How well do operational Numerical Weather Prediction configurations represent hydrology?

Ervin Zsoter, European Centre for Medium-Range Weather Forecasts, Reading, UK; Department of Geography and Environmental Science, University of Reading, Reading, UK

Current address: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK

Hannah Cloke, Department of Geography and Environmental Science, University of Reading, Reading, UK; Department of Meteorology, University of Reading, Reading UK; Department of Earth Sciences, Uppsala University, Uppsala, Sweden and Centre of Natural Hazards and Disaster Science, CNDS, Uppsala, Sweden

Elisabeth Stephens, Department of Geography and Environmental Science, University of Reading, UK

Patricia de Rosnay, European Centre for Medium-Range Weather Forecasts, Reading, UK

Joaquin Muñoz-Sabater, European Centre for Medium-Range Weather Forecasts, Reading, UK

Christel Prudhomme, European Centre for Medium-Range Weather Forecasts, Reading, UK

Florian Pappenberger, European Centre for Medium-Range Weather Forecasts, Reading, UK and College of Hydrology and Water Resources, Hohai University, Nanjing, China

Early Online Release: This preliminary version has been accepted for publication in *Journal of Hydrometeorology,* may be fully cited, and has been assigned DOI 10.1175/JHM-D-18-0086.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Corresponding author: E. Zsoter, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK (ervin.zsoter@ecmwf.int)

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Abstract

Land surface models (LSMs) have traditionally been designed to focus on providing lower boundary conditions to the atmosphere with less focus on hydrological processes. State of the art application of LSMs include land data assimilation system (LDAS) which incorporates available land surface observations to provide an improved realism of surface conditions. While improved representations of the surface variables (such as soil moisture and snow depth) make LDAS an essential component of any Numerical Weather Prediction (NWP) system, the related increments remove or add water, potentially having a negative impact on the simulated hydrological cycle by opening the water budget.

This paper focuses on evaluating how well global NWP configurations are able to support hydrological applications, in addition to the traditional weather forecasting. River discharge simulations from two climatological reanalyses are compared: one ‘online’ set which includes land-atmosphere coupling and LDAS with an open water budget, and also an ‘offline’ set with a closed water budget and no LDAS.

It was found that while the online version of the model largely improves temperature and snow depth conditions, it caused poorer representation of peak river flow, particularly in snowmelt-dominated areas in the high latitudes. Without addressing such issues there will never be confidence in using LSMs for hydrological forecasting applications across the globe. This type of analysis should be used to diagnose where improvements need to be made; considering the whole Earth System in the data assimilation and coupling developments is critical for moving towards the goal of holistic Earth System approaches.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

1 Introduction

Land surface models (LSMs) have traditionally been designed to focus on providing lower boundary conditions to the atmosphere by describing the vertical fluxes of energy and water between the land surface and the atmosphere, with less focus on predicting runoff (Mengelkamp et al. 2001). LSMs therefore maximise the quality of the atmospheric forecast, but do not necessarily bring the same benefits in the representation of the hydrological cycle (Kauffeldt et al. 2015).

There is a wide literature on assessing the hydrological capabilities of LSMs and describing various improvements in the modelling of the hydrological cycle (e.g. Balsamo et al. 2009; Wang et al. 2016; Blyth et al. 2011; Wu et al. 2014). However, there are significant limitations in the representation of hydrological fluxes and storages in LSMs, largely due to the large-scale focus of LSM applications, which has led to the neglect of some important processes for runoff generation (Overgaard et al. 2006; Le Vine et al. 2016), including inadequate snowmelt processes (Dutra et al. 2012, Zaitchik and Rodell 2009).

Data assimilation is an essential part of any Numerical Weather Prediction (NWP) system (Rabier 2005). It is designed to provide initial conditions for the Earth System by updating the model in all of the components: atmosphere, land, ocean and sea ice. State of the art NWP configurations, such as used at the European Centre for Medium-Range Weather Forecasts (ECMWF), include both an LSM and a land data assimilation system (LDAS). The objective of the data assimilation in this context is to combine the land surface model state with the available land surface observations to initialise the land surface model prognostic variables of the forecasting system (Bélair et al. 2003). The current ECMWF LDAS analyses soil moisture, soil temperature, snow mass, density and temperature (de Rosnay et al. 2014). Land

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

data assimilation was shown to contribute significantly to more skilful atmospheric forecasts, with the soil moisture data assimilation also proven essential in countering a positive precipitation/evapotranspiration feedback which can cause large positive precipitation biases (e.g. de Rosnay et al. 2013; Drusch et al. 2007, Beljaars et al. 1996).

While the improved surface conditions make LDAS an essential component of the ECMWF NWP system, by design the related increments remove or add water which can potentially have a negative impact on the representation of the hydrological cycle by opening the water budget (Zaitchik and Rodell 2009; Arsenault et al. 2013; Andreadis and Lettenmaier, 2006; De Lannoy et al. 2012; Pan and Wood, 2006). On the contrary, in a system without LDAS and coupling, the errors resulting from atmospheric forcing insufficiencies and imperfect land surface process representations are not corrected by the assimilation of land surface observations.

As an ideal configuration, an Earth System model should always maintain a closed water budget, where the amount of water in the system remains the same. By opening the water budget, river discharge biases could emerge in situations where the LSM has energy balance bias that is not corrected by the assimilation but only by accurate precipitation and snow accumulation forcing. For example, if the snow in the LSM is melting too slowly, this forces the LDAS to remove water (through snow) artificially to correct for the excessive amount of snow on the surface. If the water that is removed with the snow (and thus could not melt) is not retained within the Earth System that could lead to soil water deficit downstream, potentially causing an incorrect rate of river discharge. In such cases, LDAS could lead to replace incorrect snowmelt timing issue with incorrect snowmelt runoff amount.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Thus, an open water budget could cause problems for associated hydrological forecasting applications, which uses runoff calculated from LSMs with LDAS, such as the Global Flood Awareness System (GloFAS; Alfieri et al. 2013). As global hydrological modelling is increasingly possible with the improved realism that the state-of-the-art LSMs can nowadays offer (Overgaard et al. 2006), it is important to investigate how an LSM with LDAS can support the combined task of traditional weather forecasting and hydrology at the same time. This investigation was undertaken with this dual focus in mind, by analysing the hydrological cycle and the open water budget issues that can help the Earth System model developments with highlighting areas where the coupled system with LDAS does not yet work effectively for flood simulations.

In order to understand how well an NWP configuration with LSM and LDAS represents hydrology, and in particular to interpret the influence of the LDAS on hydrological simulations from LSMs, in this paper river discharge simulations from two climatological reanalyses of GloFAS are compared: one operational set which includes land-atmosphere coupling and LDAS with an open water budget, and also an 'offline' set with a closed water budget and no LDAS. From these two datasets, a range of hydrological and atmospheric variables will be analysed globally.

2 System Description, datasets and methods

Two hydrological experiments, ONLINE (run in operational mode with active land-atmosphere coupling and LDAS) and OFFLINE (run in offline mode without coupling and LDAS) provide time series of various surface variables (e.g. 2-metre temperature, snow depth and runoff), and also discharge after routing the runoff. Figure 1 highlights the schematic of ONLINE and OFFLINE with the main characteristics, components and data periods. In this

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

section the two experiments with the model and data aspects, and the data analysis methods will be described in detail.

2.1 Land surface model HTESSEL

The hydrological component of the analysed data sets is based on the HTESSEL land surface model (The Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land; Balsamo et al. 2009; Balsamo et al. 2011). HTESSEL is part of the ECMWF NWP system and used in coupled land-atmosphere mode on time ranges from short-range to seasonal forecasts. It includes a snow parameterisation based on a single-layer snow pack model (Dutra et al., 2010). The soil vertical diffusion solves the Richards equation using a four-layer vertical discretisation with layer depths at 7 cm, 28 cm, 100 cm and 289 cm (Balsamo et al. 2009). HTESSEL provides boundary conditions for the atmosphere (heat, moisture, and momentum) by simulating water and energy budgets on the surface and through the soil, snowpack and vegetation interception. HTESSEL generates surface (fast) and subsurface (slow) runoff components at each grid point (Balsamo et al. 2009). Surface runoff depends on the standard deviation of the orography, soil texture and soil moisture, while subsurface runoff is determined by the soil water percolation.

2.2 Land data assimilation

The ECMWF LDAS is part of the ECMWF Integrated Forecasting System (IFS). It is coupled to the atmospheric four-dimensional variational (4-D-Var) data assimilation scheme (Rabier et al. 2000), both using a 12-hour assimilation window. The upper air and land surface analyses are running separately and used to initialise a coupled land-atmosphere short-term forecast, which provides the background for the next data assimilation window. The land data assimilation relies on advanced methods to optimally combine in situ and

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

satellite observations with model background information. A schematic diagram of the ECMWF LDAS is provided in Figure 2.

Initial implementations of the ECMWF LDAS relied on simple assimilation methods for snow and soil moisture analyses (Drusch et al. 2004, Mahfouf et al. 2000), with air temperature and humidity measurements being the main input for the soil moisture analysis (Mahfouf et al. 2000, Drusch et al. 2007). The system has evolved in the past decade to use a more physically based approach and to combine satellite and in situ data in the soil analysis (de Rosnay et al. 2014, de Rosnay et al. 2013, Albergel et al. 2012).

In the current LDAS, a simplified Extended Kalman Filter (SEKF) is used to analyse soil moisture. The approach combines analysed 2-metre air temperature and humidity with satellite measurements from the ASCAT (Advanced Scatterometer) sensor on board of MetOp, as described in de Rosnay et al. (2013) and Albergel et al (2012). For snow, a two-dimensional optimal interpolation (OI) is used to analyse snow mass and snow density following the method described in Brasnett et al. (1999). In situ snow depth observations, available on the SYNOP network are used along with the 4km resolution snow cover product from the NOAA/NESDIS (National Environmental Satellite, Data, and Information Service) Interactive Multi-sensor Snow and Ice Mapping System (IMS) product (Helfrich et al. 2007).

Even though it provides significant improvements to the atmospheric forecasts and independent situ snow depth measurements (de Rosnay et al. 2015), the current ECMWF snow data assimilation follows a relatively basic method. Operational NWP configurations generally rely on simple approaches, compared to research environment, that are based on more sophisticated snow assimilation methods using in situ and remotely sensed observations

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

(e.g. Helmert et al. 2018; De Lannoy et al. 2012; Pan and Wood 2006; Slater and Clark 2006).

The ECMWF LDAS and its performance is presented and discussed in de Rosnay et al. (2014), and de Rosnay et al. (2015). A full description of the technical implementation is provided in the IFS documentation (https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation). The system used for this study is that used for the production of ERA5 (section 2.6), with IFS cycle 41r2 at a resolution of ~31 km.

2.3 CaMa-Flood river-routing

The Catchment-based Macro-scale Floodplain model (CaMa-Flood; Yamazaki et al. 2011) was applied in this study to simulate the hydrodynamics and produce river discharge from the HTESSEL runoff outputs. CaMa-Flood is a distributed global river-routing model which uses a river network map and routes runoff to oceans or inland seas. The CaMa-Flood model was chosen for the routing component as it had already been used in several similar climatological research experiments such as Emerton et al. (2017).

2.4 GloFAS

GloFAS is one of the few global scale flood forecasting systems that currently exist (Emerton et al. 2016). It is part of the Copernicus Emergency Management Service (CEMS), developed by the Joint Research Centre of the European Commission (JRC) and ECMWF. The HTESSEL runoff output is coupled to the Lisflood hydrological model over a global river network to produce river discharge with a forecast horizon of 30 days across a global river network at 0.1 degree resolution (van der Knijff et al. 2010; Alfieri et al. 2013). As part of the GloFAS configuration, the real-time river discharge forecasts are compared with

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

climatological simulations (called reanalysis) to detect the likelihood of high flow situations. These real-time and climatological datasets also present a unique opportunity for experimental analysis (Emerton et al. 2017; Stephens et al. 2015).

