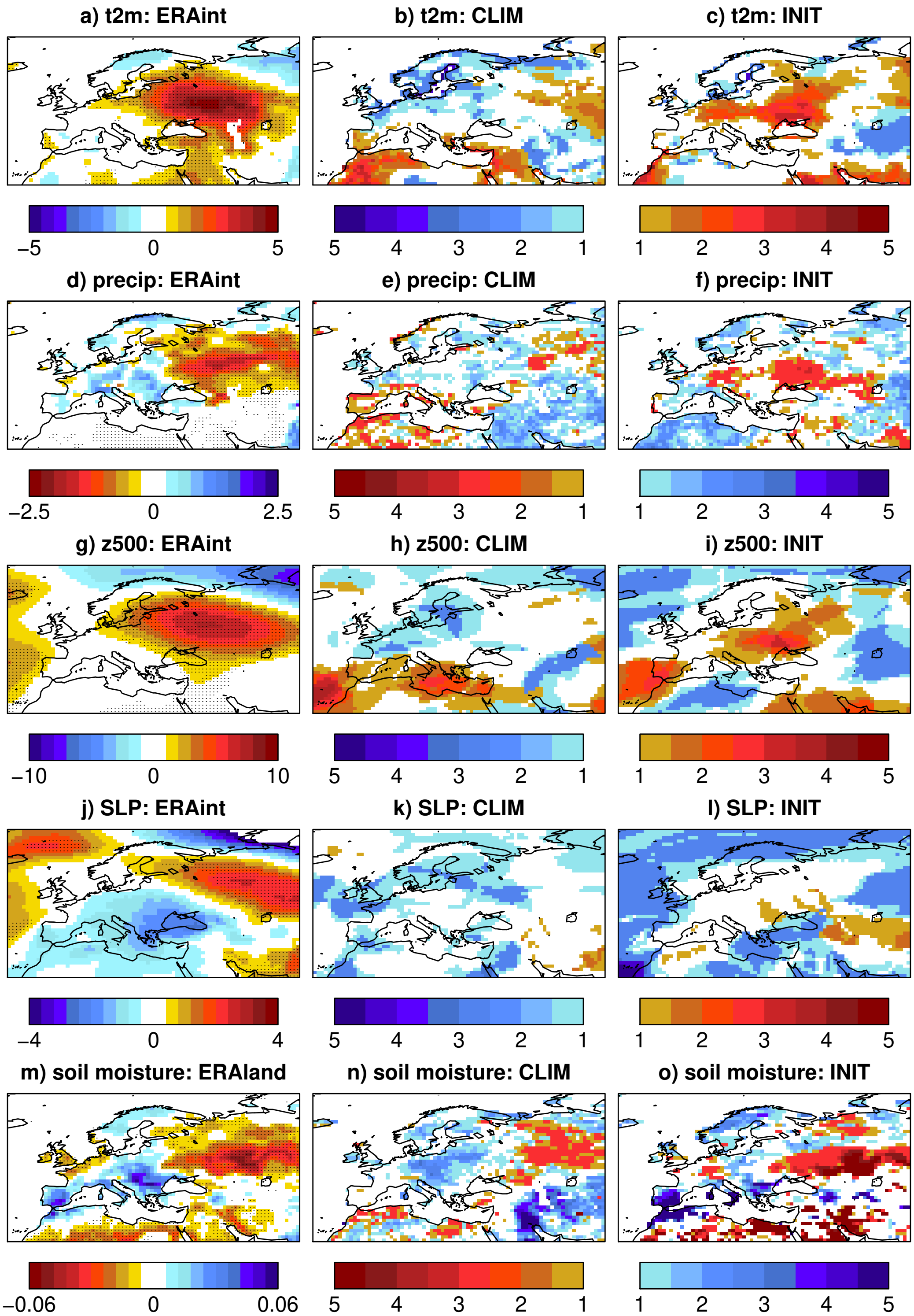

Fig. 8: Same as **Fig. 7**, but for JJA (one-month lead time) 2010.

