Multi-model assessment of the impact of soil moisture initialization on mid-latitude summer predictability

C. Ardilouze[1*], L. Batté[1], F. Bunzel[2], D. Decremer[3], M. Déqué[1], F.J. Doblas-Reyes[4,6], H. Douville[1], D. Fereday[5], V. Guemas[1,4], C. MacLachlan[5], W. Müller[2], C. Prodhomme[4]

1- CNRM-Météo France, Toulouse, France
2- Max Planck Institute for Meteorology, Hamburg, Germany
3- European Center for Medium range Weather Forecasts, Reading, United Kingdom
4- BSC-CNS, Barcelona, Spain
5- Met Office Hadley Centre, Exeter, United Kingdom
6- ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

* Corresponding Author:
Constantin Ardilouze (constantin.ardilouze@meteo.fr), phone: +33(0) 561079912

Abstract:
Land surface initial conditions have been recognized as a potential source of predictability in sub-seasonal to seasonal forecast systems, at least for near-surface air temperature prediction over the mid-latitude continents. Yet, few studies have systematically explored such an influence over a sufficient hindcast period and in a multi-model framework to produce a robust quantitative assessment. Here, a dedicated set of twin experiments has been carried out with boreal summer retrospective forecasts over the 1992-2010 period performed by five different global coupled ocean-atmosphere models. The impact of a realistic versus climatological soil moisture initialization is assessed in two regions with high potential previously identified as hotspots of land-atmosphere coupling, namely the North American Great Plains and South-Eastern Europe. Over the latter region, temperature predictions show a significant improvement, especially over the Balkans. Forecast systems better simulate the warmest summers if they follow pronounced dry initial anomalies. It is hypothesized that models manage to capture a positive feedback between high temperature and low soil moisture content prone to dominate over other processes during the warmest summers in this region. Over the Great Plains, however, improving the soil moisture initialization does not lead to any robust gain of forecast quality for near-surface temperature. It is suggested that models biases prevent the forecast systems from making the most of the improved initial conditions.

Keywords:
Land-surface initialization;seasonal forecasting;land-atmosphere coupling;multi-model;ensemble forecast

Acknowledgement :

The authors thank Jeff Knight (Met Office Hadley Centre) for his constructive comments on earlier versions of this manuscript. The research leading to these results received funding from the European Union Seventh Framework Programme (FP7/2007-2013) SPECS project (grant agreement number 308378) and H2020 Framework Programme IMPREX project (grant agreement number 641811). Constantin Ardilouze was also supported by the BSC Centro de Excelencia Severo Ochoa Programme.

1 Introduction

Human activities are affected by climate-dependent factors, such as energy demand, crop yield or disease risk management. This raises a growing demand for reliable and accurate sub-seasonal to seasonal forecasts of temperature and precipitation (Challinor et al. 2005, García-Morales et al. 2007, Thompson et al. 2006). Atmospheric predictability on these timescales is mainly driven by the coupling between the atmosphere and slowly-evolving components of the Earth system, such as the ocean, sea ice and land surfaces (Doblas-Reyes et al. 2013). Even if tropical oceans provide the major source of global interannual variability through sea surface temperature anomalies related to the El Niño Southern Oscillation (ENSO) phenomenon (Saha et al. 2006, Stockdale et al. 2011), both observational and numerical studies have highlighted the significant imprint of the continental surfaces on the climate system and their potential or effective contribution to mid-latitude sub-seasonal to seasonal predictability, particularly for near-surface temperature (T2M) and precipitation. Among these components, snowpack (Dutra et al. 2011) and soil moisture anomalies (Seneviratne et al. 2010; Seneviratne et al. 2013) have been the most investigated since they strongly affect the land surface energy budget and, hence, the energy fluxes between the surface and the atmospheric boundary layer (Hirschi et al. 2011). Land surface models (LSM), which have improved steadily in the past three decades, together with increasing computational resources have allowed for more thorough studies and a better understanding of the soil moisture and snow influence on the atmosphere at multiple spatio-temporal scales (Douville, 2010). A realistic snowpack initialization has been shown to be useful both in boreal fall (e.g. Orsolini et al. 2013) and spring (e.g. Peings et al. 2011), when the interannual variability of the Northern Hemisphere snow cover is relatively strong and has a large impact on the surface energy budget given the available incoming solar radiation even at high latitudes.

For summer predictions, the focus was mainly on soil moisture and its influence on near-surface temperature and precipitation mainly via evapotranspiration. It has been demonstrated that soil moisture content controls the evapotranspiration in regions with a semi-arid climate (“soil moisture-limited regime”). In wet regions, the evapotranspiration rate mainly depends on atmospheric control and not on soil water content (“energy-limited regime”). In the former, the evaporative fraction modulated by soil moisture affects both the local water cycle (Dirmeyer 2006) and the surface energy balance, and hence temperature and precipitation (Dirmeyer et al. 2014, Koster et al. 2004b, Seneviratne et al. 2010). Additionally, soil moisture memory has proven to last up to several months in some cases (Seneviratne et al. 2006, Orth and Seneviratne, 2012, Hagemann and Stacke, 2015). Due to these characteristics, extreme warm events can be triggered or at least amplified by dry soil initial conditions in terms of magnitude (Fischer et al. 2007, Hirschi et al. 2011, Whan et al. 2015) and persistence (Lyon and Dole, 1995, Lorenz et al. 2010).

Previous studies have highlighted a number of “hotspots” where seasonal prediction skill can be increased by realistic soil moisture initialization since they combine intense land-atmosphere coupling processes with strong soil moisture persistence (Koster et al. 2004, Seneviratne et al. 2006, Dirmeyer et al. 2011). The North-American Great Plains and the region between the Danube basin and the Mediterranean are often identified as belonging to these hotspots. Our study will focus mainly on these two regions, namely the Southern Great Plains (SGP) and the Balkan region (BKS). BKS and SGP boundaries are defined in Table 1 and highlighted by green boxes in Figure 2. The second phase of the Global Land-Atmosphere Coupling Experiment (GLACE-2, Koster et al. 2011), which consisted in a multi-model forecast quality assessment, showed that a realistic soil moisture initialization provides significantly improved skill for air temperature forecast up to two months ahead over the North American continent. More recent studies confirmed this positive impact up to seasonal timescales (Materia et al. 2014, Prodhomme et al. 2015). Prodhomme et al. (2015) described the benefits of soil initialization for the quality of temperature predictions over large parts of Eastern Europe up to four month forecast time. They could only achieve a successful hindcast of the summer of 2010 extreme heat over western Russia with a realistic soil moisture initialization.

