Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface

Article

Accepted Version

Macleod, D. A., Cloke, H. L. ORCID: https://orcid.org/0000-0002-1472-868X, Pappenberger, F. and Weisheimer, A. (2016) Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface. Quarterly Journal of the Royal Meteorological Society, 142 (694). pp. 79-90. ISSN 1477-870X doi: https://doi.org/10.1002/qj.2631 Available at https://centaur.reading.ac.uk/41336/

Published version at: http://onlinelibrary.wiley.com/doi/10.1002/qj.2631/abstract
To link to this article DOI: http://dx.doi.org/10.1002/qj.2631

Publisher: Royal Meteorological Society

Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface

David A. MacLeod[a*] Hannah L. Cloke[bc], Florian Pappenberger[d] and Antje Weisheimer[de]

[a]*Atmospheric, Oceanic and Planetary Physcis, Deptartment of Physics, University of Oxford, Oxford, UK*
[b]*Department of Geography and Environmental Science, University of Reading, UK*
[c]*Department of Meteorology, University of Reading, UK*
[d]*European Centre for Medium-Range Weather Forecasts, Reading, UK*
[e]*Department of Physics, National Centre for Atmospheric Science (NCAS), University of Oxford, Oxford, UK*

[*]Correspondence to: Department of Physics, Clarendon Lab, Parks Road, Oxford, OX1 3PU, UK E-mail: macleod@atm.ox.ac.uk

Methods to explicitly represent uncertainties in weather and climate models have been developed and refined over the past decade, and have reduced biases and improved forecast skill when implemented in the atmospheric component of models. These methods have not yet been applied to the land surface component of models. Since the land surface is strongly coupled to the atmospheric state at certain times and in certain places (such as the European summer of 2003), improvements in the representation of land surface uncertainty may potentially lead to improvements in atmospheric forecasts for such events.

Here we analyse seasonal retrospective forecasts for 1981-2012 performed with the European Centre for Medium-Range Weather Forecasts' (ECMWF) coupled ensemble forecast model. We consider two methods of incorporating uncertainty into the land surface model (H-TESSEL): stochastic perturbation of tendencies, and static perturbation of key soil parameters.

We find that the perturbed parameter approach considerably improves the forecast of extreme air temperature for summer 2003, through better representation of negative soil moisture anomalies and upward sensible heat flux. Averaged across all the reforecasts the perturbed parameter experiment shows relatively little impact on the mean bias, suggesting perturbations of at least this magnitude can be applied to the land surface without any degradation of model climate. There is also little impact on skill averaged across all reforecasts and some evidence of overdispersion for soil moisture.

The stochastic tendency experiments show a large overdispersion for the soil temperature fields, indicating that the perturbation here is too strong. There is also some indication that the forecast of the 2003 warm event is improved for the stochastic experiments, however the improvement is not as large as observed for the perturbed parameter experiment.

Key Words: Seasonal climate, climate models, forecasting, uncertainty, land surface

Received . . .

1. Introduction

Seasonal climate models are able to make predictions of average conditions for several months ahead by making use of the predictability arising from slowly evolving components of the climate system such as the ocean, the land surface and ice. Two main sources of uncertainty in seasonal forecasts come from imperfectly known initial conditions, impacting forecast error through non-linear error growth, and model uncertainty, which has its source in many places, for example: physical approximations, upscale error arising from unresolved processes and imperfectly known model parameters. Methods have been developed at numerical weather prediction centres around the world to explicitly represent this latter source of uncertainty in the atmosphere, which have consistently demonstrated positive impacts on biases, model skill and reliability, for the medium-range (Palmer 2012) and more recently at seasonal timescales (Weisheimer *et al.* 2014). However no such attempt has yet been

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.2631

made for the land surface, which in general represents an under-exploited source of seasonal predictability for the atmosphere (Seneviratne *et al.* 2012).

Land surface predictability arises in part from slow variations in soil moisture, which is a key element in the coupling between the land and atmosphere (Seneviratne *et al.* 2010). It is helpful here to consider the classical two-regime conceptual hydrological model, which describes an energy limited regime (where soil moisture content is high and evapotranspiration is limited only by available energy) and a soil moisture regime (where moisture availability constrains evapotranspiration).

In the soil moisture regime negative soil moisture anomalies limit evapotranspiration and can lead to strong coupling with atmospheric temperature via positive feedback. Reduced evapotranspiration (and therefore latent heat flux) causes an energy imbalance at the surface, which is resolved by an increase in sensible heating of the atmosphere, leading to a rise in atmospheric temperature, increased evaporative demand and further decreases in soil moisture.

The strength of land-atmosphere coupling was the focus of the GLACE multimodel intercomparison study (Koster *et al.* 2004; Seneviratne *et al.* 2006). Here it was found that derived coupling strengths between the land and the atmosphere varied widely between models, however some similarity was found in the spatial patterns, enough to identify consistent 'hotspots' of land-atmosphere coupling, with the strongest coupling located over Africa, central North America and India. Significant interannual variation in the strength of coupling has also been demonstrated (Guo and Dirmeyer 2013).

This temporal and spatial variability in coupling strength between the land and the atmosphere suggests that the impact of modifications of the land surface component of climate models may be most apparent for 'hotspot' areas, and also during periods of particularly strong coupling. One such period was the 2003 European summer, during which the positive feedback process between soil moisture and air temperature was active (Fischer *et al.* 2007a,b; Miralles *et al.* 2012). The severe impacts of this warm event have been discussed elsewhere (e.g. Robine *et al.* 2008), and it is likely that successful anticipation of similar events in future will form a key element of climate service provision. However state-of-the-art seasonal climate models have previously had difficulty simulating the 2003 event (Weisheimer *et al.* 2011a). Here we focus on forecasts for summer 2003 over Europe, considering whether improved representations of land surface uncertainty give any improvement in the forecast.

Land surface uncertainties are in part related to parameters and their high spatial variability. Heterogeneity of land cover and soil type in the land surface is not captured by the coarseness of typical climate model resolution (typically of the order of 100km). Some effort is made to deal with this, for instance in H-TESSEL a tiling system is used for the surface. However parameterization of soil moisture transport equations assumes homogeneity of parameters across each model grid box, which can lead to overly-deterministic simulations. Furthermore, parameters linking soil type to hydraulic dynamics are represented in models with too much confidence. Models use exact hydraulic parameters, yet observational studies have shown that the standard deviation of soil parameters for a particular soil type is often larger than the mean (Carsel and Parrish 1988), indicating land surface parameterization is unrealistically deterministic, with scope for improvement. Furthermore, the problem of unrepresented land surface uncertainty will be exacerbated by the trend toward higher resolution of climate models, which expose the model development process to the limits of our knowledge of what is actually occurring in the surface and subsurface hydrology (Beven *et al.* 2014).

Several methods have been suggested to incorporate uncertainties into models, based on the philosophy of ensemble forecasting. Here we explore the application of two of these methods to the land surface, for ECMWF's land surface scheme H-TESSEL. We focus on how these modifications impact simulation of the 2003 European summer, and beyond this consider their impact on mean biases, ensemble spread, deterministic and probabilistic skill. A description of the modelling framework, experimental design and verification follows in section 2. Section 3 describes results and section 4 contains a discussion.

2. Methodology

2.1. *Modelling framework*

We used the seasonal forecasting system setup that mimics the setup of System 4, the seasonal forecast system currently operational at ECMWF to perform a set of retrospective forecasts for the past (hereafter referred to as hindcasts). The atmosphere model is IFS (CY36R4) at T255 horizontal resolution (corresponding to gridboxes 80 km wide at the equator) with 91 vertical levels. The ocean model is NEMO at approximately one degree resolution, with 42 vertical levels. Initial conditions come from the ERA-Interim reanalysis (Dee *et al.* 2011) for the atmosphere, from ORA-S4 (Balmaseda *et al.* 2013) for the ocean, and the ERA-Interim Land reanalysis (hereafter ERA-Land, **?**) provides initial conditions for the land surface.

The land surface model used here is H-TESSEL, the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) with revised land surface hydrology (Balsamo *et al.* 2009), comprising a surface tiling scheme and a vertically discretized soil. The surface tiling scheme allows each gridbox a time-varying fractional cover of up to six tiles over land (bare ground, low and high vegetation, intercepted water, and shaded and exposed snow) and two over water (open and frozen water). Each tile has a separate energy and water balance, which is solved and then combined to give a total tendency for the gridbox, weighted by the fractional cover.