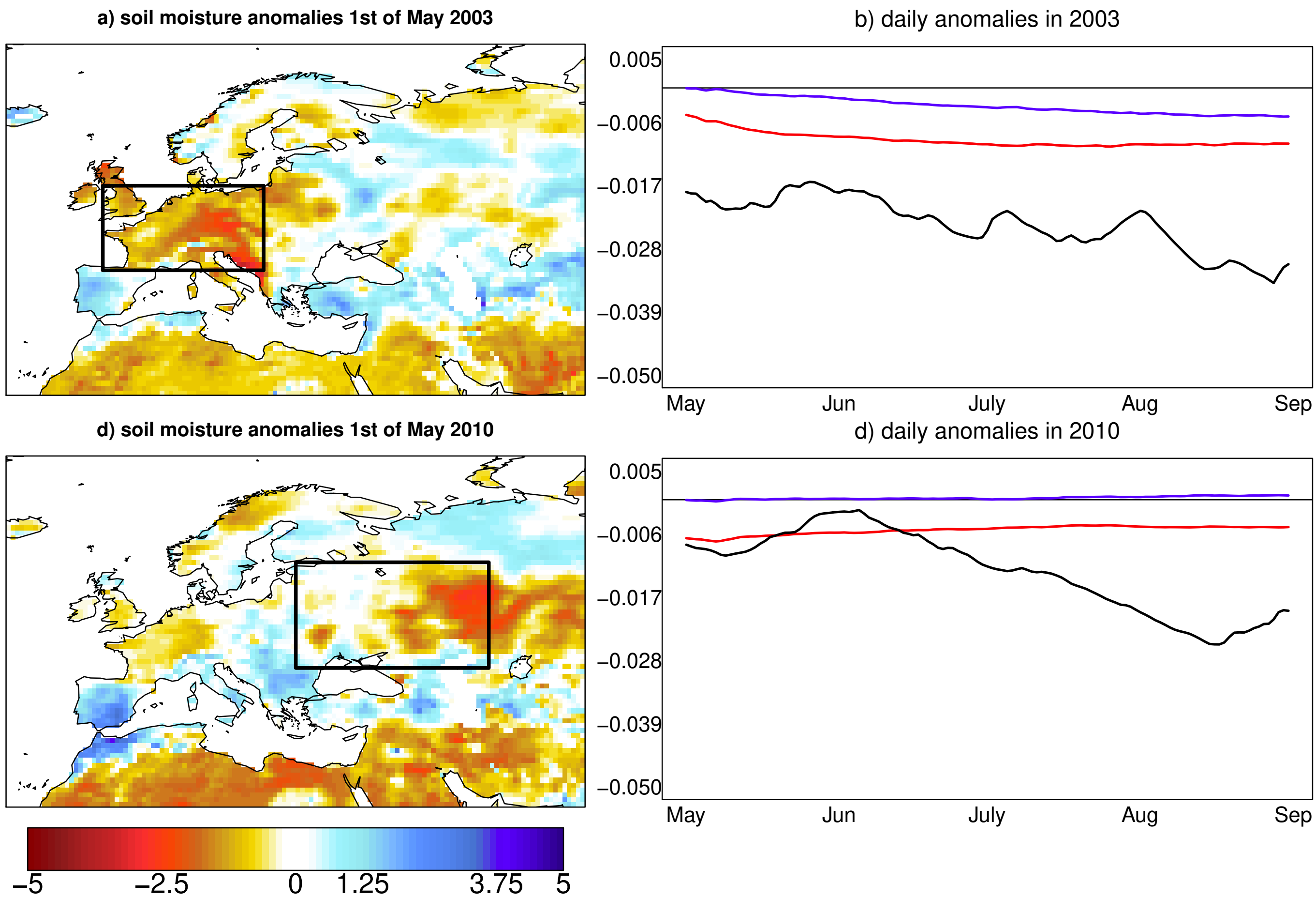

Fig. 9: **a** Standardized anomalies with respect to the daily climatology computed over 1981-2010 of ERA-Land for May 1st 2003. **b** Evolution of the daily anomalies of summer 2003 averaged in the black box of **a** (5W20E-43N55N) in black for ERA-Land, in blue for the ensemble mean of CLIM and in red for the ensemble mean of INIT. **c** Same as **a** but for May 1st 2010. **d** Same as **b** but for the box drawn on **c** (25E55E-45N60N) during summer 2010. For all the panel the unit is m^3/m^3.

Electronic Supplementary Material

Click here to access/download
Electronic Supplementary Material
sup.pdf