

Representing the Earth surfaces in the Integrated Forecasting System: Recent advances and future challenges

G. Balsamo, A. Agustì-Panareda, C. Albergel,
A. Beljaars, S. Boussetta, E. Dutra, T. Komori, S. Lang,
J. Muñoz-Sabater, F. Pappenberger, P. de Rosnay,
I. Sandu, N. Wedi, A. Weisheimer, F. Wetterhall, E. Zsoter

Research and Forecast Departments

October 2014

Special topic paper on surface processes
presented at the 43rd ECMWF Scientific Advisory Committee, Reading, UK

This paper has not been published and should be regarded as an Internal Report from ECMWF.
Permission to quote from it should be obtained from the ECMWF.



Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

<http://www.ecmwf.int/publications/>

Contact: library@ecmwf.int

© Copyright 2014

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Abstract

An improved description of the Earth surfaces, with particular focus on soil, vegetation, snow and inland water-bodies, has been driving recent research efforts within the ECMWF Integrated Forecasting System (IFS). These developments are considered together with their associated data assimilation components and the representation of inherent uncertainties. The surface fluxes and reservoirs related to energy, water, and carbon cycles predicted by the IFS are applied in an increasing number of environmental applications. This paper presents the current status and future challenges when moving towards a more complete monitoring and modelling of the Earth system, with particular focus on the surface component.

1 Introduction

Products from the ECMWF Integrated Forecasting System (IFS) are applied in an increasing number of environmental applications, ranging from river hydrology, droughts and carbon cycle to dust and fire modelling. Improving the description and realism of continental surfaces and their interactions with the atmosphere has therefore been the driver of several recent changes in the IFS.

This follows a strategy of moving towards a more comprehensive Earth system modelling by representing the surface ecosystem processes and including progressively the carbon cycle and missing components of the energy and water cycles. The aim is to enhance ECMWF's capabilities to monitor, model and ultimately predict the state of the Earth surface in conjunction with its weather and climate. The surface modelling role has therefore increased in the ECMWF prediction systems, from specifying an accurate lower boundary condition to the atmosphere, to providing accurate information to a set of cascade applications. Avoiding compensating errors has always been important, but has become a higher priority with the multitude of applications and observations that are increasingly available.

The purpose of this paper is to give an overview of the progress that has been made over the last few years in modelling, data assimilation, verification and error estimation and to give an outlook to the future. This paper is structured into four sections. In section 2, the modelling advances for different land surfaces components introduced in the ECMWF operational forecasts in the past five years are described. In section 3, the land data assimilation context and its evolution are described, with a focus on improvements in the representation of soil and snow. The representation of land surface uncertainties and recent developments connected with the atmospheric ensemble data assimilation are included in section 4. Finally in section 5 the outlook for surface related developments is discussed in the wider context of Earth system modelling. Expected developments in surface parameterizations, data assimilation and environmental applications are presented together with the associated challenges. Section 6 contains concluding remarks.

2 Land Surface Modelling Advances

The land surface is an important component of the Earth system and numerical weather prediction and climate models are evolving continuously to include more of its natural complexity. The relevance of an accurate description of the land surface for atmospheric modelling has been widely established in

the scientific community for over thirty years (Manabe, 1969; Shukla and Mintz, 1982; Delworth and Manabe, 1988; Atlas et al., 1993). Land controls the partitioning of available energy at the surface between sensible and latent heat fluxes, with a strong impact on the atmospheric heat and moisture budgets. Land also determines the partitioning of available water between evaporation, drainage and runoff. Land water reservoirs and fluxes interact with the land morphology forming lakes and rivers.

Land surfaces influence weather and climate on all time and space scales, and respond actively to modifications of weather patterns and climate change. The role of land surfaces in Earth system models is to provide a consistent description of the water, energy and carbon exchanges (between atmosphere, biosphere, hydrosphere, and cryosphere) at various time scales ranging from hours to decades. Many sensitivity studies have shown that the description of physical processes of continental surfaces can significantly affect the prediction of meteorological variables such as precipitation, wind or temperature in the lower troposphere (Beljaars et al., 1996; Koster and Suarez, 1992; Wang and Kumar, 1998; Koster et al., 2004). Evapotranspiration directly affects weather parameters such as temperature, humidity, boundary layer development and clouds (Betts, 2009; Betts et al., 2013, 2014). Furthermore, a strong feedback between evaporation and precipitation exists which appears to be negative at the convective scale (Taylor et al., 2007) and positive at the continental scale (Hohenegger et al., 2009). This feedback involves soil water in the root-zone layer, which is one of the most important variables controlling large-scale continental summer temperatures (Seneviratne et al., 2012, 2014).

The way land surfaces store and regulate water and energy fluxes is firstly controlled by the large water bodies: namely soil moisture in the unsaturated zone, snow, ground water, lakes and open water. Such water bodies are de facto reservoirs of energy and water and have a “memory” much longer than atmospheric components. Secondly, energy and water fluxes are controlled by land use and biosphere, which are complex and very heterogeneous. At ECMWF, nearly all these aspects have received considerable attention over the last few years and work has resulted in many model upgrades that can be clustered in two areas of development.

- Enhanced realism of the representation of water and energy stocks in soil, snow and inland water bodies, via parameterizations and physiography revisions.
- Improved fluxes for land-atmosphere energy and water exchanges, inclusion of natural land carbon (CO₂), and improved runoff generation for river discharges.

These developments are summarised in the following subsections.

2.1 Soil

The soil is a porous medium that can store water, energy and carbon and these can be exchanged with the atmosphere and the oceans via transport mechanisms. The amount of water in the soil and its vertical distribution in the column are important for the regulation of heat and water vapour fluxes towards the atmosphere and involves a range of time scales from minutes to months in the coupled land-atmosphere system.

In the Hydrology – Tiled ECMWF Surface Scheme for Exchanges over Land (H-TESEL, Balsamo et al., 2009), a revised soil hydrology was developed and tested for the global scale (the scheme was initially proposed by van den Hurk and Viterbo, 2003 for the Baltic basin). These model developments were a response to known weaknesses of the TESSEL hydrology (van den Hurk et al., 2000) as used in the ERA-Interim reanalysis: specifically the choice of a single global soil texture, which does not characterize different soil moisture regimes, and a Hortonian runoff scheme which produces hardly any surface runoff. Therefore, a revised formulation of the soil hydrological conductivity and diffusivity (spatially variable according to a global soil texture map) and surface runoff (based on the variable infiltration capacity approach) were introduced into the IFS in November 2007.

Balsamo et al. (2009) verified the impact of the soil hydrological revisions from field site to global atmospheric coupled experiments and in data assimilation. Figure 1 illustrates the impact of these hydrology changes on the water budget of a number of European river catchments, simulated within the framework of the Global Soil Wetness Project (Dirmeyer, 2011). H-TESEL increases the seasonal amplitude of the Terrestrial Water Storage (TWS) change due the increased water holding capacity of the soil resulting from the new hydrological parameters and soil texture. Also H-TESEL compares better than TESSEL with the Hirschi et al. (2006) dataset which has been derived as a residual of atmospheric moisture convergence and river runoff.

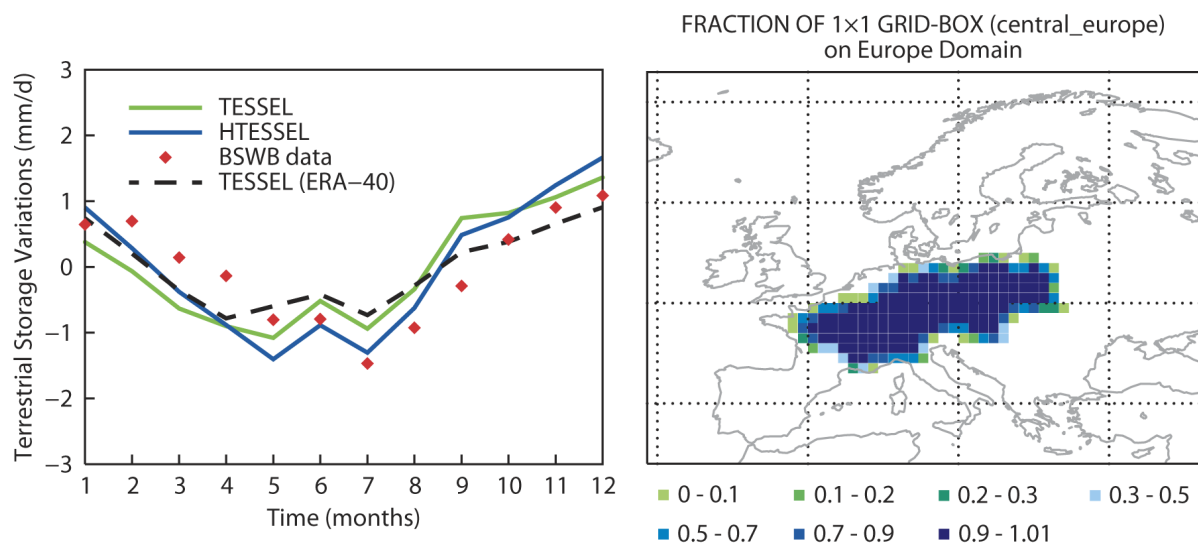


Figure 1: Monthly Terrestrial Water Storage (TWS) changes (left panel) for the Central European catchments Wisla, Odra, Elbe, Weser, Rhine, Seine, Rhone, Po, North-Danube (the coverage is shown in the right panel). The curves are for TESSEL (GSWP-2-driven, green line), H-TESEL (GSWP-2-driven, blue line), TESSEL in ERA-40 (black dashed line). The red diamonds are the Hirschi et al. (2006) monthly values derived from atmospheric moisture convergence and runoff for the years 1986–1995.

The bare ground evaporation was improved in November 2010 in conjunction with the Leaf Area Index (LAI) update (see subsection 2.3) as reported in Balsamo et al. (2011a). In this change the bare ground evaporation has been enhanced over deserts by adopting a lower stress threshold than for vegetation. Figure 2 illustrates the impact on soil moisture over the USA. The bottom panels indicate

that with the new scheme (bottom right) the August soil moisture in the western part of the USA is much lower than with the old scheme (bottom left). These changes correlate with the bare ground fraction (top left). This is clearly beneficial, as can be demonstrated by verification based on the Soil Climate Analysis Network (SCAN) over the USA (see the top left panel of Figure 2 for the location of the stations). The positive differences in the top right panel of Figure 2 indicate a reduction of the root-mean-square (RMS) error of soil moisture particularly at high bare ground fractions. Albergel et al. (2012) also demonstrate a better match with Soil Moisture and Ocean salinity (SMOS) satellite observations. These results are in agreement with previous findings by Mahfouf and Noilhan (1991).

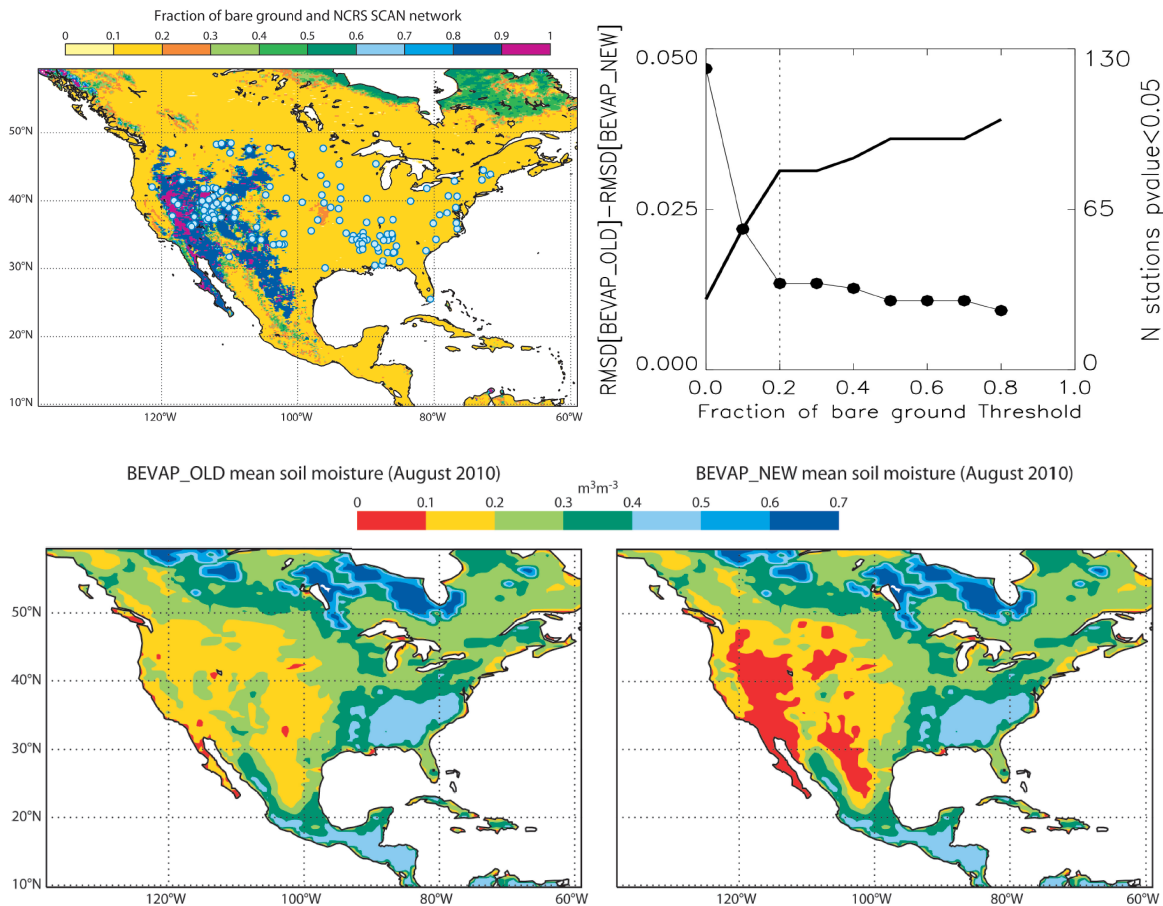


Figure 2: The top-left figure shows the fraction of bare ground and the distribution of soil moisture observation sites from the SCAN network marked by blue dots. Model errors are defined for these sites by the RMS difference with respect to observations. The top-right figure shows the soil moisture improvement in terms of the difference in RMS error (RMSD) between the ERA-Interim model version and the one used in 2010, as a function of the fraction of bare ground (thick solid line, left hand axis). A positive difference indicates that the model has been improved. The number of in situ stations with significant correlation is also presented (dots, right y-axis). The dashed vertical line represents a threshold where the sensitivity to the fraction of bare soil is less pronounced. The bottom panels illustrate the systematic change in soil moisture going from the ERA-Interim model version (bottom left) to the November 2010 scheme (bottom right), calculated as average over August 2010.

2.2 Snow and ice

The snowpack lying on top of the soil affects the evolution of atmospheric temperatures via its high albedo and its thermal insulation capacity that can create a decoupling between the Earth's surface and the atmosphere (Groisman et al., 1994; Viterbo and Betts, 1999; Beljaars et al., 2007; Cook et al., 2008; Ge and Gong, 2010). This snow insulating effect causes strong temperature inversions near the surface in winter, which represent a challenge for minimum temperature forecasting. Consequently, snow also affects the freezing of water in the soil, with an impact on hydrology in spring and on the near-surface temperatures and the stable boundary layer development (Viterbo et al., 1999). Snow cover also acts as a water reservoir, which is released by snowmelt in spring, influencing runoff, soil moisture, evapotranspiration and thus precipitation and the entire hydrological cycle (Groisman et al., 2004). Until 2009, the snow pack was parameterized in H-TESEL following Douville et al. (1995). The snow pack was represented with a single layer of dry snow (i.e. neglecting liquid water) with four snow prognostic variables: mass, albedo, density and temperature. Snow albedo decreased in time at an exponential or linear rate, for melting and normal conditions respectively, and snow density increased with time according to an exponential relaxation.

With the participation of H-TESEL in the snow models inter-comparison project 2 (SnowMIP2; Rutter et al., 2009), and after the soil hydrology revision presented in the previous subsection, several shortcomings of the snow pack representation were identified: lack of liquid water representation with freeze/thaw cycles during the melting season, unrealistic evolution of snow density, and unrealistic albedo of shaded snow (snow under high vegetation). These shortcomings were partially addressed with a full revision of the snow pack parameterization in 2009 (Dutra et al., 2010a) including: (i) a new parameterization of snow density, (ii) a liquid water reservoir and (iii) revised formulations for the sub-grid snow cover fraction and snow albedo. In offline mode, forced with near-surface observations, the revised scheme reduced the end of season ablation biases from 10 to 2 days in open areas, and from 21 to 13 days in forest areas. Figure 3 compares the evolution of snow mass, depth and density during one winter season at the Fraser forest and open stations, a research location in the Rocky Mountains, Colorado, USA. The results show the improvements of the snow scheme revision (NEW) with respect to the version used in ERA-Interim (CTR). The new snow density parameterization increased the snow thermal insulation, reduced soil freezing, improved the hydrological cycle, and substantially reduced a warm bias in winter compared to Siberian screen-level observations.

The reduction of 2-m temperature errors in winter over Siberia is also clear from long integrations as shown in Figure 4 where the DJF model climate for 2-m temperature is compared to the observation based CRU climate produced at the Climate Research Unity of the University of East Anglia. With the revised snow scheme, winter temperatures are typically 5°C lower over large parts of Siberia. Furthermore, the revised snow albedo for exposed and shaded snow reduced the systematic negative bias in surface albedo (Dutra et al., 2012). This initial revision kept the same number of prognostic variables (liquid water content is diagnosed, similarly to the soil water freezing), leading to a simple technical implementation accompanied by small changes in the land data assimilation.

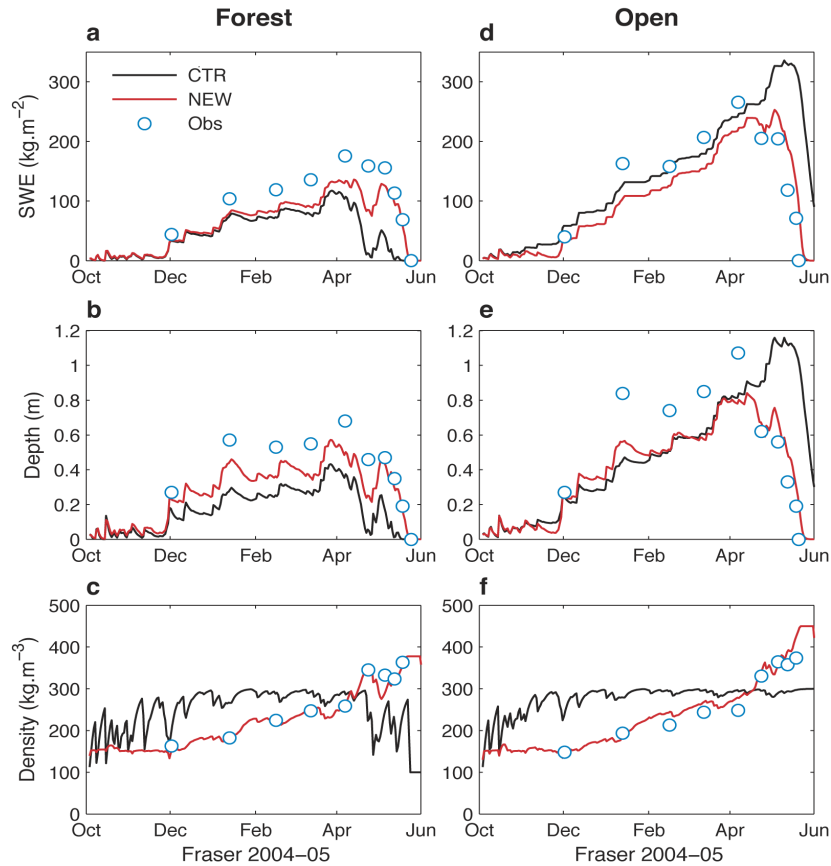


Figure 3: Offline simulation results for CTR (black line – before 2009 revision), NEW (red line – after 2009 revision) for the 2004–2005 winter season at Fraser forest (a-c) and open (d-f) sites: snow mass (a,d), snow depth (b,e) and snow density (c,f). Observations are represented by open blue circles.