The vertical discretization has soil layers below ground at 7, 21, 72 and 189cm. The soil heat budget follows a Fourier diffusion law, with net ground heat flux as the top boundary condition and zero flux at the bottom. Subsurface water fluxes are determined by Darcy's law, used in a soil water equation solved with a four-layer discretization shared with the heat budget equation.

Vertical movement of water in the unsaturated zone of the soil matrix is described by Richards' equation for flow of water in the subsurface (Richards 1931), often used in soil physics and land surface models (Hillel 1998). It is shown in equation 1

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left(\lambda \frac{\partial \theta}{\partial z} - \gamma\right) + S_\theta, \tag{1}$$

where θ is water content of the soil, γ is the hydraulic conductivity, λ is the hydraulic diffusivity, S_θ is a volumetric sink term associated with root uptake, z is the vertical height above a reference point and t is time.

Hydraulic conductivity is calculated with the van Genuchten formulation (van Genuchten 1980), introduced as part of the improved H-TESSEL model. This scheme is favoured by soil scientists as it has shown good agreement with observations in intercomparison studies (Shao and Irannejad 1999) and is given by

$$\gamma = \gamma_{sat} \frac{[(1+(\alpha h)^n)^{1-1/n}) - (\alpha h)^{n-1})]^2}{(1+(\alpha h)^n)^{(1-1/n)(l+2)}}, \tag{2}$$

where α, l and n are soil dependent soil texture parameters and h is soil water potential (the potential energy of soil water due to hydrostatic pressure). h is linked to θ by the expression

$$\theta(\phi) = \theta_r + \frac{\theta_s - \theta_r}{(1+\alpha h)^{1-1/n}}, \quad (3)$$

where θ_r and θ_s are residual and saturated soil moisture content respectively.

H-TESSEL also introduced a spatially-varying soil type map (previously in TESSEL soil parameters were spatially homogenous). This map comes from the Food and Agriculture Organization (FAO) soil type map of the world (FAO 2014), which describes a large variety of soil types at relatively high spatial resolution (around 30 arc seconds, corresponding to roughly 1km).

For inclusion in H-TESSEL, the FAO soil types are simplified to six categories, and then prepared for use with IFS resolution by taking the dominant FAO soil type within an IFS grid box as representative of the entire box. The typical seasonal climate forecast resolution is of the order of 100km, so any variability in soil type on scales lower than this is essentially ignored in the model. In reality soil properties exhibit variability on an order of magnitude at least a thousand times smaller than this, on the order of metres (e.g. Trangmar *et al.* 1986), suggesting that there is significant uncertainty associated with this approximation.

A second uncertainty arises from the fact that each H-TESSEL soil type has a single value for each of the van Genuchten parameters. Each soil type has a different van Genuchten parameter; those used in H-TESSEL are shown in table 1, calculated for simplified soil types from measurements of parameters of large soil samples (Cosby *et al.* 1984).

Texture	α (m^1)	l	n	γ_{sat} ($10^{-6} ms^{-1}$)
Coarse	3.83	1.250	1.38	6.94
Medium	3.14	-2.342	1.28	1.16
Medium-fine	0.83	-0.588	1.25	0.26
Fine	3.67	-1.977	1.10	2.87
Very fine	2.65	2.500	1.10	1.74
Organic	1.30	0.400	1.20	0.93

Table 1. Default soil parameters used in IFS/H-TESSEL for the six IFS soil types.

Again, reality exhibits more variability. The observed mean and standard deviations of hydraulic conductivity for several soil types, calculated from a large number of soil samples, are reproduced in table 2. Here the standard deviation is often larger than the mean, suggesting that the real world has a much higher spread in soil response to moisture input than the model simulates.

Soil type	μ	σ
Clay	0.56	1.17
Clay loam	0.72	1.94
Loam	2.89	5.06
Silt	0.69	0.92
Silt loam	1.25	3.42
Silt clay	0.06	0.31

Table 2. A subset of the mean (μ) and standard deviation (σ) of saturated hydraulic conductivity measured for a selection of soil types, reported in Carsel and Parrish (1988). Units are $10^{-6} ms^1$.

Land surface uncertainties are not represented in the IFS, however atmospheric uncertainties are, and a significant amount of research has gone into developing these schemes. In the IFS, model uncertainty is dealt with through the stochastically perturbed parameterized tendencies scheme (SPPT, Palmer *et al.* 2009) and the stochastic kinetic backscatter scheme (SKEB, Shutts 2005; Berner *et al.* 2009). We use the SPPT scheme as a basis for one of our experiments and so provide some further details below.

Uncertainty associated with atmospheric parameterization of sub-grid scale physical processes results in errors to the tendencies (that is, the changes in a variable from one time step to the next due to the parameterization). The SPPT scheme attempts to sample this error with an ensemble by perturbing the total tendencies for temperature, winds and humidity fields at every timestep by a dynamic spatially-correlated random number field. That is, the total parameterized tendency X for each variable is perturbed via multiplicative noise, i.e.

$$X_p = (1 + r\mu)X, \quad (4)$$

where X_p is the perturbed tendency, r is a random number and $\mu \in [0, 1]$ is a factor used for reducing the perturbation amplitude close to the surface and in the stratosphere.

The random number comes from an evolving 2-dimensional field, correlated in space and time, produced by a spectral pattern generator (SPG), evolving parallel to the main simulation. This SPG is a three-scale two-dimensional autoregressive process (AR1) and has been designed to mimic the typical scales present in the atmosphere related to the error fields. The total random field at any instant is a summation of three independent AR1 processes, each with a characteristic decorrelation length and time: 500/1000/2000km and 6/72/720hr for the small to large scales respectively. The standard deviations of the amplitudes of the perturbations for the small, medium and large scales are 0.52, 0.18 and 0.06, resulting in a total pattern in which the small scales are perturbed more strongly. These scales have been chosen as representative of the approximate scales of the atmosphere.

The impact of SPPT on model biases and forecast skill has been very positive. At the medium range the skill of tropical 850hPA temperature is much improved, with forecasts including stochastic parameterization showing the same level of skill at six days from initialization than forecasts without show after two. The reliability of precipitation forecasts over Europe is also improved for these forecasts (Palmer *et al.* 2009). At seasonal scales, SPPT helps to reduce excessively strong convection over the Maritime Continent and the tropical Western Pacific, leading to reduced biases of the outgoing longwave radiation (OLR), cloud cover, precipitation and near-surface winds (Weisheimer *et al.* 2014). The stochastic perturbation also improves statistics of the Madden-Julian oscillation and reduces errors in the forecast of El Niño.

In this paper we describe attempts to incorporate land surface uncertainties into ECMWFs seasonal forecasting model, following an SPPT-like approach where we perturb tendencies to soil fields, and a perturbed parameter approach where soil parameters are perturbed directly. A description of the control model setup and the perturbation experiments follows.

2.2. *Experimental design*

Several seasonal hindcast experiments have been run, in each case with a different method of uncertainty representation for the land surface.

We focus on hindcasts of boreal summer for the period 1981-2012, by initialising the model on the 1st May for each year and running each hindcast forward for four months, with each experiment having 25 perturbed ensemble members. The control experiment has no land surface perturbation and differences here arise only from the stochastic schemes used in the atmosphere and from the initial conditions. For the two experiment types, further differences arise from the land surface perturbed tendencies and parameters.

2.2.1. *Stochastically perturbed tendencies*

For the stochastically perturbed tendency experiments we mimic the SPPT atmospheric scheme. At each time step, the total tendency from the soil moisture and temperature fields arising from the H-TESSEL parameterization are perturbed with the instantaneous field generated by the SPG. All four layers are perturbed equally, and the pattern generated uses a different initial seed from the SPPT, so that the perturbation fields used for the atmosphere and the soil are uncorrelated.