2.5 Offline land surface modelling

The current GloFAS operational set-up uses a climatology based on the ERA-Interim/Land reanalysis of ECMWF (Balsamo et al., 2015). ERA-Interim/Land is an improved version of the ERA-Interim reanalysis (Dee et al. 2011) produced with an improved version of HTESSEL, run offline, using a rescaling of monthly precipitation totals with GPCP v2.2 (Huffman et al. 2009; Balsamo et al. 2010). “Offline” HTESSEL simulations, such as the OFFLINE experiment in this study, are uncoupled from the atmosphere, without the LDAS and forced with near-surface meteorological input data such as temperature, specific humidity, wind speed, surface pressure, radiative fluxes and water fluxes. Offline land surface only simulations are an affordable way of achieving land surface improvements and this offline research methodology has been used in numerous studies with HTESSEL in the last few decades (e.g. Agusti-Panareda et al. 2010; Dutra et al. 2010; Dutra et al. 2011; Haddeland et al. 2011).

2.6 ERA5 reanalysis

The 5th generation global climate reanalysis (succeeding ERA-Interim) at ECMWF is ERA5 (Hersbach and Dee 2016). ERA5 is a key contribution to the EU-funded Copernicus Climate Change Service (C3S). ERA5 will cover the period 1950-present and is in production with 2008-2017 already officially released. The release of the remaining period is foreseen by end of 2018. ERA5 will then continue running in (non-quality assured mode) near-real time with only a few days delay. The data is open access and free to download for all uses

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

(https://climate.copernicus.eu/).

ERA5 uses the IFS cycle 41r2 and it relies on land surface model and assimilation configuration that are consistent with those used for operational NWP with coupled land-atmosphere simulations and the latest soil moisture and snow assimilation (see sections 2.1 and 2.2 above). ERA5 has a high resolution component at ~31 km which is used in this study (hereafter called ERA5-HRES). In ERA5-HRES, variables (analysis and short range forecasts generated at 06 and 18 UTC) are available hourly. Variables that are valid for a period, e.g. precipitation or runoff with an accumulation time, are provided as hourly forecasts.

At the time of writing approximately 28 years of ERA5-HRES data was available in the ECMWF MARS data archive in three separate periods: 1985-1987, 1989-1995 and 1999-2016. The first years (1985, 1989 and 1999) were used as spin up years, so in total 25 years of daily river discharge and other surface data could be processed for the analysis (hereafter called ERA5-D25).

2.7 Experimental set-up

In the ONLINE experiment, the operational ERA5-HRES reanalysis data was used directly from all three ERA5-HRES periods for land surface variables, including runoff, produced by coupled land-atmosphere model with LDAS and an open water budget (figure 1). In the OFFLINE experiment, on the other hand, three standalone HTESSEL runs were set up, one for each of the periods, to reproduce the land surface variables in land surface only mode without the impact of coupling and LDAS, but with a closed water budget. As ERA5 has a recent model cycle (41r2), the same HTESSEL version could be used in the offline experiment as in the operational ERA5.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

In the ECMWF NWP system, there is no option currently to run the land-atmosphere coupling and LDAS separately. Either both are active as in ONLINE, or neither of them as in OFFLINE. It would be interesting to separate the impact of these two contributing modelling options, but as they are too strongly interwoven the separation would require a very large effort, which is outside of the scope of this study.

In the OFFLINE experiment, the offline HTESSEL model was forced with hourly ERA5-HRES atmospheric data, wherever it was possible on the lowest model level, with an hourly model time step. The model was run on the original horizontal resolution of ERA5-HRES (~31 km). For precipitation, temperature, specific humidity, wind speed and surface pressure the hourly analysis fields were applied, while for radiation and precipitation fluxes the first 12-hour period of the 06:00 and 18:00 UTC short-range forecasts were used to cover each 24-hour periods.

The river discharge was generated by routing the runoff using CaMa-Flood for both the ONLINE and OFFLINE datasets over the ~25 km river network. CaMa-Flood was run with a 1-hour time step and a 24-hour output frequency to match the 24-hour reporting frequency of the river discharge observations.

2.8 River discharge observations

In this study, daily river discharge observations used in the GloFAS system are selected. These are mostly from the Global Runoff Data Centre (GRDC) archive, an international depository of river discharge observations and associated metadata.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

The observations consist of a network of approximately 900 river gauging stations with upstream areas over 10,000 km^2, selected from the catchments used in Zsoter et al. (2016). After visual inspection those catchments that showed a clear non-realistic behaviour and/or influence of dams were excluded. A minimum of 9 years, with at least 330 days in each of those calendar years, were selected as criteria for the stations to be included in the river discharge analysis. This is quite a short period, but due to the limited availability in more recent years, it was accepted as a compromise. In total 590 stations could be processed globally leaving large blank areas mostly in Asia and Africa (Figure 3).

2.9 Annual peak river discharge

For the river discharge verification, the annual peak river discharges from the two ERA5-HRES simulations were determined in each calendar year as the highest value in the ±30-day window around the observed annual maximum river flow. The 30-day window was defined as a safeguard to avoid detecting high skill with similar peaks in observation and simulation of completely different flood waves at very different periods of the year.

2.10 Water budget increments

This study focuses on the impact of the water budget closure on river discharge. In order to analyse this the daily (00-00 UTC) water budget error term (dA) was computed as:

$$dA = P - E - R - dS \qquad (1)$$

where P is precipitation, E is evapotranspiration, R is runoff, all taken as the sum of the hourly forecast values (24 in total) in the ONLINE experiment from the 00-00 UTC period and dS is the change in the storage term (water content in the soil including all four layers and also in the snow cover) computed as the difference between the two subsequent 00 UTC analysis values in ONLINE (representing the change in the water content during the 24-hour

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

period). Even though the water budget error is zero in OFFLINE (the water budget is closed), the contributing variables can help identifying the behaviour of the surface processes in both the ONLINE and OFFLINE simulations.

The imbalance in the amount of water that is not accounted for in the ONLINE water budget effectively comes from the snow depth and soil moisture increments in LDAS which remove or add water in the system. The daily increments (valid for a 00-00 UTC 24-hour period) are computed as the sum of two increment values at 06 and 18 UTC (each day). Both of these increments are computed as the ERA5-HRES analysis value minus the corresponding 12-hour ERA5-HRES forecast value (initialised 12 hours earlier).

2.11 Daily 2-metre temperature and snow depth

The in situ surface synoptic observations (SYNOP) were used to verify 2-metre temperature and snow depth for both the OFFLINE and ONLINE experiments. The observing stations were filtered according to the station altitude difference to the model orography and only those were used which had less than 150 metres discrepancy, as orography has control on both variables and large differences would make the comparison unreliable. This maximum orography difference value was chosen in accordance with the general practice at ECMWF, where 100 metres is used to filter stations in the 2-metre temperature verification. For our study, a less stringent compromise value was preferred in order to increase the sample size and still guarantee good match between model and real orography.

2-metre temperature was verified for around local noon (Table 1), while for snow depth the first measurement of the calendar day was evaluated in case of sub-daily records. In total, observations from about 4000 stations for 2-metre temperature and 1500 stations for snow

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

depth were available for verification. For each catchment, a representative daily observation was also determined for both variables. For catchments with more than one SYNOP station available, these were calculated as the arithmetic average of the stations within the catchment. It has to be acknowledged that the observation network available was not dense enough to represent the full spatial variability of these surface variables, specially snow depth, which vary dramatically in space from one point to another (Molotch and Bales, 2005). However, for a global study on the hydrological impacts it is expected to be sufficient.

2.12 Climatologies

Daily climatologies were used for river discharge and other surface variables in this work for both observations and the two simulations. These data sets were produced with all potentially available 25 years of data in ERA5-D25, always matching the number of available nearly complete calendar years (with minimum 330 river discharge observations) for all the catchments. For each day of the year a 21-day window, centred over the day, was used which provided a minimum of about 180 values in the climate sample (with the 9 years minimum criteria). The only exception are 2-metre temperature and snow depth, where a fixed shorter period of 2000-2007 was used without the criteria of nearly complete years. As the 2-metre temperature and snow depth observation availability is much better in more recent periods and also less prone to missing values than river discharge, a shorter fixed period (when ERA5-HRES was available) is sufficient.

2.13 Verification statistics

A number of statistics were applied to evaluate the overall performance of the two climatological simulations in ERA5-D25 (Table 2). Several scores were selected in order to give a more representative description of the general behavior including the differences

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

between the ONLINE and OFFLINE experiments. This is recommended e.g. by Legates and McCabe (1999) as different scores demonstrate different aspects of the model attributes ultimately providing a more complete picture.

The climatological daily time series were compared to the observed data using mean error (ME), mean absolute error (MAE), Nash-Sutcliffe model efficiency (NSE; Nash and Sutcliffe 1970) and also Pearson correlation coefficient (R; Pearson 1896) in order to measure the fit between model and observations. In addition, the mean and standard deviation of the observed and modelled values were analysed with four additional indices, the percentage sample mean error, the percentage sample mean absolute error, the percentage sample standard deviation error and the percentage sample standard deviation absolute error.

Another very important aspect of hydrological model verification is the ability of the systems to correctly predict the extremes, as these events can cause the highest impact. To measure this, the timing and magnitude errors of the annual peaks were considered. Both the ME and MAE measures (mean of all years in the sample) were computed for the timing and for the percentage magnitude errors using the annual peaks over the 25 analysed years (for details on how the annual peaks were computed see Section 2.9). For the analysis of the data assimilation impact on 2-metre temperature and snow depth the ME and MAE scores were used. In this study verification was conducted on homogeneous samples across all compared scores for all the verified surface variables.

3 Results

The river discharge behaviour provides a useful indication of the hydrological differences between the ONLINE and OFFLINE simulations. However, in order to understand the

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

underlying processes better, the coupling and LDAS impact was also analysed globally and regionally based on the water budget and the related surface variables.

3.1 Snow depth and 2-metre temperature impact

The LDAS is designed to provide adequate initial surface conditions to the NWP forecasts. The impact on the hydrology could be demonstrated on two important surface variables: 2-metre temperature and snow depth (at least in snow impacted areas) which are relatively well observed variables and can be used to analyse the impact of the land-atmosphere coupling and LDAS on the surface globally in the two experiments. For details on how the observations were used please see Section 2.11.

The picture for 2-metre temperature is rather mixed geographically with an overall MAE improvement in ONLINE of around 0.3-0.4 °C as a global average up to 1-2 °C locally (not shown). This corresponds to about 20-30% decrease in MAE on average in ONLINE, with the impact of coupling and LDAS, compared to OFFLINE.

The improvement in the snow depth, which has much larger direct impact on the hydrology, is more pronounced, based on the stations used in this study. The errors in ONLINE are significantly reduced with most stations showing below ±1-2 cm ME (not shown), and decrease of MAE by as much as 10-20 cm in some of the snow dominant locations in the 50-70 latitude band (Figure 4). This is a very large improvement in ONLINE by removing 70-80% (as global average) of the errors found in the OFFLINE experiment. Countries of Central America, including Mexico, Venezuela and Columbia, tend to provide snow information in their SYNOP observations. In these regions both the model and the in situ stations mostly indicate snow free conditions, leading to very low MAE as shown in Figure 4.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Although the improvements are large, this does not necessarily mean that the simulation is generally better. In situ snow observations are associated to potential representativeness issues, particularly in mountainous areas. When assimilating a non-representative dataset at a coarse special scale, the results can potentially degrade, even though the match to the actual observations is better (Molotch and Bales, 2015). As the 2-metre temperature and snow depth observations used in this study for verification were also assimilated in ERA5, the result will favour to some extent the ONLINE experiment.