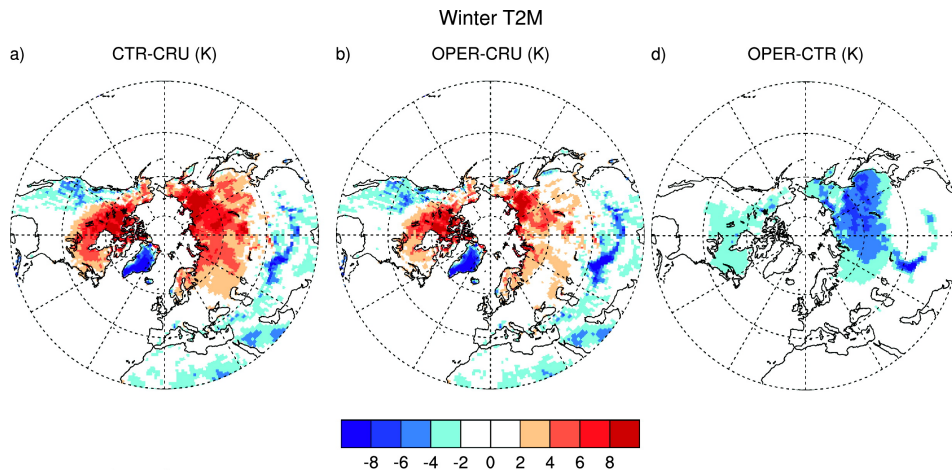


Figure 4: Impact of snow scheme revisions on the 2-m temperature model climate for DJF. The left panel shows the bias of the old model (with respect to the observation based CRU climate). The middle panel shows the bias of the new model. The right hand panel shows the difference between the two model versions.

2.3 Inland water bodies

Lakes and other inland water-bodies need to be represented in a land surface model because they influence the weather on local to regional scales. Their characteristics differ substantially from the surrounding land primarily due to the differences in albedo, roughness and heat storage (Mironov et al., 2010). Although, until recently neglected in most models, community efforts in NWP centres have triggered advances in the understanding of the importance of lakes (e.g. Salgado et al., 2010; Samuelsson et al., 2010).

Research aimed at introducing inland water bodies (lakes, rivers and coastal waters) into the operational model at ECMWF has started by considering a medium-complexity scheme that satisfies the operational constraint of having a low computational cost. FLake (Mironov et al., 2010), a shallow-water scheme originally applied to lakes, was introduced into the IFS in progressive steps. This model is a particularly appropriate choice as it predicts the vertical temperature structure and mixing conditions in lakes of various depths on time scales from a few hours to a few years, while maintaining a relatively low number of prognostics variables (7 in total). The model is intended for use as a lake parameterization scheme in NWP, climate modelling and other prediction systems for environmental applications. FLake has been implemented in the operational regional weather forecast model of Deutscher Wetterdienst (the German weather service) and is used for research at several meteorological services across Europe including Météo-France, UK Met Office and the Swedish Meteorological and Hydrological Institute.

Before implementation in the IFS, FLake has been evaluated in a series of preparatory studies: first in an offline experimental framework by Dutra et al. (2010b) and Balsamo et al. (2010), and then extended to fully-coupled lake-atmosphere simulations by Balsamo et al. (2012). More recently the possibility of treating sub-grid water bodies (lakes and coastal waters) using the land surface tiling methodology has been included. With this approach, each grid box is divided into fractions of different land use, each with their own tile. Manrique Suñén et al. (2013) have assessed the merits and limitations of the tiling methodology when accommodating lakes and forest within the same model gridbox.

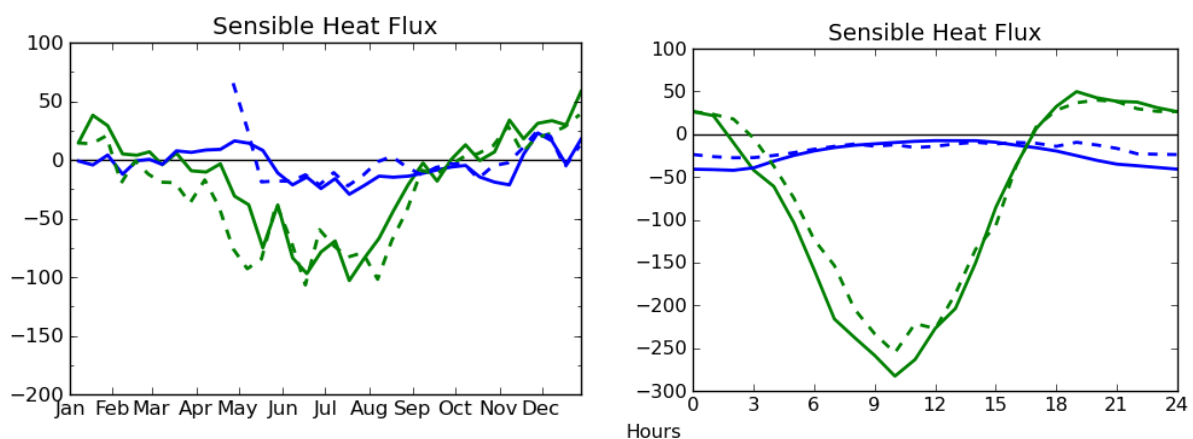


Figure 5: Sensible heat-flux over a lake (blue line) and a near-by forest (green line) in Finland. An annual cycle (left) and a mean July diurnal cycle (right) are shown for the model (solid line) and the flux-tower observations (dashed line).

They conclude that the tiling method captures very well the influence on surface fluxes of the contrast between the “lake” and “forest” surface boundary conditions, on time scales from diurnal to seasonal (see Figure 5). However, the contrast in aerodynamic coupling between the atmosphere and surface is not well represented because the tiles are blended at the lowest model level and horizontal advection between tiles is not represented.

The behaviour of temperature and ice duration over large inland water bodies has been verified using a satellite-based product for the lake surface temperature and ice cover (see Balsamo et al., 2012; Balsamo, 2013). The impact of introducing interactive inland water bodies in the IFS has been examined by a set of dedicated analysis experiments.

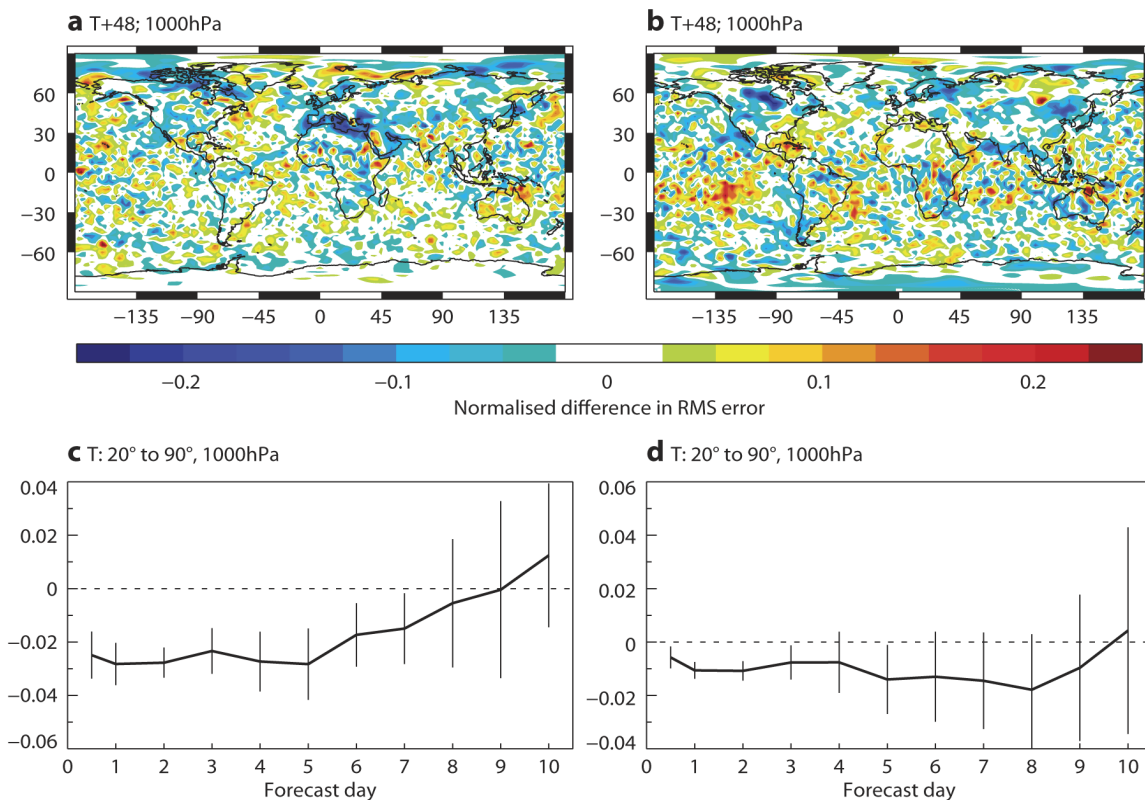


Figure 6: Forecast impact on near-surface temperature (at 1000 hPa) in terms of relative RMS error reduction, as a result of activation of the interactive lake and shoreline parameterization in (a) summer (15 June to 5 July 2013) and in (b) winter (1–31 December 2013). The bottom panels show the mean northern hemisphere relative RMS error difference for temperature at 1000 hPa with 95% significant improvements achieved over the first 7 days of forecast in summer (c) and over the first 3/5 days in winter(d). A value of -0.02 corresponds to a 2% relative improvement in the forecast.

Figure 6 shows a significant reduction of RMS errors compared to the own analysis in near-surface temperature (at 1000 hPa) up to day 8. The signal in summer is stronger than in winter and is also associated with shorelines that are consistently treated as sub-grid inland water bodies.

The lake model does not yet consider a mass balance of water (only for the ice components) and lakes are not allowed to dry out or extend. This constitutes a limitation for temporary or seasonal lakes in monsoon climate areas. Research efforts to include a mass balance and water hydrodynamics for rivers and flooded areas are discussed in section 5.

2.4 Vegetation and carbon cycle

The biosphere plays a prominent role in regulating fluxes of mass, energy and momentum into the atmosphere, so it is important to properly represent the biosphere in models. Parameterizations of the biosphere are simplified representations of the natural processes (spatial scales such as the plant, field or watershed will remain largely sub-grid in the foreseeable future) and can only describe the main feedback mechanisms sometimes marred by sizeable systematic errors. A key characteristic for water vapour and carbon fluxes is the so-called canopy resistance, which is a bulk representation of stomatal resistance, vegetation type and leaf area. The stomata are the leaf pores through which the plants absorb carbon and transpire water vapour.

In H-TESSSEL, the Leaf Area Index (LAI), which expresses the phenological phase of vegetation (growing, mature, senescent, dormant), was kept constant (van den Hurk et al., 2000) and assigned by a look-up table depending on the vegetation type. Thus vegetation appeared to be fully developed throughout the year. In November 2010, Boussetta et al. (2013a) introduced the seasonality of vegetation via a LAI monthly climatology based on the MODIS (collection 5) satellite product by Myneni et al. (2002). The new monthly LAI climatology is shown to affect particularly spring seasons when the radiative forcing is already strong, but the vegetation not yet fully developed. The sensitivity generally indicates a warming for spring as a consequence of lower LAI and reduced evaporation (and consequently more sensible heat flux). This results in a reduction of the systematic 2-m temperature errors in spring (Boussetta et al., 2013a).

More recently (in November 2011), H-TESSSEL has been extended with a carbon dioxide module based on the A-gs model (Calvet et al., 1998). This is a relatively simple vegetation and carbon dioxide model, which is highly suitable for the NWP environment where environmental factors are controlled by meteorological forcing and constrained by data assimilation. The model relates photosynthesis to radiation, atmospheric carbon dioxide (CO₂) concentration, soil moisture and temperature. Ecosystem respiration is based on empirical relations dependent on temperature, soil moisture, snow depth and land use (as detailed in Boussetta et al., 2013b). The CO₂ module parameters are optimized by vegetation type considering the Gross Primary Production (GPP) and Ecosystem Respiration (Reco), the CO₂ fluxes composing the Net Ecosystem Exchange (NEE) between biosphere and atmosphere.

The FLUXNET-based benchmarking system (<http://www.fluxdata.org/>), with flux tower observations in different climate regimes, was used for parameter optimization by minimizing flux errors. Subsequently, a different year of the FLUXNET data was used for verification. The seasonal cycle of NEE is illustrated in Figure 7 for six sites with different biomes. Two model configurations are shown: the first uses a stomatal resistance formulation for evaporation that is controlled by the photosynthesis module (C-TESSSEL), and the second uses the Jarvis-based stomatal resistance for evapotranspiration (CH-TESSSEL). Also the CASA climatology (Carnegie-Ames-Stanford-Approach, Potter et al., 1993) is shown because it is extensively used in the community and it was previously used as a boundary

condition for atmospheric CO₂ in the MACC (Monitoring Atmospheric Composition and Climate) project.

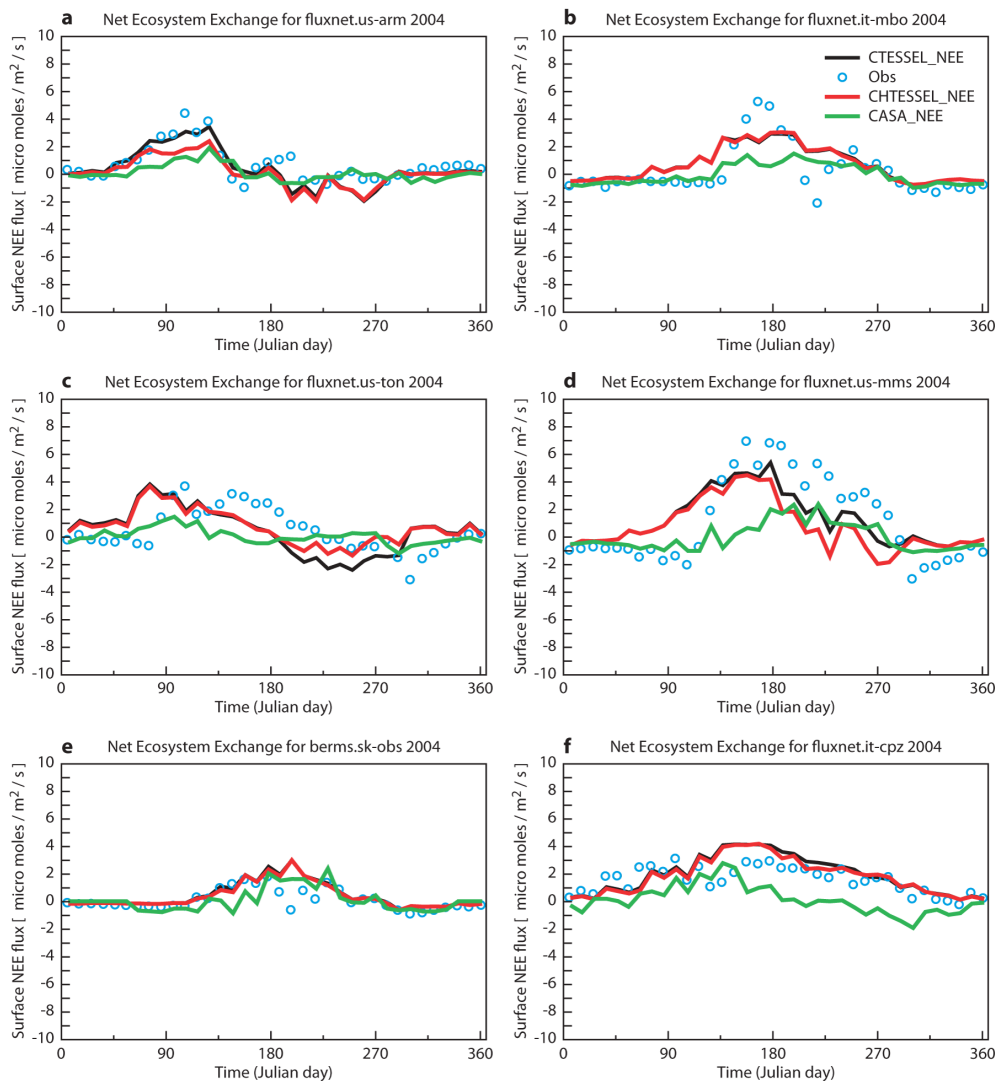


Figure 7: Seasonal cycle (2004) of 10-day averaged offline simulated (lines) and observed (blue dots) Net Ecosystem Exchange [$\mu \text{ mol m}^{-2} \text{ s}^{-1}$] for C-TESSSEL (with A-gs, black line), CH-TESSSEL (with Jarvis-type evaporation, red line) and CASA-GFED3 (green line) at different observation sites with different biomes.

Although it is difficult to draw firm conclusions, it is clear that errors in NEE are large, errors vary a lot from site to site, and differences between C-TESSSEL and CH-TESSSEL are small compared to the errors. The correlation between model NEE and observations averaged over 34 flux tower sites is 0.37 for CASA, 0.68 for C-TESSSEL and 0.65 for CH-TESSSEL. Both TESSSEL versions have a correlation of about 0.80 for sensible and latent heat fluxes (Boussetta et al., 2013b). The substantial improvement of C-TESSSEL/CH-TESSSEL with respect to the CASA climatology is significant because it suggests that the real-time meteorological variability is a key driver of the NEE variability.

Correlations coefficients of the two components of NEE, GPP and Reco, compared with tower observations (0.8 for both fluxes, on average over 34 sites) indicate much higher skill and confirm the robustness of these results. NEE is a small residual of GPP and Reco, therefore a correlation coefficient above 0.6 is highly relevant. Again, C-TESSSEL and CH-TESSSEL show similar performance, with site to site variability attributed to representativity of model grid-box (see Boussetta et al. 2013b). CH-TESSSEL has also been evaluated in coupled integration mode for the 2003 to 2008 period.

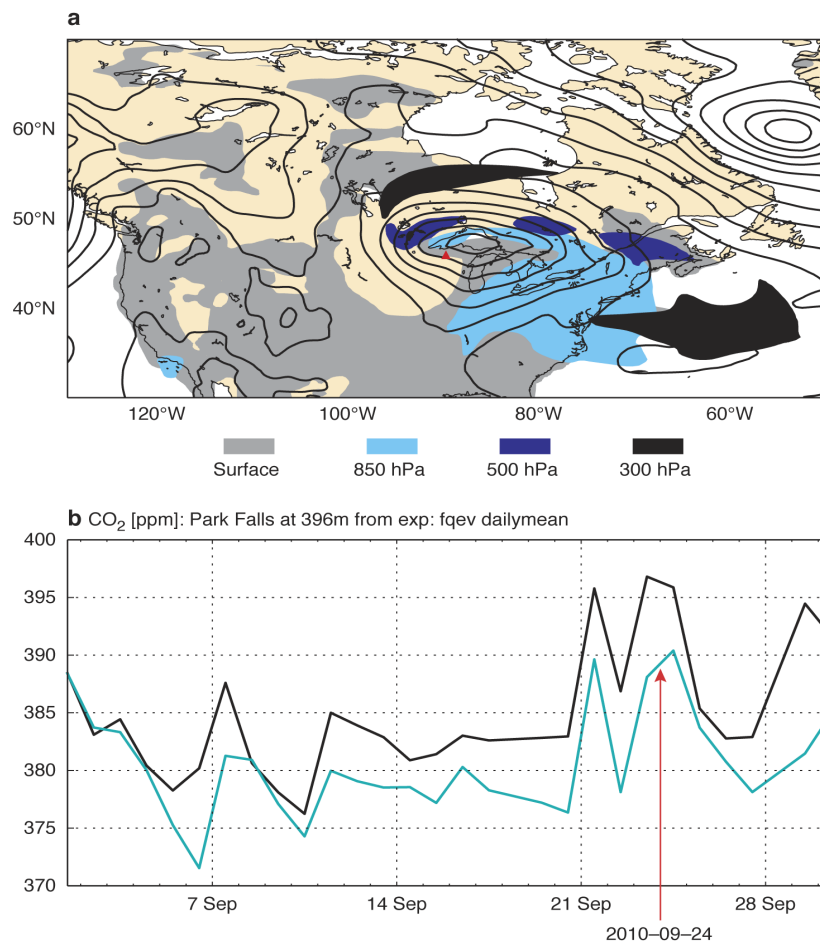


Figure 8: Atmospheric CO₂ anomalies associated with the passage of a low pressure system over N. America: (a) CO₂ anomalies above the well-mixed background CO₂ at different vertical levels: grey near the surface, cyan at 850 hPa, blue at 500 hPa and dark grey at 300 hPa on 24 September 2010. The anomalies are defined as CO₂ dry molar fraction above the background value of 392 ppm for both near the surface and at the 850 hPa levels, and above the background value of 388 for the 500 and 300 hPa levels. The locations of the observing site of the NOAA/ESRL tall tower at Park Falls (Wisconsin, USA) is depicted by a red triangle. The black contours of mean sea level pressure show the location of the centre of low pressure systems. (b) Daily mean dry molar fraction [ppm] of CO₂ from measurements (black) and forecasts (cyan) at Park Falls in September 2010. The observations are courtesy of NOAA/ESRL (Andrews et al., 2013).