This study aims at exploring to what extent previous results are robust across a variety of forecast systems. Its originality lies in being the first multi-model assessment of soil moisture initialization impact on atmospheric predictability on seasonal timescales with ocean-atmosphere coupled models over a nearly two-decade period. We use a highly comprehensive database of seasonal prediction experiments produced within the framework of the European FP7 SPECS (Seasonal-to-decadal climate Prediction for the improvement of European Climate Services) project and covering the 1992 to 2010 period. The following section describes the forecast systems and datasets used to perform the experiments and to assess their output. Section 3 focuses on the model systematic errors and on the predictive skill related to soil moisture initialization. Section 4 explains how the models respond to the soil moisture initialization over the two regions of interest (BKS and SGP) and precedes the discussion and conclusions to this study in section 5.

2 Experimental design and methodology

2.1 Overview of the experiments

Five forecast systems (Table 2) have been used to perform twin sets of boreal summer season hindcasts over the 1992-2010 period. These simulations start at the beginning of May and span 4 months, including the June-August trimester (JJA).

For each system, the twin experiments consist of one control and one sensitivity experiment differing only by their land-surface initialization. The former is initialized with climatological surface fields while the latter is performed with initial conditions closer to observed interannual variations in soil moisture (hereafter ‘realistic’ initialization). The different strategies adopted to derive these initial conditions are detailed in the following subsection. All the experiments consist of 10-member ensemble simulations. The methods applied for the generation of the ensembles as well as the experimental design are summarized in Table 2.

The five twin experiments allow the comparison of two fifty-member grand ensembles. They are named ALL-CLIM and ALL-INIT hereafter. We refer similarly to CLIM and INIT experiments when discussing individual forecast system results. The multi model approach diminishes the impact of individual model errors and thus leads to more reliable seasonal predictions (Palmer et al. 2004, Hagedorn et al. 2005).

2.2 Land-surface initial conditions

Different methods were used to generate the so-called ‘realistic’ initial conditions of soil moisture used in the ALL-INIT ensemble:

- Atmosphere-Ocean General Circulation Model (AOGCM) simulation relaxed towards reanalyses:

For MPI-ESM, divergence, vorticity, temperature and surface pressure were assimilated into the atmospheric component (ECHAM6) and temperature, salinity and sea-ice concentration into the ocean component (MPIOM). For data assimilation, ERA-Interim (hereafter ERAI, Dee et al. 2011) is used for the atmosphere, ORAS4 for the ocean and NSIDC/Bootstrap for sea ice. No assimilation was performed in the LSM (JSBACH).

- Standalone LSM simulation forced by atmospheric reanalysis

This method was applied for the LSM component (JULES) of HADGEM3 applying WFDEI atmospheric forcing.

- Land surface reanalysis dataset

The last three models used the pre-existing daily dataset of land surface pseudo-reanalysis ERA-Interim/Land (hereafter ERALand, Balsamo et al. 2013). It results from a stand-alone run of the HTESSEL LSM, forced by ERA-Interim atmospheric fields and bias-corrected precipitation using the GPCP monthly climatology (Huffman et al. 2009) for precipitation.
The two AOGCMs using the HTESSEL land component (namely EC-Earth and ECMWF System 4) were initialized with May the 1st ERALand reanalyses, horizontally interpolated over the model grid. For CNRM-CM5, ERALand data was additionally interpolated onto the SURFEX vertical soil layers (which differ from the ERALand vertical distribution), while preserving the soil wetness index for each soil layer (Boisserie et al. 2015).

These initial conditions were computed for the 1st of May start dates of each of the nineteen years of the seasonal re-forecast experiments, e.g. 1992 through 2010. The land-surface initial conditions for each of the five CLIM ensembles are obtained by averaging the initial conditions for the 1st of May from the corresponding INIT initial conditions.

Snow initial conditions are also considered realistic with the described techniques to generate INIT initial conditions. However, different choices have been made for CLIM : snow fields were averaged for BSC-CLIM and MF-CLIM, similarly to soil moisture, while their yearly variability was preserved in the other three CLIM simulations. This experimental set-up inhomogeneity might affect the conclusions since significant snow-atmosphere coupling occurs during and after snowmelt over snow transition zones of the Northern hemisphere (Xu and Dirmeyer, 2011). However, this impact is considered limited in our regions of interest where the influence of snow in boreal summer is lower than in other seasons.

2.3 Reference data and forecast quality assessment

The monthly-mean precipitation observations used are the Global Precipitation Climatology Center (GPCC) (Schneider et al. 2008) gridded gauge analysis products, available at a 1º resolution, while monthly mean T2M reference data are provided by the CRU TS v.3.23 analysis (Harris et al. 2010). The ERA-Interim (Dee et al. 2011) dataset is used for daily averaged two-meter temperature as well as daily-mean precipitation and daily maximum and minimum temperature (Tmax and Tmin, respectively) references as no other global daily precipitation or temperature data spans the full hindcast period. Both observational and model outputs were regridded onto a T85 Gaussian grid and only land surface grid points are considered for score computations.

The bias is computed as the mean difference between the model and the observed climatologies. We assume that the individual model drift does not depend on the start dates, meaning that no distinction between the different hindcast years is required to compute the model climatologies. Removing the bias is equivalent to considering observed and re-forecast anomalies relative to their respective climatologies. Thus, the skill of the simulation is evaluated by means of the correlation coefficient (r) between the predicted and the observed anomalies of a given variable. The difference r_{INIT} minus r_{CLIM} is computed at every grid point and then mapped to highlight regions impacted by the land-surface initialization.

A confidence interval for correlations is provided by a 2-sided 95% confidence level t-test. The assessment of correlation differences between the CLIM and INIT simulations must take into account the degree of dependence between the two experiments as both are run over the same time period. To that end, the Hotelling-Williams t-test is computed (Steiger, 1980).
In addition to correlation, the comparison of the root mean square error (RMSE) of each experiment through the root mean square skill score (RMSSS) helps in assessing how the soil moisture initialization affects the interannual departure from observations. The RMSSS, contrary to the RMSE, is positively-oriented so that a negative (positive) score means the INIT ensemble has lower (higher) skill than the CLIM ensemble.

RMSSS=1 - RMSE(INIT)/RMSE(CLIM)

The RMSSS is considered to be significantly different from 0 if RMSE(INIT) is not included into the confidence interval of RMSE(CLIM) computed through a 95% confidence level chi2 test.

3 Results

3.1 Bias analysis

A preliminary analysis of the surface bias can provide insight on both individual and multi-model climatological limitations, as well as an overview of the ensemble consistency. Biases are estimated as the forecast-time dependent difference (temperature) or ratio (precipitation) between ensemble mean and reference data. The bias analysis can also contribute to understanding model differences in forecast skill.

This analysis reveals almost indistinguishable differences in pattern and amplitude between the CLIM (Fig. S1) and INIT (Fig. 1) experiments for both T2M and precipitation fields. As expected, soil initialization used in these experiments does not alter the model climate in the seasonal re-forecasts.