Three stochastically perturbed tendency experiments are carried out, each using a different weight for the three scales in the SPG. The first uses the same weighting as SPPT (standard deviations of small/medium/large set as 0.52/0.18/0.06), the second uses a field with all scales weighted equally (0.32/0.32/0.32), whilst the third uses a mirrored version of SPPT scales (0.06/0.18/0.52), such that the largest scales are perturbed most. For each experiment the standard deviation of each pattern is chosen such that the root of the sum of the squared value (i.e. the standard deviation of the total pattern) is kept constant.

The SPG is in a general way intended to have a spectrum roughly similar to the scales present in the atmosphere (thus, the small/fast scales dominate). Since the decorrelation time scales associated with soil anomalies are somewhat larger than the atmosphere we chose to experiment with different SPG scales, as we consider that a larger scale pattern may be more appropriate for the land surface.

2.2.2. *Perturbed parameters*

The perturbed parameter experiment focuses on two key soil parameters: α and γ_{sat} (see equation 2, 3 and the associated discussion for details). These particular parameters have been chosen as previous studies found them to be particularly sensitive (Cloke *et al.* 2008). A static perturbation is used, where each ensemble member takes a different combination of perturbation to the two parameters, which remains constant for the length of the integration. Perturbations are taken from the set {-80%, -40%, 0, 40%, 80%}, where the perturbation percentage applies to the default parameter for the soil type at a particular gridpoint. A 25 member ensemble is then generated by applying all the perturbation combinations to the two parameters.

The magnitude of the maximum perturbation was guided by some initial experiments, which indicated that using lower maxima of 20% and 40% had little observable impact on the hindcasts (N.B. a perturbation of 80% is still well within the observed variation in parameter values from the mean for most soil types, see table 2).

An unexplored element of the PP experiment is the question of model spin-up and initialisation. The initial state from ERA-Land used was created by running the unperturbed model in an offline run with ERA-Interim forcing, generating initial land surface states. However, the ensemble members of the PP experiment are each essentially a different version of the model, and so presumably would calculate each a different initial state when run in the same way as the ERA-Land initial state generation. This means there is likely to be some spin-up time in the PP experiment, where ensemble members move toward their individual model climatologies, different from the unperturbed climatology.

For the top soil level this spin-up is only likely to be a few days and deeper levels will take longer to relax. It is unclear to what extent this may impact results as the process of a model adjusting to its forcing can bias land surface simulations (Rodell *et al.* 2005; Shrestha and Houser 2010). Work is in progress to explore this question in the current experimental setup, including the creation of and experiments with initial conditions using this perturbed ensemble. However here we assume the impact on results is minimal.

Table 3. A list of the short names and descriptions of land surface perturbation experiments carried out.

ID	Description
ST1	Stochastic tendencies, default scales of the SPG
ST2	Stochastic tendencies, all scales weighted equally
ST3	Stochastic tendencies, largest scales weighted most (mirror scales)
PP	Perturbed parameters α and γ_{sat}

Hereafter experiments are referred to by a short name, these are summarized in table 3.

2.3. *Assessing model performance*

We first investigate the 'hotspots' of strong land-atmosphere coupling, following the definition of Seneviratne *et al.* (2006), where a negative correlation between air temperature and evaporation indicates a strong coupling region. We diagnose this coupling for the ERA-Interim reanalysis and all the experiments, using the Pearson's correlation coefficient (Wilks 2011) between June/July/August (JJA) 2m air temperature and evaporation. For the experiments the correlation is calculated between the ensemble mean of the forecast and the observations.

Following this we consider the predicted anomalies corresponding to the 2003 summer, and subsequently the average model performance over all hindcast years. Here we describe the methodology followed for verification, the reference data used and the metrics used.

The soil fields considered are temperature and moisture in the top level. As a reference for these soil fields throughout, we verify against the reanalysis dataset ERA-Land, which was created by forcing H-TESSEL with high quality atmospheric data (**?**).

To consider the potential impact on the atmosphere, we consider 2m air temperature, precipitation, surface and latent heat fluxes. We use the ERA-Interim reanalysis as a reference for these (Dee *et al.* 2011), except for precipitation, for which we use the GPCP monthly precipitation analysis (Adler *et al.* 2003).

For the analysis of summer 2003, we use the probability of summer 2m air temperature falling in the upper quintile as an indicator of the forecast probability. That is, we calculate the percentage of ensemble members in which month 2-4 (corresponding to the forecast for June-August for a May start date) falls in the upper quintile of the climatological distribution. The upper quintile threshold is calculated across all ensemble members and years, on a grid point basis. We also consider the ensemble mean anomalies, which are calculated with reference to the hindcast period 1981-2012.

In addition to this we calculate the climatological Probability Density Function (PDF) of 2m air temperature for the experiments over Southern Europe (10W-40E, 30-48N, following Weisheimer *et al.* 2011a) and the forecast PDF for 2003. To convert the discrete ensemble members into a smooth probability distribution, we use affine kernel dressing (Bröcker and Smith 2008), where kernel parameters are fitted by minimum continuously ranked probability score estimation (Wilks 2011).

For the analysis across the hindcasts, we focus on predictions for JJA, looking at the impact of the perturbations on the ensemble mean biases in land and atmospheric variables.

To measure the impact of the perturbation on ensemble dispersion, we consider the spread/error ratio. In a perfect system the ensemble spread gives a measure of forecast uncertainty and so ideally would be equal to the error of the forecast. To quantify this, we calculate the standard deviation of the ensemble divided by the root mean square error of the ensemble mean, averaged across all hindcast years. A spread/error ratio greater than 1 indicates that the field tends to overdispersion and issues underconfident forecasts, whilst a ratio less than one indicates a tendency for underdispersion and overconfident forecasts.

Finally to consider probabilistic skill we use the Brier skill score, which measures the skill at forecasting events (Brier 1950). In our case the event we consider is a particular season falling in either the upper or lower tercile of the climatological distribution, interpreting this as an anomalously warm/cold or wet/dry season. The Brier score is given by

$$BS = \frac{1}{n}\sum_{i=1}^{n}(p_i - x_i)^2, \tag{5}$$

where p_i is the forecast probability of occurrence of the event i from a set of n forecast-observation pairs, and x_i is a corresponding binary value indicating occurrence/non-occurrence in the observations (1/0 respectively). The Brier score takes a value in the interval $[0, 1]$ and is negatively biased when calculated from a smaller ensemble; we employ here a correction which estimates the score obtained by an infinite ensemble, described in Ferro and Fricker (2012). Furthermore we present the score as a skill score relative to climatology, as

$$BSS = 1 - \frac{BS_{for}}{BS_{clim}}, \tag{6}$$

where BS_{for} is the Brier score for our forecast system and BS_{clim} is the Brier score obtained by a climatological forecast, always issuing a forecast probability equal to the climatological frequency. A BSS equal to 1 indicates a perfect forecast system, and BSS equal to or less than 0 indicates one with a Brier score equal to or worse than climatology.

A description of these results follows. Where presented, confidence intervals are based on a 1000 member bootstrap resampling. All model and reference data has interpolated to a common 2.5 degree grid before analysis.

3. Results

3.1. *Hotspots*

Correlation between evaporation and 2m air temperature in JJA is shown in figure 1, for ERA-Interim, the control run and the PP experiment. Negative correlations here represent strong coupling regions and potential hotspots, following Seneviratne *et al.* (2006). Large negative correlations indicate regions where high moisture availability leads to increased latent heat flux through evapotranspiration and so lower levels of sensible heat flux and air temperature. Conversely in these regions of negative correlation, dry soil leads to a decrease in latent heat flux and a corresponding increase in sensible heat flux and air temperature. For ERA-Interim the hotspots are located over central North America and Brazil, the Sahel, southern Africa, central Asia, China, India and Australia.