The MACC configuration was used where daily forecasts are initialized with the analysed meteorological fields and atmospheric CO₂ with the previous 24-hour forecast (open loop). It shows that the global CO₂ atmospheric inter-annual variability is well simulated. The correlation of global CO₂ with observationally based estimates is 0.70.

CH-TESSSEL is now incorporated in the new global atmospheric real-time CO₂ forecasts which is now available as part of the pre-operational MACC system. A detailed documentation of the configuration and evaluation of the CO₂ forecast product is provided in Agustí-Panareda et al. (2014). The global CO₂ forecast has high skill in simulating day-to-day synoptic variability, which is crucial in order to be able to assimilate atmospheric CO₂ observations.

Figure 8 illustrates the spatial and temporal CO₂ synoptic anomalies associated with the passage of synoptic weather systems over North America. The CO₂ forecast can represent the peaks of CO₂ observed at Park Falls (Wisconsin, USA) originating mainly from the advection of high CO₂ anomalies generated at the surface within the warm conveyor belt of synoptic low-pressure systems. Modelling day-to-day variability of the CO₂ fluxes from vegetation compared to using equivalent monthly mean fluxes with a diurnal cycle enhances significantly the atmospheric CO₂ variability and skill. Again, this illustrates the advantage of modelling the CO₂ fluxes inside the IFS with real-time meteorology.

2.5 Physiography and resolution

Land surface processes and parameters strongly depend on land use, vegetation cover and soil type, and therefore the climatological fields describing these characteristics are a key part of any land surface scheme. For every grid point, the ECMWF model has values for surface elevation (orography), sub-grid orography statistics, land cover (used as land/sea mask), lake cover and depth, glacier cover, low and high vegetation type, low and high vegetation cover, albedo, LAI, and soil texture. These fields are derived from different external sources (e.g. albedo and LAI from MODIS, and vegetation type and cover from GLCC/AVHRR).

The global datasets come in different resolutions, data formats and projections and need to be interpolated, or merged in the case of non-global coverage, to a suitable reduced Gaussian grid at all the relevant resolutions required by ECMWF forecasting suites. The severe limitations of existing software to test and explore the many emerging datasets from Earth Observation missions, the need to revise the 20–30 year old datasets used in operations in view of the changing climate, and the need for higher-resolution data information required to initialize ultra-high resolution experiments important for ECMWF's future strategy, led to the multi-year climatological data revision project. The initial aim was to start a global data repository in grib2 format at 1-km resolution for all climatological fields and to establish reproducible procedures that would allow an easier incorporation of future datasets.

A 1-km resolution represents a challenge for global NWP and Earth system models but it is relevant to allow direct comparison with present and future Sentinel missions monitoring the Earth system. This is also highly relevant for the future Copernicus services. In phase 1 of the project, which concludes with the operational implementation in December 2014 (see example in Figure 9), the underlying data sources for the land-sea mask, lake mask, mean and sub-grid orography fields, glacier information, and surface albedos have been changed. Figure 9 shows a merged orography and bathymetry as an

illustration. The lake depth is based on Kourzeneva (2010) and the ocean bathymetry is based on ETOPO1 by Amante et al. (2009).

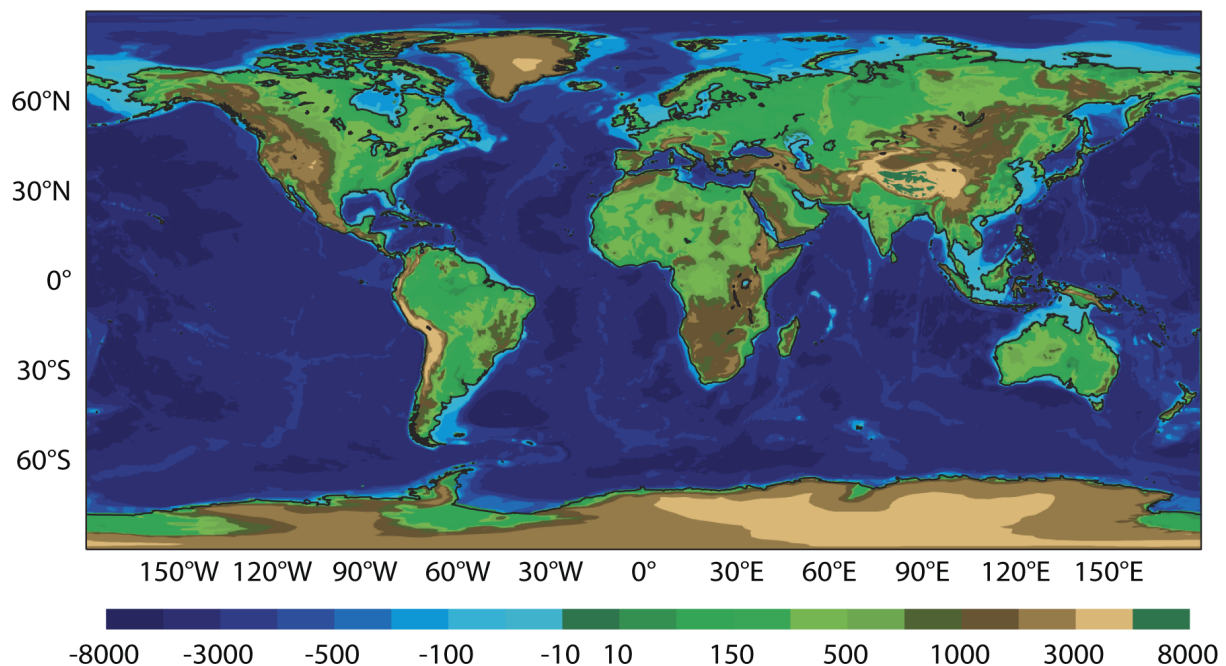


Figure 9: Combined land orography and ocean+lakes bathymetry in m (asl and bsl respectively) at T1279 resolution as produced in the new climate fields.

The land-sea mask and orography are based on the following raw data information: ESA's Globcover V2.2 based on Envisat MERIS (300 m resolution) mapping 2005/2006 (ESA, 2010), the Shuttle Radar Topography Mission (SRTM30) dataset provided by the US geological survey (SRTM, 2004) at about 90 m resolution, the Global Land One-kilometer Base Elevation (GLOBE, 1999) (only north of 60°N and south of 60°S), and specialised DEMs of Greenland (BPRC, 2002), Iceland (IMO, 2013), and Antarctica (RAMP2, Liu et al., 2001) replacing the corresponding data points on the 1 km latitude/longitude grid. The lake mask has been created from the land sea mask and complemented by consistency algorithms. The surface albedos representative of different spectral bands are based on a 0.05° (approx. 5 km) MODIS 5-year gap-filled and snow free product provided by the University of Massachusetts Boston. The albedos are specified in parametric form but for consistency with the approach in current operations, they are pre-calculated at local solar noon.

New data has not yet been introduced for lake depth, soil type, and vegetation type and cover. While new datasets and procedures exist for these fields, the large differences and consequences for the hydrological cycle must be assessed in the future. The aim of phase 1 is to minimize the impact at existing coarser resolutions without the need for additional model developments. However, the experiments clearly showed substantial sensitivities to the procedures employed when processing the underlying physiographic fields, particularly the filtering of the sub-grid orography fields.

A future upgrade is planned where the parametric information (isotropic, volumetric, geometric) will be used directly in the IFS. In that case the model will calculate the surface albedos dependent on the

actual solar zenith angle. All spatial resolutions that are used by the IFS have been prepared and can be run in fully coupled forecasts (Wedi et al., 2014).

2.6 Land-atmosphere coupling

Producing accurate forecasts of near-surface parameters becomes increasingly important as near-surface temperature, humidity and winds are used in a widening range of applications and are themselves highly relevant during extreme events (e.g. wind-storms, droughts and heat-waves).

The representation of land-atmosphere interactions depends on a large number of processes. The exchange of energy or moisture between the land surface and the atmosphere plays an important role for near-surface temperatures and humidity; while the roughness of the surface exerts the main control on near-surface winds. Yet, the parameterization of these interactions is hindered by difficulties in estimating the land-atmosphere coupling strength from theory or observations. Thus the parameterization relies on a number of parameters, set to poorly constrained values, that depend at most on the vegetation type or sub-grid information (tile fractions). The skin layer conductivity, the roughness length for heat or momentum, the minimum canopy resistance, the dependence of the canopy resistance on the water vapour pressure deficit and the root distribution over the soil layers are just a few examples.

In the past four years, efforts have been made to (i) evaluate the sensitivity of near-surface variables to the values chosen for such parameters, and (ii) find ways in which to better constrain them (Sandu et al., 2011, 2013). The first obvious candidate was the roughness length for momentum, which has a strong influence on the representation of 10-m wind speeds. For many years, ECMWF forecasts overestimated the near-surface (10 m) wind speed over land. The mean forecast errors, with respect to routine SYNOP observations, range roughly between 0.5 and 1 m/s for various regions/times of the day (see for example the errors for Europe at 00 and 12 UTC up to November 2011 shown in the bottom panel of Figure 10). A stratification of the 10-m wind speed forecast errors by vegetation type showed that these errors depend on the vegetation type, or more precisely on the momentum roughness length of each vegetation type (see top panel of Figure 10). The overestimation of the near-surface wind speed for most of the vegetation types suggests that the values used for the momentum roughness length were too low. This motivated the revision of these values based on theoretical considerations and SYNOP observations of wind speed at 10 m. The basic idea was to search, for each vegetation type, for a new value of the momentum roughness length for which the mean 10-m wind speed forecast error with respect to SYNOP observations drops to zero. This calibration showed that the momentum roughness length values should be increased for nine and decreased for one of the 18 vegetation types characterizing land areas. The newly derived values were introduced in IFS cycle 37r3 (November 2011). As the roughness length for momentum was on average increased, the roughness length for heat was decreased in order to account for terrain heterogeneity.

The use of the new momentum roughness length values reduced significantly the 10-m wind speed forecast errors, especially for the predominant vegetation types, but also on average over continental areas. The improvement for Europe was clear both in the pre-operational testing and in the operational verification afterwards (see Figure 10). The reduction of the heat roughness length had also a positive impact on the 2-m temperatures, leading to a small warming during night-time and cooling during daytime over the continental regions for both winter and summer, which reduced the forecast biases over Europe (Sandu et al., 2011).

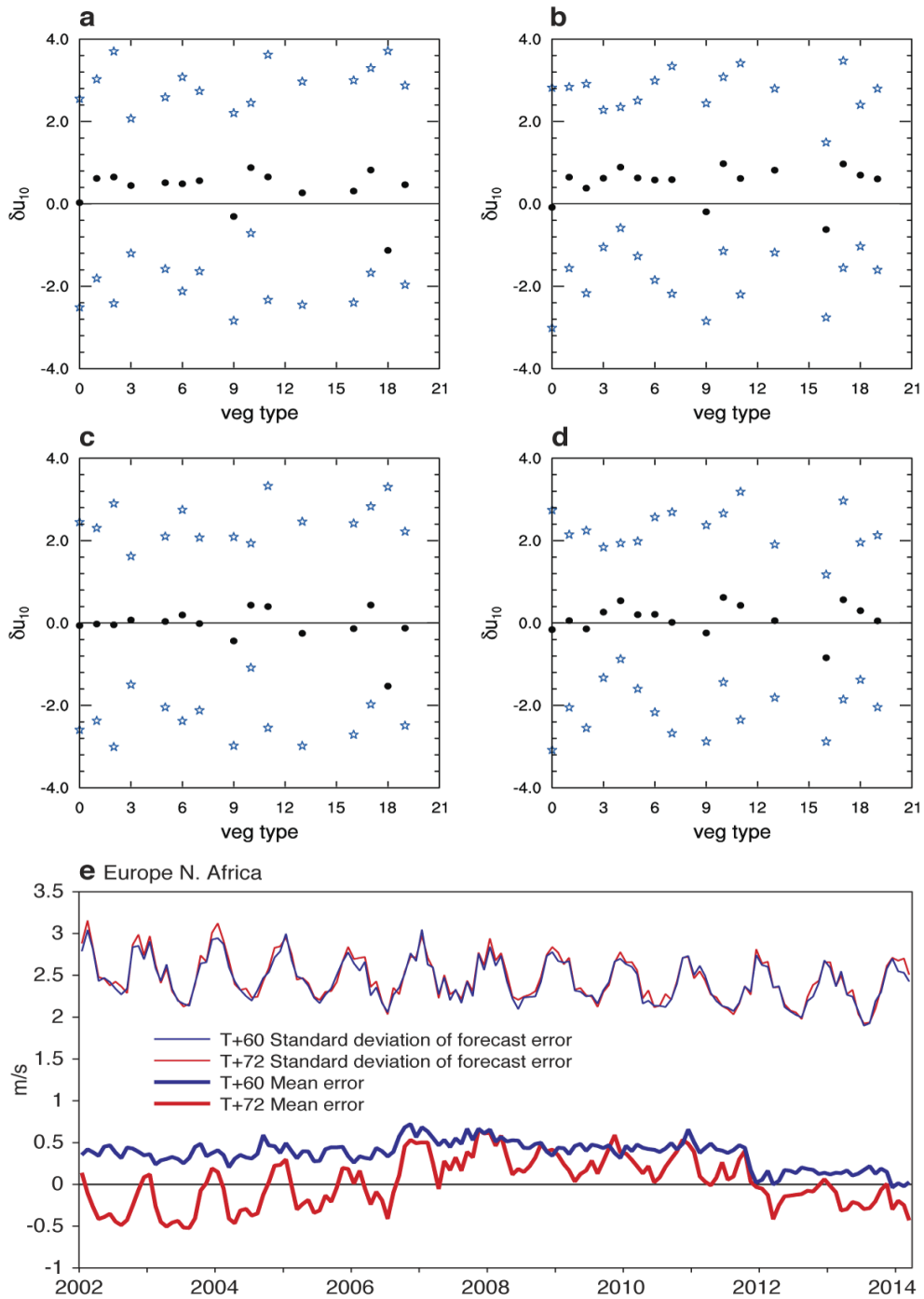


Figure 10: Mean bias and standard deviation with respect to routine SYNOP observations for daytime 10-m winds (m/s) in snow free locations during February 2010 (left), and August 2010 (right), as a function of the vegetation type of the respective locations. The upper/middle panels show results from step 36 of T511 forecasts initialized at 00 UTC from their own analysis performed with the control and the new roughness length table. The vegetation type associated to each of these pairs is the dominant vegetation type for the closest grid point to the SYNOP station. The bottom panel shows the near-surface wind speed bias and standard deviation with respect to routine SYNOP observations, for Europe at 00 (blue) and 12 UTC (red).

3 Land Surface Data Assimilation Advances

Accurate initialisation of the prognostic variables of the land surface model is of crucial importance for the quality of near-surface weather forecasts. A number of data assimilation studies have shown a significant impact of soil moisture conditions on weather forecast skill at short and medium ranges (de Rosnay et al., 2011b, 2013; van den Hurk et al., 2008; Drusch and Viterbo, 2007; Mahfouf et al., 2000). Drusch et al. (2008) demonstrated that soil moisture analysis improved prediction in the monthly and seasonal forecasts due to its long memory. Initialisation of snow conditions was also shown to have a large impact on the accuracy of atmospheric forecasts in both the medium (de Rosnay et al., 2011a; de Rosnay et al., 2014; Drusch et al. 2004; Brasnett, 1999) and the sub-seasonal forecasting range (see SAC paper ECMWF/SAC/43(14)6).

Research into data assimilation for continental surfaces has made considerable progress in going towards the adoption of schemes based on optimal estimation theory. The main advantage of this family of data assimilation techniques is that they preserve the physical internal consistency of the prognostic quantities (in the land analyses at ECMWF: soil moisture, soil temperature, snow temperature and snow depth). The Land Data Assimilation System (LDAS) combines information from the H-TESEL model background, the meteorological situation, and available observations (directly or indirectly informative about land quantities) taking into account their statistical error characteristics.

Advances in atmospheric data assimilation have provided inspiration for the land developments but with simplified and adjusted methods that are pertinent to land surface specificities and needs. Surface analyses are characterized by two factors distinguishing them from the atmospheric analysis: (i) lack of direct observations over large parts of continental areas, and (ii) high spatial variability of surface parameters due to natural heterogeneities.

Due to these specifics, most continental surface assimilation techniques applied in operational NWP use the following three assumptions:

- Land surface variables are independent from atmospheric variables within the assimilation window (i.e. the assumption of truncated variable space) that permits an analysis of the surface state separate from the four-dimensional variational atmospheric analysis. This is based on the separation of time scales.
- Land surface points are horizontally decoupled. This is related to the land surface model representation, with the vertical soil column and time as the only dimensions along which prognostic variables evolve. For data assimilation this reduces the four-dimensional analysis problem to a two-dimensional problem.
- The tangent linear approximation for model and observation operators is sufficiently accurate to relate observed atmospheric/superficial quantities to water and temperature in the soil column.

These assumptions can be experimentally verified (e.g. Hess, 2001; Balsamo et al., 2004, 2007; Mahfouf et al., 2008) and permit the surface analysis to be treated separately from the atmospheric analysis and as a collection of independent low dimensional problems rather than as a global-scale analysis problem.

Advanced methods using static approaches (such as the Optimal Interpolation as implemented by Mahfouf et al., 2000) or dynamic approaches (Variational or Kalman filters) have been the subjects of active research and development at ECMWF. A big advantage of dynamic approaches is that they evolve with the rest of the system and permit an easier extension to the assimilation of satellite observations, in combination with conventional observations (English et al., 2013).

Different methods have been tested (Mahfouf, 1991; Hess, 2001) and are used operationally in other NWP centres: optimal interpolation (OI), variational assimilation (VAR), and Kalman filters (KF, in several variants). Two types of assimilation methods are currently applied to land surfaces at ECMWF: OI and a Simplified Extended Kalman Filter (SEKF).

The subsections below present recent advances in the ECMWF land surface data assimilation, which is being continuously improved to provide an accurate initial state for the land variables.

3.1 Near-surface atmospheric fields

A key source of information for the land surface analysis lies in the near-surface atmospheric observations of 2-m temperature and relative humidity. These observations are available from the SYNOP network and they are assimilated to analyse the corresponding model diagnostic variables. The screen-level parameters analysis is conducted using a two-dimensional Optimal Interpolation (2D-OI), as developed by Mahfouf et al. (2000).