JJA precipitation and temperature biases from individual models show relatively inconsistent patterns over Eurasia (Fig. 1). Over Eastern Siberia, the five models overestimate the amount of rainfall, although the very limited number of rain gauges available in that region (Fig. S2b) suggests that reference data may have a substantial level of uncertainty. Biases partly cancel out in the multi-model over Central Europe, but a notable dry and warm bias over the Steppes east of the Caspian Sea, and a strong wet bias over Eastern Russia and the Iberian Peninsula tend to stand out of the multi-model ensemble average. For the latter region as well as for the Steppes, since the observed amount of JJA precipitation is very low (Fig. S2a), small differences between these values can result in a strong relative bias. Over North America, in contrast, all models present fairly similar patterns of wet and slightly cold bias over Alaska and pronounced dry and warm bias over the Central Plains. This warm bias was also found in many models of the Coupled Model Intercomparison Project Phase 5 (CMIP5) and would stem from excessive incoming shortwave radiation combined to a lack of evaporative fraction (Cheruy et al. 2014). We will discuss further how this could impact the seasonal forecast quality with respect to soil moisture initialization in section 4. This preliminary analysis confirms the interest of the multi-model approach since the individual model climatologies show a number of similarities with each other and the multi-model biases are not excessively influenced by any one of the contributing models.

Soil moisture biases are far more difficult to assess due to the scarcity of in-situ observations to be assimilated in any soil moisture reanalysis. Furthermore, remote sensing can only reflect the superficial soil layer state, without taking into account the deeper root-layer soil moisture, and do not necessarily provide a sufficient sampling for deriving reliable monthly mean values. Root-zone soil moisture controls the plants' transpiration and thereby plays a major influence on total evapotranspiration in vegetated areas. Finally, the limited knowledge of soil depth and global scale physical processes at stake leads to a large variety of land surface modelling techniques and parameters, which somewhat hampers the inter-model comparison of soil moisture as well as the comparison of simulated versus observed data. However, a straightforward way to gain insight on the simulated soil moisture is to consider the total soil water content of the entire soil depth averaged over specific regions for each model and to assess the relative evolution in time of its daily climatology. This evolution can be compared with that of ERALand.The assessment of the mean soil moisture over the SGP and BKS regions (Fig. S3) shows that the soil dries faster than the reference for four models out of the five analysed over both regions, although none of them shows any obvious abnormal evolution. However, for the SGP region, according to ERALand, there is little evolution in the soil water content during the first third of the forecast period, followed by a drying phase starting in mid-June. Only one forecast system evolves similarly to ERALand during the steady stage but retains somewhat too much water afterwards. The drying tendency occurs too early for the other systems. This suggests that in addition to the JJA precipitation bias discussed earlier, these models simulate either a deficit of rain in May and early June, or an excessive evapotranspiration, or both simultaneously. These results suggest that understanding not just the model bias, but also the forecast drift is essential to have a chance to correctly interpret the quality of a forecast system.

3.2 Summer skill over boreal mid-latitudes

Figure 2 shows the JJA seasonal anomaly correlations of ALL-CLIM and ALL-INIT for near surface temperature. Large parts of continents south of 50º N show significant T2M correlation in all the experiments. This feature could be attributed to the correct representation of ENSO teleconnections by the models, but also to the warming trend over the recent period, especially over Europe (Doblas-Reyes et al. 2013). These hypotheses are assessed by computing for each grid point the temporal correlation of JJA simulated T2M with respectively JJA observed T2M averaged over the Niño 3.4 region defined in Table 1 and JJA observed global T2M averaged over land. ENSO teleconnections, if present, do not seem to impact greatly the skill south of 50°N (Fig. S4a). Observations suggest that the models over-estimate the link between Niño 3.4 and Eastern Canada T2M. However, T2M over Eastern Canada, Southern Greenland and the Middle-East is significantly correlated with global T2M, with correlation values of similar amplitude to the hindcast skill (Fig. S4b). This is supported by observations over the same period (not shown) in addition to the longer 1979-2013 period (Fig S4d). On the contrary, the interannual simulated T2M over BKS and SGP is not significantly correlated to the global T2M during the hindcast period, meaning that the global warming trend does not account for most of the skill found over these regions. This is further confirmed by removing a linear trend from both experimental and reference data, which does not affect greatly the correlation pattern nor its values (Fig. 3).

An overall increase of skill is found over Europe in the T2M correlation differences between INIT and CLIM (Fig. 4a). ALL-INIT is only outperformed by ALL-CLIM over the Iberian Peninsula, although not significantly, whereas the effect is either positive or neutral anywhere else. This skill enhancement is significant over Scandinavia, Ukraine and most of the Balkans peninsula. The assessment of the RMSSS computed with respect to the CLIM experiments (Fig. 4b) confirms these improvements. Over North America, soil initialization leads to a limited score improvement. The model even exhibits a significant decrease in skill over Central Canada. However, it should be kept in mind that this region has a poor temperature skill in the first place. Such upper latitude regions are considered to be in an energy-limited regime where the evaporative fraction of the surface energy budget is not controlled by soil moisture. Moreover, snow melting - soil freezing interactions within the HTESSEL model seem to generate too much and early runoff, which could have implications on soil moisture storage after the melting season (E. Dutra, personal communication). If this were the case, the May 1st land surface initial conditions derived from ERALand, which are used for three models out of five, could then be locally unsuitable.

The multi-model ALL-CLIM (Fig. S5) and ALL-INIT (Fig. 5) display almost no skill for precipitation, except for Western North America. This could be related to the great influence of the ENSO activity on the local atmospheric circulation, although evidence of this teleconnection has been found mainly during the winter season (Quan et al. 2006, Yoon et al. 2015). This skill pattern should be considered with caution as the region receives limited amounts of precipitation during summer (Fig. S2), implying that correlation values may be influenced by extremely small differences in precipitation amounts. The difference of skill computed between INIT and CLIM for precipitation (Fig. 6a) is quite patchy over the Northern Hemisphere mid-latitudes. Moreover, the Iberian Peninsula, which results as one of the very few regions where the increase of correlation leads to significant predictive skill, receives limited amounts of rain in summer as mentioned earlier. Hence, small changes in simulated precipitation may greatly impact correlation values. The negligible improvement of RMSSS tends to support this hypothesis (Fig. 6b) although models have already exhibited skill for precipitation over this region in past coordinated experiments (Diez et al. 2005)

The results described above suggest that the BKS region is one of the most positively impacted by soil moisture initialization in terms of predictive skill for temperature. Furthermore, the multi-model ensemble displays relatively weak temperature and precipitation biases over BKS (Fig. 1), although one should keep in mind that some of the contributing models have pronounced biases of opposite signs. On the other hand, SGP was previously identified as a region with a high potential for seasonal predictability due to its sensitivity to soil moisture. This set of experiments did not show any skill increase over SGP associated to improved land surface initialization. A possible reason for this lack of sensitivity may be related to the common dry and warm bias of the five individual models.