For the control seasonal hindcast experiment the pattern is quite similar, with a slight increase in the extent and strength of the hotspot over North America, and an extension of the strong coupling over South America such that it covers most of the continent. Coupling is also present over Europe compared to ERA-Interim, where it is absent, and similarly for the highest latitudes in the east of Asia. Results for the PP experiment are shown in figure 1c. The only difference between this experiment and the control is the extent of the coupling in North America, which in the PP experiment is closer to that seen from ERA-Interim. This result is representative for the ST experiments (not

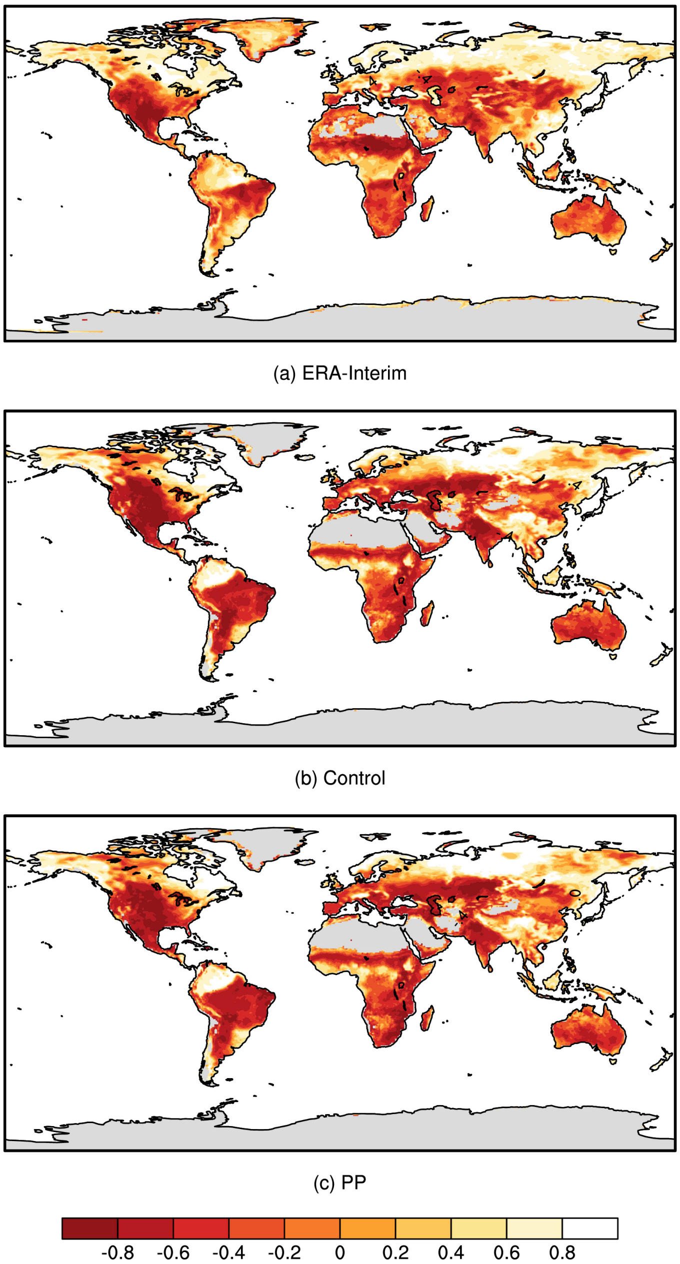

Figure 1. Correlation between JJA evaporation and 2m temperature, for (a) ERA-Interim, (b) the control experiment and (c) the PP experiment. Large negative correlations indicate regions of strong land-atmosphere coupling, following Seneviratne *et al.* (2006). Areas of low climatological evaporation have been masked.

shown), which also show a reduction in the strength of coupling over North America.

The spatial pattern of the hotspots agrees with previous results from the GLACE project (Koster *et al.* 2004) and from the IPCC AR4 models (Seneviratne *et al.* 2006), which indicated the strongest coupling over North and South America, the Sahel, southern Africa, Central Asia and India. Despite the fact that the average coupling strength shown here for Europe is low, the magnitude of the coupling is variable in time and for this region it has been shown that coupling was particularly high for the 2003 extreme warm event over Europe. It is to this event we turn now.

3.2. *Impact on simulation of summer 2003 over Europe*

Figure 2 shows the probability of 2003 summer 2m air temperature falling in the highest quintile indicated by the experiments. The percentiles in which JJA 2003 fell according to the ERA-Interim reanalysis are indicated in figure 2a. The highest quintile in the reanalysis extends over nearly all of Western

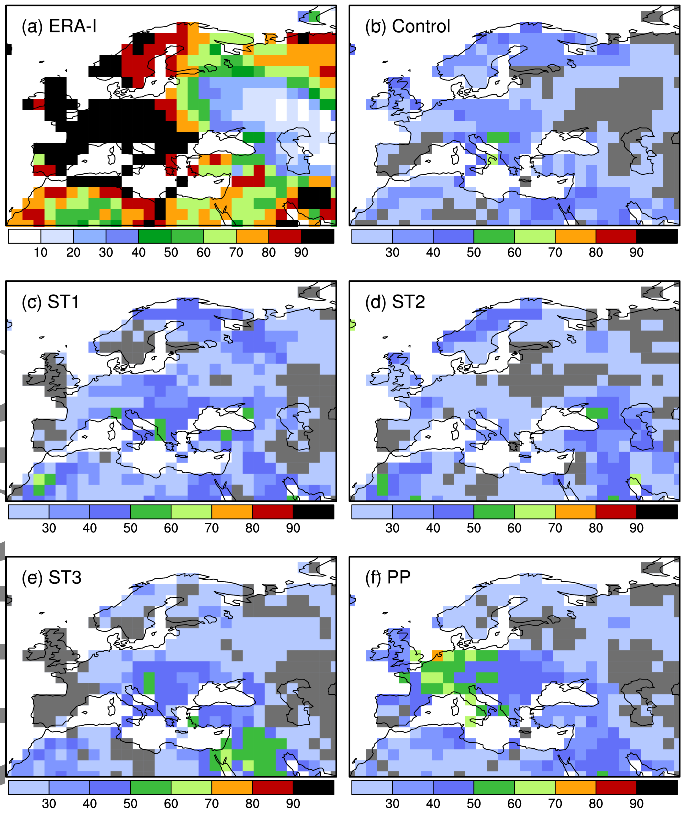

Figure 2. 2m JJA average air temperature in 2003: (a) percentile for the season from ERA-Interim, relative to its climatology, (b-f) probability of temperature for the same period falling in the highest quintile for the experiments. In plots b-f the areas where the probability is below the climatological frequency (20%) have been masked.

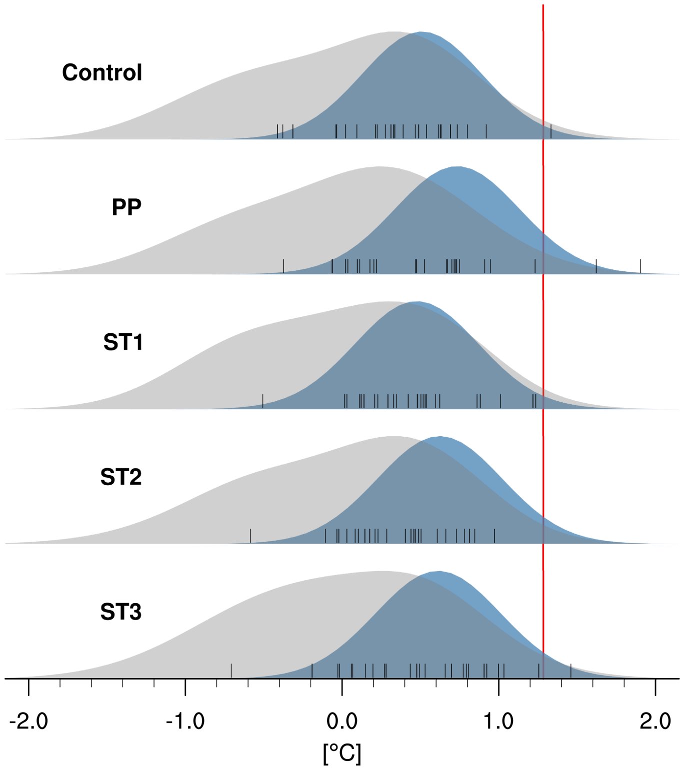

Figure 3. Probability density functions for 2003 JJA 2m air temperature over Europe (10W-40E, 30-48N), represented by the blue curve (small black lines indicate individual ensemble members). The grey curve indicates the climatological distribution, and the red line indicates the observed temperature from ERA-Interim.