3.2 Global water budget analysis

The water budget is closed in OFFLINE by design, while in ONLINE the LDAS increments can add or remove water, which could potentially lead to large errors in the budget over a long period. The first aspect that was important to check is the amount of water that is lost or gained in a day on average in the hydrological cycle.

Figure 5 shows the average daily water budget errors (Figure 5a) and the related snow water equivalent (Figure 5b) and soil water content (Figure 5c) increments (for the definition of these terms please see Section 2.10). In Figure 5, negative values (red) indicate water removal by LDAS, while positive values (blue) show where water is added to the hydrological cycle.

The three figures highlight significant biases in the ONLINE experiment as these water budget errors represent generally ±10-25% of the total precipitation with locally even higher ratios (not shown). In addition, at latitudes higher than 50 degrees North the dominant pattern is a negative water budget error (Figure 5a). The major contributing factor to the clearly negative errors in this area is the correction of snow pack with LDAS removing snow to

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

account for possible inaccuracies in the HTESSEL snow scheme (Figure 5b). On average snow water increments are negative almost everywhere where snow is present. The only notable exception is in Canada where some central areas have positive water budget errors which could possibly come from a negative precipitation bias that needs to be compensated by LDAS.

Other areas of the world, the central USA, most of Amazonia, Africa, south Asia with India and also large parts of Australia show positive errors in Figure 5a, where extra water is added by LDAS. However, the positive errors are not exclusive as large parts of China, southeast US and areas in central South-America experience negative water budget errors in these mostly warm climatic conditions. Most of these increments come from the soil moisture assimilation impact (Figure 5c). The soil moisture assimilation can generally compensate for precipitation or 2-metre temperature biases. For example, if the 2-metre temperature is too low, the assimilation will remove water, therefore reducing evaporative cooling which subsequently increase the temperature in general.

3.3 Catchment-level process examination

To demonstrate how HTESSEL handles the land surface processes with and without coupling and LDAS, an in-depth case study analysis of the annual water budget cycle was performed for an example catchment on the Amur river in east Russia (see Figure 6, catchment no.13). This catchment is heavily snow impacted during winter and can demonstrate nicely the important aspects of the hydrological cycle behaviour with the LDAS in action.

In the HTESSEL hydrological cycle representation the input precipitation combined with the melted part of the snowpack (snowmelt) is distributed into evapotranspiration, runoff (as sum

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

of surface and sub-surface runoffs), snow water storage (falling snow part of the precipitation) and soil water storage (soil moisture in the four soil layers). The daily water budget error, computed as in Eq. 1 (without the snowmelt separated), is zero in OFFLINE, while ONLINE can show errors due to the increments adding or removing water. Figure 7 summarises the annual cycle of all the water budget contributing variables.

The displayed variables are daily climatological means calculated as described in Section 2.12. The following variables are shown in Figure 7: simulated precipitation (same for both experiments), evapotranspiration, runoff, soil water and snow water storage terms (in Eq. 1) for both ONLINE and OFFLINE; snow and soil water content increments for ONLINE; simulated snowmelt, snow depth and river discharge for both the ONLINE and OFFLINE experiments, and finally the corresponding river discharge and snow depth observations.

Figure 7 shows that for the Amur the ONLINE simulation significantly improves the representation of snow depth, but as consequence, by the snow assimilation removing a lot of snow, it drastically reduces the river discharge peak seen during the snowmelt season. The explanation of this conclusion with detailed analysis of the evolution of the different surface variables in the different seasons is given in the following:

- **Winter**: During December to February there is relatively little activity. The little amount of precipitation falls mostly as snow, building the snowpack. Some snow is removed by the assimilation through the small negative snow increments. Water leaves the bottom of the soil as sub-surface runoff with hardly any surface runoff. The OFFLINE simulation is generally similar to ONLINE, but snow depth bias shows increasingly positive values in OFFLINE due to the extra amount of water going into

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

the snowpack in the OFFLINE experiment from snowfall (especially during first half of the winter).

- **Spring**: From March, there is a pronounced snowmelt period in the model, peaking at the end of April, lasting until middle of June (with virtually zero snowpack in catchment average after middle of May). The increased precipitation in this spring period with the large amount of snowmelt increases the soil water content and also results in larger surface runoff output in both experiments. However, the snowmelt is much smaller in ONLINE during April-May as a direct consequence of the large negative snow increments (peaking early April) removing snow in the ONLINE experiment. Similarly, due to the smaller amount of available water in ONLINE, the surface runoff is also significantly smaller mainly in April/May. The snow depth errors peak in middle of March by about 5 cm in OFFLINE with no errors in ONLINE (as catchment average). The data assimilation rightly corrects this substantial positive snow bias, however, the removed snow will be missing from the water cycle as is highlighted by the unnoticeable spring peak river flow, which is higher in the OFFLINE simulation mainly due to the extra snowmelt.

- **Snowmelt problem**: This behaviour of HTESSEL with LDAS is rather surprising and at first it might sound as a contradiction. How can the correct snow conditions lead to such poor river discharge in the ONLINE experiment? A possible explanation could be the representativeness issue of some of the snow observations, which can potentially cause local degradation in some of the catchments. It can also be explained by the HTESSEL tendency to melt the snow too slowly (Dutra et al. 2012). In its simple, single layer snow scheme, too much snow accumulates into the snow pack

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

and then that snow melts too slowly. For example, during a 20 mm mixed snow/rain forecast event (10 mm liquid and 10 mm solid) the snow scheme will accumulate most of the 10 mm solid (snow) part of the precipitation into the snowpack regardless of the temperature conditions and melt only a little of this 10 mm. However, in reality a lot of that rain, sleet or wet snow would not accumulate on the ground, and instead most of it would melt straightaway. It seems the OFFLINE simulation gets the river discharge right mainly for the wrong reasons. Although the snowpack is clearly more poorly represented, the better timing with the delayed snowmelt (through the too slow melting) and the extra water in the snowpack, the OFFLINE experiment gets the runoff peak more correct.

- **Summer**: The water budget is balanced between precipitation and evapotranspiration with some soil water increments. During early summer water is taken out of the soil to cover the higher evapotranspiration. In OFFLINE more water leaves the soil which increases the runoff and also evapotranspiration. By August, however, the excess water from precipitation over evapotranspiration goes again into the soil which is more pronounced in ONLINE where the soil is drier. The end of summer river discharge peak is present in both simulations with the OFFLINE showing a better peak due to more water in the soil and subsequently higher surface and sub-surface runoff during all summer. The OFFLINE river discharge exceeds the ONLINE values all summer and the two will level out by September, when the runoffs become similar in the two experiments.

- **Autumn**: From the middle of September there is another smaller snowmelt period starting with the falling temperatures and bringing some negative snow increments in

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

the ONLINE simulation. The snow accumulates into the snowpack in both experiments, but again with a higher rate in OFFLINE, and also with larger snowmelt amounts in OFFLINE.

3.4 Regionally representative catchments

In the previous section the LDAS response was highlighted for an important weakness of HTESSEL with significant consequences on river discharge. In the following, the land-atmosphere coupling and LDAS impact is now demonstrated with a simplified representation of the annual water cycle in different geographical areas and also various climatic conditions for a selection of the world's catchments in Figure 8. The displayed variables are simulated snowmelt, evapotranspiration and river discharge in both the ONLINE and OFFLINE experiments, the snow and soil water increments for ONLINE and finally the river discharge observations. All values are daily climatological mean values as in Figure 7. The location of the catchments is provided in Figure 6.

In Figure 8, twelve catchments are selected to represent all main areas of the world where river discharge observations are available. Many of them are very large rivers, some of the catchments are dominated with mixed snow and soil moisture influence from the Northern Hemisphere while others, mainly in the tropics, are only soil moisture impacted. In table 3, the main catchment details are provided, complemented with the NSE and the percentage peak magnitude ME and MAE values for the catchments. The scores favouring the ONLINE experiment are displayed with bold numbers.

Figure 8 suggests that the decreased snowmelt is a general feature in ONLINE across the Northern Hemisphere as predicted already by Figure 5b. All displayed catchments have

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

generally lower river discharge in ONLINE, either concentrated over the high river discharge season (e.g. Ob (no. 1) and Yukon (no. 2)), or elongated over most of the year (e.g. Danube (no. 3) and Rhine (no. 4)). The snowmelt is universally smaller in the ONLINE simulation, with the LDAS removing snow at different periods of the year, which seems to be the driving force behind the river discharge differences.

The decreased amount of water has a mixed river discharge skill impact. For some catchments (Ob (no. 1), Yukon (no. 2), Columbia (no. 6), and the case study catchment on the Amur (no. 13)) the change during the high river discharge season is disadvantageous in ONLINE, confirmed by mostly negatively impacted scores, such as the NSE and the percentage peak magnitude MAE values in Table 3. On the other hand, for the Mississippi (no. 5), Danube (no. 3) and Rhine (no. 4) it is rather beneficial as the daily climatological mean river discharge is closer to the corresponding observations during the high season, accompanied with mainly positive skill changes in the ONLINE experiment as both NSE and percentage peak magnitude MAE improves (Table 3), except the Rhine catchment (no. 4) where the percentage peak magnitude MAE deteriorates.

In the warm climate, however, where soil water dominates the land surface processes (Xingu, Amazon, Hadejia, Ubangi, Zambesi and Flinders (no. 7-12)), the land-atmosphere coupling and LDAS impact on river discharge seems to be smaller than for the snow influenced catchments, and on evapotranspiration it tends to be larger. There are large biases over five of the six highlighted tropical catchments (the only exception of the Flinders river in Australia), where both the ONLINE and OFFLINE experiments show significant mismatch with the observed values for the total river discharge volume and also for the annual peaks. For example, as displayed in Table 3, on the Hadejia river in Nigeria the percentage peak

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

magnitude ME is 297% (the simulation is almost three time higher than the observation) in ONLINE which is significantly better than OFFLINE (the improvement is 139% in the percentage peak magnitude MAE). This points to the fact that even though the river discharge differences are smaller in relative terms, it can still lead to noticeable change in the scores for some of these highlighted catchments (Table 3).

Even though there is no clear systematic difference between the exclusively soil moisture and the mixed (snow and soil moisture) catchments in terms of river discharge skill impact, the snow clearly looks to carry a more direct influence on the river discharge volume and also on the river discharge skill.

3.5 Global river discharge analysis

In the previous sections it could be shown that the water budget is out of balance in the ONLINE simulation over large parts of the world leading to significant impact on the river discharge for the analysed list of catchments. As an extreme example, it was demonstrated that the snowmelt driven spring river discharge peak was almost completely missed in a large catchment in east Russia in ONLINE. After the individual catchment examples, a systematic analysis of the river discharge quality in the ONLINE and OFFLINE experiments is provided based on all available catchments globally.

Although a large number of scores were computed in this study, this section will focus only on the annual peak flow scores. The timing and magnitude of the high river discharges are both crucial aspects of river discharge simulations in any flood prediction system such as GloFAS. The accurate simulation of the river discharge peaks is essential to get the best possible guidance for the potentially most damaging floods. The analysed performance of the

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

annual peak river flows should give a good indication on the general ability of the two experiments to predict peaks.