The next section of this paper therefore aims at providing insights on the reasons for such contrasted results over SGP and BKS. This is achieved by comparing the relationship for these two regions between the realistic initial soil moisture and the subsequent simulation of temperature and precipitation during the hindcast period. The next section intends to shed light on the link between the multi-model skill and the systematic error analysed so far.

4 Preliminary understanding of the models response to realistic soil moisture initialization

This section focuses on the two previously defined regions, namely BKS and SGP, to better understand the response of seasonal predictions to soil moisture initial conditions.

The standard deviations of simulated JJA T2M anomalies over BKS and SGP are enhanced with realistic initial conditions, especially over SGP (Table 3) confirming the sensitivity of the models' response to soil moisture conditions in summer. They also get closer to the observed standard deviation value in each region. To assess this sensitivity more closely, temporal correlations between detrended ERALand total soil water content at start dates and observed or simulated JJA T2M have been computed (Table 4). The time series of these anomalies are represented on Fig. 7 where the blue and red envelopes feature the temperature anomaly spread between individual model ensemble means for respectively CLIM and INIT simulations. In the following sections, both regions are analyzed separately.

4.1 SGP region

Over SGP, unlike in the observations, the simulated JJA T2M is significantly anticorrelated with the initial soil moisture for the five models. This is well illustrated in Fig. 7 where prevailing dry initial conditions in the early 2000's coincide with warm simulated summers according to ALL-INIT, which does not match observations. This implies that models tend to overestimate either the land-atmosphere coupling processes or their contribution among other factors that could explain interannual near-surface temperature variability.

In order to provide further insight on the models' response, 31-day running means of daily-averaged simulated fields are correlated with the initial soil water content on May 1st over the re-forecast period. Results for temperature, precipitation and soil moisture according to the forecast time throughout the four months of simulation are presented in Fig. 8. The initial soil moisture is very persistent in the simulations, with a correlation coefficient close to 1 and barely decreasing throughout the summer. This persistence is also present in the reference soil moisture data, although less pronounced. This implies that initial dry (wet) anomalies in the models rarely turn into wet (dry) anomalies during the summer, while such changes in sign are marginally more likely in the reference data. When considering the INIT-ALL ensemble, initial soil moisture is correlated with both simulated precipitation and Tmax over SGP from the beginning of the period. This correlation grows stronger in time for a few days before reaching a plateau for Tmax at about 0.9, i.e. about 80% of variance explained, while it is about 0.6, about 35% of variance explained, right from the start for precipitation and persists throughout the whole summer. On the other hand, in the reference data, the correlations are of the same sign as in the simulations but they are not significant and tend to zero after the first month for temperature. This suggests a larger amount of intraseasonal variability in the observational dataset that is not reproduced by the models. The latter tend to simulate a smoother evolution of the variables.

Based on Seneviratne et al. (2010), the following mechanism could explain the simulated tendencies. Years with initial dry soils lead to reduced evapotranspiration, which inhibits precipitation and in turn increases soil dryness. As soil moisture decreases due to this positive feedback loop, it fails to respond to the evaporative demand, permitting the role of the sensible heat flux to grow in the surface energy budget, at the expense of the latent heat flux. This leads to higher daily Tmax, which triggers another positive feedback loop by increasing evaporative demand and thus reducing soil moisture content. At night, however, this mechanism is weakened by the development of a stable boundary layer decoupling the land surface from the atmosphere aloft. Based on an observational campaign over Kansas, Ha and Mahrt (2003) highlighted the development of a surface inversion primarily due to radiative cooling when turbulent fluxes collapse in the early evening. This could explain why simulated Tmin is not significantly anticorrelated to initial soil moisture during the first days, unlike Tmax. However, the anticorrelation becomes significant about two weeks later than for Tmax, ultimately reaching values comparable to those of Tmax. This feature of INIT-ALL is supported by three individual models but not by observations. The Tmin values are generally reached at the end of the night, when the diurnal soil moisture-temperature feedback loop is still off. This lagged co-variability of Tmin and soil moisture in the simulations could result from a progressive overall warming of the surface-boundary layer system. Depending on the stability regime of the nocturnal boundary layer over grassland (Mahrt 1999), turbulence due to wind shear at the top of the stable layer may redistribute downward the heat stored in the residual layer aloft. This mechanism competes with the suppression of turbulence by thermodynamic stability that favours nocturnal radiative cooling of the surface (McNider et al. 2010). However, the representation of such complex subgrid scale phenomena in large-scale GCMs is likely to be inadequate and a source of model error.

It is beyond the scope of this study to determine the reasons for the discrepancies between the coupled model simulations and the observations. However, the similarities between forecast systems in terms of correlation between initial soil moisture and summer variables likely relate to their similarities in terms of biases. If the simulated climate over SGP is too dry, as suggested in section 3.1, the models' evapotranspiration remains strongly controlled by soil moisture but its absolute value and variations are too small to impact climate variability (Seneviratne et al. 2010). An additional explanation can be provided by the development of the biases over SGP during the forecast (Fig. 9). The simulated climatologies look smoother than for the reference data because they result from a ten-member averaging. The comparison of the precipitation daily climatologies (Fig. 9a) show that for four models out of five, the deficit of daily rainfall establishes at the beginning of June and persists throughout summer. On the contrary, the Tmax biases (Fig. 9b) develop at a different rate and reach different amplitudes among forecast systems. Nonetheless, all of them switch from neutral or cold biases during the first month to warm by the end of summer. In some cases, this warm systematic error starts to grow up to forty days after the appearance of the precipitation bias. The contrast between simultaneous precipitation biases and asynchronous temperature biases supports, albeit without confirming it, the hypothesis that the majority of models have a limited capacity to represent accurate precipitation in summer over this region. A number of studies suggest that summer precipitation regime in that region has particular features that makes it very challenging to model properly. These particularities are the atypical diurnal cycle of precipitation with a nocturnal maximum in summer (Klein et al. 2006), the meso-scale systems that account for much of the warm season precipitation (Mearns et al. 2012), or the atmospheric low-level jet that substantially contributes to the moisture budget of this region and influences nocturnal convection triggering (Bellprat et al., 2016). If confirmed, this dry bias would trigger the excessive soil drying and its reduced ability to respond to the evaporative demand, eventually leading to the aforementioned feedback loop with the atmosphere that amplifies temperature biases.

Tackling this bias issue seems to be a prerequisite for the forecast systems to make the most out of the soil moisture initial conditions and thus to improve the prediction skill over SGP Nonetheless, a dedicated study would be required to disentangle the role of the biases from that of potential shortcomings in the simulated surface processes.

