



LAND MODEL EO DATA INTEGRATION

Deliverable 1.2

**Authors: P.C. McGuire, E. Dutra, A. Hurduc, and
T. Quaife**

Document information

Grant agreement	
Project title	EXPECT: Towards an Integrated Capability to Explain and Predict Regional Climate Changes
Project acronym	EXPECT
Project start date	1st April 2024
Related work package	WP1
Related task(s)	1.2
Lead organisation	University of Reading
Authors	Patrick C. McGuire, Emanuel Dutra, Alexandra Hurduc, and Tristan Quaife
Submission date	31st March 2026
Dissemination level	

History

Date	Submitted by	Reviewed by	Notes
25/2/2026	Tristan Quaife		Initial version for internal review
13/3/2026	Marlene Kretschmer		Internally reviewed
20/3/2026	Stefano Materia		Internally reviewed
23/3/2026	Patrick C. McGuire		Updated version after internal review

Please cite this report as: P.C. McGuire, E. Dutra, A. Hurduc, and T. Quaife (2026), Land model EO data integration, D1.2 of the EXPECT project.

Disclaimer: Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Table of contents

1 About	5
2 Executive summary	6
3 Acronyms and abbreviations	7
4 Introduction	9
5 JULES	11
5.1 JULES forcing data	12
5.2 JULES ancillary files	12
6 Data Assimilation	14
6.1 Data assimilation within the Rose/Cylc suite	16
7 Other Methods	16
7.1 Automated plot generation	16
7.2 Calculation of the analysis increment and the RMSE reduction	17
7.3 Calculation of Weeks of Max Wetting and Max Drying	18
7.4 Calculation of SSI hydrological drought index	19
8 Results	19
8.1 Change of soil PTF parameters after data assimilation	19
8.2 Reduction of soil moisture RMSE after data assimilation	20
8.3 Re-estimation of soil moisture after data assimilation	24
8.4 Re-estimation of SSI hydrological drought index	25
8.5 Re-estimation of the Week of Maximum Wetting	26

8.6 Re-estimation of Drying during the Week of Maximum Drying	27
8.7 Re-estimation of the net primary production (NPP)	28
8.8 Re-estimation of the carbon use efficiency (CUE)	29
8.9 Re-estimation of the water use efficiency (WUE)	31
8.10 Comparing analysis increments of WUE, CUE, GPP, NPP, latent heat flux, and soil moisture	32

9 Land Surface Temperature evaluation 34

10 Conclusions 38

10.1 Conclusions from evaluation with LST	39
10.2 Links to other parts of EXPECT	39

11 References 39

List of figures

Figure 1: Data assimilation gives posteriori estimates of Tóth Pedotransfer-Function (PTF) soil parameters.	20
Figure 2: The Root Mean Square Error (RMSE) of soil moisture (SM) for 2003 (units = None, plot limits of ± 0.07), before (a) and after (b) DA.	22
Figure 3: For the years 1980-2019, the European average ($\pm$ the spatial standard deviation) of the reduction of soil moisture RMSE after the data assimilation is plotted.	23
Figure 4: The ensemble mean of soil moisture (SM) for 2003 (units = None, plot limits of 0.0-0.40), before (a) and after (b) DA.	24
Figure 5: The annual standardized soil-moisture index (SSI) for the Central Europe drought year 2003 (units = standard deviations), before (a) and after (b) DA. In (c), the absolute analysis increment of SSI is shown for 2003. In (d), the absolute hindcast increment of SSI is shown for 2003.	25
Figure 6: For the year 2003: (a-b) "Week of max wetting" (WMW),	26
Figure 7: For the year 2003, which is a drought year in Central Europe: (a-b) "Drying during week of max drying" (DWMD),	28
Figure 8: Analysis increments (a) and Hindcast increments (b) of annual net primary production (NPP) for 1999-2003 (units = %).	29
Figure 9: The annual carbon use efficiency (CUE) for 2003 (units = None), before (a) and after (b) DA.	30
Figure 10: The global spatial pattern of average NPP/GPP (carbon use efficiency, CUE) from a MODIS analysis of 2000-2003 data. Reprinted from Zhang et al. (2009).	31
Figure 11: The annual water use efficiency (WUE) for 2003 (units = $\text{gC}/\text{mmH}_2\text{O}$), before (a) and after (b) DA, with plot limits of 0-1.5.	32