Figure 9a highlights a large systematic percentage peak magnitude ME in the ONLINE simulation. Many catchments show over 50% error (either positive or negative) of the annual river discharge peaks on average. The majority of the Northern Hemispheric higher latitudes is overwhelmingly under predicted, while Amazonia, western USA and also many catchments in Africa are over predicted in the ONLINE experiment. The geographical pattern in Figure 9a is rather similar to the one seen in Figure 5a. Most of the catchments with significant negative values over the Northern Hemisphere and positive ones mainly in lower latitudes, do resemble well the water budget error pattern seen in Figure 5a.

The water budget imbalance, caused by the increments in LDAS, is only one of the many potential contributing factors to peak river flow errors (and in fact to general river discharge errors); atmospheric forcing biases, imperfect river routing and observation errors could also lead to large inaccuracies (Zhao et al. 2017).

The impact of the land-atmosphere coupling and LDAS seems to decrease the amount of water overwhelmingly in the rivers (decreased sample mean river discharge, not shown). The sample average river discharge increased only in the southern half of Brazil, in the central part of Canada and one or two catchments in Africa, East Asia and South Australia (not shown). It is expected that the decreased average river discharge in ONLINE should generally also result in lower annual peak river flows over most of the globe. Figure 9b shows that this decreasing tendency of the annual peaks in the ONLINE experiment coincides with widespread, quite large deterioration in the percentage peak magnitude MAE score

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

(increase of the annual peak magnitude errors) especially in Asia and Europe and the north western part of North America, where the majority of the catchments show significant negative bias in Figure 9a. On the other hand, quite a few catchments seem to benefit from the coupling and LDAS as the annual peak errors decrease especially in the western parts in North America, where there is a large cluster of catchments with noticeably smaller percentage peak magnitude MAE.

The river discharge peak timing bias in the ONLINE simulation is dominantly positive (peaks are too late) in the Northern Hemisphere and mainly negative (peaks too early) in the Tropics (not shown). However, the coupling and LDAS do not seem to have any systematic impact on this aspect of the peak river flows. There are noticeable differences but they have no distinguishable geographical pattern (not shown). It seems the short time series (9-25 annual values only) were not sufficient to extract any representative timing differences between the two experiments.

In addition to the analysis of the annual river discharge peak performance, the general fit between modelled and observed daily river discharge time series is also extensively measured by several scores. Table 4 shows a global summary giving an indication on the overall performance of the two experiments. The scores are calculated as global averages weighted by the square root of the catchment area size. This way a more representative picture can be provided by giving more emphasis on the larger catchments.

The generally decreasing amount of water leads to larger differences for most of the volume related bias scores. The percentage sample ME, the percentage sample standard deviation error and also the percentage peak magnitude ME scores all decrease significantly in the

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

ONLINE simulation, bringing the global biases closer to zero. The only exception is the discharge ME score which changes from a positive value to a negative one with similar magnitude. The better biases, however, do not necessarily help improving the river discharge skill globally; the scores presented in Table 4 provide a mixed picture, with some favouring the ONLINE while others the OFFLINE simulation. This agrees with the mixed scores shown in Table 3 for the regional example catchments. In general, the MAE, R, the percentage sample MAE and the percentage peak magnitude MAE values are all slightly better for OFFLINE, while the NSE and percentage sample standard deviation absolute error show improvement for ONLINE. And finally, the peak timing ME is slightly better for the OFFLINE experiment, while there is no difference in the global average peak timing MAE.

4 Discussion

In Section 3, the land-atmosphere coupling and LDAS impact on hydrology, including river discharge and the related water budget variables was analysed. The river discharge scores showed a mixed picture between the ONLINE and OFFLINE simulations with relatively similar global performance. Larger differences could be highlighted in certain regions, such as many of the snow dominant catchments in the Northern Hemisphere, where over many areas a large amount of water is missing from the hydrological cycle and causing downstream issues in river discharge especially during the snowmelt season in ONLINE.

The general decrease in the volume of water in the ONLINE experiment, mainly coming from the snow dominated areas where the assimilation removes snow, seems to be the primary impact on the hydrology. In soil moisture dominated areas the river discharge seems to be less impacted by the increments and the evapotranspiration rate holds a more important role.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Data assimilation is a very important component of any NWP system with a lot of effort and research concentrated on the use of observations to correct for random (day-to-day) errors. Data assimilation systems are not there to correct for systematic biases. The fact that LDAS produces consistent negative increments in snow covered areas in this study is pointing towards an apparent snow model bias. In contrast, a model affected by random errors only, would lead to data assimilation increments of both signs with close to zero annual mean values.

Other studies have also highlighted significant snow assimilation impacts on the water balance. For example, De Lennoy et al. (2012) showed that on a small catchment in Colorado (USA) the season averaged snowpack water content is largely decreased by the snow water equivalent assimilation in the Noah land surface model, and could only be overcome by scaling applied (to anomalies) to the observations prior to assimilation. Similarly, Arsenault et al. (2013) found that assimilating MODIS snow cover fraction observations into the CLM land surface model by a simple rule-based direct insertion and the one-dimensional ensemble Kalman filter methods, lead to substantial snowpack removal (without melting, thus causing negative bias in runoff), by both methods in Colorado and Washington.

In the ECMWF system, the snow increments are correcting for the systematic overestimation of the current HTESSEL snow scheme which melts the snow too slowly. Dutra et al. (2012) highlighted that although the current snow scheme provides a significant improvement over the previous one, it does not yet improve on the short-duration melting events during late winter and spring. They argued that the experimental multi-layer snow scheme was able to

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

reproduce, at least partially, those snowmelt episodes thanks to the top snow layer having a reduced thermal inertia.

The findings in this work are specific to the NWP configuration at ECMWF with the HTESSEL land surface model and the processes within. However, any LSM's ability to support hydrological simulations can be limited by inadequate handling of the processes, potentially causing a similar problem downstream in the hydrology. The areas highlighted here for ECMWF's HTESSEL in supporting the flood forecasting activities can be improved by some potential developments in the future. Some of the areas where substantial improvements could be achieved are described in the following below:

- A new multi-layer snow scheme is currently being tested at ECMWF which is similar to the one evaluated in Dutra et al. (2012). This improved snow scheme is expected to represent better the snow melt processes and therefore reduce the snow increments that currently remove a significant amount of water from the hydrological cycle. The hydrological context developed in this study will be used to aid this development of the new scheme.

- Another potential way of improving HTESSEL performance for hydrological applications would be to modify the LDAS by special handling of the snow increments in order to retain the water in the hydrological cycle during the data assimilation. For example, Zaitchik and Rodell (2009) proposed an interesting approach using near-future, snow-covered area observations to adjust the air temperature and precipitation forcing data in order to preserve the local hydrological balance. In another study, Pan and Wood (2006) developed a constrained ensemble

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Kalman filter method to assure closure of the water balance when assimilating hydrological observations. These types of studies rely on uncoupled systems and they would be difficult to implement in operational, real-time environment. However, they provide some insight on water budget closure in data assimilation, and they should be further investigated and adapted to coupled land-atmosphere NWP systems. On the longer term, further coupling between NWP and hydrological forecasting systems will be considered, opening thereby the possibility for coupled land-hydrology data assimilation. In this context, joint assimilation of land surface and river discharge observations will consistently correct the different components of the Earth System.

- In addition, the land surface development methodology including data assimilation techniques and process representation is continuously improved at ECMWF. The future inclusion of the LDAS scheme in the offline HTESSEL is in development. It will create an environment where the offline research work, including the reanalysis improvements (e.g. ERA5), could be done in a consistent way with the real-time forecast generation. In parallel to these developments, addressing the water budget closure in land-atmosphere data assimilation systems should be a priority in the future to ensure consistent high quality coupled NWP and hydrological forecasts.

GloFAS is one of the few existing flood forecasting systems that utilises an LSM (HTESSEL) for representing the hydrology (Emerton et al. 2016). Although we acknowledge that in some cases a simple routing model, initialised from observed upstream river levels (either from river gauges or satellite measurements), could be a simpler alternative to simulate downstream discharge on large rivers a few days in advance, e.g. in Hossain et al. (2014); in other cases where forecasts are required further in advance or where observations

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

are unavailable or of too low quality, a more complex modelling configuration, which represents hydrological fluxes, becomes essential. Regardless of some limitations (e.g. the one highlighted in the ECMWF NWP configuration), these complex models play crucial roles in harnessing the available predictability in the land-atmosphere system.

5 Conclusions

Understanding the impacts of both the data assimilation and land surface process representation in land surface models on simulated hydrological variables is very important, not only for improving the weather and climate forecasts, but specifically for supporting flood forecasting and other hydrological applications such as drought forecasting, and also for giving feedback about the Earth System. In this paper, the influence of land-atmosphere coupling and land data assimilation on global hydrological simulations from LSMs was evaluated. Two river discharge simulations from two climatological reanalyses (based on ERA5) were compared: one operational set which includes land-atmosphere coupling and LDAS with an open water budget, and also an offline HTESSEL set with a closed water budget and no LDAS.

It was found that while the ONLINE version of the model largely improves the 2-metre temperature and snow depth conditions, it is causing poor representation of peak river flow in snowmelt-dominated areas, particularly in the high latitudes. However, there are localised improvements to peak river flow, such as in the western United States. The LDAS increments remove or add water even on an annual average scale which inevitably leads to systematic water budget errors and subsequently contribute to significant errors in river discharge during times of peak flow downstream, something that is critical during times of flooding.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Implications for hydrological forecasting: This study has highlighted the impact of using land data assimilation in reanalysis products. Where data assimilation is adjusting snowpack in forecasting mode then there will also be important implications for hydrological predictions. Future studies should address how far ahead the impact of data assimilation propagates in hydrological forecasts. In addition, hydrological forecasting systems often use initial river conditions derived from climatology. In these circumstances using climatological products derived using data assimilation methodologies could lead to issues with the hydrological forecasts. There are also related issues for forecasting systems such as GloFAS which compare model output to climatology to provide early awareness of extreme events – consistency between operational and climatological configurations goes some way to bypass this problem, and this conclusion has directly influenced the design of the new GloFAS-seasonal system (Emerton et al. 2018).

Implications for land surface modelling and data assimilation: Data assimilation is designed to compensate for noise errors and not systematic bias. In the case of the current HTESSEL snow assimilation scheme it is doing the latter; compensating for system deficiencies such as the slow snowmelt process. This paper has discussed potential ways of addressing water budget deficiencies in land surface approaches, for example including multiple layers within the HTESSEL snow scheme or moving towards data assimilation that conserves the water budget.

Without addressing such issues there will never be confidence in using LSMs for hydrological forecasting applications across the globe. This type of analysis should be used to diagnose where improvements need to be made; considering the whole Earth System in data

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

assimilation and coupling developments is critical for moving towards the goal of holistic Earth System approaches.

6 Acknowledgements

Ervin Zsoter‘s PhD is supported by the Wilkie Calvert Co-Supported PhD Studentships at the University of Reading. Ervin Zsoter was supported by the Copernicus Emergency Management Service - Early Warning Systems (CEMS-EWS (EFAS)). Hannah Cloke is supported by the TENDERLY project: Towards END-to End flood forecasting and a tool for ReaL-time catchment susceptibility UK NERC Flooding From Intense Rainfall (FFIR) programme, NE/K00896X/1. Elisabeth Stephens and Hannah Cloke are supported by the FATHUM project: Forecasts for AnTicipatory HUManitarian Action funded by UK NERC as part of their Science for Humanitarian Emergencies & Resilience (SHEAR) programme, NE/P000525/1. We are also grateful to The Global Runoff Data Centre, 56068 Koblenz, Germany for providing the observation dataset for our river discharge analysis.