4.2 BKS region

Over BKS, the two hottest summers of the period, namely 2003 and 2007, had both drier initial soil moisture conditions than average. These are correctly predicted only with the INIT ensemble (Fig. 7). Similar results are found with the cooler than average summers of 1996, 1997 and 2006 despite wet initial anomalies of relatively low amplitude. Observations, as well as the INIT multi-model ensemble, show significant correlation between the initial soil moisture and summer T2M for the BKS region (Table 4). Yet, when considering the individual forecast systems, no relationship could be established between this correlation and the gain of skill permitted by land surface initialization over BKS (as shown in Figure S6). Hence, the increase in T2M correlation related to land surface initialization in this region does not result from local linear processes - such as persistence - derived from initial soil moisture anomalies.

A correlation analysis similar to that performed for the SGP region (Fig. 8) is displayed for the BKS region on Figure 10. It shows very distinct correlation features among forecast systems. The different systems do not highlight any common process that would help explaining the gain of skill in this region. It is likely that a wider range of processes related to soil moisture coupling with the atmosphere with contradictory effects are at play. As opposed to the SGP region, the BKS region is characterized by a steep topography and the proximity of the sea. Based on regional meso-scale simulations over France, Stéfanon et al. (2014) highlighted different soil moisture-temperature responses over low-elevation plains, mountains and coastal regions during heat waves. Over plains, the dominant mechanism is consistent with the positive feedback loop described earlier. Over mountains, on the other hand, enhanced heat fluxes due to dry anomalies can reinforce upslope winds and favor convective precipitation with a subsequent cooling effect, hence a negative feedback. Dry anomalies can also enhance the gradient of diurnal near surface temperature between the air above coastal land and sea. This could trigger anomalous moist advection from the sea through the breeze process, resulting in a negative feedback on T2M over land. These last two meso-scale mechanisms may compete with the first one over BKS, in spite of the relatively low resolution of the models used. Since the five forecast systems have quite distinct spatial resolutions, it is likely that the impact of these mesoscale processes, if represented, differs greatly.

What could therefore explain the successful prediction of the hottest summers of 2003 and 2007 conditioned to realistic soil moisture initialization, as indicated by Fig.7 ? The study from Conil et al. (2008) based on a single AGCM showed that the benefit of a realistic land surface initialization for summer predictions appears when widespread and strong soil moisture anomalies are observed at the beginning of the season. This result was found over typical land-atmosphere coupling hotspots, namely central North America and Eastern Europe. The present work tends to generalize this result for the latter region when initial anomalies are negative. Furthermore, Quesada et al. (2012) showed observational evidence of an asymmetry in hot day predictability over Europe. Wet springs lead to a reduced number of hot summer days regardless of the dominant large-scale weather pattern during summer, while dry springs precede a greater number of hot days only if anticyclonic weather types prevail during the summer. From these studies and our results, we can infer that initializing soil moisture realistically is a necessary condition for models to predict abnormally warm summers, but not a sufficient one. We hypothesize here that in the case of pronounced dry initial anomalies over the BKS region, forecast systems agree on the dominant process of positive feedback between low soil moisture, reduced fraction of latent heat flux and warmer temperature. However, as mentioned earlier, verifying this statement would require additional studies with a dedicated experiment framework.

5 Conclusion and Discussions

A set of multi-model seasonal prediction experiments aiming at assessing the impact of land surface initial conditions on boreal summer predictability has been carried out in the framework of the FP7-SPECS European project. Five distinct global coupled ocean-atmosphere forecast systems were run with 10 members each, initialized on May 1st over the period 1992 to 2010 with climatological soil moisture conditions for the reference experiment, and realistic ones for the sensitivity experiment. For both experiments, the 50 resulting members have been considered together as a large multi-model ensemble. This is the first multi-model experiment assessing the added-value of initializing the land surface in a ‘real’ prediction context, as opposed to potential predictability and/or purely AGCM frameworks. It therefore provides the most robust assessment of land surface initialization impact on boreal summer prediction quality to date. The comparison of precipitation and near surface temperature scores show evidence of an enhanced predictive skill over large parts of Europe for realistically versus climatologically initialized simulations, although mainly for temperature and with a significant increase limited to a few regions. No such conclusion can be drawn for Asia and North America.

Previous studies had identified several mid-latitude regions with a high summer prediction potential a few months in advance, stemming from intense land-atmosphere coupling combined with long-lasting soil moisture memory. Among them, the Balkans proved to actually gain predictability from a more accurate soil moisture initialization, unlike the Southern Great Plains of North America where no improvement was achieved.Over the latter region, the five models show very similar overestimates of the correlation between initial soil moisture anomalies and summer daily maximum temperature (Tmax) and daily mean precipitation with respect to the correlation estimated from reference data. A locked positive feedback settles between dry (wet) soil moisture anomalies leading to increased (decreased) Tmax and precipitation deficit, which favours in turn an increase of the soil moisture anomaly. This overestimated feedback over SGP is likely related to the systematic errors for temperature and precipitation, and in the excessive decrease of soil water content during the early stage of the summer simulated by the majority of forecast systems. Thus, biases appear as potential culprits in the lack of predictive skill enhancement with respect to soil moisture initialization over SGP. Previous studies based on CMIP experiments pointed out at model deficiencies in both cloud physics and evapotranspiration processes that should be addressed over the Great Plains to reduce systematic biases (Cheruy et al. 2014).

For the BKS region, the coupling of soil moisture with temperature and precipitation could be driven by various processes with opposite feedbacks. Nonetheless, for some years with a pronounced dry initial anomaly, summer predictions from distinct models agree on a warm JJA T2M anomaly. It is likely that in the case of dry soil moisture anomalies combined with prevailing anticyclonic weather regimes during summer such as Blocking or Atlantic Low (Quesada et al. 2012), the land-atmosphere coupling processes simulated by different models over BKS converge towards a similar dominant process or feedback loop.

Previous studies suggested a potential remote impact of soil moisture initialization on summer temperature prediction (Van den Hurk et al. 2012, Koster et al. 2014), that could be related to an alteration of the atmospheric circulation either locally or remotely (Fischer et al. 2007). The correlations between JJA T2M averaged over BKS and initial soil moisture computed on every grid point for OBS and INIT (Fig. S7a) do not rule out such a hypothesis, since a few common patterns appear such as high positive correlations over Northern Europe and negative correlations East of the Black Sea. However these patterns are not large or significant enough to conclude on this potential remote influence.