Figure 11 illustrates the ECMWF 2-m temperature (left) and relative humidity (right) monthly mean analysis increments for June 2013 at 12 UTC from the operational high-resolution suite. Over Europe, systematic positive increments of temperature at 12 UTC indicate a model cold bias in the afternoon as discussed in subsection 3.4, whereas over North America negative increments indicate a warm bias at night in the western part of the USA and positive increments over Eastern USA indicate a cold bias. Although the near-surface analysis does not have any direct feedback on the forecasting system, because diagnostic variables are analysed, it provides best estimates of actual 2-m temperature and relative humidity. They are used as such in the subsequent soil moisture and soil temperature analyses.

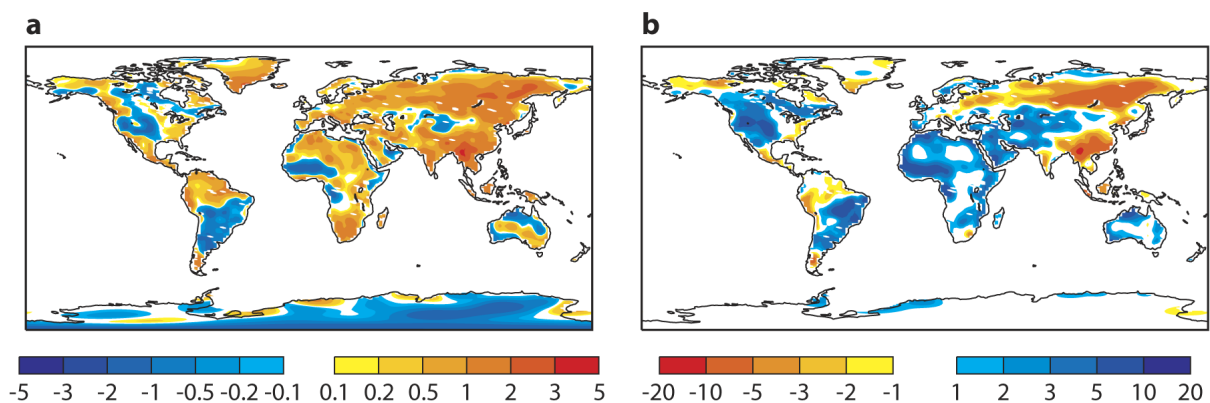


Figure 11: ECMWF 2-m temperature (a, in K) and relative humidity (b, in %) monthly mean analyses increments obtained at 12 UTC from the operational high resolution suite for June 2013.

3.2 Soil moisture analysis

The ECMWF soil moisture analysis is based on a point-wise SEKF (Seuffert et al., 2003; Drusch et al., 2009; de Rosnay et al., 2011b, 2013). This data assimilation scheme is particularly suitable for low-dimensional systems, such as the ECMWF soil moisture analysis. At each ECMWF 12-hour assimilation cycle and for each model grid point, the SEKF computes the soil moisture increments vector that includes the top three soil layers of the land surface scheme, with the objective of slightly adjusting the model first-guess soil moisture values. The soil moisture increment vector ($\Delta\theta$) is computed at analysis time (t_a) based on the difference between the observations vector (\mathbf{y}^o) and the model equivalent of the observations $H_i(\theta^b)$ at the time of the observations (t_o):

$$\Delta\theta(t_a) = \mathbf{K}[\mathbf{y}^o(t_o) - H_i(\theta^b)]$$

with the Kalman gain matrix \mathbf{K} expressed as:

$$\mathbf{K} = [\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}^T]^{-1} \mathbf{H}^T \mathbf{R}^{-1}$$

Here \mathbf{B} is the background error covariance matrix, \mathbf{R} is the observation error matrix and \mathbf{H} is the linearized observation operator bringing together the observations and the model equivalents to the same space. \mathbf{H} is computed by finite differences (i.e. by perturbing the model state vector by small amounts $\delta\theta$). For each element of the control state vector, perturbed model integrations are required; this is the main driver of the computational cost of the simplified EKF (de Rosnay et al., 2013). In contrast, the main advantage of the finite difference approach is that no adjoint of the land surface model is required. The \mathbf{B} and \mathbf{R} are currently fixed and static in time and space, with error standard deviations of background soil moisture, 2-m temperature analysis and 2-m relative humidity analysis being $\sigma^\theta = 0.01\text{m}^3\text{m}^{-3}$, $\sigma^T = 2\text{K}$ and $\sigma^{\text{RH}} = 10\%$, respectively. The greater the confidence in the observations, the more weight the assimilated observations will have in the computed soil moisture increments.

The SEKF was operationally implemented in the IFS in 2010 (de Rosnay et al., 2013). It replaced the previous one-dimensional OI (1D-OI) approach that was used from 1999 to 2010 for the deterministic analysis, and which is still used in ERA-Interim. Similar to the 1D-OI system, the SEKF uses the 2-m temperature and relative humidity analysis as input for the observation vector. They are available twice per 12-hour assimilation window at synoptic times. The SEKF implementation represented a major improvement to the operational land data assimilation system. In contrast to the 1D-OI, the SEKF makes dynamical estimates based on perturbed simulations, and allows the optimisation of soil moisture increments at different depths to match screen-level observations according to the strength of the local and current soil-vegetation-atmosphere coupling. Figure 12: Annual cycle of soil moisture increments in the first metre of soil (global mean value) in mm of water per month from January 2009 to November 2009, produced by the 1D-OI (red) and the SEKF (blue) soil moisture analysis scheme. Figure 12 shows the annual cycle of the global mean soil moisture increments for 1D-OI and SEKF experiments. It shows that the soil moisture increments of the 1D-OI scheme systematically add water to the soil, as discussed in the past by van den Hurk et al. (2008). The global mean value of the 1D-OI analysis increments is 5.5 mm per month, which represents a substantial and unrealistic contribution to the global water cycle. In contrast the simplified EKF global mean soil moisture analysis increments are much smaller, representing global mean increments of 0.5 mm per month. The reduction in soil moisture increments is particularly important for deeper layers. The SEKF computes Jacobians for

each soil layer separately and therefore it accounts for all the physical processes represented in the model. Compared to the 1D-OI, the SEKF allows for more controls on the vertical and temporal distribution of the increments due to meteorological forcing and soil moisture conditions. So, it prevents undesirable and excessive soil moisture corrections.

Figure 13 shows the overall impact of the SEKF soil moisture analysis, compared to the ‘Open Loop’ (which is without soil moisture analysis), on the 1000 hPa temperature and humidity forecasts for the extra-tropical northern hemisphere in June–July 2012. It shows a positive impact (error reduction), significant until day 3.5, in the near-surface temperature and humidity forecasts.

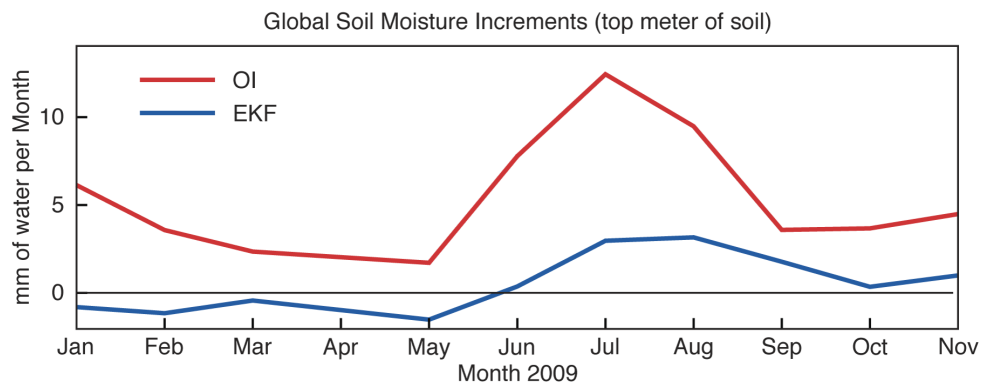


Figure 12: Annual cycle of soil moisture increments in the first metre of soil (global mean value) in mm of water per month from January 2009 to November 2009, produced by the 1D-OI (red) and the SEKF (blue) soil moisture analysis scheme.

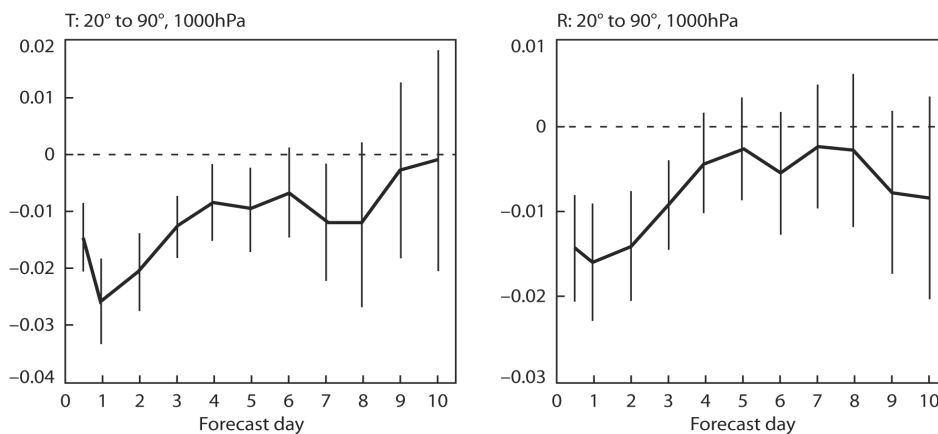


Figure 13: SEKF soil moisture analysis impact on the 1000 hPa temperature (left) and humidity (right) forecasts in the northern hemisphere for June–July 2012. The impact is expressed in terms of relative RMS error difference between T511 SEKF and soil moisture open loop experiments. A value of -0.02 corresponds to a 2% relative improvement in the forecast.

The SEKF soil moisture analysis was also developed because it can handle satellite data from microwave sensors, such as soil moisture from the Advanced Scatterometer (ASCAT) active system (Scipal et al., 2008; de Rosnay et al., 2013) and L-band brightness temperatures from the Soil

Moisture and Ocean Salinity (SMOS) passive system (Muñoz-Sabater et al., 2011). Both ASCAT and SMOS observations are currently monitored operationally and ASCAT soil moisture data assimilation will be implemented in operations end of 2014.

3.3 Snow depth analysis

The ECMWF snow data assimilation has been continuously improved in the past five years. In 2010 a 2D-OI snow analysis was implemented to replace the previous Cressman (1969) interpolation, and the use of snow cover data from NOAA (National Oceanic and Atmospheric Administration) NESDIS (National Environmental Satellite, Data, and Information Service) Interactive Multi-sensor Snow and Ice Mapping System (IMS) was revised to use high resolution 4 km data (de Rosnay et al., 2011a; de Rosnay et al., 2014). The snow OI at ECMWF has followed the implementation proposed by Brasnett (1999).

Figure 14 shows snow depth analysis fields in north-east Asia on 30 October 2010, obtained from the IFS when using a Cressman snow analysis (top) and an OI snow analysis (bottom). A qualitative comparison shows that the Cressman analysis produces disk-shaped spurious patterns of snow in northern Asia related to the Cressman interpolation. The OI presents a smoother and more correct snow analysis without spurious patterns. The OI analysis makes a better use of SYNOP snow depth data than Cressman. The difference between the two analyses mainly results from differences in the structure functions between OI and Cressman.

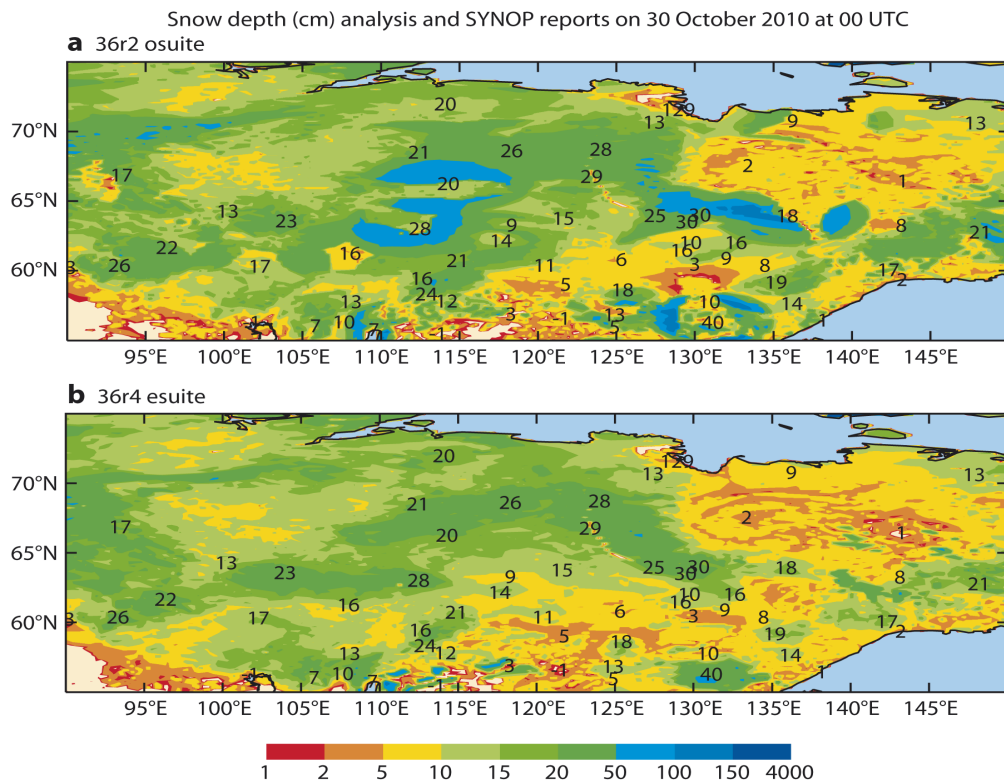


Figure 14: Snow depth fields obtained using a Cressman snow analysis (a) and the new OI snow analysis (b), in northern Asia on 30 October 2010. SYNOP snow depth measurements are printed in black.

In parallel to these system and scientific developments, significant efforts were dedicated to improve in situ snow depth data availability. In situ snow depth observations constitute a very reliable source of information for the snow data assimilation system, which largely influences the accuracy of the snow analysis and the quality of near-surface weather parameter forecasts. However large regions show extremely sparse SYNOP stations reporting snow depth. For some of these areas, national networks exist and provide near real-time data that could be made available through the GTS. In Europe, ECMWF took the initiative to develop a BUFR template for snow observations and encouraged member states to report their national network snow depth data at 06 UTC. So far six countries have put their national snow depth data on the GTS in addition to traditional SYNOP data: Sweden, the Netherlands, Denmark, Romania, Hungary and Norway. The data is now used in the operational ECMWF snow depth analysis. In 2013 this European initiative was extended to WMO in the context of the Snow Watch project of the Global Cryosphere Watch programme. Snow depth data providers are now encouraged by WMO to use the BUFR template to report their snow depth data through the GTS for NWP and other near-real-time applications. This initiative is expected to considerably improve the availability of snow depth data, particularly over the Americas, Russia and China where the majority of available in situ data is not reported onto the GTS.

3.4 Soil temperatures analysis

Soil temperature is an important variable for the representation of many physical processes in NWP. It is the key driver for all surface emissions of energy, carbon dioxide and water. Surface temperature is also needed in the assimilation of satellite data over land, such as SMOS. Yet, there is little information on the quality of this variable in NWP systems.

A one-dimensional optimum interpolation (1D OI) technique as described in Mahfouf (1991), Mahfouf et al. (2000) and Douville et al. (2001) is used to analyse the temperature of the uppermost layer of soil. The analysis increments from the screen-level temperature analysis are used to produce increments for the first layer of soil temperature and the model propagates the information downward and in time. Verification of short-range forecasts will therefore reflect the quality of the analysis as well as the forecast model. In a recent study (Albergel et al., 2014), soil temperature measurements from nearly 700 stations belonging to four networks across the USA and Europe were used to assess ECMWF soil temperature during 2012. To investigate the soil temperature diurnal cycle it was decided to use hourly forecasts instead of the analysis (available four times each day). Forecasts of soil temperature were initialized at 00 UTC with lead times from 0 to 23 hours. Evaluation of the time series shows a good performance of the day-1 forecasts in capturing both soil temperature annual and diurnal cycles with a very high level of correlation (0.92 and over), averaged root mean square differences ranging from 2.5°C to 3.9°C and averaged biases ranging from -0.5°C to 0.9°C. Results suggest that the IFS has good skill in predicting soil temperature, particularly in early morning. Over the USA, an overestimation of the diurnal amplitude was found, associated with a warm bias in the afternoon, while the opposite signal was observed over Europe (Ireland, Germany, Hungary and Czech Republic) presenting a cold bias in the afternoon. Figure 15 shows the diurnal cycles of mean soil temperature and RMS error, as average for the year 2012 and as averages for January to March (JFM) and July to September (JAS). The observations are from the USCRN (US Climate Reference Network). All data is transposed to local solar time (LST) before averaging. Figure 15 highlights ECMWF's warm bias at daytime when the RMS error is larger, particularly from 12 to 15 LST. As expected during daytime JAS higher temperatures present also higher RMS error comparing to JFM.

The orography dataset used to specify elevation was found to have a strong impact on results, as the difference between elevation of a station and that of the corresponding grid cell in the ECMWF model may lead to larger temperature differences (linked to linear processes resulting in a constant bias, as well as non-linear processes related to, for example, snow melt in spring). Based on results from one network spanning the entire United States, a clear relationship between biases of soil temperature and 2-m temperature was found. This verification study aims to contribute to a better understanding of the near-surface forecast errors, highlighting key processes that need to be better represented in future.

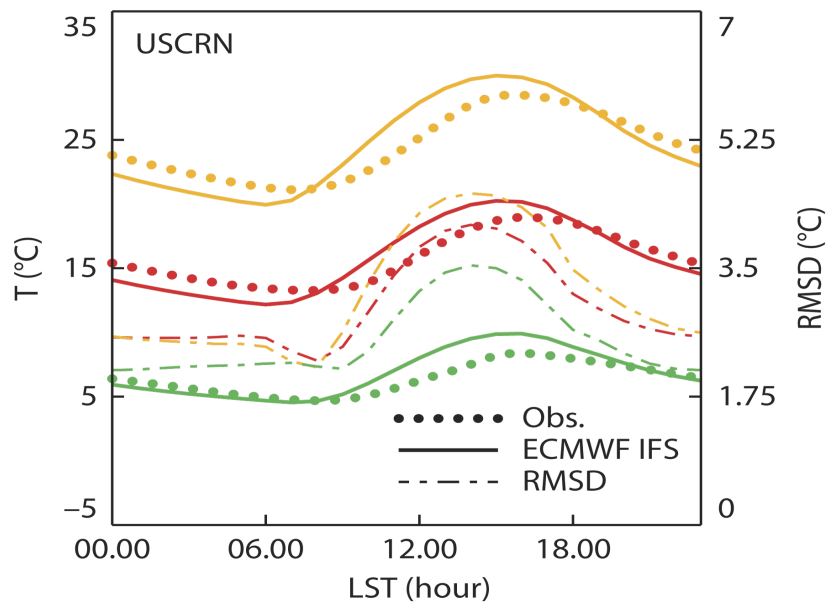


Figure 15: Mean diurnal soil temperature cycle for the USCRN spanning the entire USA for the year of 2012 (All in red), January to March (JFM in green) and July to September (JAS in orange). Dashed lines represent the mean diurnal RMS error between ECMWF 0 to 24 hour forecasts and observations.

4 Accounting for Land Surface Uncertainties

Under-dispersion and unreliability are well known issues (Hamill and Colucci, 1997) in ensemble forecasting and several methods have been adopted to adjust the spread-error ratio, including stochastic perturbation of initial conditions, perturbations of model tendencies and spectral backscattering (e.g. Leutbecher and Palmer, 2008; Berner et al., 2009). In order to improve the forecast reliability and ensemble dispersion, it is important to properly take into account the various categories of uncertainties, including those related to land surfaces. However, land surface uncertainties were not taken into account until recently due to non-linearity and non-gaussianity of errors, which often hinder the application of traditional methods of representing uncertainty in ensemble forecasting.