7 Author contributions

Ervin Zsoter designed the experiment, produced the ERA5-D25 data sets, carried out the river discharge data analysis and led the writing of the manuscript. Hannah Cloke and Liz Stephens assisted with posing the research question, and designing the analysis, Patricia and Joaquin helped with the scientific analysis of data assimilation and coupling issues, Christel Prudhomme and Florian Pappenberger helped design the research methodology. All authors assisted with writing the manuscript.

8 References

Albergel C., P. de Rosnay, G. Balsamo, L. Isaksen and J. Muñoz Sabater, 2012: Soil moisture

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

analyses at ECMWF: evaluation using global ground-based in situ observations, J. Hydrometeor., 13, 1442-1460, https://doi.org/10.1175/JHM-D-11-0107.1.

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger, 2013: GloFAS - global ensemble streamflow forecasting and flood early warning, Hydrology and Earth System Sciences, **17**, 1161–1175, https://doi.org/10.5194/hess-17-1161-2013.

Agustí-Panareda, A., G. Balsamo, and A. Beljaars, 2010: Impact of improved soil moisture on the ECMWF precipitation forecast in West Africa, Geophys, Res. Lett., 37, L20808, https://doi.org/10.1029/2010GL044748.

Andreadis, K., and D. Lettenmaier, 2006: Assimilating remotely sensed snow observations into a macroscale hydrology model, Advances in Water Resources, 29(6), 872-886, https://doi.org/10.1016/j.advwatres.2005.08.004.

Arsenault, K. R., P. R. Houser, G. J. M. De Lannoy, and P. A. Dirmeyer, 2013: Impacts of snow cover fraction data assimilation on modeled energy and moisture budgets, J Geophys Res - Atmos, https://doi.org/10.1002/jgrd.50542.

Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts, 2009: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, J. Hydrometeor., **10**, 623–643. http://dx.doi.org/10.1175/2008JHM1068.1.

Balsamo, G., S. Boussetta, P. Lopez, and L. Ferranti, 2010: Evaluation of ERA-Interim and

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

ERA-Interim-GPCP-rescaled precipitation over the U.S.A., Era report Series 01/2010, 10 pp, ECMWF, Reading, UK.

Balsamo, G., F. Pappenberger, E. Dutra, P. Viterbo, and B. van den Hurk, 2011: A revised land hydrology in the ECMWF model: a step towards daily water flux prediction in a fully-closed water cycle, Hydrol. Proc., **25**, 1046–1054, https://doi.org/10.1002/hyp.7808.

Balsamo, G., and Coauthors, 2015: ERA-Interim/Land: a global land surface reanalysis data set, Hydrol. Earth Syst. Sci., **19**, 389–407, https://doi.org/10.5194/hess-19-389-2015.

Bélair S., L.-P. Crevier, J. Mailhot, B. Bilodeau, Y. Delage, 2003: Operational Implementation of the ISBA Land Surface Scheme in the Canadian Regional Weather Forecast Model. Part I: Warm Season Results, J. Hydrometeor., 4, 352-370, https://doi.org/10.1175/1525-7541(2003)4<352:OIOTIL>2.0.CO;2.

Beljaars A. C. M., P. Viterbo, M. Miller, A. K. Betts, 1996: The anomalous rainfall over the United States during July 1993: Sensitivity to land surface parameterization and soil anomalies. Mon. Weather Rev. 124, 362–383.

Blyth, E., D. B. Clark, R. Ellis, C. Huntingford, S. Los, M. Pryor, M. Best, and S. Sitch, 2011: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale, Geosci. Model Dev., 4, 255-269, https://doi.org/10.5194/gmd-4-255-2011.

Brasnett B., 1999: A global analysis of snow depth for numerical weather prediction. J Appl

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Meteorol 38:726–740, https://doi.org/10.1175/1520-0450(1999)038<0726:AGAOSD>2.0.CO;2.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., **137**, 553–597, https://doi.org/10.1002/qj.828.

Drusch M, D. Vasiljevic P. Viterbo, 2004: ECMWF s global snow analysis: assessment and revision based on satellite observations. J Appl Meteorol 43:1282–1294, https://doi.org/10.1175/1520-0450(2004)043<1282:EGSAAA>2.0.CO;2.

Drusch M., and P. Viterbo, 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, https://doi.org/10.1175/MWR3309.1.

Dutra, E., G. Balsamo, P. Viterbo, P. Miranda, A. Beljaars, C. Schär, and K. Elder, 2010: An improved snow scheme for the ECMWF land surface model: description and offline validation, J. Hydrometeor., **11**, 899–916, https://doi.org/10.1175/JHM-D-11-072.1.

Dutra, E., C. Schär, P. Viterbo, and P. M. A. Miranda, 2011: Land-atmosphere coupling associated with snow cover, Geophys. Res. Lett., **38**, L15707, https://doi.org/10.1029/2011GL048435.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

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

Emerton, R. E., and Coauthors, 2016: Continental and global scale flood forecasting systems., WIREs Water, 3, 391–418, https://doi.org/10.1002/wat2.1137.

Emerton R., H. Cloke, E. Stephens, E. Zsoter, S. Woolnough, F. Pappenberger, 2017: ‘Complex picture for likelihood of ENSO-driven flood hazard’. Nature Communications. https://doi.org/10.1038/NCOMMS14796.

Emerton R., E. Zsoter, L. Arnal, H. Cloke, D. Muraro, C. Prudhomme, E. Stephens, P. Salamon, and F. Pappenberger, 2018: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, Geosci. Model Dev., 11, 3327-3346, https://doi.org/10.5194/gmd-11-3327-2018, 2018.

Haddeland, I., and Coauthors, 2011: Multimodel estimate of the global terrestrial water balance: Setup and first results., J. Hydrometeor., 12, 869–884, https://doi.org/10.1175/2011JHM1324.1.

Helfrich SR., D. McNamara, B. Ramsay, T. Baldwin, and T. Kasheta, 2007: Enhancements to, and forthcoming developments in the interactive multisensor snow and ice mapping system, (IMS). Hydrol Process 21:1576–1586. doi:10.1002/hyp.6720

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Hersbach H., and D. Dee, 2016: ERA5 reanalysis is in production, ECMWF Newsletter No. 147, pp. 7.

Hossain, F., A. H. M. Siddique-E-Akbor, S. Biancamaria, H. Lee, and C. K. Shum, 2014: Proof of Concept of an Altimeter-Based River Forecasting System for Transboundary Flow Inside Bangladesh, IEEE J. Sel. Topics Appl. Earth Observ., vol. 7, no. 2, pp. 587-601, https://doi.org/10.1109/JSTARS.2013.2283402

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the Global Precipitation Record: GPCP Version 2.1., Geophys. Res. Lett., **36**, https://doi.org/10.1029/2009GL040000.

Kauffeldt, A., S. Halldin, F. Pappenberger, F. Wetterhall, C.-Y. Xu, and H. L. Cloke, 2015: Imbalanced land surface water budgets in a numerical weather prediction system, Geophys. Res. Lett., 42, 4411–4417, https://doi.org/10.1002/2015GL064230.

De Lannoy, G. J. M., R. H. Reichle, K. R. Arsenault, P. R. Houser, S. Kumar, N. E. C. Verhoest, and V. R. N. Pauwels, 2012: Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, Water Resources Research, 48(1), https://doi.org/10.1029/2011WR010588.

Helmert J., A. Sensoy, R. Alvarado Montero, C. de Michele, P. de Rosnay, M. Dumont, D. C. Finger, M. Lange, G. Picard, V. Potopova, S. Pullen, D. V. Schuler, and A. Nadir Arslan, 2018: Review of Snow Data Assimilation Methods for Hydrological, Land Surface,

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Meteorological and Climate Models: Results from a COST HarmoSnow Survey, Geosciences, 8(12), 489, https://doi.org/10.3390/geosciences8120489.

Legates, D. R., and G. J. McCabe Jr., 1999: Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35(1), 233–241, https://doi.org/10.1029/1998WR900018.

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.

Mengelkamp, H.-T., K. Warrach, C. Ruhe, and E. Raschke, 2001: Simulation of runoff and streamflow on local and regional scales. Meteor. Atmos. Phys., **76**: 107–117.

Molotch, N. P. and R. C. Bales, 2005: Scaling snow observations from the point to the grid element: Implications for observation network design, Water Resour. Res., 41, W11421, https://doi.org/10.1029/2005WR004229.

Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology. **10** (3), 282–290, https://doi.org/10.1016/0022-1694(70)90255-6.

Overgaard, J., D. Rosbjerg, and M. B. Butts, 2006: Land-surface modelling in hydrological perspective—A review, Biogeosciences, **3**(2), 229–241, https://doi.org/10.5194/bg-3-229-2006.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Pan, M., and E. Wood, 2006: Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter, Journal of Hydrometeorology, 7(3), 534-547, doi:10.1175/JHM495.1.

Pearson, K., 1896: Mathematical contributions to the theory of evolution: III. Regression, heredity and panmixia. Philosophical Transactions of the Royal Society of London Series A, 187, 253–318.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four dimensional variational assimilation. Part I: Experimental results with simplified physics, Q. J. R. Meteorol. Soc., 126, 1143–1170, https://doi.org/10.1002/qj.49712656415.

Rabier, F., 2005: Overview of global data assimilation developments in numerical weather prediction centres. Q. J. R. Meteorol. Soc., 131, 3215 – 3233, https://doi.org/10.1256/qj.05.129.

de Rosnay P., M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel, and L. Isaksen, 2013: A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF, Q. J. R. Meteorol. Soc., 139, 1199-1213, https://doi.org/10.1002/qj.2023.

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), pp 607-621,

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

de Rosnay P., L. Isaksen, M. Dahoui, 2015: Snow data assimilation at ECMWF, ECMWF Newsletter no 143, pp 26-31.

Slater, A. G., and M. P. Clark, 2006: Snow data assimilation via an ensemble Kalman filter, Journal of Hydrometeorology, 7(3), 478-493, https://doi.org/10.1175/JHM505.1.

Stephens, E., J. J. Day, F. Pappenberger, and H. Cloke, 2015: Precipitation and floodiness, Geophys. Res. Lett., 42, 10,316–10,323, https://doi.org/10.1002/2015GL066779.

van der Knijff, J. M., J. Younis, and A. P. J. de Roo, 2010: LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation, Int. J. Geogr. Inf. Sci., 24, 189–212, https://doi.org/10.1080/13658810802549154.

Le Vine, N., A. Butler, N. McIntyre, and C. Jackson, 2016: Diagnosing hydrological limitations of a land surface model: application of JULES to a deep-groundwater chalk basin, Hydrol. Earth Syst. Sci., 20, 143-159, https://doi.org/10.5194/hess-20-143-2016.

Wang, L. L, D. H. Chen, H. J. Bao, 2016: The improved Noah land surface model based on storage capacity curve and Muskingum method and application in GRAPES model. Atmospheric Science Letters, 17, 190–198. https://doi.org/10.1002/asl.642.

Wu, H., R. F. Adler, Y. Tian, G. J. Huffman, H. Li, and J. Wang, 2014: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model, Water Resour. Res., 50, 2693–2717, https://doi.org/10.1002/2013WR014710.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Yamazaki, D., S. Kanae, H. Kim, and T. Oki, 2011: A physically based description of floodplain inundation dynamics in a global river routing model. Water Resour. Res., 47, W04501, https://doi.org/10.1029/2010WR009726.