Figure 12: The ensemble-mean maps of the analysis increment in 2003 for a) GPP, b) Soil Moisture, c) NPP, d) Latent Heat Flux (can be used as evapotranspiration), e) WUE, and f) CUE. _____	33
Figure 13: Mean differences of land surface temperature daily maximum (lst_max, left panels), daily mean (lst, middle panels) and daily minimum (right panels) _____	35
Figure 14: Standard deviation of monthly anomaly time series (removing the mean annual cycle, 2004-2019) of land surface temperature diurnal amplitude (left panel), _____	36
Figure 15: Top panels: Difference of standard deviation of monthly anomaly time series _____	37

List of tables

Table 1 Table of the 15 parameters that we use for data assimilation with the Tóth et al. (2015) pedotransfer function (PTF).	14
---	----

1 About

Water held in soil can be returned to the atmosphere via two main processes – evaporation from the soil surface, and transpiration via plants which can extract moisture at lower soil levels. Combined, this flux is referred to as evapotranspiration and is the dominant component of the latent heat flux from the surface. The magnitude and timing of this flux can influence the evolution of the state of the atmosphere in such a way that it exacerbates extremes. Dry soils, in particular, are known to modulate heat waves.

This report details the results of experiments assimilating satellite measurements of soil moisture into the land surface model JULES. We used a novel, ensemble-based data assimilation methodology to tune the soil hydraulics in JULES so that the model predictions of soil moisture more closely match those of the European Space Agency Climate Change Initiative (ESA CCI) soil moisture data set. We test the model predictions outside of the assimilation window to show that the improvement in model skill persists in time.

We show that after data assimilation (DA) of the ESA CCI soil moisture within the data-assimilation window of 2014-2018, the soil moisture modelled by JULES when averaged across Europe for example for the year 2003 is about 25% lower than before the data assimilation. These modelled drier soils also match the ESA CCI soil-moisture observations better after assimilation than before assimilation for all modelled years 1980-2019. After the data assimilation, the ensemble averages of most (12) of the 15 soil-pedotransfer parameters are shifted significantly, to values outside the assumed distributions of the prior soil-pedotransfer parameters. The data assimilation of soil moisture across Europe certainly affects other aspects of the JULES model, including reducing gross primary productivity (GPP) by about 20% on average and reducing the

evapotranspiration by about 6% on average (both numbers are for the selected year of 2003 and for continental Western Europe excluding the Nordic countries and Iberia). Furthermore, the data assimilation reduces some of the extreme values of the weekly change in soil moisture to more modest values, with the effect of changing drought-related statistics like the week of maximum drying and flood-related statistics like the week of maximum wetting. Overall, this technique of data assimilation of soil moisture into the JULES land model via the soil-pedotransfer parameters appears to adjust both the soil hydrology and the vegetation growth rates towards more reasonable behaviour.

The comparison of the model simulations with satellite land surface temperature (LST) and reanalysis data (ERA5) using monthly mean diurnal cycles, showed that ERA5 underestimates the diurnal amplitude (cold daytime / warm nighttime bias), while the JULES model before DA ("JULES PRIOR") exhibits region-dependent biases; the JULES model after DA ("JULES POST") warms the daily minimum, improving Mediterranean performance but degrading central/northern Europe. The interannual variability analysis showed the strongest variability in the diurnal maximum and the weakest in the diurnal amplitude; ERA5 underestimates amplitude variability, whereas JULES PRIOR aligns more closely with satellite LST. The temporal correlations with satellite LST are similar for ERA5 and JULES PRIOR, with slightly lower agreement for JULES PRIOR in southern regions affected by data assimilation impacts on nighttime temperatures.