This has changed in recent years, and land surface uncertainties are starting to be accounted for both in local and global ensemble forecasting. A strategy for perturbing surface initial conditions designed in the context of local area forecasting (Wang et al., 2010) was shown to have beneficial impact in terms of near-surface weather parameters and precipitation, but also of temperature and humidity in the

lower troposphere. In global ensemble forecasting, land surface uncertainties can be accounted for in different ways. One way is by using Ensemble Kalman Filters, which implicitly take into account the uncertainty in the entire system (e.g. Houtekamer et al., 2014) or only the land surface component (e.g. Zhou et al., 2006; Reichle et al., 2008). Another way is by using an Ensemble of Data Assimilations (EDA) framework, such as that used at ECMWF (Isaksen et al., 2010).

The EDA system is run operationally at ECMWF to provide analysis uncertainty estimates for the high-resolution forecast (HRES) four-dimensional data assimilation (4DVAR) and the ensemble forecast (ENS). Within the EDA the observations as well as the model physics are perturbed to account for observation uncertainty and an imperfect forecast model.

The analysis uncertainty estimates from the EDA are combined with singular vector perturbations (Buizza and Palmer, 1995) to perturb the initial conditions of the ENS starting from the HRES 4DVAR analysis (Buizza et al., 2008, 2010). However, until IFS cycle 40r1 was implemented in November 2013, the surface observations in the EDA and the surface fields in the ENS remained unperturbed. Thus, all ENS members shared the same surface analysis at the initial time. In some circumstances this was detrimental in terms of the 2-m temperature ensemble spread (defined as the ensemble standard deviation).

In cycle 40r1, operational since 19 November 2013, land surfaces uncertainties were accounted for in the ENS by perturbing the surface fields used as initial condition for the ENS members (Lang et al., 2013). These perturbations are constructed using the EDA, similarly to what was already done for the atmospheric fields (subsection 4.3). By construction, they represent uncertainties related to the atmospheric forcing and to the upper-air observations, which are both represented in the EDA. An additional feature introduced in cycle 40r1, is that the screen-level observations of 2-m temperature and relative humidity, used as input to the LDAS system (see previous section), are also perturbed in each EDA member. This allows further increases in the spread of the EDA surface fields that are used for creating the perturbations of the ENS surface initial conditions. The contributions of the atmospheric forcing and of observation uncertainties to the uncertainties in land surface fields are discussed in the following two subsections.

4.1 Impact of upper-air meteorological forcing and SST uncertainties

One way to estimate the uncertainties induced by the meteorological forcing in the land surface is to consider an ensemble of land surface model runs driven by an ensemble of meteorological forcings produced by a long reanalysis.

One such reanalysis is ERA-20C (Poli et al., 2013). This pilot reanalysis, produced by the ERA-CLIM project (<http://www.era-clim.eu>) covers the years 1900–2013, features a horizontal resolution of 125 km (T159), and assimilates surface observations of pressure and marine winds. The uncertainties in this reanalysis are estimated by a 10-member EDA, where the following sources of error are represented: observations (perturbed), model (stochastic physics), and SST (10-member ensemble HadISST2.1.0.0).

Each member of ERA-20C can then be used as atmospheric forcing to run the offline version of the land surface model CH-TESSSEL at a horizontal resolution of 25 km (T799). This produces a set of 10 land surface reconstructions for the 20th century. This ensemble bears the name ERA-20CL and is due

to be released to users. These offline runs were produced by adapting the suite that was used for ERA-Interim/Land (Balsamo et al., 2012; Albergel et al., 2013). In ERA-20CL, the meteorological forcings at T159 are downscaled to the higher horizontal resolution (T799), adapting to orography and adjusting the lowest model level temperature, humidity, and pressure.

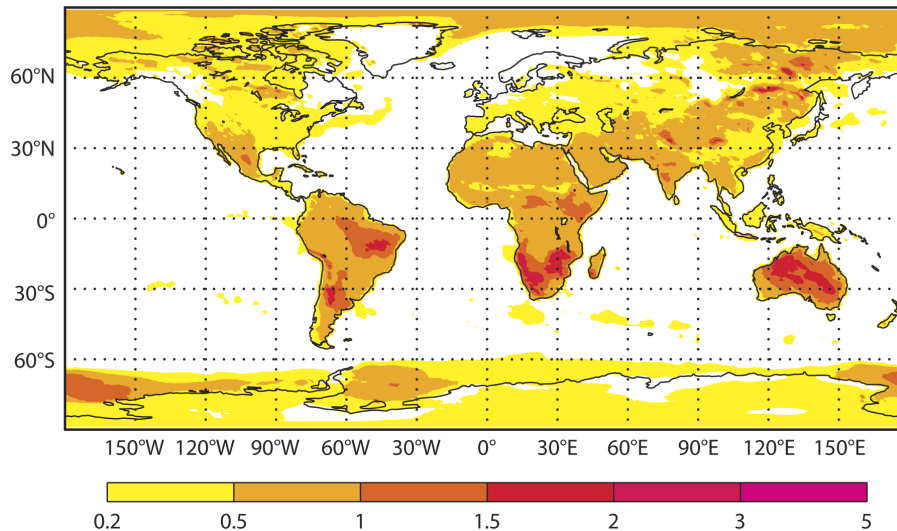


Figure 16: Soil temperature ensemble spread (stdev in K) obtained from the 10-member ERA-20CL at T799 resolution (November 2009-January 2010), representative of the meteorologically induced land surface spread.

Figure 16 shows that the spread in soil temperature caused by the meteorological forcing is below 0.5 K in well-observed regions, and up to 2 K over less observed continental areas.

However, this spread results from a reanalysis that assimilates only a fraction of the observations used by the current ECMWF operational suite (which assimilates also upper-air and satellite observations). Consequently, this spread can be considered as an upper bound of the land surface uncertainties induced by the meteorological forcing in the ECMWF operational suite. With the offline land surface model being used under the “perfect model” assumption the ensemble land surface estimates are likely to be still under-dispersive and a more complete uncertainty estimate of the land surface fields, should also represent model error for instance associated with parameters (see subsection 4.4).

4.2 Impact of land surface observation uncertainties

Another source of uncertainties in the land surface estimates are the observations entering in the analysis. The observations assimilated by the LDAS serve to update the land surface background fields so that, when and where such observations are present, the spread of these fields is reduced. A perturbation added to the observations assimilated by the LDAS is useful to account also for uncertainties in the analysis and contributes to a small increase in the ensemble spread. Isaksen et al. (2010) have shown that the analysis error is best estimated when all sources of uncertainties in the information used are represented. In the EDA the observation uncertainties were represented for upper-air observations, but were not represented for the observations assimilated in the LDAS until November 2013. Recent developments of the LDAS, related to a merge of the Observation Data Bases

(ODB) used by the upper-air and land surface analysis, enabled the use of the upper-air EDA observations perturbation framework for the land surface component. This development has therefore permitted the introduction of a consistent perturbation of the surface observations.

In IFS cycle 40r1, 2-m temperature and relative humidity observations were included in the perturbation system. These observations are used to produce the 2-m analyses of temperature and relative humidity which are used as input to the soil moisture, soil temperature and snow temperature analyses (de Rosnay et al., 2014). So, perturbing the observations in each EDA member is expected to increase the spread of the EDA surface fields. This is illustrated in Figure 17, which shows the increase in the soil temperature spread due to perturbation of the screen-level observations assimilated in LDAS. This indicates that an increase of soil temperature spread of about 0.5 K is obtained. In the future, it is expected that more perturbations of land surface observations (e.g. satellite soil moisture and snow depth) in the LDAS will be included in the EDA members.

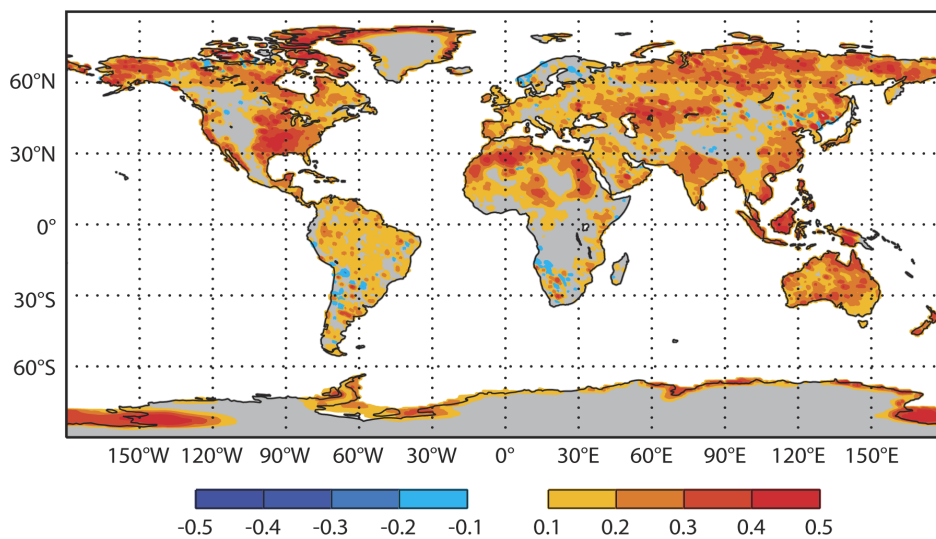


Figure 17: Mean differences of analysed surface (first soil layer, top 7cm) soil temperature ensemble standard deviation (in K) of an EDA with and without perturbed surface observations (average between 24 Nov. and 31 Dec. 2012). Positive values indicate larger spread for the experiment with perturbed surface observations.

4.3 Transferring EDA uncertainties to ENS uncertainties

Since November 2013, the EDA surface fields account not only for uncertainties related to the atmospheric forcing and upper-air observations, but also for uncertainties due to ground based observations. The EDA surface analyses fields then are used to derive perturbations for the land surface fields used as initial conditions for the ENS members, using the same perturbation strategy applied to the upper-air model variables (with the exception of singular vector perturbations, not applicable to surface fields). In the operational ENS, perturbations are generated from 6-hour forecasts (see Buizza et al., 2008) started from the 25 perturbed EDA members. The perturbations are added to the high-resolution (TL1279) analysis with a plus-minus symmetry:

$$x = x_{AN} \pm (x_{EDA,k} - \overline{x_{EDA}})$$

with the high-resolution analysis x_{AN} , the k^{th} EDA member $x_{EDA,k}$, the EDA mean $\overline{x_{EDA}}$ and the resulting perturbed analysis x . The perturbed surface fields are: the volumetric soil water content and soil temperature within the four soil layers, the sea ice temperatures in all four sea ice layers, the snow mass, temperature, density and albedo. These perturbed fields are constrained in a post-processing step to be consistent with the HTESSEL IFS surface parameterisation schemes to avoid unphysical values (e.g. negative values for variables like snow depth that could result from the perturbation method).

The impact of using perturbed surface initial conditions in ENS was assessed during the experimental phase that led to the operational implementation of IFS cycle 40r1 by comparing two experiments: an experiment with perturbed surface fields (pert-sfc) and one without perturbations to the surface fields (ref). The two experiments were run for 17 initialisation dates within the period from 12 November 2012 to 16 January 2013. Both experiments were run with cycle 38r1 at TL639 resolution, 62 vertical levels and 50 perturbed ensemble members and a 10-member EDA, which was the operational EDA configuration prior to cycle 40r1.

The perturbations to the land surface fields increase the 2-m temperature spread of the experiment pert-sfc in comparison to the ref one. The experiment pert-sfc yields an improved match between the ensemble spread and the ensemble mean error. The difference between the experiments is largest during the early forecast range and becomes very small after 72 hours forecast-time. On average the impact is small, but it can be large in some situations (e.g. if there is uncertainty associated with snow cover which can lead to large differences in 2-m temperature between the ensemble members). To assess how the tails of the predicted 2-m temperature distribution change due to the land surface perturbations, we define the ensemble range as difference between the maximum and minimum value of all ensemble members of one experiment at a specific grid point for a specific initialisation date

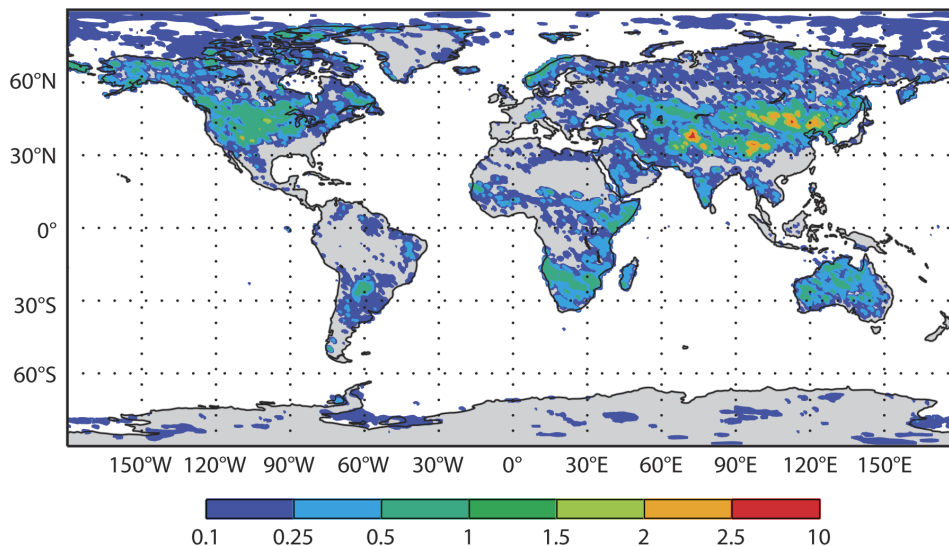


Figure 18: Mean difference of ENS range for 2-m temperature between the experiments with and without surface perturbations after 12 hours forecast-time. The positive values (average between 12 November 2012 and 16 January 2013, see text) indicate larger spread of the experiment with surface perturbations.

In Figure 18 the mean difference between the two ENS experiments is shown in terms of their ensemble ranges calculated over a number of initialisation dates. The surface perturbations have a large impact in regions with some variation in snow cover, especially over parts of the USA, East Asia and Siberia. However, the surface perturbations also increase the ensemble spread in the tropics and subtropics (e.g Australia, South and East Africa and over some parts of South America).

Verification against in situ observations, or HRES analysis, can provide guidance to the added value of improved representation of uncertainties. The impact of introducing perturbations for the snow cover and soil moisture in the initial conditions of the ENS is highlighted in Figure 19, for three locations during winter.

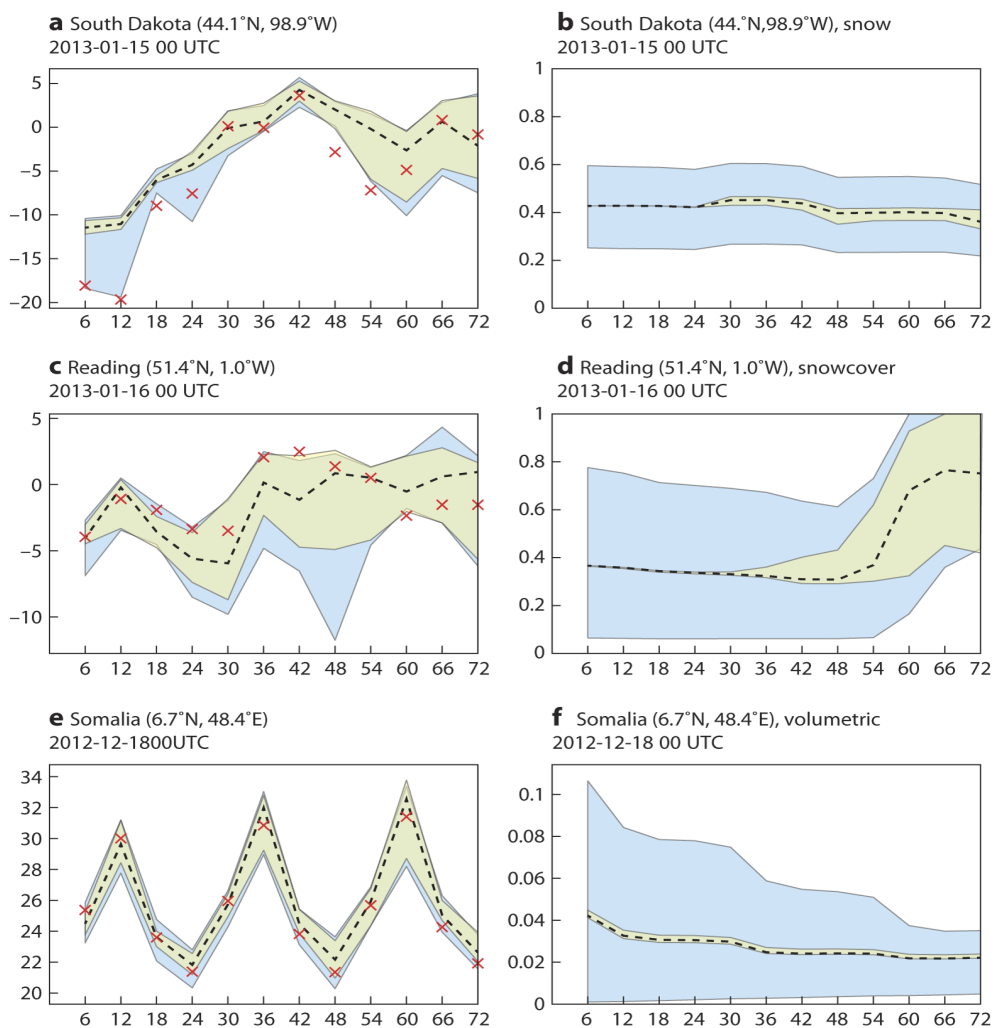


Figure 19: Forecasts shown with their ensemble spread for 2-m temperatures (in °C; a, c, and e), snow cover fraction (b and d) and volumetric soil water in layer 1 (in m³m⁻³; f) for selected locations and initialization dates from the ENS with perturbed surface initial conditions (pert-sfc, blue shading) and without surface perturbation (ref, yellow shading). Depicted are the first 72 hours of the forecast. Values are taken from the grid point closest to the given location. The black dashed curve shows the values from the control forecast. The red crosses indicate the grid-point temperature values from the analysis.

The impact of including snow cover uncertainties in the ENS initial conditions (between 25% and 60% of a grid-box in South Dakota, USA) on the 2-m temperature forecasts (between -10°C and -20°C) is maximal in the short range (first 12-hours of the forecast, Figure 19, a and b). At mid-latitudes, an increase in spread is visible up to 48 hours. (Figure 19, c and d). Soil moisture uncertainties have a widespread influence and an example is provided for dry areas of Somalia (Figure 19 e and f) although impact is expected to be larger in areas with high variance in evaporation (Koster et al., 2010).

4.4 Parameter uncertainty

Hydrological processes are often complex and non-linear at scales that are beyond reach in large-scale models. They are characterised by a high level of spatio-temporal variability, which makes the representation of hydrological processes in coarse resolution global land surface models a challenging task. The model error associated with deep soil hydrological processes has to be characterized by statistical and ensemble techniques.

For example, the flow of water in the subsurface in H-TESEL is based on the Richards equation (Richards, 1931; Hillel, 1998), which describes the evolution of soil water content over time as a function of the pressure head and the hydraulic conductivity. The impact of uncertainty in hydraulic conductivity on hydrological predictions has been clearly demonstrated in land surface modelling (e.g., Schulze-Makuch et al., 1999; Christiaens and Feyen, 2002; Cloke et al., 2008; Williams and Maxwell, 2011).

The compound problem of land surface models being sensitive to parameters whose allocated values are difficult to infer from measurements will become a focus in future work with the H-TESEL land surface model. ECMWF, University of Oxford and other partners, through the EU FP7 project SPECS (<http://www.specs-fp7.eu>), started collaborative research on the representation of model uncertainty for land surface processes through perturbed parameters and stochastic perturbations to the soil temperature and soil moisture tendency equations.