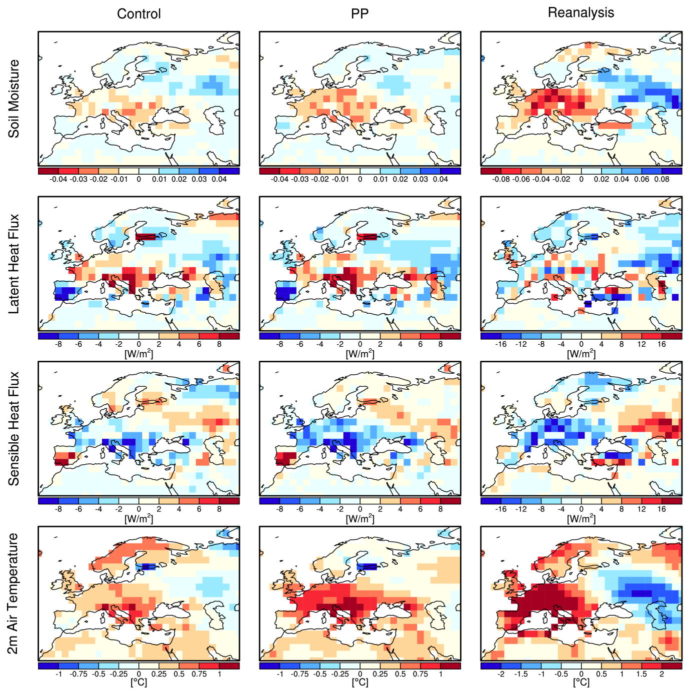

Figure 4. 2003 JJA ensemble mean anomalies for the control (left column) and PP experiment (central column). Reanalysis for JJA 2003 is shown in the right column. Top-bottom rows for variables top level soil moisture, latent heat flux, sensible heat flux and 2m air temperature. Reanalysis is ERA-Land for soil moisture, ERA-Interim otherwise. N.B. Flux anomalies are measured downwards, and the contour scale is different between the models and reanalysis.

Europe, from Ireland to the Black Sea, and from the north coast of Africa to Scandinavia. Concurrently the temperature for regions around the Caspian sea were equally extreme, falling in the lowest deciles relative to the reference period.

The control (figure 2b) indicates slightly enhanced probability of upper quintile temperature (compared with the climatological frequency of 20%) over most of Europe, with a maximum of around 40% over the Balkans. The stochastic perturbation experiments follow a similar pattern of probability, with slight increases, particularly for ST3 (figure 2e). The PP experiment (figure 2f), indicates a much higher probability over Western Europe, particularly over France, where it gives a probability of upper quintile temperatures of 60-70%. The pattern of anomalously high air temperature is more consistent with reanalysis and reports of extreme summer temperatures (where the largest impacts occurred over France).

Figure 3 shows the PDFs of temperature over Southern Europe for all four experiments and the control, in each case comparing the forecasted to the climatological PDF. In the control experiment there is a clear shift in probability mass in 2003 toward higher temperatures, and in the PP experiment this shift is even more pronounced, with a much stronger indication of higher temperatures. The ST1 experiment shows a slight decrease in probability mass at high temperatures, whilst the ST2 and ST3 both show stronger probabilities, though with smaller magnitude compared to the PP experiment.

These results imply that the PP experiment is simulating the key processes leading to the extreme temperatures more realistically than the control and the stochastic tendency experiments. One key process, (described previously in section 1) involves negative soil moisture anomalies, decreased latent and increased sensible heat flux from the surface, and positive air temperature anomalies. Anomalies for these fields are shown in figure 4, comparing 2003 JJA anomalies for these variables for the control and the PP experiment to reanalysis. Though the latent heat flux anomaly for JJA shown in figure 4 does not show a large difference between the PP experiment and the control, this is based on an ensemble mean average for JJA. Looking closer at the individual members for seperate months reveals an improvement in the ensemble distribution for July, though no difference in June and August (not shown).

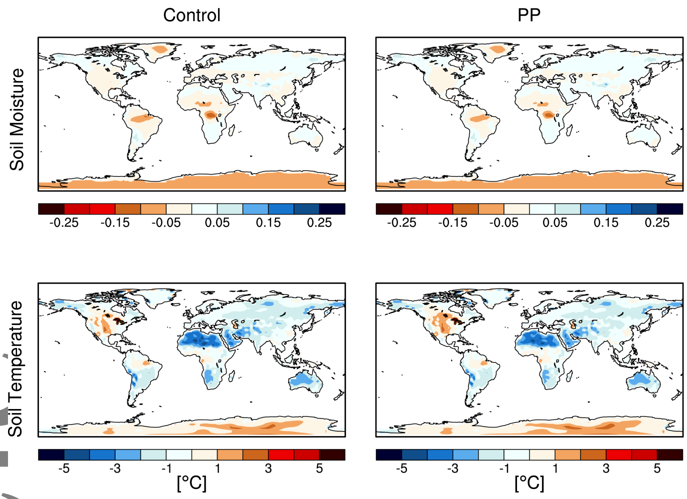

Figure 5. JJA biases for top level soil moisture and temperature, in the control and PP experiments. The minimal impact of land surface perturbations on model biases demonstrated here is representative of the results for other lead times, experiments and atmospheric variables (2m air temperature and precipitation).

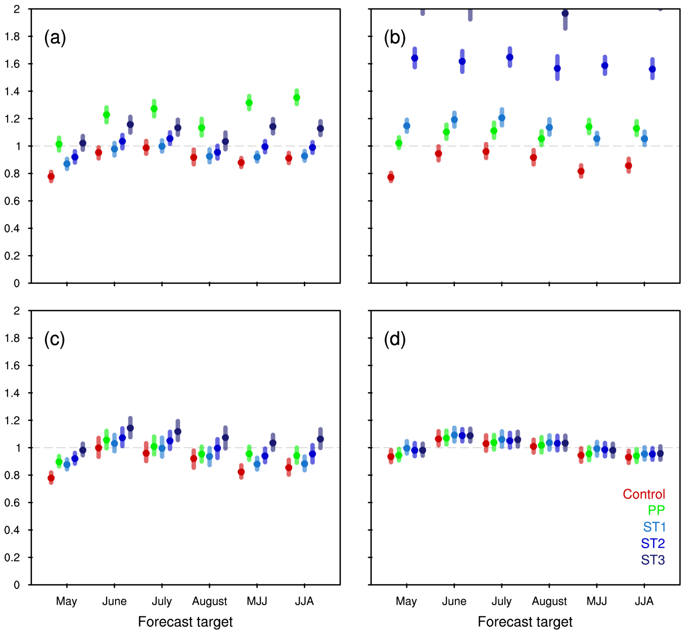

Figure 6. Ratio of spread to RMSE over tropical land points for (a) soil moisture, (b) soil temperature, (c) 2m air temperature and (d) precipitation, for all experiments, monthly and seasonal targets. Bars indicate 95% confidence intervals from bootstrapping.

The PP experiment has a much clearer negative soil moisture anomaly than the control, in the region that the reanalysis indicates the driest soil. This improved simulation of negative moisture anomalies is also seen for lower soil levels (not shown).. There does not seem to be a corresponding decrease in latent heat flux, this is also not present in the reanalysis. The large upward anomaly for sensible heat flux for the PP experiment is more in agreement with the reanalysis than the control, as is the 2m air temperature. Improvement is not uniformly positive, as the simulation of cold air and wet soil east of Europe, and warm air in Scandinavia are worse in the PP experiment. However in the region of maximum impact (over France), the experiment better simulates the extreme event, consistent with soil moisture and sensible heat flux anomalies.

3.3. *Impact on model climatology and forecast skill*

Mean summer biases are shown in figure 5 for the control and PP experiment, for top level soil moisture and temperature. The atmospheric variables 2m temperature and precipitation are also disccussed but are not shown for brevity.

In general the control is slightly too cold, particularly over the tropics, Australia, Russia and South America. This is reflected in both the soil and air temperature bias. Over most of North America the air temperature has a slight warm bias, whilst the soil does not, which suggests that this results from some error in the simulated atmospheric processes which is not transmitted to the soil. The control is generally dry in JJA over the tropics, which is also reflected in the soil moisture bias.