A limitation of this study stems from the discrepancies between experimental protocols for each participating forecast system. For instance, it does not clearly disentangle the potential impact of snowpack initial conditions as two contributors out of five averaged out snow cover parameters in addition to soil moisture parameters to produce climatological initial conditions. According to Xu and Dirmeyer (2011), the snow-atmosphere coupling strength can be considerable during snowmelt and up to several weeks after that, due to the albedo and subsequent soil moisture states. Even if the similarity of the models' response in this study suggests a limited impact in our regions of interest, this pleads for a more careful assessment of snow cover and snow water equivalent in the initial conditions of subseasonal to seasonal summer predictions. The diversity of spatial resolution also hampers the investigation of potential physical processes at play. Furthermore, our study does not take into account the proportion of the total soil water content in models and in the reference data that is prone to imprint the atmosphere at seasonal scale by means of evapotranspiration. A focus on the soil wetness index of the root layer instead of the total soil water content is required to further disentangle the processes involved in the soil-moisture surface climate interplays and the associated predictability. The use of ERALand for soil moisture initialization and as a reference data might be a source of uncertainties since no in-situ nor remote-sensed soil observations are assimilated in this product. Nonetheless, state-of-the-art global remote sensed soil moisture products usually estimate superficial soil wetness. Hirschi et al (2012) pointed out the limitations of a mere extrapolation of observed superficial soil moisture to the root-zone and suggests an assimilation of these data in a land-surface model to obtain a more realistic product. These limitations should be addressed when defining the set-up of the predictability experiment of the Land Surface, Snow and Soil moisture Model Intercomparison Project (LS3MIP, van den Hurk et al. 2016).

In the light of our results, two main topics would require future research and attention in the community. The first one is that of the initialization technique, a potential caveat of this study. The climatology and variability of distinct AOGCM land components may differ greatly because of the diversity of parametrizations and the limited constraints with respect to the atmospheric component. This questions the technique of initializing a model with data derived from another model. However, even if the land initial conditions are computed from an offline simulation of the same LSM that is then used in the coupled model simulation, initial shocks and spin-up may occur due to inconsistencies at the land-atmosphere interface and ultimately degrade the prediction skill. A cleaner initialization would imply to perform either a coupled data assimilation or a coupled nudging towards observational data for each forecast system individually. However, this technique does not explicitly correct the simulated precipitation, which can remain biased and thus lead to an unrealistic soil water content. A correction of precipitation in this case might jeopardize the water balance of the model. Therefore, the best initialization strategy is still an open question, and may very well be model-dependent.

The role of vegetation and land-use on continental climate predictability is the second issue that could be of great interest in future works. Previous studies have demonstrated that the use of interactive vegetation affects precipitation variability (Alessandri and Navarra, 2008) as well as T2M seasonal predictability over the continents (Weiss et al. 2012, Alessandri et al. 2016). The extensive use of irrigation and crop growing practices can affect water fluxes between the soil and the atmosphere. Mueller et al (2015) showed evidence that agricultural intensification - and to a lesser extent increased irrigation - over the past century led to cooler temperature extremes and enhanced rainfall during the growing season in the North American Midwest. These features are not taken into account in the coupled models used in this paper whereas they affect atmospheric observations assimilated in the reference data. The results of the present study plead for a coordinated seasonal prediction effort aiming at enlightening the impact of vegetation and land-use on summer predictive skill over mid-latitudes.

References

Alessandri A, Navarra A (2008) On the coupling between vegetation and rainfall inter-annual anomalies: Possible contributions to seasonal rainfall predictability over land areas. Geophys Res Lett. doi: 10.1029/2007gl032415

Alessandri A, Catalano F, De Felice M, Van Den Hurk B, Doblas Reyes FJ ,Boussetta S, Balsamo G, Miller P (2016) Multi-scale enhancement of climate prediction over land by increasing the model sensitivity to vegetation variability in EC-Earth. Clim Dyn, In Press.

Baehr J, Piontek R (2014) Ensemble initialization of the oceanic component of a coupled model through bred vectors at seasonal-to-interannual timescales. Geosci Model Dev 7:453–461. doi: 10.5194/gmd-7-453-2014

Balsamo G, Albergel C, Beljaars A et al (2015) ERA-Interim/Land: a global land reanalysis dataset. Hydrol Earth Syst Sci 19:389-407. doi: 10.5194/hess-19-389-2015

Bellprat O, Kotlarski S, Lüthi D et al. (2016). Objective calibration of regional climate models: Application over Europe and North America. J Climate, 29(2), 819-838. doi: 10.1175/JCLI-D-15-0302.1

Best MJ, Pryor M, Clark DB, et al (2011) The Joint UK Land Environment Simulator (JULES), Model description – Part 1: Energy and water fluxes. Geosci Model Dev Discuss 4:595–640. doi: 10.5194/gmdd-4-595-2011

Boisserie M, Decharme B, Descamps L, Arbogast P (2016) Land surface initialization strategy for a global reforecast dataset. QJR Meteorol Soc 142:880–888. doi: 10.1002/qj.2688

Boussetta S, Balsamo G, Beljaars A et al (2012) Natural land carbon dioxide exchanges in the ECMWF integrated forecasting system: implementation and offline validation. European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading

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

Cheruy F, Dufresne JL, Hourdin F, Ducharne A (2014) Role of clouds and land-atmosphere coupling in midlatitude continental summer warm biases and climate change amplification in CMIP5 simulations. Geophys Res Lett, 41, 6493–6500, doi:10.1002/2014GL061145.

Conil S, Douville H, Tyteca S (2008) Contribution of realistic soil moisture initial conditions to boreal summer climate predictability. Clim Dyn 32:75–93. doi: 10.1007/s00382-008-0375-9

Dee DP, Uppala SM, Simmons AJ et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. QJR Meteorol Soc 137:553–597. doi: 10.1002/qj.828

Díez E, Primo C, García-Moya JA, Gutiérrez JM et al (2005), Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts. Tellus A, 57: 409–423. doi:10.1111/j.1600-0870.2005.00130.x

Dirmeyer PA (2006) The Hydrologic Feedback Pathway for Land–Climate Coupling. J Hydrometeor 7:857–867. doi: 10.1175/jhm526.1

Dirmeyer PA (2011) The terrestrial segment of soil moisture-climate coupling. Geophys Res Lett. doi: 10.1029/2011gl048268

Dirmeyer PA, Wang Z, Mbuh MJ, Norton HE (2014) Intensified land surface control on boundary layer growth in a changing climate. Geophys Res Lett 41:1290-1294, doi: 10.1002/2013GL058826

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

Douville H (2009) 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

Dutra E, Schär C, Viterbo P, Miranda PMA (2011) Land-atmosphere coupling associated with snow cover. Geophys Res Lett. doi: 10.1029/2011gl048435

Fischer EM, Seneviratne SI, Vidale PL et al (2007) Soil Moisture–Atmosphere Interactions during the 2003 European Summer Heat Wave. J Climate 20:5081–5099. doi: 10.1175/jcli4288.1

García-Morales MB, Dubus L (2007) Forecasting precipitation for hydroelectric power management: how to exploit GCM's seasonal ensemble forecasts. Int J Climatol 27:1691–1705. doi: 10.1002/joc.1608

Ha KJ and Mahrt L (2003). Radiative and turbulent fluxes in the nocturnal boundary layer. Tellus A, 55(4), 317-327.doi:10.1034/j.1600-0870.2003.00031.x

Hagedorn R, Doblas-Reyes FJ, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. Tellus A 57:219–233. doi: 10.1111/j.1600-0870.2005.00103.x