The impact of simultaneously perturbing two of the most sensitive hydrological parameters (hydraulic conductivity K and the van Genuchten soil moisture release parameter) for seasonal forecasts has been tested. Perturbed parameters (by $\pm 40\%$ and $\pm 80\%$ of their standard H-TESEL values) in the range of uncertainties have the potential to improve the forecast reliability of land surface and atmospheric variables in regions of strong land surface–atmosphere coupling. The perturbed parameter experiment showed a promising skill increase over the control experiment for the individual forecast months as well as for the seasonal means (not shown). Work is in progress in collaboration with Oxford University to examine a range of perturbed land surface parameters together with a more traditional stochastically perturbed tendency scheme (similar to SPPT in the atmosphere).

Similarly, the perturbation of snow variables can add onto predictability of winter season anomalies due to the link between snow cover and large-scale dynamical patterns such as the Arctic Oscillation (e.g. Orsolini et al., 2013, and SAC paper ECMWF/SAC/43(14)6).

5 Outlook for Land Surface Developments

Land surface developments will support ECMWF's long-term strategy to provide users with the best possible forecasts of severe weather and near-surface weather parameters across the medium range. Also monthly and seasonal forecasts will require special attention to land surfaces because of the enhanced predictability skill coming from variables with long memory.

Traditionally, land surface models in NWP focused on the question: how does the land surface partition the available energy into sensible and latent heat flux to support forecast accuracy? The land data assimilation is also designed to optimize this requirement by ingesting information on atmospheric temperature and moisture errors and inferring soil moisture increments, with the risk of aliasing several sources of model errors into soil moisture. With new observing systems for soil moisture and new applications such as river flow forecasts and carbon flux modelling, accuracy on all components of the water cycle is required. Not only evaporation is important, but the model should also provide high quality soil moisture, runoff and carbon fluxes. Such a comprehensive Earth system modelling approach will be adopted in future developments. The variety of applications and observations will also help to reduce compensating errors and to improve the realism of representation of land surface processes. The following subsections describe in more detail the options for future work at ECMWF.

5.1 Upcoming developments for improving land surface processes

Three aspects are discussed here: (i) the land surface model with its various components, (ii) horizontal and vertical resolution including the supporting physiographic data, and (iii) anthropogenic changes (urbanisation and irrigation).

The basic land surface model with its hydrology, vegetation, carbon and snow components will evolve further to more comprehensive schemes where the bio-geo-chemical processes of vegetation are considered and where the cycles of carbon, water and energy are mutually interacting (C-TESSSEL as described in Boussetta et al., 2013b). The use of a CO₂ model for stomatal control has a positive impact on evapotranspiration, and even with a climatological leaf area index (LAI), the inter-annual variability of the global atmospheric CO₂ budget is well reproduced due to the inter-annual variability in the meteorological forcing (i.e. radiation, precipitation, wind, temperature, humidity and soil moisture) despite the simplified or missing processes. The stomatal control by CO₂ is currently not activated in the IFS because the H-TESSSEL positive bias in evaporation compensates for an atmospheric near-surface dry bias, probably associated with too strong vertical transport in the atmosphere.

The above-ground biomass (AGB) and interactive prognostic LAI are already existing modelling components of C-TESSSEL in tested research versions and their impact will be explored further in the IMAGINES European project (<http://fp7-imagines.eu>) and compared with agricultural modelling systems. Use of LAI and albedo satellite products produced by the Copernicus Global Land Service (<http://land.copernicus.eu>) will be studied to detect the inter-annual variability of the vegetation state in response to weather and climate disturbances. Accounting for the inter-annual variability of vegetation is key to enhance the effects of extreme events such as droughts in the forecasts. Biomass is also an ESA priority that led to the selection of the BIOMASS satellite mission as the future seventh Earth Explorer.

All model components will benefit from verification and parameter optimization in which a variety of datasets will play a role: flux towers, in situ observations of soil moisture and temperature, satellite observations of soil moisture, satellite skin temperature and vegetation characteristics, river runoff and SYNOP observations. The offline modelling and land data assimilation systems will also play a key role because they allow fast and cost effective experiments covering diurnal to seasonal and multi-annual time scales.

A vertical resolution increase in soil and snow will be necessary to support the satellite observations that are sensitive to shallow near-surface layers and to improve the model capability to represent multiple time scales. The current model has four soil layers only (7, 21, 72, 189 cm) and a single snow slab. Participation to the GEWEX Global Land Atmosphere System Study (GLASS) coordinated experiments such as DICE (<http://appconv.metoffice.com/dice/dice.html>) (Diurnal land/atmosphere coupling experiment) over the US Southern Great Plains and the GABLS4 (<http://www.cnrm.meteo.fr/aladin/meshtml/GABLS4/GABLS4.html>), the GEWEX Atmospheric Boundary Layer Study-4th edition) over Dome Concordia in Antarctica, will allow the study of new vertical discretizations for soil and snow respectively. Those experiments involve the use of OpenIFS Single Column Software (<https://software.ecmwf.int/wiki/display/OIFS/OpenIFS+Home>), which facilitates more collaboration with external partners. A finer vertical discretization is expected to be not only beneficial for the land/atmosphere interaction but it will extend the memory from the deeper soil layers.

The currently available high-resolution physiographic data (discussed in subsection 2.5) and the recently developed processing software make it possible to support model resolutions between 1 km up to hundreds of kilometres. Exploratory experiments were already performed with the CH-TESEL and C-TESEL model versions at 5 km and 1 km resolution in the framework of the AMMA (African Monsoon Multidisciplinary Analysis) Land surface Model Intercomparison Project phase 2 (ALMIP2, Boone et al., 2009a,b). Increasing horizontal resolution separately from the atmospheric model has become a serious alternative to the surface tiling. In this scenario the terrain heterogeneity is not represented by tiles, but by a number of surface grid points (e.g. 10 to 100 land points for every atmospheric grid point). The advantage over the tiling method is that the “sub-areas” are geo-located and can be supported by satellite vegetation information.

Although urban areas occupy only 1% of the land area, they host 50% of the global population, so forecasts in these areas are particularly relevant. The surface energy balance in urban areas is quite different from the one over vegetated surfaces and therefore a special treatment for heating processes related to urban canyons is needed (e.g. Oleson et al., 2008). An urban model can be implemented as an additional tile or as a geo-located high-resolution subsurface. Inter-comparison studies involving several urban parameterizations (Grimmond et al., 2010, 2011) with simplified approaches to urban modelling were shown to be well suited for NWP and would permit to represent the heat-island effect that characterize the modified diurnal cycle in large cities.

Irrigated areas have also been considered in preliminary tests, following the approach by Wisser et al. (2008) and de Rosnay et al. (2003). Puma and Cook (2010) have demonstrated the relevance of including irrigation in a 20th century reanalysis.

The global distribution of urban and irrigated land fractions is shown Figure 20.

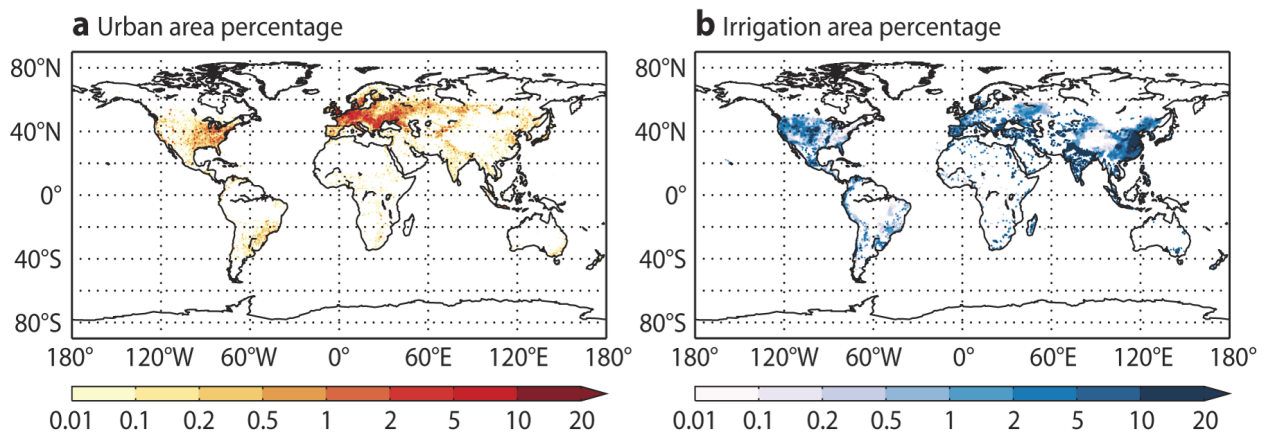


Figure 20: Urban area (a, in %, from ECOCLIMAP, Masson et al., 2003) and irrigated area (b, in %, from Döll and Siebert, 2002).

The highest percentages of urban areas are over Europe (Figure 20a) while irrigated areas reach quite high percentages in China and on the Indian sub-continent, and to a lesser extent at mid-latitudes (Figure 20b). Model development accounting for anthropogenic surfaces would be particularly relevant for extreme events associated with droughts and heat waves therefore having high relevance for forecast users.

5.2 Upcoming developments for improving land data assimilation

The analysis of surface variables in the Land Data Assimilation System (LDAS) at ECMWF plays an important role in the quality of forecasts from short range to medium range and beyond, thanks to the memory effect of slowly evolving surface quantities. Soil moisture and snow are among the most important variables with memory, and LDAS has concentrated on the introduction of new observations, and on improvements in the assimilation methods, the land observation operators, and operational monitoring (see <https://software.ecmwf.int/wiki/display/LDAS/LDAS+Home>).

Future developments in the area of land data assimilation will include (i) a complete technical overhaul to reduce the costs and allow fast offline experimentation over seasonal time scales, (ii) inclusion of data assimilation components related to new modelling components, e.g. lakes and vegetation, (iii) introduction of new observations e.g. skin temperature, and (iv) optimization of the data assimilation system.

A significant issue with the current LDAS is its computational cost and infrastructural overhead that makes testing of new developments a costly activity. Future land surface data assimilation activities will rely on a modular system that includes different components for the screen-level and snow analyses (OI) and the soil moisture analysis (SEKF). The OI scheme is already a stand-alone data assimilation system, which runs independently from 4DVAR. The SEKF soil moisture analysis also runs independently from the 4DVAR, however it is embedded in the IFS. This structure was necessary for the SEKF implementation in 2010 to enable the Jacobians computation and it also provides the interface to the Observations Data Base. However, recent developments of the model make it possible to run a stand-alone version of H-TESSEL on the supercomputer, which opens the possibility to compute the SEKF Jacobians without running the IFS, with a sizeable saving in computational burden

(a factor 60). On-going work to externalise the Jacobians computation of the SEKF and further steps for the SEKF externalisation will also include the observations interface and link to the COPE activities (Continuous Observation Processing Environment).

These major developments require revising completely the surface analysis and the SEKF structures as well as the surface analysis observations interface. Recent developments in the conventional observations interface to the OI constitute a first step toward these revisions. They also include the EDA screen-level temperature and relative humidity observations perturbations as discussed in subsection 4.2.

Current efforts to externalise the SEKF will make it possible to run the LDAS in a fully offline mode, at a reduced computing cost. So, it will enable the inclusion of the LDAS tasks between 4DVAR trajectories, thereby improving the coupling between the surface and the upper-air analyses. It will also make it possible to extend, in the next four years, the use of the SEKF to analyse more surface variables such as soil, lake, and snow temperature. Another advantage is that the use of more satellite observations can be explored (e.g. NASA Soil Moisture Active Passive mission, SMAP, EUMETSAT Land surface temperatures).

Participation in the GEWEX/GLASS Project for Intercomparison of Land Data Assimilation Systems (PILDAS) will use the offline developments, and it will compare a range of land assimilation methods. PILDAS will also include sensitivity studies of assimilation input parameters and it will provide guidance and priorities for future land assimilation research and applications.

After the introduction of new modelling components, such as the lake model, further improvement of forecast accuracy can be expected from the initialization of lake temperature and ice cover with observations. Surface temperature from instruments such as the Moderate Resolution Imaging Spectroradiometer (MODIS), and from geostationary satellites such as Meteosat Second Generation (MSG) with the Spinning Enhanced Visible and Infrared Imager (SEVIRI) are not yet sufficiently exploited and can provide valuable information on surface temperature and thermal properties.

Future improvements of the global land data assimilation will broaden the observation usage to more satellite data such as NASA SMAP, ESA Sentinels and EUMETSAT Meteosats, collecting remote sensing data informative for soil moisture, snow and lake initialisation. Land surface emissivity and temperature assimilations studies as explored by Masiello et al. (2013) based on SEVIRI, or Karbou et al. (2010) using microwave sensors will be relevant to support atmospheric data assimilation over land surfaces.

The quality of the error statistics and bias correction in the assimilation system controls the LDAS optimality, especially following the introduction of new parameterization schemes or new observations. Diagnostics based on optimality and information content criteria play an important role in the optimization procedures (Desroziers and Ivanov, 2001; Desroziers et al., 2009).

A more complete account of land surface model error as developed in the EDA and ENS and described in Chapter 3, may allow the future exploitation of the background error co-variances calculated within EDA surface to provide a spatially variable model error with flow dependency on the meteorological and land surface conditions used in the LDAS for both the OI and the SEKF.

The analysis increments will continue to provide information also on the systematic patterns of corrections, which can guide research for missing parameterization components. A distinction between systematic errors (e.g. bias in the seasonal or daily annual cycle) that should be the subject for modelling efforts via calibration (or data assimilation in the parameter space) and random errors due to the limits of predictability, and subject to daily analyses, has to be made.

Data assimilation applications to model parameter optimisation will be equally explored as they were already performed in Boussetta et al. (2013b) for land carbon parameters or in Balsamo et al. (2010) and Dutra et al. (2010) for lake parameters.

5.3 Supporting environmental and hazard predictions

Global demand for the modelling of water resources in lakes, river and reservoirs is very high and ECMWF already supports the forecasting of river flow and inundated areas. The fresh water input to oceans is also provided. The river discharge forecasts are the basis for flood early warning systems or water management tools and offer at the same time a very powerful verification for the water cycle of the IFS.

Floods, droughts, wild and forest fires, landslides, desert locusts, dust storms and thermal hazards can generate disasters. Hazards originating from weather and climate disturbances are responsible for the largest number of fatalities and economic damage over the last century. WMO objectives are to reduce average fatalities by 50% in 2019 compared to 1994–2003.

At ECMWF, development of the capability to model river discharge in real-time has reinforced the link between meteorology and river hydrology and it allowed to use river-discharge observations for verification of the water cycle. The river discharge observations from the Global Runoff Data Centre (GRDC, http://www.bafg.de/GRDC/EN/Home/homepage_node.html) provide reliable integrated and representative information on the water budget in a catchment area. Although indirect, discharge is strongly related to precipitation in extreme events, so it provides valuable verification information about precipitation and does not suffer from the representativeness issues with rain-gauges.

Figure 21 shows the 2013 example of floods in central Europe, in which ECMWF forecasts were run through the European Flood Awareness System (EFAS), to evaluate the effects of different resolutions (16 km and 5 km) and to test a new development in the cloud physics scheme (see Haiden et al., 2014). In this case EFAS is driven by forecast precipitation and temperature to help diagnose the forecast performance of extreme precipitation.

Another example is the Global Flood Awareness system (GloFAS), which was developed jointly by ECMWF and the EC's Joint Research Centre (JRC). It routes the runoff from H-TESSSEL to forecast discharge for the major rivers globally (Alfieri et al., 2013). GloFAS is a research initiative, designed to run daily in a semi-operational manner. The GloFAS system has already provided information to organizations such as the Red Cross and World Food Programme during flood events, showing skill in many parts of the world.

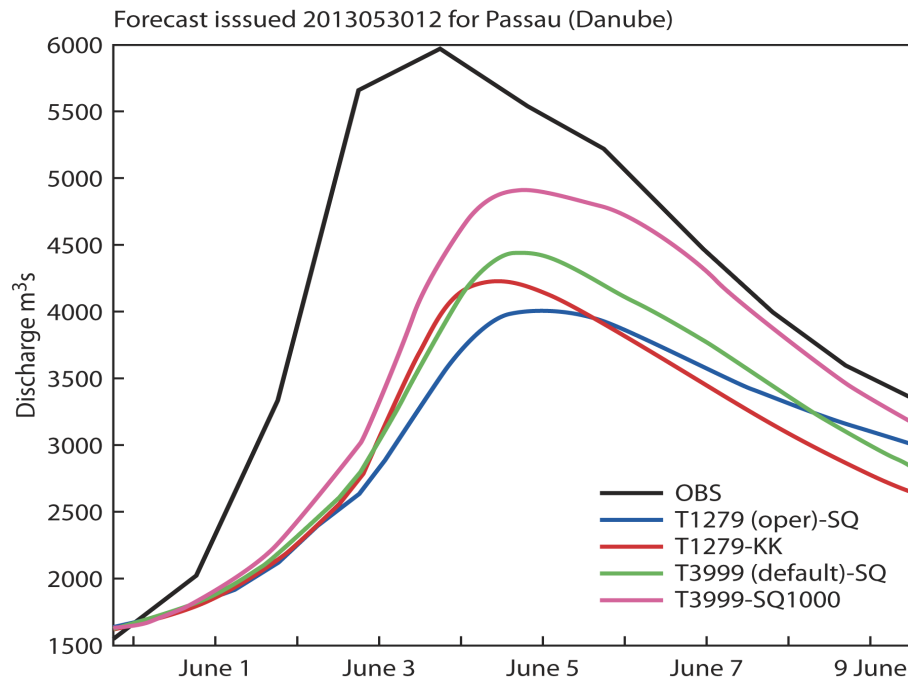


Figure 21: Forecast discharge levels with two different model resolutions (T1279 in blue and T3999 in green) for the station Passau at the river Danube. The black line denotes the simulated discharge using observed precipitation, interpreted as the forecast potential. The red and magenta lines indicate the improvement in river discharge forecasts with improved cloud physics for the T1279 and T3999 resolution respectively.

Figure 22 illustrates the usefulness of GloFAS in predicting river flow particularly over large mid-latitude rivers. The system is continuously developed. Currently the routing in GloFAS is done by a component from LISFLOOD, which is the operational hydrological model of EFAS. Modular coupling of river routing models such as TRIP2 (Pappenberger et al., 2010; Balsamo et al., 2011a) or CaMaflood (Yamazaki et al., 2010, 2011, 2012) has also been tested with direct output from H-TESSSEL.

The flexible coupling between models has the advantage of being able to explore new applications. For instance, floodplains were simulated in an attempt to derive global flood hazard maps, driven by runoff from ERA-Interim (Pappenberger et al., 2012). CaMaflood has also been used to route the runoff from the TIGGE archive (Zsoter et al., 2014). A new setup with CaMaFlood coupled to H-TESSSEL will be further tested in operational mode alongside the existing GloFAS system and in research projects such as Earth2Observe (<http://www.earth2observe.eu>).

The coupling of H-TESSSEL to hydrology can also be extended by giving flood plain and ground water information back to the evaporation formulation of H-TESSSEL. Similarly, fresh water flow into the ocean could be coupled to salinity in an ocean model.