2 Executive summary

Water held in soil plays an important role in the exchange of energy between the atmosphere and land, and it is an important feedback mechanism for both weather and climate. Heatwaves can be exacerbated by dry soils and, as a consequence, being able to more accurately model the amount of water in soil is important for correctly modelling the progress of extreme heat. This, in turn, is critical for making well founded predictions about the impact of heat on human health by making better forecasts of heatwave intensity and duration.

A key issue, which we address here, is the representation of soil hydraulics in the land component of weather and climate models. The hydraulics determine, for example, how much water can be held and how quickly it drains away. However, the data used to prescribe these values is typically based on sparse and uncertain observations. Here we use an alternative approach, which is to tune the hydraulics of a land model so that the modelled soil moisture better matches observations from the European Space Agency Climate Change Initiative.

We show that our approach improves the model fit to the satellite data (1979-2019) and that, for a selected drought year, the resulting model predicts much drier soils more closely matching observations, than when using its original parameters. We also show likely impacts on vegetation, with a marked reduction in carbon uptake for large parts of Europe.

3 Acronyms and abbreviations

Acronym	Definition
4DVAR	4-Dimensional VARIational data assimilation
AI	analysis increment
CEC	cation exchange capacity
CINEA	European Climate, Infrastructure and Environment Executive Agency
COSMOS	COsmic-ray Soil Moisture Observing System
CRU-NCEP	Climatic Research Unit - National Centers for Environmental Prediction
CUE	carbon use efficiency
DA	data assimilation
dSMCL	weekly change in soil-moisture
DWMD	Drying during the Week of Maximum Drying
ECMWF	European Centre for Medium Range Forecasting
EO	Earth observation
ERA5	ECMWF ReAnalysis5
ESA CCI	European Space Agency Climate Change Initiative
ET	evapotranspiration
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
EXPECT	(project name) towards an integrated capability to EXplain and PrEdiCt regional climate changes
GL	Global Land
GPP	gross primary production
HI	hindcast increment
IPMA	Instituto Português do Mar e da Atmosfera
JASMIN	The UK's data analysis facility for environmental science
JULES	Joint UK Land Environment Simulator
LHF	latent heat flux

LST	Land Surface Temperature
LSA SAF	Satellite Applications Facility on Land Surface Analysis
LSA	short name for LSA SAF LST
MODIS	MODerate-resolution Imaging Spectroradiometer
MOSRS	Met Office Science Repository Service
MSG	Meteosat Second Generation
NPP	net primary production
NRT	near real-time
PNG	Portable Network Graphics filename extension
POST	posterior
PRIOR	prior
PTF	pedotransfer function
RA	autotrophic respiration
RMS	Root Mean Square
RMSE	Root Mean Square Error
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SM	soil moisture
SMAP	Soil Moisture Active Passive mission
SSI	standardized soil-moisture index
TXT	Text filename suffix
UKESM	UK Earth System Model
WFDE5	Watch Forcing Data ECMWF ReAnalysis5
WFDEI	Watch Forcing Data ECMWF ReAnalysis Interim
WMD	Week of Maximum Drying
WMW	Week of Maximum Wetting
WUE	water use efficiency
WP	work package

4 Introduction

Soil moisture is a key environmental variable needed for correctly diagnosing land-atmosphere exchanges of both mass and energy. As such, correctly representing soil moisture in the land surface schemes of Numerical Weather Prediction models and Climate models is important for making well-founded predictions of the state of the atmosphere and the weather we experience. Water arrives in the soil column via precipitation or lateral transport (e.g., rivers or run-off from adjacent land) and then is lost either by drainage, further lateral transport, or evapotranspiration. Evapotranspiration is the dominant component of the latent heat flux and, along with the sensible heat flux, represents the turbulent exchange of energy from the land to the atmosphere. The balance between these two fluxes is directly controlled by the amount of water in the soil available to be evaporated or transpired by plants. When soils are dry, turbulent exchange with the atmosphere will be primarily in the form of sensible heat and result in direct warming of the air. In contrast, when soils are wet, turbulent exchange with the atmosphere will be primarily in the form of latent heat fluxes, which can ultimately affect cloud formation through the water vapor transport.