The change in bias from the control for the PP experiment is not significant. This is also true for the ST experiments (not shown), and for each of the four forecast months individually, suggesting that of the land surface perturbations considered, none make any noticeable impact on the mean state of land surface fields or the atmosphere. This suggests that perturbations of at least this magnitude can be applied to the land surface without degradation of the model climate. Anomaly correlations between the ensemble mean and observations were also considered for soil and atmospheric variables (not shown). For these variables the experiments show no significant change from the control.

Results for the spread/error are shown in figure 6 for tropical land points (30°S to 30°N). These scores are shown with 95% confidence intervals from bootstrapping, for targets corresponding to all monthly and three-monthly averages corresponding to four-month forecasts issued in May. All of the perturbation experiments generally increase the spread of the soil variables, which tend toward underdispersion in the control. However the PP experiment increases the spread in soil moisture too much after the first month, resulting in overdispersion (figure 6a). For soil moisture, the ST experiments also increase the spread, though not to the same extent as the PP experiment. The ST experiments with more weighting on longer scales (i.e. ST2 and ST3) show more of a tendency toward overdispersion.

This overdispersion of the experiments when the perturbation is weighted toward larger scales is also seen for soil temperature (figure 6b). The effect is quite significant, increasing the spread/error ratio for ST3 from around 0.9 in the control to over 2. This is clearly undesirable, suggesting that this perturbation is not appropriate for soil temperature. Results for PP and ST1 are more reasonable, with only slight increases in the spread.

The spread is also increased for 2m air temperature (figure 6c) following the same pattern for the experiments as seen in soil temperature. This is likely driven by the large dispersion in soil temperature being transmitted to the atmosphere through sensible heating. Results for precipitation are also shown (figure 6d), but there is no noticeable impact on the spread here.

Brier skill scores for wet soil events are shown in figure 7. There are limited areas where the score is significantly above zero in the control: part of the Middle East, Western USA, central South America, Southern Africa, Indonesia and East Australia. The PP experiment shows improvement over the control for Central and South America and the Middle East (figure 7e) with slight increases in the Brier skill score, consistent with the hotspot regions indicated in figure 1, though it remains below significance in these areas. For the ST3 experiment the score slightly increases for South America, but for other regions and the other two ST experiments there is little impact. A similar pattern of skill is seen in the control for dry soil events (not shown), and again the PP experiment shows similar improvements in South America and the Middle East.

Results for warm soil events are shown in figure 8. Scores for the control are higher than for soil moisture, with largest scores over South America, equatorial Africa, Greenland, the Middle East, Central Asia and Indonesia. All experiments seem to show an improvement in skill over Africa, particularly the ST3

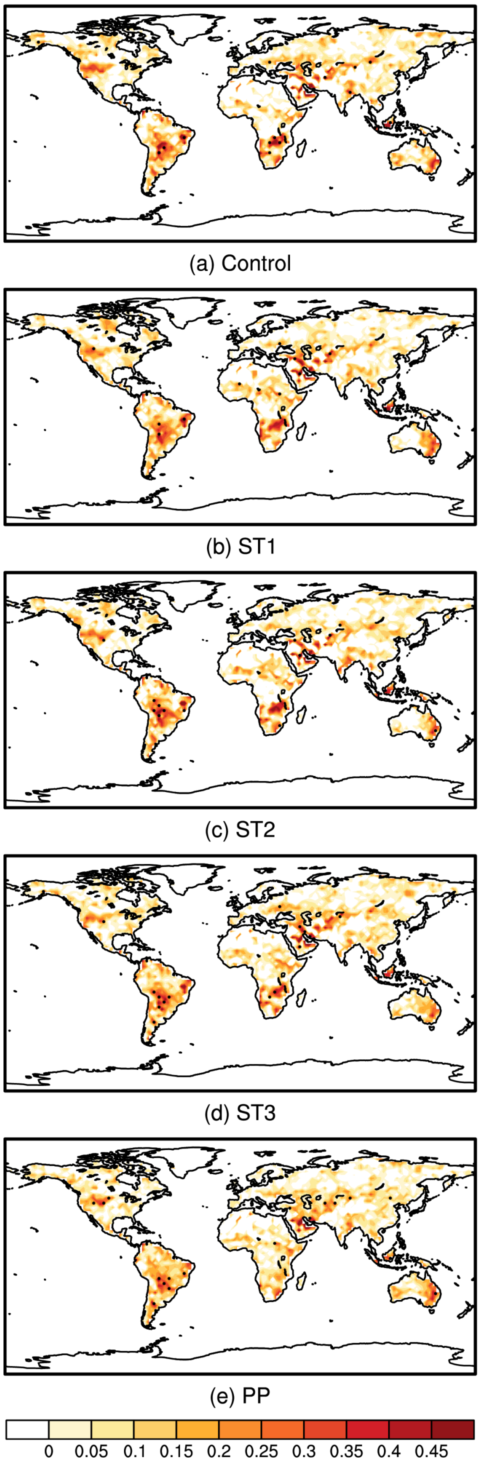

Figure 7. Brier skill score for upper tercile soil moisture JJA for the control and experiments. Stippling indicates where score is significant at 95% level.

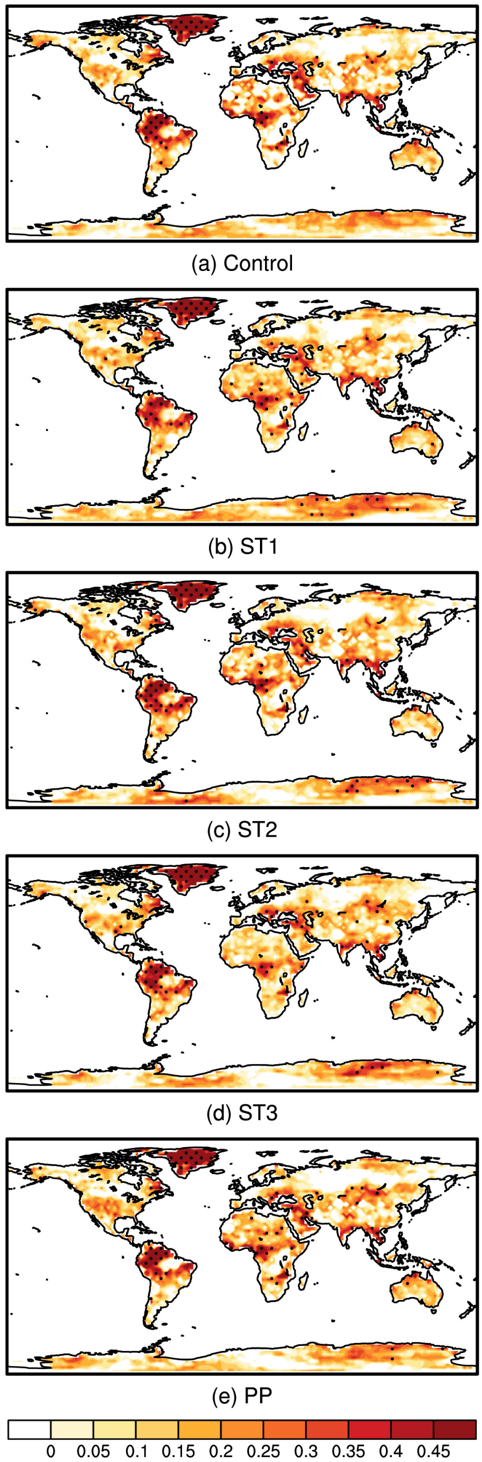

Figure 8. As figure 7, for upper tercile soil temperature.

experiment (figure 8d), which shows a score greater than zero across the continent, whilst the control has many areas where the score is below climatology. The ST3 experiment also shows a slight increase across the whole of Asia. Results for lower tercile soil temperature are not shown, but are similar to those for upper tercile events.

Scores for upper tercile 2m air temperature are not shown, but indicates skill in the control in similar places to soil temperature, with scores significant at the 95% level over much of the Middle East, Central and Southeast Asia, Greenland, South America and parts of Africa. The largest improvement seems to be with the PP experiment for North/Central America, where the small area with scores significantly above zero in the control increases, and the area with scores below zero decreases. Results for precipitation have also been analysed, and indicate that the perturbations have a negligible impact on the probabilistic skill (not shown).

4. Discussion and Conclusions

The land is an important component of the climate system and is a key player in the development of extreme warm events. However there are uncertainties in the land surface, which are not normally represented in global coupled models.