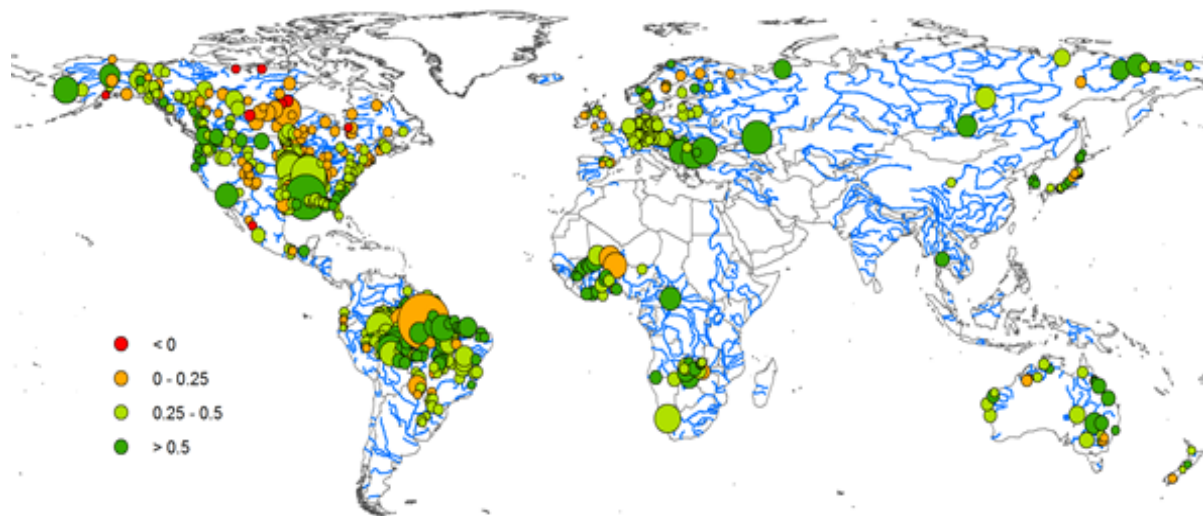


Figure 22: Peirce's skill score of the Global Flood Forecasting System (GloFAS) based on observed discharge (GRDC) at each of the 620 stations considered. Circle size is proportional to the upstream area of each river station. Higher values indicate larger skills.

5.4 A strategy for surface-model-error characterisation and error reduction

Land-atmosphere interactions are complex with strong links between energy, water and carbon cycles. Therefore improvements in one physical variable have the potential to translate into improvements of other components (e.g. a soil moisture simulation in closer agreement with observation may lead to more skilful carbon dioxide emissions). On the other hand it also implies that extended verification is required when changes are introduced in a system that has to improve several processes and applications at the same time.

State-of-the-art land surface schemes in large-scale models are necessarily highly empirical as it is currently impossible to describe the relevant processes in all their complexity and detail. However, these complex processes, occurring over a wide range of spatial and temporal scales, govern the energy and water cycles at global scale. The large-scale budgets are obviously a priority in global models. The water budget is a clear example, as precipitation falls on the ground in a non-uniform way, some of it is intercepted by the vegetation and evaporates again, another part falls through and can either run off on the surface or infiltrate in the soil. Soil texture, being highly heterogeneous, affects vertical transport and horizontal water transport to rivers. Evaporation is another important component of the water budget. It is linked to the available energy, but is also highly regulated by vegetation through root distribution and plant physiological processes. The consequence of all the complexity is that it is difficult to build a land surface scheme from the smallest scale of individual leaves and plants and the smallest elements of soil heterogeneity and to integrate such a description to effective scales of the order of 10 km. The difficulty is also that it is impossible to characterise the surface vegetation and soil with sufficient accuracy over the whole globe; accurate datasets to support such characterisation do not presently exist. This is the reason that a bulk parameterization of the land surface processes is necessary and inevitably there are strong elements of inverse modelling (i.e. parameters have to be optimised on the basis of the results). Having observational information

strongly related to the processes that are modelled is important; otherwise introduction of compensating errors is very likely.

We therefore identify two major ways of considering model errors associated with land surface processes.

- Model errors that can be identified by observations. Such errors may be due to missing parameterizations (e.g. the model misses a lake or urban component and from satellites we can just see that lakes or urban areas have a different surface temperature than vegetated areas). The observability in this case supports the parameterization design. The observations help parameter optimization and support cyclic corrections of the land prognostic variables in data assimilation.
- Model errors that are largely unobservable and for which only the effects are measurable. Such errors may be related to inevitable simplifications in models and therefore statistical perturbation and ensemble techniques are best-suited approaches to characterize the model error associated to those land surface aspects (e.g. representativity errors such as those associated with soil texture, hydraulic conductivity, vegetation properties and soil carbon stocks).

The development of more sophisticated land surface schemes in a global NWP context will necessarily be driven by remote sensing that is informative about the land surface and covers large scales, combined with in-situ observations to verify physical consistency of the parameterizations. Priorities in error reduction and error characterization will rely heavily on availability of observations particularly on routinely available observations. The use of research quality observations from in-situ and remote sensing will help in developments that aim at improving both fluxes and reservoirs (of water, energy and carbon). They will allow going beyond simple “calibration” of models. A large variety of observations enable a coherent framework for assessing the increased complexity of a given parameterization.

A more complex land surface modelling and data assimilation system will need more rigorous testing and evaluation and a process-oriented benchmarking of different model versions against a range of diverse and independent observational datasets to ensure that parameterizations are fit for their purpose. Using in-situ flux measurements of latent and sensible heat and carbon fluxes, river flows, conventional meteorological data, will allow developments towards enhanced ecosystem modelling for operational monitoring and forecasting with particular attention to the land/atmosphere coupling.

Integrating diverse observational sources will require a more holistic approach to parameterization testing and it should be combined with innovative applications of data assimilation techniques extending into parameters space. Improved understanding of surface process mechanisms will rely on the combination of optimal estimation techniques and modelling, to allow optimal parameters and also to better identify certain limits within the adopted schemes.

6 Conclusions

The land surface components of the IFS at ECMWF have received considerable attention over the last few years and work has resulted in many system upgrades. Changes in the land surface parameterizations, associated assimilation schemes, assimilated observations, and characterization of land surface errors, have all contributed to medium-range forecast accuracy and reliability.

The increased complexity in the treatment of soil, snow, and vegetation has been supported by updated physiographic fields and by dedicated data assimilation advances to constrain errors in the main water and energy reservoirs. Results have been verified using multiple observational sources characterizing the energy, water and carbon cycle. A multi-source observation approach has been identified as key for avoiding over-fitting and compensating errors, both in model development and data assimilation.

Sizeable challenges are foreseen when extending further the representation of land surface processes and embracing environmental applications that involve the use of higher spatial and vertical resolutions and consideration for both natural and anthropogenic components.

Observation availability, anticipated forecast impact, and relevance for users, will provide guidance in defining development priorities when moving towards a more complete representation of the Earth's surface.

Acknowledgements

The authors are grateful for comments and support to this review paper as provided by Peter Bauer, Roberto Buizza, Dick Dee, Lars Isaksen, Erland Källén, Adrian Simmons, Jean-Noël Thépaut and Alan Thorpe. We acknowledge editorial support by Bob Riddaway, Anabel Bowen and Simon Witter.

References

- Agustí-Panareda, A., S. Massart, F. Chevallier, S. Boussetta, G. Balsamo, A. Beljaars, P. Ciais, N. M. Deutscher, R. Engelen, L. Jones, R. Kivi, J.-D. Paris, V.-H. Peuch, V. Sherlock, A. T. Vermeulen, P. O. Wennberg, and D. Wunch: Forecasting global atmospheric CO₂, *Atmos. Chem. Phys. Discuss.*, 14, 1–54, doi:10.5194/acpd-14-1-2014, 2014.
- Albergel C., G. Balsamo, P. de Rosnay P., J. Muñoz-Sabater, and S. Boussetta., 2012: A bare ground evaporation revision in the ECMWF land-surface scheme: evaluation of its impact using ground soil moisture and satellite microwave data, *Hydrol. Earth Syst. Sci., HESS*, 16, 3607-3620, also as ECMWF Technical memorandum 685, September 2012.
- Albergel, C., W. Dorigo, R. H. Reichle, G. Balsamo, P. de Rosnay, J. Muñoz-Sabater, L. Isaksen, R. de Jeu, W. Wagner, 2013: Skill and global trend analysis of soil moisture from reanalyses and microwave remote sensing. *J. Hydromet.*, doi: <http://dx.doi.org/10.1175/JHM-D-12-0161.1>, also available ECMWF Tech. memo. 695.
- Albergel C., E. Dutra, J. Muñoz-Sabater, T. Haiden, G. Balsamo, A. Beljaars, L. Isaksen, P. de Rosnay, I. Sandu and N. Wedi, 2014: Soil temperature at ECMWF: an assessment using ground based observations. ECMWF Technical Memorandum, in prep.
- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS – global ensemble streamflow forecasting and flood early warning, *Hydrol. Earth Syst. Sci.*, 17, 1161-1175, doi:10.5194/hess-17-1161-2013, 2013.
- Amante, C. and B.W. Eakins, 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M
- Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier, C. O., de Wekker, S. F. J., Wolfe, D. E., Munger, J. W., and Tans, P. P.: CO₂, CO and CH₄ measurements from the NOAA Earth System Research Laboratory's Tall Tower Greenhouse Gas Observing Network: instrumentation, uncertainty analysis and recommendations for future high-accuracy greenhouse gas monitoring efforts, *Atmos. Meas. Tech. Discuss.*, 6, 1461–1553, doi:10.5194/amtd-6-1461-2013, 2013.
- Atlas R., N. Wolfson, and J. Terry, 1993: The effect of SST and soil moisture anomalies on GLA model simulations of the 1998 U.S. summer drought. *J. Climate*, 6, 2034–2048.
- Balsamo, G., 2013: Interactive lakes in the Integrated Forecasting System. ECMWF Newsletter 137, 30-34.
- Balsamo, G., R. Salgado, E. Dutra, S. Boussetta, T. Stockdale, M. Potes, 2012: On the contribution of lakes in predicting near-surface temperature in a global weather forecasting model, *Tellus-A*, 64, 15829, DOI: 10.3402/tellusa.v64i0.15829

- Balsamo, G., C. Albergel, A. Beljaars, S. Boussetta, E. Brun, H. Cloke, D. Dee, E. Dutra, F. Pappenberger, P. de Rosnay, J. Muñoz Sabater, T. Stockdale, F. Vitart, 2012: ERA-Interim/Land: A global land-surface reanalysis based on ERA-Interim meteorological forcing, ERA Report Series, 13, 1-25.
- Balsamo, G., Pappenberger, F., Dutra, E., Viterbo P., van den Hurk, B., 2011a: A revised land hydrology in the ECMWF model: a step towards daily water flux prediction in a fully-closed water cycle, *Hydrological Processes*, 25(7) DOI : 10.1002/hyp.7808
- Balsamo, G., S. Boussetta, E. Dutra, A. Beljaars, P. Viterbo, B. Van Den Hurk, 2011b: Evolution of land-surface processes in the IFS, *ECMWF Newsletter*, 127.
- Balsamo G., E. Dutra, V.M. Stepanenko, P. Viterbo, P.M. Miranda, D. Mironov, 2010: Deriving an effective lake depth from satellite lake surface temperature data: A feasibility study with MODIS data, *Bor. Env. Res.*, 15, 178-190, also available as ECMWF Tech. memo. 609.
- Balsamo G, Viterbo P, Beljaars ACM, van den Hurk B, Hirschi M, Betts AK, Scipal K., 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System.” *J. Hydrol.* 10: 623–643.
- Balsamo, G., J.-F. Mahfouf, S. Bélair, and G. Deblonde, 2007: A land data assimilation system for soil moisture and temperature: An information content study. *J. Hydromet.*, 8, 1225-1242.
- Balsamo, G., F. Bouyssel and J. Noilhan, 2004: A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather prediction model, *Quart. J. Roy. Meteor. Soc.*, 130, 895-915.
- Beljaars A., Viterbo P., Miller M., Betts A., 1996: Sensitivity to land surface parameterization and soil anomalies. *Mon Weather Rev* 124:362–38.
- Beljaars, A., G. Balsamo, A. Betts and P. Viterbo, 2007: Atmosphere/surface interactions in the ECMWF model at high latitudes. *Proc. of ECMWF Seminar on Polar meteorology*, 4-8 September 2006, ECMWF, 153-168.
- Berner, J., G. J. Shutts, M. Leutbecher, T. N. Palmer, 2009: A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System. *J. Atmos. Sci.*, 66, 603–626. doi: <http://dx.doi.org/10.1175/2008JAS2677.1>.
- Betts, A. K., 2009: Land-surface-atmosphere coupling in observations and models. *J. Adv. Model Earth Syst.*, Vol. 1, Art. #4, 18 pp., doi: 10.3894/JAMES.2009.1.4.
- Betts, A.K., R. Desjardins and D. Worth, 2013: Cloud Radiative Forcing of the Diurnal Cycle Climate of the Canadian Prairies. *J. Geophys. Res.* 118, 8935–8953, doi:10.1002/jgrd.50593.
- Betts, A.K., R. Desjardins, D. Worth, S. Wang and J. Li, 2014: Coupling of winter climate transitions to snow and clouds over the Prairies. *J. Geophys. Res. Atmos.*, 119, 1118-1139, doi:10.1002/2013JD021168.

- Boone, A., P. de Rosnay, G. Basalmo, A. Beljaars, F. Chopin, B. Decharme, C. Delire, A. Ducharne, S. Gascoin, M. Grippa, F. Guichard, Y. Gusev, P. Harris, L. Jarlan, L. Kergoat, E. Mougin, O. Nasonova, A. Norgaard, T. Orgeval, C. Ottlé, I. Pocard-Leclercq, J. Polcher, I. Sandholt, S. Saux-Picart, C. Taylor, and Y. Xue, 2009: The AMMA Land Surface Model Intercomparison Project. *Bull. Amer. Meteor. Soc.*, 90(12), 1865-1880, doi:10.1175/2009BAMS2786.1
- Boone, A., A. C. V. Getirana, J. Demarty, B. Cappelaere, S. Galle, M. Grippa, T. Lebel, E. Mougin, C. Peugeot and T. Vischel, 2009: The African Monsoon Multidisciplinary Analyses (AMMA) Land surface Model Intercomparison Project Phase 2 (ALMIP2). *GEWEX News*, November, 19(4), 9-10.
- Boussetta, S., G. Balsamo, A. Beljaars, T. Kral and L. Jarlan, 2013a: Impact of a satellite-derived Leaf Area Index monthly climatology in a global Numerical Weather Prediction model. *Int. J. Rem. Sens.*, 34, 3520-3542, doi:10.1080/01431161.2012.716543.
- Boussetta, S., Balsamo G., Beljaars A., Agustí-Panareda A., Calvet J.-C., Jacobs C., van den Hurk B., Viterbo P., Lafont S., Dutra E., Jarlan L., Balzarolo M., Papale D., and van der Werf, G., 2013b: Natural land carbon dioxide exchanges in the ECMWF Integrated Forecasting System: Implementation and offline validation, *J. Geophys. Res.* 118, Issue 12, 5923–5946. doi: 10.1002/jgrd.50488.
- Brasnett B, 1999: A global analysis of snow depth for numerical weather prediction. *J Appl Meteorol* 38:726–740.
- Buizza, R., and T.N. Palmer, 1995: The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, 52, 9, 1434-1456.
- Buizza, R., M. Leutbecher and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System, *Q.J.R. Meteorol. Soc.*, 134, 2051–2066.
- Buizza, R., M. Leutbecher, L. Isaksen and J. Haseler, 2010: The use of EDA perturbations in the EPS. *ECMWF Newsletter*, 123, 22-27 (available online at: <http://www.ecmwf.int/publications/>)
- Calvet J.C., J. Noilhan, J.-L. Roujean, P. Bessemoulin, M. Cabelguenne, A. Olioso, J.-P. Wigneron, 1998: An interactive vegetation SVAT model tested against data from six contrasting sites. *Agric. For. Meteorol.*, 92, 73-95.
- Christiaens, K. and Feyen, J., 2002: Constraining soil hydraulic parameter and output uncertainty of the distributed hydrological MIKE SHE model using the GLUE framework. *Hydrological Processes*, 16: 373–391. doi: 10.1002/hyp.335
- Cloke H. L., Pappenberger F. and Renaud J.P., 2008: Multi-method Global Sensitivity Analysis (MMGSA) for Modelling Floodplain Hydrological Processes. *Hydrological Processes*, 22((11) 1660-1667.
- Cook, B. I., G. B. Bonan, S. Levis, and H. E. Epstein, 2008: The thermoinsulation effect of snow cover within a climate model. *Climate Dyn.*, 31, 107–124.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, 87, 367-374.

de Rosnay, P., J. Polcher, K. Laval and M. Sabre, 2003: Integrated parameterization of irrigation in the land surface model ORCHIDEE. Validation over Indian Peninsula, *Geophys. Res. Lett.*, 30(19), 1986, doi:10.1029/2003GL018024.

de Rosnay P., G Balsamo, J Muñoz Sabater, C Albergel and L Isaksen., 2011a: "Land surface data assimilation", Proceeding of the ECMWF Seminar on Data Assimilation for atmosphere and ocean, 6-9 September 2011.

de Rosnay P., Drusch M., Balsamo G., Albergel C. and Isaksen L., 2011b: Extended Kalman Filter soil-moisture analysis in the IFS ECMWF Newsletter no 127, 12-16.

de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., & Isaksen, L., 2013: A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. *Q. J. R. Meteorol. Soc.*, 139(674):1199-1213 , also available as ECMWF Technical Memorandum 662

de Rosnay P., G. Balsamo, C. Albergel J. Muñoz-Sabater and L. Isaksen, 2014: Initialisation of land surface variables for Numerical Weather Prediction, *Surveys in Geophysics*, 35(3), 607-621.

Dee D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Köllberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. J. Morcrette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavalato, J. N. Thépaut, F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137 : 553–597. DOI:10.1002/qj.828

Delworth T. L., and S. Manabe, 1988 : The influence of potential evaporation on the variability of simulated soil wetness and climate. *J. Climate*, 1, 523–547.

Desroziers, G. and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Q. J. R. Meteor. Soc.*, 127, 1433–1452.

Desroziers, G., L. Berre, V. Chabot, and B. Chapnik, 2009: A posteriori diagnostics in an ensemble of perturbed analyses. *Mon. Wea. Rev.*, 137:3420–3436.

Dirmeyer, Paul A., 2011: A History and Review of the Global Soil Wetness Project (GSWP). *J. Hydrometeorol*, 12, 729–749. DOI: <http://dx.doi.org/10.1175/JHM-D-10-05010.1>.

Döll, P., and S. Siebert, 2002: Global modeling of irrigation water requirements, *Water Resour. Res.*, 38(4), 8.1–8.10, 2002.

Douville, H., Royer, J. F., and Mahfouf, J. F., 1995: A New Snow Parameterization for the Meteo-France Climate Model .1. Validation in Stand-Alone Experiments, *Climate Dyn.*, 12, 21-35.

Douville H., Viterbo P., Mahfouf J.-F. and Beljaars A., 2001: Evaluation of the optimum interpolation and nudging techniques for soil moisture analysis using FIFE data. *Mon. Weather Rev.* 128: 1733–1756.

Drusch, M., D. Vasiljevic and P. Viterbo, 2004: ECMWF's global snow analysis: Assessment and revision based on satellite observations. *J. Appl. Met.*, 43(9), 1282-1294.

Drusch M. and Viterbo P., 2007: Assimilation of screen-level variables in ECMWF's Integrated Forecast System: a study on the impact on the forecast quality and analyzed soil moisture. *Mon Weather Rev* 135:300–314

Drusch, M., K. Scipal, P. de Rosnay, G. Balsamo, E. Andersson, P. Bougeault and P. Viterbo, 2008: Exploitation of satellite data in the surface analysis, ECMWF Technical Memorandum 576, October 2008, also SAC paper 2008 (ECMWF/SAC/37(08)6)

Drusch M, Scipal K, de Rosnay P, Balsamo G, Andersson E, Bougeault P, Viterbo P., 2009: “Towards a Kalman Filter-based soil moisture analysis system for the operational ECMWF Integrated Forecast System.” *Geophys. Res. Lett.* 36: L10401, DOI: 10.1029/2009GL037716

Dutra, E., Balsamo, G., Viterbo, P., Miranda, P. M. A., Beljaars, A., Schar, C., and Elder, K., 2010a: An Improved Snow Scheme for the ECMWF Land Surface Model: Description and Offline Validation, *J. Hydrometeor.*, 11, 899-916, doi: 10.1175/2010jhm1249.1.

Dutra, E., V.M. Stepanenko, G. Balsamo, P. Viterbo, P.M. Miranda, D. Mironov, C. Schär, 2010b: Global offline lake simulations: Validation and Impact on ERA-Interim, *Bor. Env. Res.*, 15, 100-112, also available as ECMWF Tech. Memo. 608.