Hagemann S, Stacke T (2014) Impact of the soil hydrology scheme on simulated soil moisture memory. Clim Dyn 44:1731–1750. doi: 10.1007/s00382-014-2221-6

Harris I, Jones P, Osborn T, Lister D (2013) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. Int J Climatol 34:623–642. doi: 10.1002/joc.3711

Hazeleger W, Wang X, Severijns C et al (2011) EC-Earth V2.2: 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 (2010) Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. Nature Geosci 4:17–21. doi: 10.1038/ngeo1032

Hirschi M, Mueller B, Dorigo W, Seneviratne S (2014) Using remotely sensed soil moisture for land–atmosphere coupling diagnostics: The role of surface vs. root-zone soil moisture variability. Remote Sensing of Environment 154:246–252. doi: 10.1016/j.rse.2014.08.030

Huffman GJ, Adler RF, Bolvin DT, Gu G (2009) Improving the global precipitation record: GPCP Version 2.1. Geophys Res Lett. doi: 10.1029/2009gl040000

Hurk BVD, 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

Hurk BVD, Kim H, Krinner G et al (2016) The Land Surface, Snow and Soil moisture Model Intercomparison Program (LS3MIP): aims, set-up and expected outcome. Geosci Model Dev Discuss 1–41. doi: 10.5194/gmd-2016-72

Klein SA, Jiang X, Boyle J, Malyshev S, Xie S (2006) Diagnosis of the summertime warm and dry bias over the U.S. Southern great plains in the GFDL climate model using a weather forecasting approach. Geophys. Res. Lett. 33 :L18805.doi:10.1029/2006gl027567

Koster RD (2004) Regions of Strong Coupling Between Soil Moisture and Precipitation. Science 305:1138–1140. doi: 10.1126/science.1100217

Koster RD, Sud YC, Guo Z et al (2006) GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. J Hydrometeor 7:590–610. doi: 10.1175/jhm510.1

Koster RD, Chang Y, Schubert SD (2014) A Mechanism for Land–Atmosphere Feedback Involving Planetary Wave Structures. J Climate 27:9290–9301. doi: 10.1175/jcli-d-14-00315.1

Ma H-Y, Xie S, Klein SA et al (2014) On the Correspondence between Mean Forecast Errors and Climate Errors in CMIP5 Models. J Climate 27:1781–1798. doi: 10.1175/jcli-d-13-00474.1

Maclachlan C, Arribas A, Peterson KA et al (2014) Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. QJR Meteorol Soc 141:1072–1084. doi: 10.1002/qj.2396

Mahrt L (1999) Stratified Atmospheric Boundary Layers. Boundary-Layer Meteorology 90: 375-396. doi:10.1023/A:1001765727956

Masson V, Moigne PL, Martin E et al (2013) The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. Geosci Model Dev 6:929–960. doi: 10.5194/gmd-6-929-2013

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 Climate 27:9253–9271. doi: 10.1175/jcli-d-14-00163.1

Lyon B, Dole RM (1995) A Diagnostic Comparison of the 1980 and 1988 U.S. Summer Heat Wave-Droughts. J Climate 8:1658–1675. doi: 10.1175/1520-0442(1995)008<1658:ADCOTA>2.0.CO;2

McNider RT, Christy JR, Biazar A (2010) A stable boundary layer perspective on global temperature trends IOP Conf. Ser.: Earth Environ. Sci. 13 012003 doi: 10.1088/1755-1315/13/1/012003

Mearns LO, Arritt R, Biner S, et al.(2012) The North American Regional Climate Change Assessment Program: Overview of Phase I Results. Bull. Amer. Meteor. Soc., 93, 1337–1362, doi: 10.1175/BAMS-D-11-00223.1.

Mueller ND, Butler EE, Mckinnon KA et al (2015) Cooling of US Midwest summer temperature extremes from cropland intensification. Nature Climate Change 6:317–322. doi: 10.1038/nclimate2825

Orsolini YJ, Senan R, Balsamo G et al (2013) Impact of snow initialization on sub-seasonal forecasts. Clim Dyn 41:1969–1982. doi: 10.1007/s00382-013-1782-0

Orth R, Seneviratne SI (2012) Analysis of soil moisture memory from observations in Europe. J Geophys Res. doi: 10.1029/2011jd017366

Peings Y, Douville H, Alkama R, Decharme B (2010) Snow contribution to springtime atmospheric predictability over the second half of the twentieth century. Clim Dyn 37:985–1004. doi: 10.1007/s00382-010-0884-1

Prodhomme C, Doblas-Reyes F, Bellprat O, Dutra E (2015) Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe. Clim Dyn. doi: 10.1007/s00382-015-2879-4

Quan X, Hoerling M, Whitaker J et al (2006) Diagnosing Sources of U.S. Seasonal Forecast Skill. J Climate 19:3279–3293. doi: 10.1175/jcli3789.1

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

Raddatz TJ, Reick CH, Knorr W et al (2007) Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century? Clim Dyn 29:565–574. doi: 10.1007/s00382-007-0247-8

Saha S, Nadiga S, Thiaw C et al (2006) The NCEP Climate Forecast System. J Climate 19:3483–3517. doi: 10.1175/jcli3812.1

Schneider U, Fuchs T, Meyer-Christoffer A, Rudolf B. (2008). Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation, 112. ftp://ftp.dwd.de/pub/data/gpcc/PDF/GPCC_intro_products_2008.pdf. Accessed May 25th 2016

Seneviratne SI, Lüthi D, Litschi M, Schär C (2006) Land–atmosphere coupling and climate change in Europe. Nature 443:205–209. 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 Reviews 99:125–161. doi: 10.1016/j.earscirev.2010.02.004

Seneviratne SI, 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

Stéfanon M, Drobinski P, D'Andrea F et al (2013) Soil moisture-temperature feedbacks at meso-scale during summer heat waves over Western Europe. Clim Dyn 42:1309–1324. doi: 10.1007/s00382-013-1794-9

Steiger JH (1980) Tests for comparing elements of a correlation matrix. Psychological Bulletin 87:245–251. doi: 10.1037/0033-2909.87.2.245

Stevens B, Giorgetta M, Esch M et al (2013) Atmospheric component of the MPI-M Earth System Model: ECHAM6. J Adv Model Earth Syst 5:146–172. doi: 10.1002/jame.20015

Stockdale TN, Anderson DLT, Balmaseda MA et al (2011) ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. Clim Dyn 37:455–471. doi: 10.1007/s00382-010-0947-3

Thomson MC, Doblas-Reyes FJ, Mason SJ et al (2006) Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature 439:576–579. doi: 10.1038/nature04503

Voldoire A, Sanchez-Gomez E, Mélia DSY et al (2012) The CNRM-CM5.1 global climate model: description and basic evaluation. Clim Dyn 40:2091–2121. doi: 10.1007/s00382-011-1259-y