Since improvements in seasonal forecast skill and reliability have resulted from explicit representation of uncertainty in the atmosphere (Weisheimer *et al.* 2014), we have attempted to follow a similar approach for the land surface component of the ECMWF model. Three stochastic tendency experiments have been carried out, based on the SPPT scheme used operationally in the ECMWF atmospheric model, with each having a different weighting of the scales used in the pattern generating the stochastic fields. As well as this, one further experiment has been run, where key soil parameters have been perturbed.

Regions with strong atmospheric coupling (hotspots) have been estimated in the experiments, demonstrating quite similar pattern of coupling in the control experiment as observed for the ERA-Interim reanalysis, which itself is similar to results from modelling systems reported previously (Koster *et al.* 2004; Seneviratne *et al.* 2006), with hotspots reported over central North and South America, the Sahel, India, central and east Asia. For most areas the perturbations show little difference in coupling from the control, except over North America, where the perturbations are closer to the reanalysis, with a slightly decreased extent negative correlation compared to the control.

Hotspot regions in models must be interpreted with caution however, as it has been demonstrated that land-atmosphere feedbacks can be highly dependent on parameterization. For instance, results demonstrated a positive soil moisture-precipitation feedback with convective parameterization was used and negative feedback when convection was explicitly simulated (Taylor *et al.* 2013). This raising questions about the accuracy of land-atmosphere feedback and coupling assessed with models using convective parameterizations, including those here and elsewhere (Koster *et al.* 2004; Seneviratne *et al.* 2006).

Considering now the impact of explicit representations of uncertainty on the forecast of the extreme European summer of 2003, results here suggest an improvement in the forecast for the perturbed parameter experiment. The result is mechanistically consistent, arising from improved simulation of negative soil moisture anomalies and increased sensible heat flux from the surface. The same improvement in the forecasted probability of higher temperatures is observed to a lesser extent for the larger-scale stochastic tendency experiments.

Some improvement in probabilistic skill is also detected for the perturbed parameter experiment when measured across the hindcast, though not for Europe, for which skill is low. Note that this lack of skill is entirely consistent with the improvement for the individual year. Since the land-atmosphere coupling was especially strong for this year we would expect improvements in the land to impact the atmosphere here, whilst for other years when coupling is weaker or absent, this impact would be lower or non-existent. Improvements may not therefore be visible when looking across the entire hindcast range.

Improvements in probabilistic skill for the perturbed parameter experiment however are visible for soil moisture in regions of strong land-atmosphere coupling. That is, over Central/South America and the Middle East and for air temperature in North/Central America. The only improvement for the stochastic tendency experiments occurs for the largest scale experiment (ST3), which shows some improvement over Africa and Asia for soil temperature, though the score is still below 95% significance in these regions.

We observe no significant impact on the mean state of soil or atmospheric fields, or any impact on deterministic scores, though the stochastic tendency experiments slightly decrease the anomaly correlation for soil temperature. In terms of the spread/error ratio, all experiments act to increase the spread of the soil fields, which are generally underdispersive in the control. However they go too far and cause overdispersion. For soil moisture this is largest with the perturbed parameter experiment, though the stochastic tendency experiments also increase the spread, particularly for experiment ST3.

For soil temperature the perturbed parameter experiment increases the spread so that the spread/error ratio is close to 1 (compared to an underdispersed control) for the first month, though it is slightly overdispersed after this. The stochastic tendency experiments also have an impact on the spread of soil temperature, and introduce overdispersion from the first month. This is dependent on the weighting of the scales used - with the most dispersion occurring again for ST3, for which the spread is increased to over twice the error.

It is not clear why the ST3 experiment, in which the largest temporal and spatial scales are the most prominent, shows the most dispersion, particularly so for soil temperature. This may be related to the interaction between the timescale of the evolution of the tendency and the decorrelation timescale of the stochastic forcing. For example, in a situation where both are equal it would be possible for the tendencies and the stochastic forcing to 'lock', so that occasionally the sign of the tendency and the forcing act in complement for a period of time, resulting in a large dispersion of the ensemble. Conversely, in a situation when the decorrelation timescale of the forcing and the autocorrelation of the tendency act on different timescales, periods in which the sign of the tendency acts in one direction would be perturbed over time in contradictory ways by a more-quickly varying forcing. This would act over time to dampen the effect of the perturbation and create relatively less dispersion than the case in which the timescales are equal.

Whilst this hypothesis will be tested theoretically in future work, practically this overdispersion suggests that the perturbation of soil temperatures directly is somewhat unrealistic. Furthermore, a consideration of the approximations in the land surface models used in climate models suggests that the main uncertainties lie in the hydraulic characteristics of the soil, i.e. the way moisture interacts with different soil types. This indicates that representation of uncertainty in the hydraulic equations directly is more consistent with the nature of model imperfections.

These results can be contrasted with previous work assessing representations of model uncertainty in monthly and seasonal forecast ensembles, which looked at the impact of atmospheric perturbed parameter and stochastic parameterization schemes (Weisheimer *et al.* 2011b). These results suggest that stochastic parameterization gives the most improvement in model skill for precipitation, particularly in the first month. Whilst this is not quite consistent with results in this paper, the stochastic physics scheme used for comparison in the atmosphere has been developed over the past decade, whilst the stochastic schemes tested here are new and relatively ad hoc. It is likely then that there is potential for this stochastic land surface scheme to be refined.

Furthermore, as these results show, incorporating uncertainty directly into the land surface hydrology equations has serious potential to improve seasonal forecasts, for regions and periods of strong land-atmosphere coupling. This is demonstrated here for the 2003 European summer, with the perturbed parameter experiment giving an improved anticipation of this high-impact societally-relevant event. We plan to continue this work toward explicitly represented land surface uncertainties by developing more sophisticated methods (e.g. stochastic parameters) and considering their impact on seasonal forecasts for Europe and beyond.

Acknowledgements

The authors would like to acknowledge valuable comments and discussion with the Tim Palmer, Gianpaolo Balsamo, Emanuel Dutra, Tim Stockdale and others at ECMWF and the University of Reading. DM and AW would like to acknowledge funding from the EU-FP7 project SPECS (grant agreement 308378) and also to thank the developers of the freely-available R package 'SpecsVerification' produced for SPECS, and used here for ensemble dressing routines. HC would like to acknowledge funding from the NERC project IMPETUS (NE/L010488/1).

References

Adler RF, Susskind J, Huffman GJ, Bolvin D, Nelkin E, Chang A, Ferraro R, Gruber A, Xie PP, Janowiak J, Rudolf B, Schneider U, Curtis S, Arkin P. 2003. The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979 - Present). *Journal of Hydrometeorology* **4**(6): 1147–1167, doi:10.1175/1525-7541(2003)004⟨1147:TVGPCP⟩2.0.CO;2.

Albergel C, Dorigo W, Balsamo G, Muñoz Sabater J, de Rosnay P, Isaksen L, Brocca L, de Jeu R, Wagner W. 2013. Monitoring multi-decadal satellite earth observation of soil moisture products through land surface reanalyses. *Remote Sensing of Environment* **138**: 77–89, doi:10.1016/j.rse.2013.07.009, URL `http://linkinghub.elsevier.com/retrieve/pii/S0034425713002216`.

Balmaseda MA, Mogensen K, Weaver AT. 2013. Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society* **139**(674): 1132–1161, doi:10.1002/qj.2063, URL `http://doi.wiley.com/10.1002/qj.2063`.

Balsamo G, Beljaars A, Scipal K, Viterbo P, van den Hurk B, Hirschi M, Betts AK. 2009. A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *Journal of Hydrometeorology* **10**(3): 623–643, doi:10.1175/2008JHM1068.1, URL `http://journals.ametsoc.org/doi/abs/10.1175/2008JHM1068.1`.

Berner J, Shutts GJ, Leutbecher M, Palmer TN. 2009. A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System. *Journal of the Atmospheric Sciences* **66**(3): 603–626, doi:10.1175/2008JAS2677.1, URL `http://journals.ametsoc.org/doi/abs/10.1175/2008JAS2677.1`.