Zaitchik, B.F. and M. Rodell, 2009: Forward-looking assimilation of MODIS-derived snow-covered area into a land surface model. J. Hydrometeor., 10, 130–148, https://doi.org/10.1175/2008JHM1042.1.

Zhao, F., T. I. Veldkamp, K. Frieler, J. Schewe, S. Ostberg, S. Willner, B. Schauberger, S. N. Gosling, H. M. Schmied, F. T. Portmann, and G. Leng, 2017: The critical role of the routing scheme in simulating peak river discharge in global hydrological models, Environ. Res. Lett., 12, 075003, https://doi.org/10.1088/1748-9326/aa7250.

Zsoter, E., F. Pappenberger, P. Smith, R. E. Emerton, E. Dutra, E. Wetterhall, D. Richardson, K. Bogner and G. Balsamo, 2016: Building a multi-model flood prediction system with the TIGGE archive. J. Hydrometeor., http://dx.doi.org/10.1175/JHM-D-15-0130.1.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Longitude band	30W - 60E	60-150E	150-180E	120-180W	30-120W
~ local noon	12	6	00	00	18

Table 1. Criteria for selecting daytime 2-metre temperature

Score	**Description**	**Used for**
ME	Mean error	Daily river discharge, snow depth and 2-metre temperature
MAE	Mean absolute error	Daily river discharge, snow depth and 2-metre temperature
NSE	Nash-Sutcliffe efficiency	Daily river discharge time series
R	Pearson correlation coefficient	Daily river discharge time series
PMnE	Percentage sample mean error	Whole river discharge sample
PMnAe	Percentage sample mean absolute error	Whole river discharge sample
PStE	Percentage sample standard deviation error	Whole river discharge sample
PStAe	Percentage sample standard deviation absolute error	Whole river discharge sample
PkTiMe	Peak timing mean error	Annual river discharge peaks

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

PkTiMae	Peak timing mean absolute error	Annual river discharge peaks
PPkMgMe	Percentage peak magnitude mean error	Annual river discharge peaks
PPkMgMae	Percentage peak magnitude mean absolute error	Annual river discharge peaks

Table 2. List of verification scores used in the analysis with a short description and also the areas where they were applied.

No.	Station	River	Area (*1000 km2)	NSE		PPkMgMe (%)		PPkMgMae (%)	
				ONLINE	OFFLINE	ONLINE	OFFLINE	ONLINE	OFFLINE
1.	Salekhard	Ob	2541	0.40	**0.52**	-55.0	**-40.7**	55.0	**40.7**
2.	Pilot station	Yukon	865	0.31	**0.64**	-64.7	**-50.7**	64.7	**50.7**
3.	Boogojevo	Danube	257	**0.47**	-0.43	**-3.5**	29.1	**19.8**	32.4
4.	Lobith	Rhine	163	**0.45**	0.05	-39.1	**-14.8**	39.1	**18.5**
5.	Viicksburg	Mississippi	2963	**-0.02**	-2.69	**1.6**	31.4	**17.7**	43.5
6.	Quincy	Columbia	663	0.25	**0.54**	-24.0	**-7.6**	27.5	**20.2**
7.	Boa Sorte	Xingu	207	-1.53	**-0.85**	159.0	**147.9**	159.0	**147.9**
8.	Obidos-Linigrafo	Amazon	4664	**-0.17**	-0.21	**26.6**	26.9	**26.6**	26.9
9.	Hadejia	Hadejia	22	**-9.01**	-11.85	**297.1**	436.1	**297.1**	436.1
10.	Bangui	Ubangi	496	**-5.72**	-6.17	162.8	**159.1**	162.8	**159.1**
11.	Katima Mulilo	Zambesi	331	-7.97	**-6.70**	196.6	**183.0**	196.6	**183.0**

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

12.	Walkers bend	Flinders	106	**0.66**	0.62	-24.5	**-11.4**	46.9	**45.9**
13.	Komsomolsk	Amur	1846	0.43	**0.68**	-33.5	**-18.7**	33.5	**18.7**

Table 3. Details of the 13 catchments analysed in Figure 7 (no. 13) and Figure 8 (no. 1-12) with the NSE, PPkMgMe (percentage peak magnitude ME) and PPkMgMae (percentage peak magnitude MAE) score values for the ONLINE and OFFLINE experiments based on the ERA5-D25 dataset. Bold scores denote better performance. For further details on the scores see Section 2.13.

Score	ME (m3/s)	MAE (m3/s)	NSE	R	PMnE (%)	PMnAe (%)	PStE (%)	PStAe (%)	PkTiMe (day)	PkTiMae (day)	PPkMgMe (%)	PPkMgMae (%)
ONLINE	-264	3017	**-0.29**	0.67	**-2.6**	29.0	**9.6**	**48.3**	-0.95	11.8	**6.3**	61.3
OFFLINE	**236**	**2954**	-0.53	**0.70**	16.9	**27.2**	34.2	52.1	**-0.81**	11.8	27.3	**59.2**

Table 4. List of global average scores for the ONLINE and OFFLINE experiments based on the ERA5-D25 dataset. Each value is a mean of scores from 590 catchments (where minimum of 9 years of river discharge observations was available) weighted by the square root of the catchment area sizes. For further details on the scores see Section 2.13. Bold numbers denote the better score of ONLINE and OFFLINE. The following scores are displayed: ME, MAE, NSE, R, Percentage sample mean error (PMnE), Percentage sample mean absolute error (PMnAe), Percentage sample standard deviation error (PStE), Percentage sample standard deviation absolute error (PStAe), Peak timing ME (PkTiMe), Peak timing MAE (PkTiMae), Percentage peak magnitude ME (PPkMgMe), Percentage peak magnitude MAE (PPkMgMae).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

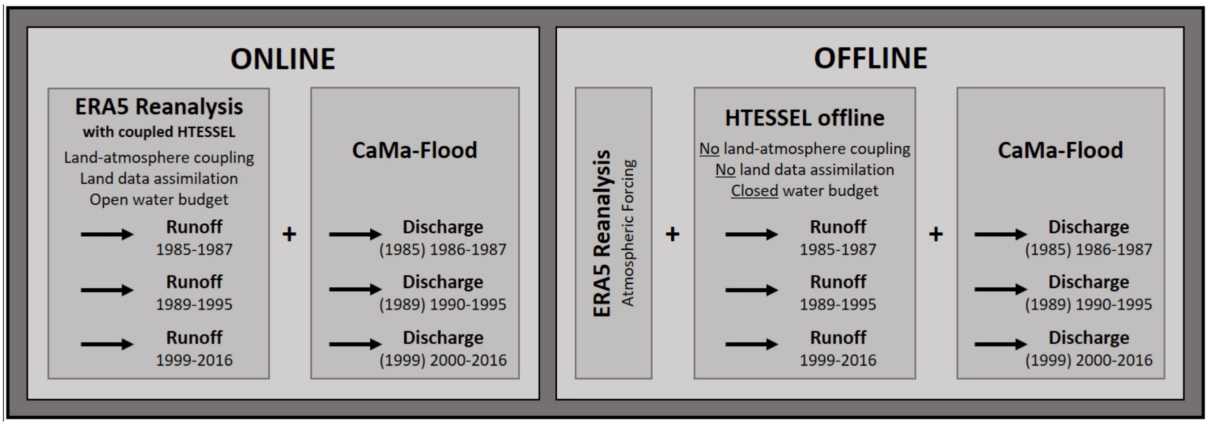


Figure 1. Schematic of the ONLINE and OFFLINE experiments that were carried out to produce the ERA5-D25 dataset. The years in brackets for the discharge indicate the first spin-up year in each period that were excluded from the analysis.

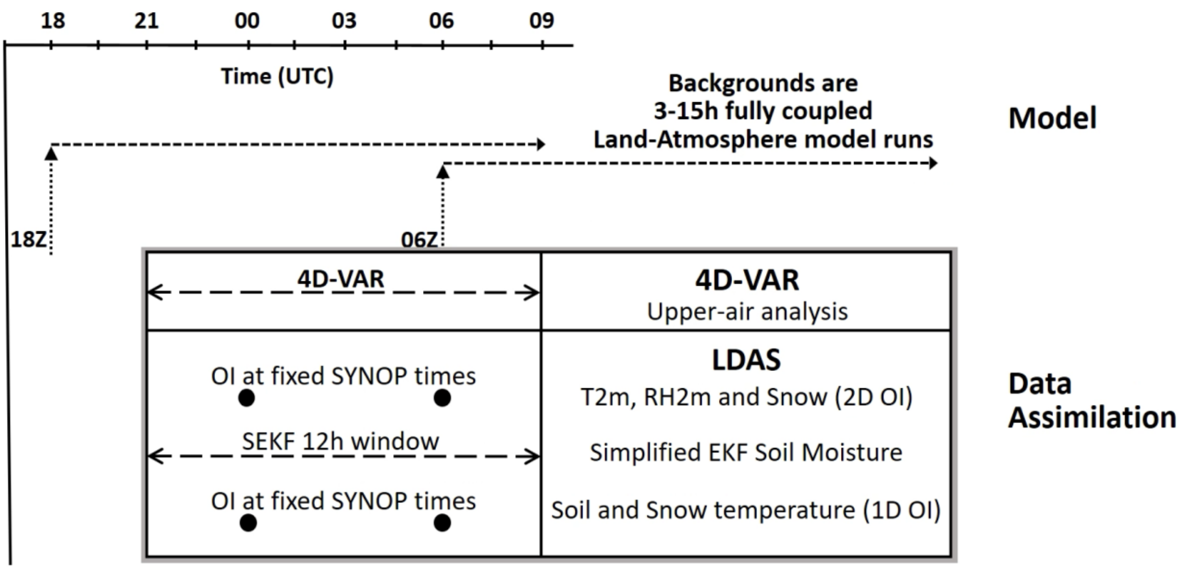


Figure 2. Schematic diagram of the land data assimilation system at ECMWF.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

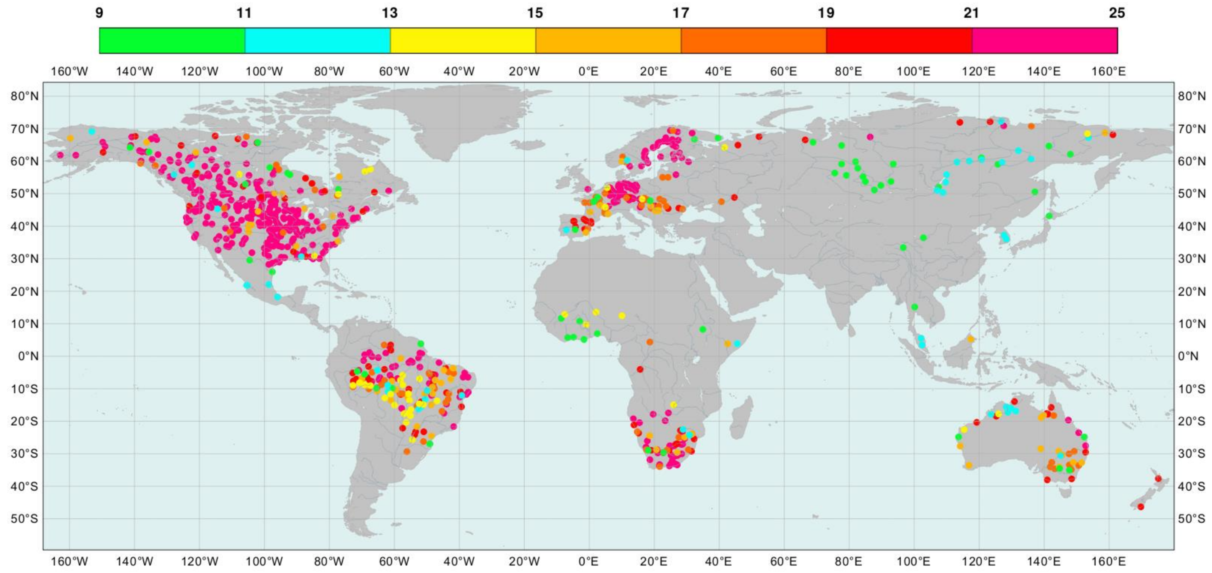


Figure 3. Geographical distribution of river discharge observations with sufficient record length selected for the analysis. Colours indicate the length of the available data in years (from 9 to 25).