Waters J, Lea DJ, Martin MJ et al (2014) Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. QJR Meteorol Soc 141:333–349. doi: 10.1002/qj.2388

Weedon GP, Balsamo G, Bellouin N et al (2014) The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. Water Resour Res 50:7505–7514. doi: 10.1002/2014wr015638

Weiss M, Hurk BVD, Haarsma R, Hazeleger W (2012) Impact of vegetation variability on potential predictability and skill of EC-Earth simulations. Clim Dyn 39:2733–2746. doi: 10.1007/s00382-012-1572-0

Whan K, Zscheischler J, Orth R et al (2015) Impact of soil moisture on extreme maximum temperatures in Europe. Weather and Climate Extremes 9:57–67. doi: 10.1016/j.wace.2015.05.001

Xu L and Dirmeyer P (2011) Snow-atmosphere coupling strength in a global atmospheric model, Geophys Res Lett 38, L13401, doi:10.1029/2011GL048049

Yoon J-H, Leung LR (2015) Assessing the relative influence of surface soil moisture and ENSO SST on precipitation predictability over the contiguous United States. Geophys Res Lett 42:5005–5013. doi: 10.1002/2015gl064139

Tables

	Coordinates
BKS	15°E-25°E 40°N-50°N
SGP	105°W-95°W 35°N-45°N
Niño 3.4	120°W-170°W 5°S-5°N

Table 1: Boundary coordinates of the BKS, SGP and Niño 3.4 boxes

Exp. Name	Model	Horizontal Resolution	Vertical levels	Ensemble generation	Land surface component	Land surface initialization	Atmosphere, ocean and sea-ice initializations
MPI-CLIM MPI-INIT	MPI-ESM v.1.1.00 (Stevens et al., 2013)	Atm/Land:T63 (~300 km) Ocean: GR15 (two poles in Greenland / Antarctica, 1.5 degree resolution)	Atm: 47 Ocean: 40	Atm: slight disturbance of stratospheric diffusion Ocean: breeding vectors (Baehr & Piontek, 2014)	JSBACH (Raddatz et al., 2007)	GCM run with nudging of the atmosphere, superficial ocean and sea-ice towards reanalyses (resp. ERAI, ORAS4 and NSIDC)	atm:ERAI ocean:ORAS4 sea-ice:NSIDC
EC-CLIM EC-INIT	ECMWF Sys4	Atm/Land: N128 (TL255, ~80km) Ocean: NEMO ORCA 1° L42	Atm: 91 Ocean: 42	singular vectors	CHTESSEL-Lakes (Boussetta et al. 2012)	ERALand horizontal interpolation (same model)	Atm: ERAI Ocean: ORAS4
MF-CLIM MF-INIT	CNRM-CM5 (Voldoire et al. 2013	Atm/Land: Tl127 (~150 km) Ocean: NEMO Orca 1º L42	Atm: 91 Ocean: 42	Initial atmospheric perturbations	SURFEX V7.2 (Masson et al. 2012)	ERALand horizontal and vertical interpolation with conservative Total Soil Wetness Index (different model)	Atm: ERAI Ocean: ORAS4 Sea-ice: restarts from a nudged run
BSC-CLIM BSC-INIT	EC-Earth V2.3 (Hazeleger et al. 2012)	Atm/Land: T106 (~120km) Ocean: NEMO Orca 1°	Atm: 91 Ocean: 46	Singular vectors in the atmosphere; different members of ORAS4 reanalyses for the ocean	HTESSEL	ERALand horizontal interpolation (same model)	Atm: ERAI Ocean: ORAS4 Sea-ice: IC3 analysis
MO-CLIM MO-INIT	GloSea5 (Maclachlan et al., 2015)	Atm/Land: N216 (~50km) Ocean: ORCA 0.25°	Atm : 85 Ocean:75	Lagged start dates and SKEB stochastic physics scheme	JULES (Best et al., 2011)	JULES offline run driven with WFDEI atmospheric data (Weedon et al., 2014)	Atm: ERAI Ocean and sea-ice: GloSea5 reanalysis (Waters et al., 2015

Table 2: Summary of the simulations

	BKS	SGP
OBS	0.69	1.01
ALL-CLIM	0.40	0.51
ALL-INIT	0.50	0.88

Table 3: Standard deviation of JJA area-averaged T2M anomaly (K)

	BKS	SGP
OBS	-0.58*	0.18
ALL-INIT	-0.50*	-0.64*
MPI-INIT	-0.46*	-0.53*
MO-INIT	-0.71*	-0.6*
MF-INIT	-0.35	-0.53*
EC-INIT	-0.23	-0.48*
BSC-INIT	-0.20	-0.55*

Table 4: Anomaly correlations of detrended ERALand May 1st total soil moisture with detrended area-averaged June-to-August T2M. 95% confidence significant values are marked by a star

Figure captions

Fig 1: Biases for June-to-August average near-surface temperature in K with respect to CRU TS v.3.23 (left panel) and relative biases for accumulated precipitation in % with respect to GPCC (right panel). The right-hand side large map corresponds to the multi-model ALL-INIT, small left-hand side maps correspond to each individual forecast system.

Fig 2: Anomaly correlation between the reference data and the June-to-August average near-surface temperature for ALL-CLIM (a) and ALL-INIT (b). Dots mark those points where the correlations are significantly different from zero with a 95% confidence level

Fig 3: Same as Fig 2b with linearly detrended anomalies

Fig4: (a) Anomaly correlation difference ALL-INIT minus ALL-CLIM and **(b)** Root Mean Square Skill Score ALL-INIT vs. ALL-CLIM for detrended June-to-August average near-surface temperature. Dots mark those points where the difference (the skill score) is significantly different from zero with a 95% confidence level

Fig 5: Anomaly correlation between the reference data and the June-to-August average accumulated precipitation for ALL-INIT

Fig 6: Same as Fig 4 for precipitation

Fig 7: Top: detrended June-to-August near-surface temperature anomaly inK. ERAInt (black solid line), ALL-CLIM and CLIM multimodel spread (blue solid line and blue envelope, respectively), ALL-INIT and INIT multimodel spread (red solid line and red envelope, respectively) for SGP (left) and BKS (right)
Bottom: detrended ERALand soil water content anomaly on May 1st for SGP (left) and BKS (right) in $m^3.m^{-3}$

Fig 8: Correlation between May 1st total soil water content and 31-day running mean of daily maximum temperature (red), minimum temperature (blue), precipitation (green) and total soil water content (gray) for individual model ensemble mean (left), multi-model ensemble mean (top right) and observations (bottom right) over the SGP region.Significant correlations are displayed with circles

Fig 9 : Individual model ensemble mean and observations daily climatologies of (a) maximum temperature in K and (b) cumulated precipitation in mm over the SGP region.

Fig 10: Same as Fig 8 over the BKS region