Beven K, Cloke H, Pappenberger F, Lamb R, Hunter N. 2014. Hyperresolution information and hyperresolution ignorance in modelling the hydrology of the land surface. *Science China Earth Sciences* doi:10.1007/s11430-014-5003-4, URL `http://link.springer.com/10.1007/s11430-014-5003-4`.

Brier GW. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78**(1): 1–3, doi:10.1175/1520-0493(1950)078⟨0001:VOFEIT⟩2.0.CO;2, URL `http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2`.

Bröcker J, Smith LA. 2008. From ensemble forecasts to predictive distribution functions. *Tellus, Series A: Dynamic Meteorology and Oceanography* **60 A**: 663–678, doi:10.1111/j.1600-0870.2008.00333.x.

Carsel RF, Parrish RS. 1988. Developing joint probability distributions of soil water retention characteristics. *Water Resources Research* **24**(5): 755–769, doi:10.1029/WR024i005p00755, URL `http://doi.wiley.com/10.1029/WR024i005p00755`.

Cloke HL, Pappenberger F, Renaud JP. 2008. Multi-method global sensitivity analysis (MMGSA) for modelling floodplainhydrological processes. *Hydrological Processes* **22**(11): 1660–1674, doi:10.1002/hyp.6734, URL `http://doi.wiley.com/10.1002/hyp.6734`.

Cosby BJ, Hornberger GM, Clapp RB, Ginn TR. 1984. A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. *Water Resources Research* **20**(6): 682–690, doi:10.1029/WR020i006p00682, URL `http://doi.wiley.com/10.1029/WR020i006p00682`.

Dee DP, Uppala SM, Simmons aJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda Ma, Balsamo G, Bauer P, Bechtold P, Beljaars aCM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer aJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kå llberg P, Köhler M, Matricardi M, McNally aP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137**(656): 553–597, doi:10.1002/qj.828.

FAO. 2014. Digital soil map of the world. URL `http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/`.

Ferro CaT, Fricker TE. 2012. A bias-corrected decomposition of the Brier score. *Quarterly Journal of the Royal Meteorological Society* **138**(668): 1954–1960, doi:10.1002/qj.1924, URL `http://doi.wiley.com/10.1002/qj.1924`.

Fischer EM, Seneviratne SI, Lüthi D, Schär C. 2007a. Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophysical Research Letters* **34**(6): L06 707, doi:10.1029/2006GL029068, URL `http://doi.wiley.com/10.1029/2006GL029068`.

Fischer EM, Seneviratne SI, Vidale PL, Lüthi D, Schär C. 2007b. Soil MoistureAtmosphere Interactions during the 2003 European Summer Heat Wave. *Journal of Climate* **20**(20): 5081–5099, doi:10.1175/JCLI4288.1, URL `http://journals.ametsoc.org/doi/abs/10.1175/JCLI4288.1`.

Guo Z, Dirmeyer PA. 2013. Interannual Variability of LandAtmosphere Coupling Strength. *Journal of Hydrometeorology* **14**(5): 1636–1646, doi:10.1175/JHM-D-12-0171.1, URL `http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-12-0171.1`.

Hillel D. 1998. *Environmental Soil Physics*. Academic Press, ISBN 9780123485250.

Koster RD, Dirmeyer PA, Guo Z, Bonan G, Chan E, Cox P, Gordon CT, Kanae S, Kowalczyk E, Lawrence D, Liu P, Lu CH, Malyshev S, McAvaney B, Mitchell K, Mocko D, Oki T, Oleson K, Pitman A, Sud YC, Taylor CM, Verseghy D, Vasic R, Xue Y, Yamada T. 2004. Regions of strong coupling between soil moisture and precipitation. *Science (New York, N.Y.)* **305**: 1138–1140, doi:10.1126/science.1100217.

Miralles DG, van den Berg MJ, Teuling aJ, de Jeu RaM. 2012. Soil moisture-temperature coupling: A multiscale observational analysis. *Geophysical Research Letters* **39**(21): n/a–n/a, doi:10.1029/2012GL053703, URL `http://doi.wiley.com/10.1029/2012GL053703`.

Palmer TN. 2012. Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction. *Quarterly Journal of the Royal Meteorological Society* **138**(665): 841–861, doi:10.1002/qj.1923, URL `http://doi.wiley.com/10.1002/qj.1923`.

Palmer TN, Buizza R, Jung T, Leutbecher M, Shutts GJ, Steinheimer M, Weisheimer A. 2009. Stochastic Parametrization and Model Uncertainty. Technical report, ECMWF technical memorandum.

Richards LA. 1931. Capillary conduction of liquids through porous mediums. *Physics* **1**(5): 318, doi:10.1063/1.1745010, URL `http://scitation.aip.org/content/aip/journal/jap/1/5/10.1063/1.1745010`.

Robine JM, Cheung SLK, Le Roy S, Van Oyen H, Griffiths C, Michel JP, Herrmann FR. 2008. Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes rendus biologies* **331**(2): 171–8, doi:10.1016/j.crvi.2007.12.001, URL `http://www.ncbi.nlm.nih.gov/pubmed/18241810`.

Rodell M, Houser PR, Berg AA, Famiglietti JS. 2005. Evaluation of 10 Methods for Initializing a Land Surface Model. doi:10.1175/JHM414.1.

Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, Orlowsky B, Teuling AJ. 2010. Investigating soil moistureclimate interactions in a changing climate: A review. *Earth-Science Reviews* **99**(3-4): 125–161, doi:10.1016/j.earscirev.2010.02.004, URL `http://linkinghub.elsevier.com/retrieve/pii/S0012825210000139`.

Seneviratne SI, Lüthi D, Litschi M, Schär C. 2006. Land-atmosphere coupling and climate change in Europe. *Nature* **443**(7108): 205–9, doi:10.1038/nature05095, URL `http://www.ncbi.nlm.nih.gov/pubmed/16971947`.

Seneviratne SI, Mueller B, Koster RD, Orth R. 2012. Role of soil moisture for (sub-) seasonal prediction. In: *ECMWF Seminar on Seasonal prediction: science and applications 3-7 September*. URL `http://old.ecmwf.int/publications/library/do/references/list/3072013`.

Shao Y, Irannejad P. 1999. On the choice of soil hydraulic models in land-surface schemes. *Boundary-Layer Meteorology* **90**(1): 83–115.

Shrestha R, Houser P. 2010. A heterogeneous land surface model initialization study. *Journal of Geophysical Research: Atmospheres* **115**(19), doi:10.1029/2009JD013252.

Shutts G. 2005. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society* **131**(612): 3079–3102, doi:10.1256/qj.04.106, URL `http://doi.wiley.com/10.1256/qj.04.106`.

Taylor CM, Birch CE, Parker DJ, Dixon N, Guichard F, Nikulin G, Lister GMS. 2013. Modeling soil moisture-precipitation feedback in the Sahel: Importance of spatial scale versus convective parameterization. *Geophysical Research Letters* **40**: 6213–6218, doi:10.1002/2013GL058511.

Trangmar BB, Yost RS, Uehara G. 1986. Application of Geostatistics to Spatial Studies of Soil Properties. *Advances in Agronomy* **38**(C): 45–94.

van Genuchten MT. 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils1. *Soil Science Society of America Journal* **44**(5): 892, doi:10.2136/sssaj1980.03615995004400050002x, URL `https://www.soils.org/publications/sssaj/abstracts/44/5/SS0440050892`.

Weisheimer A, Corti S, Palmer T, Vitart F. 2014. Addressing model error through atmospheric stochastic physical parametrizations: impact on the coupled ECMWF seasonal forecasting system. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **372**(2018): 20130 290–20130 290, doi:10.1098/rsta.2013.0290, URL `http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.2013.0290`.

Weisheimer A, Doblas-Reyes FJ, Jung T, Palmer TN. 2011a. On the predictability of the extreme summer 2003 over Europe. *Geophysical Research Letters* **38**(5): n/a–n/a, doi:10.1029/2010GL046455, URL http://doi.wiley.com/10.1029/2010GL046455.

Weisheimer A, Palmer TN, Doblas-Reyes FJ. 2011b. Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophysical Research Letters* **38**, doi:10.1029/2011GL048123.

Wilks DS. 2011. *Statistical Methods in the Atmospheric Sciences*. International Geophysics, Academic Press, 3rd edn, ISBN 9780123850232.