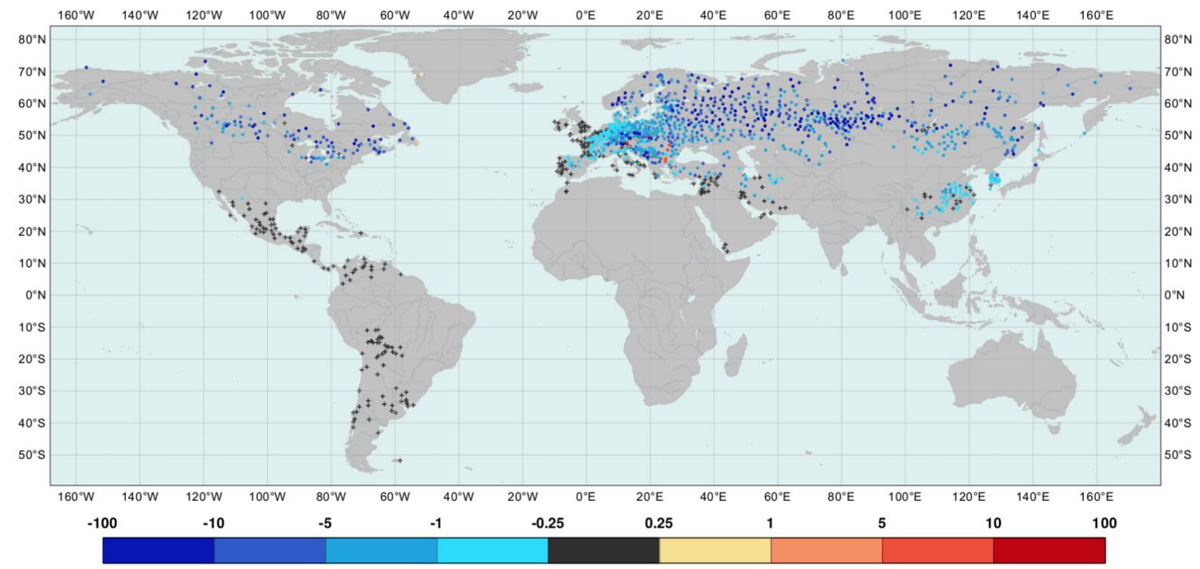


Figure 4. Difference in the snow depth mean absolute errors between ONLINE and OFFLINE for January based on observations in 2000-2007 (in cm). Points are shown where observations are available. Blue colours indicate lower errors in the ONLINE experiment.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

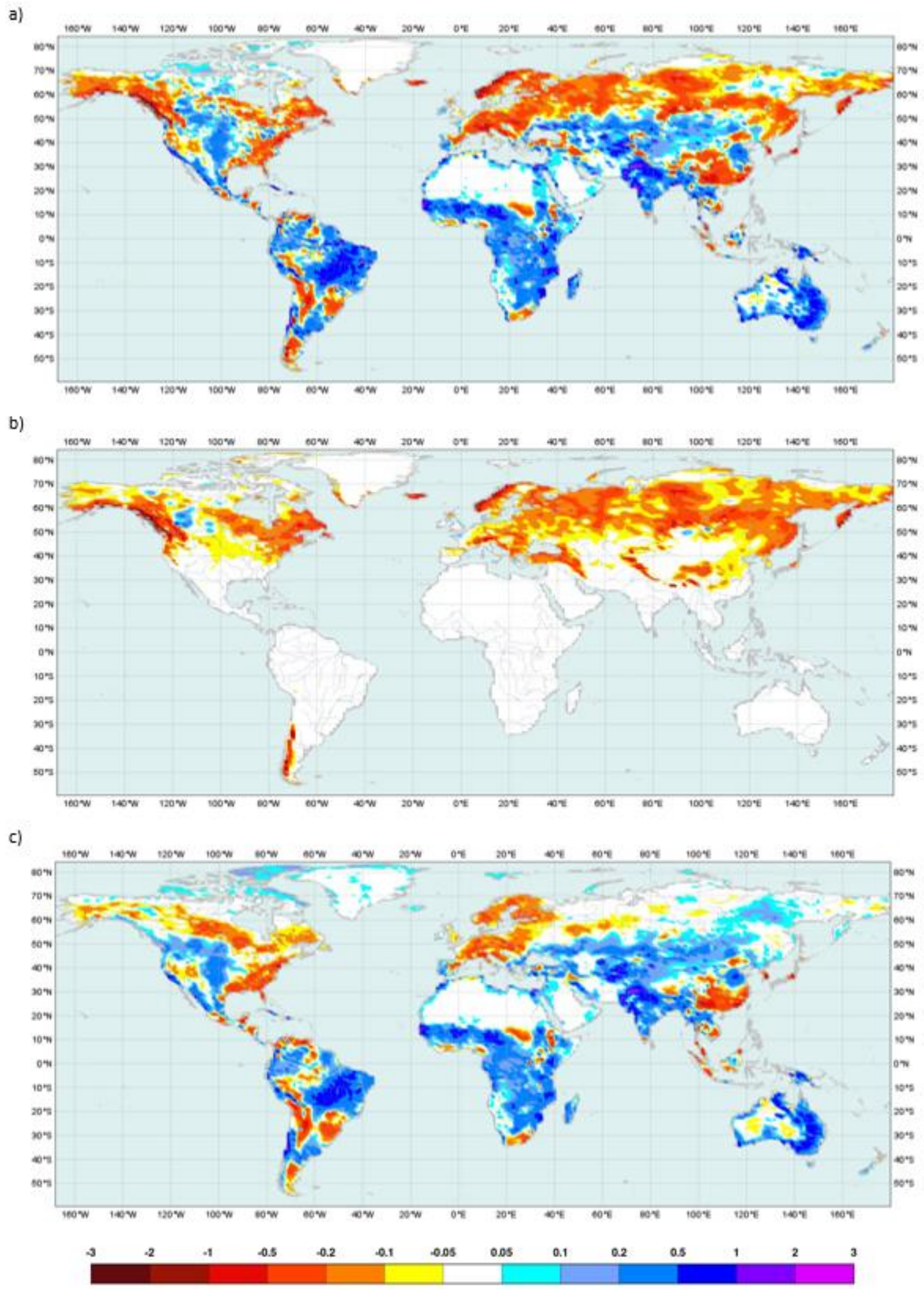


Figure 5. Average daily water budget analysis (mm/day) of the ONLINE experiment based on the ERA5-D25 dataset for (a) the total 24-hour water budget errors, (b) the 24-hour snow water equivalent increments and (c) the 24-hour soil water content increments. Negative values (red) indicate water removal by LDAS, while positive values (blue) show where water is added to the hydrological cycle.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

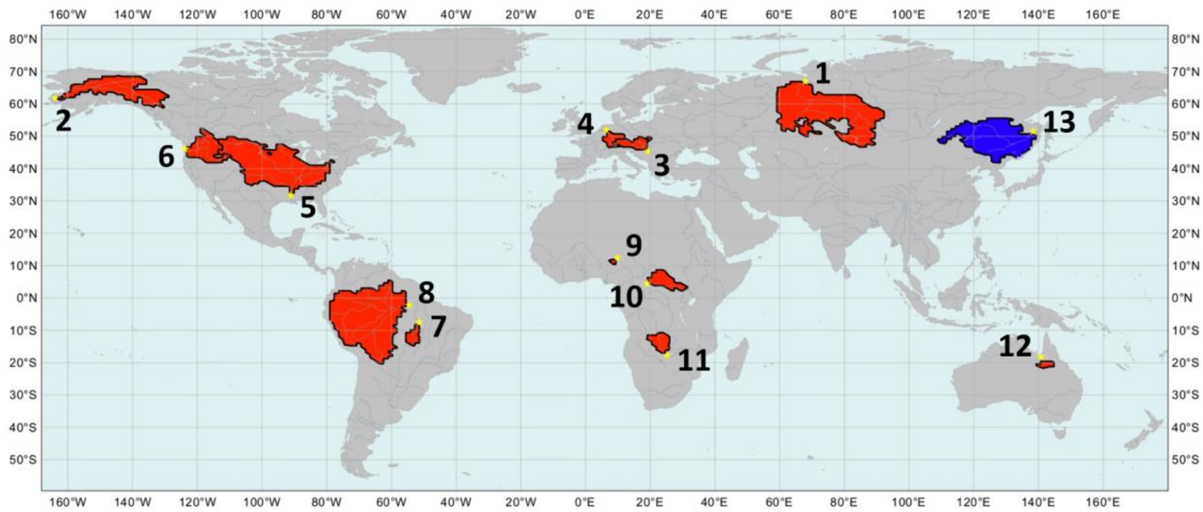
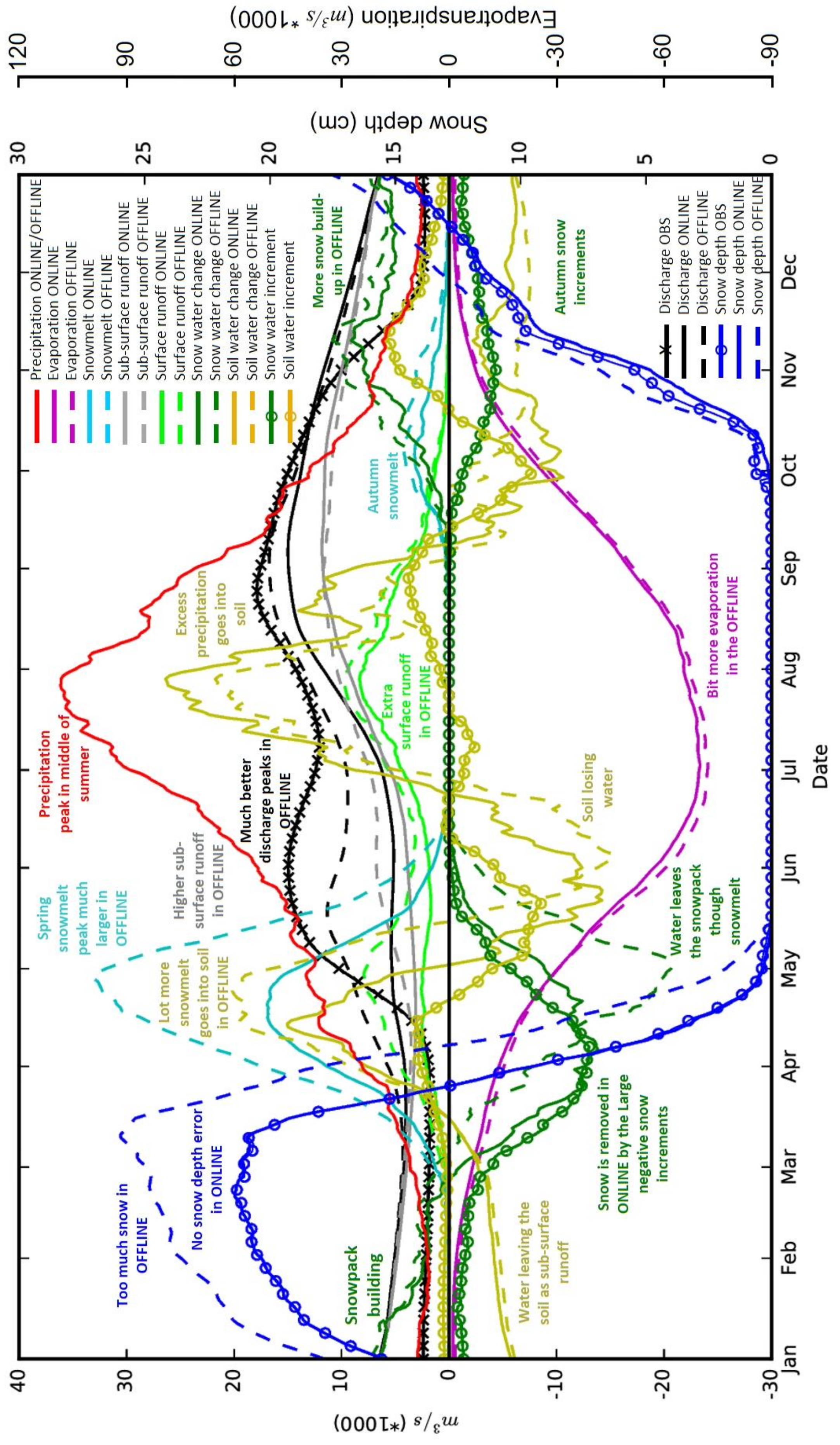