Dutra, E., P. Viterbo, P. M. A. Miranda, G. Balsamo, 2012: Complexity of snow schemes in a climate model and its impact on surface energy and hydrology. *J. Hydrometeor.*, 13, 521-538. doi:10.1175/JHM-D-11-072.1.

English, S, T. McNally, N. Bormann, K. Salonen, M. Matricardi, A. Horanyi, M. Rennie, M. Janisková, S. Di Michele, A. Geer, E. Di Tomaso, C. Cardinali, P. de Rosnay, J. Muñoz Sabater, M. Bonavita, C. Albergel, R. Engelen and J.-N. Thépaut., 2013: Impact of satellite data. ECMWF Technical Memorandum 711, also SAC paper 2013 (ECMWF/SAC/42(13)6)

Ge, Y., and G. Gong, 2010: Land surface insulation response to snow depth variability. *J. Geophys. Res.*, 115, D08107, doi:10.1029/2009JD012798.

Grimmond, C. S. B., M. Blackett, M. J. Best, J. Barlow, J.-J. Baik, S. E. Belcher, S. I. Bohnenstengel, I. Calmet, F. Chen, A. Dandou, K. Fortuniak, M. L. Gouvea, R. Hamdi, M. Hendry, T. Kawai, Y. Kawamoto, H. Kondo, E. S. Krayenhoff, S.-H. Lee, T. Loridan, A. Martilli, V. Masson, S. Miao, K. Oleson, G. Pigeon, A. Porson, Y.-H. Ryu, F. Salamanca, L. Shashua-Bar, G.-J. Steeneveld, M.

Grimmond, C. S. B., M. Blackett, M. J. Best, J.-J. Baik, S. E. Belcher, J. Beringer, S. I. Bohnenstengel, I. Calmet, F. Chen, A. Coutts, A. Dandou, K. Fortuniak, M. L. Gouvea, R. Hamdi, M. Hendry, M. Kanda, T. Kawai, Y. Kawamoto, H. Kondo, E. S. Krayenhoff, S.-H. Lee, T. Loridan, A.

Martilli, V. Masson, S. Miao, K. Oleson, R. Ooka, G. Pigeon, A. Porson, Y.-H. Ryu, F. Salamanca, G.J. Steeneveld, M. Tombrou, J. A. Voogt, D. T. Young, N. Zhang. (2011) Initial results from Phase 2 of the international urban energy balance model comparison. *International Journal of Climatology* 31:2, 244-272.

Groisman, P. Y., T. R. Karl, R. W. Knight, and G. L. Stenchikov (1994), Changes of snow cover, temperature, and radiative heat-balance over the Northern Hemisphere, *J. Clim.*, 7(11), 1633–1656, doi:10.1175/1520-0442(1994)007<1633:COCTA>2.0.CO;2.

Groisman, P. Y., Knight, R. W., Karl, T. R., Easterling, D. R., Sun, B. M., and Lawrimore, J. H., 2004: Contemporary changes of the hydrological cycle over the contiguous United States: Trends derived from in situ observations, *J. Hydrometeor.*, 5, 64-85.

Haiden, T. L. Magnusson, I. Tsonevsky, F. Wetterhall, L. Alfieri, F. Pappenberger, P. de Rosnay, J. Muñoz-Sabater, G. Balsamo, C. Albergel, R. Forbes, T. Hewson and S. Malardel, 2014, ECMWF forecast performance during the June 2013 flood in Central Europe, ECMWF Technical Memo 723, May 2014.

Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta/RSM Short-Range Ensemble Forecasts. *Mon. Wea. Rev.*, 125, 1312-1327.

Hess, R., 2001: Assimilation of screen-level observations by variational soil moisture analysis, *Meteorol. Atmos. Phys.*, 77, 145–154.

Hillel D., 1998: *Environmental soil physics*. Academic Press. 771 pages.

Hirschi, M., P. Viterbo, and S. Seneviratne, 2006: Basin-scale water-balance estimates of terrestrial water storage variations from ECMWF operational forecast analysis. *Geophys. Res. Letters*, 33, L21 401.

Hohenegger C., P. Brockhaus, C. S. Bretherton, and C. Schär, 2009: The soil-moisture precipitation feedback in simulations with explicit and parameterized convection. *J. Climate*, 22, 5003-5020.

Houtekamer, P. L., X. Deng, H. L. Mitchell, S.-J. Baek, N. Gagnon, 2014: Higher Resolution in an Operational Ensemble Kalman Filter. *Mon. Wea. Rev.*, 142, 1143–1162. doi: <http://dx.doi.org/10.1175/MWR-D-13-00138.1>.

Isaksen, L., M. Bonavita, M. Fisher, J. Haseler, M. Leutbecher, L. Raynaud, 2010: Ensemble of data assimilations at ECMWF, ECMWF Tech. Memo, 636.

Karbou, F., E. Gérard, F. Rabier (2010), Global 4D-Var assimilation and forecast experiments using AMSU observations over land. Part-I: Impact of various land surface emissivity parameterizations, *Weather and Forecasting*, 25, 5–19, doi : 10.1175/2009WAF2222243.1

Koster R. D. and Suarez, M. J., 1992 : Relative contributions of land and ocean processes to precipitation variability *J. Geophys. Res.*, 100(D7), 13775-13790

Koster R. D., Dirmeyer PA, Guo Z, Bonan G, Cox P, Gordon C. Kanae S, Kowalczyk E, Lawrence D, Liu P, Lu C, Malyshev S, McAvaney B, Mitchell K, Mocko D, Oki T, Oleson K, Pitman A, Sud Y, Taylor C. Verseghy D, Vasic R, Xue Y and Yamada T., 2004 : Regions of strong coupling between soil moisture and precipitation. *Science*, 305.

Kourzeneva, E. ,2010 : External data for lake parameterization in Numerical Weather Prediction and climate modeling. *Boreal Env. Res.* 15, 165–177.

Lang, S., G. Balsamo, R. Buizza, E. Dutra and M. Leutbecher, 2013: Using the ensemble of 4D-Var data assimilations to generate perturbations to the land surface initial conditions of the ensemble, ECMWF Res. Memo., RD13-065.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, 227, 3515–3539.

Mahfouf, J.-F. and Noilhan, J., 1991: Comparative Study of Various Formulations of Evaporations from Bare Soil Using in situ Data, *J. Appl. Meteorol.* 9, 351–362.

Mahfouf, J.-F.: Analysis of soil moisture from near surface parameters: A feasibility study, *J. Appl. Meteorol.*, 30, 506–526, 1991.

Mahfouf, J.-F., P.Viterbo, H. Douville, A.C.M. Beljaars and S. Saarinen, 2000: A revised land-surface analysis scheme in the Integrated Forecasting System. ECMWF newsletter, No 88, 8-13.

Mahfouf J-F, Bergaoui K, Draper C, Bouyseel F, Taillefer F and Taseva L. 2008: A comparison of two off-line soil analysis schemes for assimilation of screen level observations. *J. Geophys. Res.*, 114, D08105, doi:10.1029/2008JD011077.

Manabe S., 1969 : Climate and ocean circulation: 1. The atmospheric circulation and the hydrology of the earth's surface. *Mon. Wea. Rev.*, 97, 739–774.

Manrique-Suñén, A., A. Nordbo, G. Balsamo, A. Beljaars, and I. Mammarella, 2013: Representing Land Surface Heterogeneity: Offline Analysis of the Tiling Method. *J. Hydrometeor.* 14, 850–867. doi: <http://dx.doi.org/10.1175/JHM-D-12-0108.1>.

Masiello, G., Serio, C., De Feis, I., Amoroso, M., Venafrà, S., Trigo, I. F., and Watts, P. 2013: Kalman filter physical retrieval of surface emissivity and temperature from geostationary infrared radiances, *Atmos. Meas. Tech.*, 6, 3613-3634, doi:10.5194/amt-6-3613-2013.

Masson, V. J.-L. Champeaux, F. Chauvin, C. Meriguet and R. Lacaze, 2003 : A global database of land surface parameters at 1km resolution in meteorological and climate models, *J. Climate*, 16, 1261-1282.

Mironov, D., Heise, E., Kourzeneva, E., Ritter, B., Schneider, N. and co-authors. 2010. Implementation of the lake parameterisation scheme FLake into the numerical weather prediction model COSMO. *Boreal Env. Res.* 15, 218–230.

Muñoz-Sabater J., de Rosnay P. and Fouilloux A., 2011: Use of SMOS data at ECMWF; ECMWF Newsletter no 127, 23-27.

Myneni, R. B., S. Hoffman, Y. knyazikhin, J. L. Privette, J. Glassy, Y. Tian, Y. Wang, X. Song, Y. Zhang, G. R. Smith, A. Lotsch, M. Friedl, J. T. Morisette, P. Votava, R. R. Nemani, S.W. Running. 2002. "Global Products of Vegetation Leaf Area and Fraction Absorbed PAR from Year One of MODIS Data." *Remote Sensing of Environment* 83: 214–31.

Oleson, K. W., G. B. Bonan, J. Feddema, M. Vertenstein, C. S. B. Grimmond, 2008: An Urban Parameterization for a Global Climate Model. Part I: Formulation and Evaluation for Two Cities. *J. Appl. Meteor. Climatol.*, 47, 1038–1060.

Orsolini, Y. J., Senan, R., Balsamo, G., Doblas-Reyes, F. J., Vitart, F., Weisheimer, A., Carrasco, A., and Benestad, R. E.: Impact of snow initialization on sub-seasonal forecasts, *Climate Dyn.*, 41, 1969-1982, doi: 10.1007/s00382-013-1782-0, 2013.

Pappenberger, F., Cloke, H.L., Balsamo, G. Oki, P., Ngo-Duc, T., 2010. Global routing of surface and subsurface runoff produced by the hydrological component of the ECMWF NWP system, *International Journal of Climatology*, 30(4), 2155-2174

Pappenberger, F., Dutra, E., Wetterhall, F., Cloke, H.L, 2012, Deriving global flood hazard maps of fluvial floods through a physical model cascade, *Hydrology and Earth System Sciences*, 16, 4143-4156

Poli, P., H. Hersbach, D. Tan, D. Dee, J.-N. Thépaut, A. Simmons, C. Peubey, P. Laloyaux, T. Komori, P. Berrisford, R. Dragani, Y. Trémolet, E. Holm, M. Bonavita, L. Isaksen and M. Fisher, 2013: The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C), *ERA Report Series*, 14, 1-59.

Potter, C.S., J.T. Randerson, C.B. Field, P.A. Matson, P.M. Vitousek, H.A. Mooney, and S.A. Klooster, 1993: Terrestrial ecosystem production: A process model based on global satellite and surface data. *Glob. Biogeochem. Cyc.*, 7, 811-841.

Puma, M. J., and B. I. Cook (2010), Effects of irrigation on global climate during the 20th century, *J. Geophys. Res.*, 115, D16120, doi:10.1029/2010JD014122.

Richards L. A., 1931: Capillary conduction of liquids through porous mediums. *Physics*, 1 (5) 318-333.

Reichle, R. H., W. T. Crow, and C. L. Keppenne, 2008: An adaptive ensemble Kalman filter for soil moisture data assimilation, *Water Resources Research*, 44, W03423, doi:10.1029/2007WR006357.

Rutter, N., Essery, R., Pomeroy, J., Altimir, N., Andreadis, K., Baker, I., Barr, A., Bartlett, P., Boone, A., Deng, H. P., Douville, H., Dutra, E., Elder, K., Ellis, C., Feng, X., Gelfan, A., Goodbody, A., Gusev, Y., Gustafsson, D., Hellstrom, R., Hirabayashi, Y., Hirota, T., Jonas, T., Koren, V., Kuragina, A., Lettenmaier, D., Li, W. P., Luce, C., Martin, E., Nasonova, O., Pumpanen, J., Pyles, R. D., Samuelsson, P., Sandells, M., Schadler, G., Shmakin, A., Smirnova, T. G., Stahli, M., Stockli, R.,

- Strasser, U., Su, H., Suzuki, K., Takata, K., Tanaka, K., Thompson, E., Vesala, T., Viterbo, P., Wiltshire, A., Xia, K., Xue, Y. K., and Yamazaki, T., 2009: Evaluation of forest snow processes models (SnowMIP2), *J. Geophys. Res.*, 114, D06111, doi: 10.1029/2008JD011063.
- Salgado, R. and Le Moigne, P., 2010: Coupling of the FLake model to the Surfex externalized surface model. *Boreal Env. Res.* 15, 231–244.
- Samuelsson, P., Kourzeneva, E. and Mironov, D., 2010: The impact of lakes on the European climate as simulated by a regional climate model. *Boreal Env. Res.* 15, 113–129.
- Sandu, I., A. Beljaars, P. Bechtold, T. Mauritsen, G. Balsamo, 2013: Why is it so difficult to represent stably stratified conditions in Numerical Weather Prediction (NWP) models? *J. Adv. Model. Earth Syst.*, 5, 117-133.
- Sandu, I., A. Beljaars, G. Balsamo, A. Ghelli, 2011: Revision of the surface roughness length table, *ECMWF Newsletter*, 130, 8-10.
- Sandu, I., A. Beljaars, G. Balsamo, 2014: Improving the representation of stable boundary layers, *ECMWF Newsletter No. 138 – Winter 2013/2014*, 24-29.
- Schulze-Makuch, D., Carlson, D. A., Cherkauer, D. S. and Malik, P. (1999), Scale Dependency of Hydraulic Conductivity in Heterogeneous Media. *Ground Water*, 37:904–919. doi: 10.1111/j.1745-6584.1999.tb01190.x
- Scipal K, Drusch M, Wagner W., 2008: “Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system.” *Adv. Water Resources* 31(8): 1101–1112, DOI: 10.1016/j.advwatres.2008.04.013
- Seneviratne, S.I., N. Nicholls, D. Easterling, C.M. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, and X. Zhang, 2012: Changes in climate extremes and their impacts on the natural physical environment. In: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Edited by Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC), pp. 109-230.
- Seneviratne, S.I, M. Donat, B. Mueller, and L.V. Alexander, 2014: No pause in the increase of hot temperature extremes. *Nature Climate Change*, 4, 161-163.
- Seuffert, G., Wilker, H., Viterbo, P., Mahfouf, J.-F., Drusch M. and Calvet J.-C., 2003: Soil moisture analysis combining screen-level parameters and microwave brightness temperature: A test with field data. *Gephy. Res. Lett.*, 30, DOI: 10.1029/2003GL017128.
- Shukla J and Mintz Y., 1982: Influence of land-surface evaporation on the earth’s climate, *Science* 215, 1498-1501.

Taylor, C.M., D.J. Parker and P.P. Harris, 2007: An observational case study of mesoscale atmospheric circulations induced by soil moisture, *Geophys. Res. Lett.* 34, L15801, doi:10.1029/2007GL030572.

van den Hurk B., Ettema J., Viterbo P., 2008: Analysis of soil moisture changes in Europe during a single growing season in a new ECMWF soil moisture assimilation system. *J Hydrometeorol* 9:116–131. doi: 10.1175/2007JHM848.1.

van den Hurk, B. and P. Viterbo, 2003: The Torne-Kalix PILPS 2(e) experiment as a test-bed for modifications to the ECMWF land surface scheme. *Global and Planetary Change*, 38, 165-173.

Van den Hurk, B., P. Viterbo, A. Beljaars and A. K. Betts, 2000: Offline validation of the ERA40 surface scheme. ECMWF Tech Memo, 295, 43 pp., European Centre. for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, England, UK.

Viterbo, P., and A. K. Betts, 1999: Impact on ECMWF forecasts of changes to the albedo of the boreal forests in the presence of snow. *J. Geophys. Res.*, 104 (D22), 27 803–27 810.

Viterbo, P., A. Beljaars, J.-F. Mahfouf, and J. Teixeira, 1999: The representation of soil moisture freezing and its impact on the stable boundary layer. *Quart. J. Roy. Meteor. Soc.*, 125, 2401– 2426, doi:10.1002/qj.49712555904.

Wang W., and A. Kumar, 1998: A GCM assessment of atmospheric seasonal predictability associated with soil moisture anomalies over North America. *J. Geophys. Res.*, 103, 28637–28646.

Wang, Y., A. Kann, M. Bellus, J. Pailleux and C Wittmann, 2010: A strategy for perturbing surface initial conditions in LAMEPS, *Atmospheric Science Letters*, 11(2), 108–113, DOI: 10.1002/asl.260.

Wedi, N., 2014: Increasing horizontal resolution in NWP and climate simulations — illusion or panacea?, *Philos Trans A Math Phys Eng Sci.* 2014 Jun 28;372(2018):20130289. doi: 10.1098/rsta.2013.0289.

Williams J.L. and Maxwell R.M., 2011: Propagating Subsurface Uncertainty to the Atmosphere Using Fully Coupled Stochastic Simulations. *Journal of Hydrometeorology*, 12, 690-701. DOI: 10.1175/2011JHM1363.1.

Wisser, D., S. Frohling, E. M. Douglas, B. M. Fekete, C. J. Vörösmarty, and A. H. Schumann, 2008: Global irrigation water demand: Variability and uncertainties arising from agricultural and climate data sets, *Geophys. Res. Lett.*, 35, L24408, doi:10.1029/2008GL035296.

Yamazaki, D., S. Kanae, H. Kim, and T. Oki, 2010: A physically-based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.* 47, W04501, doi:10.1029/2010WR009726.

Yamazaki, D., Kanae, S., Kim, H. & Oki, T. A physically based description of floodplain inundation dynamics in a global river routing model. *Wat. Resour. Res.* 47, W04501 (2011)

Yamazaki, D., H. Lee, D.E. Alsdorf, E. Dutra, H. Kim, S. Kanae, and T. Oki, 2012: Analysis of the water level dynamics simulated by a global river model: A case study in the Amazon River. *Water Resour. Res.*, 48, W09508, doi:10.1029/2012WR011869.

Zhou, Y. D. McLaughlin, and D. Entekhabi, 2006: Assessing the performance of the ensemble kalman filter for land surface data assimilation, *Monthly Weather Review*, 134(8), 2128-2142.

Zsoter, E., F. Pappenberger, D. Richardson, 2014: Sensitivity of model climate to sampling configurations and the impact on the Extreme Forecast Index, *Met. Applic.*, DOI: 10.1002/met.1447.

Dataset references

ESA, GlobCover, 2010: <http://dup.esrin.esa.int/prjs/prjs68.php>.

SRTM, The Shuttle Radar Topography Mission, *Rev. Geophys.*, 45, RG2004, doi:10.1029/2005RG000183, 2014: <http://www.dgadv.com/srtm30/>.

GLOBE, GLOBE Task Team and others (Hastings, David A., Paula K. Dunbar, Gerald M. Elphinstone, Mark Bootz, Hiroshi Murakami, Hiroshi Maruyama, Hiroshi Masaharu, Peter Holland, JohnPayne, Nevin A. Bryant, Thomas L. Logan, J.-P. Muller, Gunter Schreier, and John S. MacDonald), eds., 1999. The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80305-3328, U.S.A. Digital data base on the World Wide Web (URL: <http://www.ngdc.noaa.gov/mgg/topo/globe.html>).

Greenland DEM, Byrd Polar Research Center (BPRC), Ohio State University, http://bprc.osu.edu/rsl/greenland_data/dem/index.html.

Iceland DEM, Icelandic Meteorological Service (IMO) and National Land Survey of Iceland (NLSI).

RAMP2, Liu, H., K. Jezek, B. Li, and Z. Zhao, Radarsat Antarctic Mapping Project digital elevation model version 2. Boulder, Colorado USA: National Snow and Ice Data Center. Digital media.

University of Massachusetts, Boston, [http://www.umb.edu/spectralmass/terra\(aqua\)modis/v006/v004](http://www.umb.edu/spectralmass/terra(aqua)modis/v006/v004)

GLCC2, US geological survey (USGS), http://edc2.usgs.gov/glcc/globdoc2_0.php.