Figure 6. Map of the catchments analysed in Section 3.3 (Figure 7), where the catchment-level process is examined over the Amur river (blue area, no. 13), and in Section 3.4 (Figure 8), where the simplified representation of the annual water cycle is shown for some selected regional catchments of the world (red areas, no. 1-12). The catchment details are provided in Table 3.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.



Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

Figure 7. Average daily water budget cycle for a catchment on the Amur river in Russia at Komsomolsk. It includes the following parameters: precipitation (red line), snow (green line with markers) and soil (mustard line with markers) water content increments for the ONLINE simulation; surface runoff (light green), subsurface runoff (grey), evapotranspiration (magenta), snowmelt (cyan) and soil (mustard) and snow (green) water storage daily changes for both ONLINE (solid lines) and OFFLINE (dashed lines); snow depth (blue) and also river discharge (black) for the ONLINE (solid lines) and OFFLINE (dashed lines) experiments and also observations (lines with markers). The snow depth values are based on 2000-2007 while all other displayed daily climatological means are based on the ERA5-D25 dataset (for more detail on the computation of these values see Section 2.11 and 2.12).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

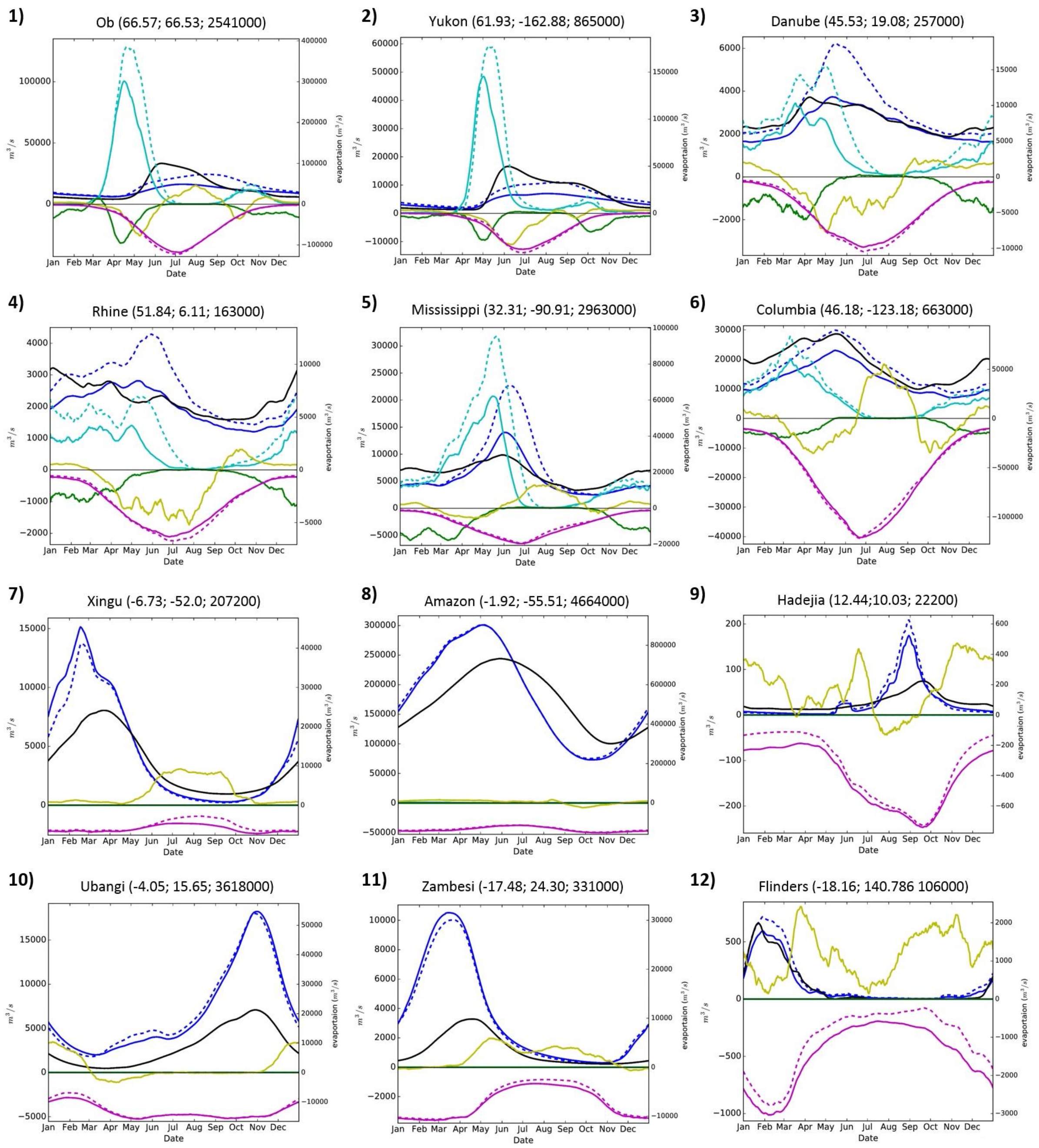


Figure 8. The annual cycle of water budget variables for a selection of catchments worldwide numbered from 1 to 12 (see Figure 6). The displayed variables are the snowmelt (cyan), evapotranspiration (magenta) and river discharge (blue) for both the ONLINE (solid lines) and OFFLINE (dashed lines) experiments, the snow (green) and soil (mustard) increments for ONLINE and the river discharge observations (black line). All values are daily climatological averages based on the ERA5-D25 dataset (for details on the computation of these values see Section 2.12). The river names, the gauge coordinates and the upstream area values are

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

displayed in the subplot titles. The catchment descriptions with the main verification score values for the ONLINE and OFFLINE simulations are provided in Table 3. In addition, the catchment area contours are provided in Figure 6. The evapotranspiration scale is provided on the secondary vertical axis while the scale for all other parameters is shown on the main vertical axis.

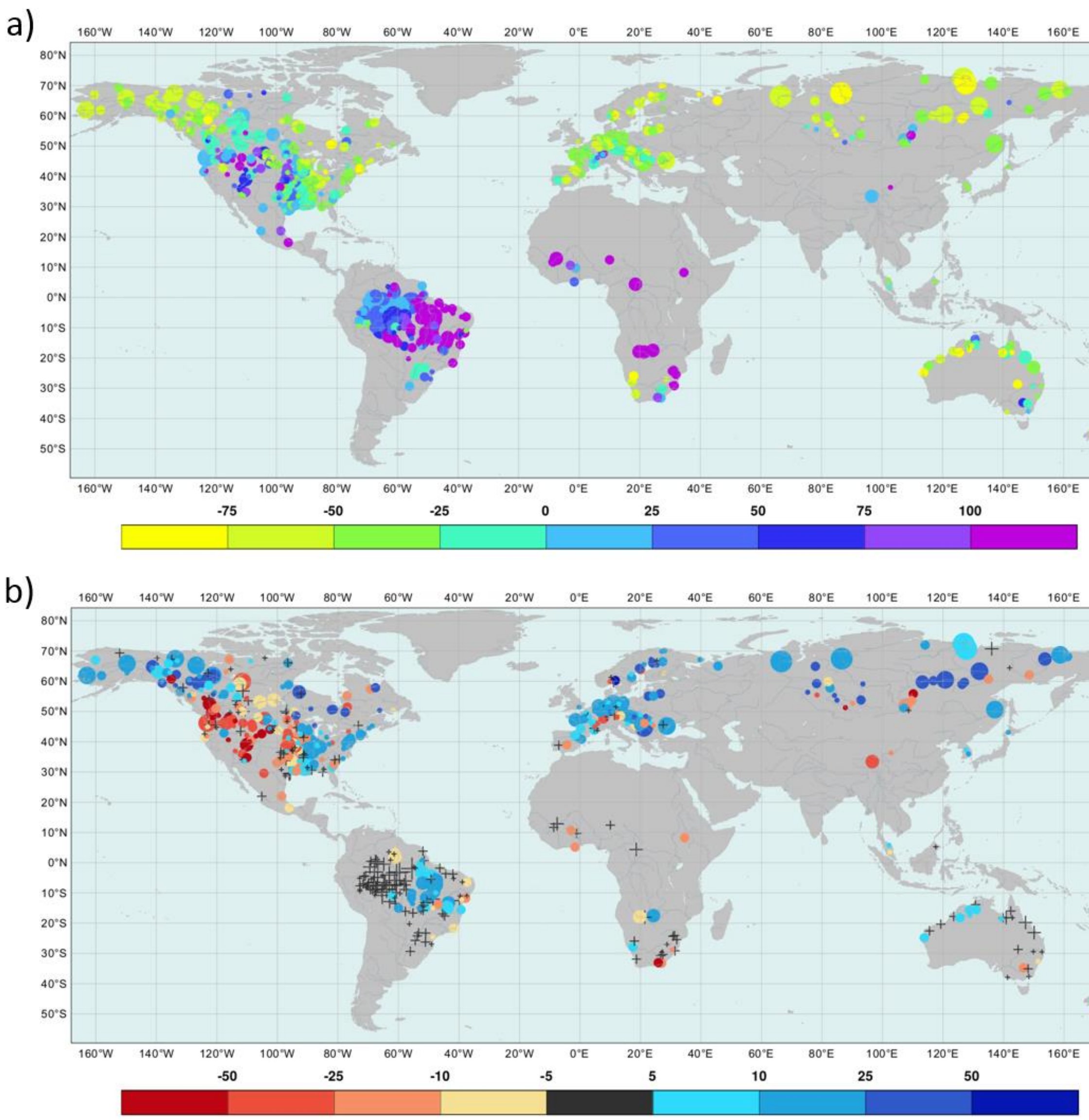


Figure 9. River discharge percentage peak magnitude (a) ME (in %) of the ONLINE experiment and (b) change in the percentage peak magnitude MAE (in %) between ONLINE and OFFLINE based on the ERA5-D25 dataset. Positive error differences in b) indicate

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.

deterioration (blue) while negative changes show improvement (red) in the ONLINE simulation compared with OFFLINE. The catchments are displayed with different marker sizes representing the size of the catchment area. Near zero differences are shown by black crosses, while all other categories are displayed by circles.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-18-0086.1.