The interaction between soil moisture and heatwaves has been well documented. A comprehensive review is provided by Seneviratne *et al.* (2011) and examples from the past few years include studies by Gurung and Chen (2024), Chagnaud *et al.* (2025) and Tripathy *et al.* (2025). A recent study by Maraun *et al.* (2025), using an ensemble of climate models, suggests a future increase in strong coupling between soil moisture and near-surface air temperature in areas where evapotranspiration is currently energy limited (i.e. the latent heat flux to the atmosphere is controlled by the amount of energy available at the surface, as opposed to the amount of soil moisture). Their results highlight several regions across the globe, including large parts of Europe, where extreme heatwaves are likely to become more common because of this mechanism. This points to the need to more accurate modelling of soil moisture, especially for those regions likely to transition away from being in an energy limited state.

In land models the vertical transport of water in the soil column is controlled by hydrological parameters such as the saturation point and hydraulic conductivity. Further details on the parameters used in the model in this project are provided in the report below. Typically, these are defined on the basis of knowledge about soil texture, using spatially coarse data sets that are usually extrapolated on the basis of a relatively small number of observations. Consequently, these parameters are a key source of uncertainty in determining the behaviour of water in the soil column; they control how much water can be held, how easily it percolates to deeper depths, and so on. To

reduce this uncertainty, one option is to assimilate relevant observations. In the case of this project, we have assimilated satellite-derived estimates of soil moisture and assimilated them into the JULES model to tune the soil hydraulic parameters. In some senses this is more akin to model calibration rather than traditional state-based data assimilation (which in this case, would involve changing the instantaneous value of the modelled soil moisture). However, the mathematical basis is identical to that of state assimilation and, because soils properties evolve slowly in time, it means that the additional skill provided by the assimilation also persists in time (see results, Figure 3). By contrast, updates to the soil-moisture model *state*, tend only to persist for a few days in the upper layers, which is where typical satellite observations are able to provide information. For Numerical Weather Prediction applications, such short time scales are adequate, but for examining the impacts of seasonal and longer-term events it is better to remove persistent biases in the model – in this case those caused by inadequate hydrological parameterisation.

Although it is not the focus of the EXPECT project per se, changes in the soil moisture have consequences for the terrestrial carbon budget. Below a certain volumetric moisture content, often referred to as the critical point, evaporation shifts from becoming energy limited to water limited. In land models, including JULES, this concentration is also used as the point at which photosynthesis is down-regulated in response to the lack of water – in effect, simulating the closure of stomata to prevent water loss at the expense of reducing carbon uptake. At a lower moisture concentration, the wilting point defines the point where photosynthesis shuts down entirely. Consequently, improving the modelling of soil moisture in droughts and heatwaves will improve the representation of impact of these events on vegetation. Some of the analysis in this report addresses this, examining the impact on carbon uptake by the land surface, for example via Net Primary Productivity predicted by JULES.

Land Surface Temperature (LST) from satellite observations provides an independent, spatially consistent benchmark for evaluating Earth system models due to its direct link with the surface energy balance, which are simulated in the models. Unlike in-situ measurements, satellite LST offers almost global coverage, high temporal frequency (sub-daily in the case of retrievals from geostationary), and sensitivity to land-atmosphere interactions, with the potential to evaluate a model's systematic biases and also inter-annual variability issues (Fróis *et al.*, 2026; Johannsen *et al.*, 2019; Nogueira *et al.*, 2021).

In this work, the satellite observations of LST are obtained from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) sensor on board the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)'s Meteosat Second Generation (MSG) geostationary satellites, provided by EUMETSAT's Satellite Applications Facility on Land Surface Analysis (LSA SAF) (Trigo *et al.*, 2011). Land surface temperature is derived from top-of-atmosphere SEVIRI measurements using two bands in the thermal infrared region (Freitas *et al.*, 2010). SEVIRI land surface temperature retrievals are

available only in clear sky conditions. The retrievals are performed every 15 min with a 3 km spatial sampling at nadir, from 2004 to present.

In this report to the EXPECT project, we have methods sections on: (a) the JULES model (including the WFDE5 forcing data and the associated ancillary files); (b) the data assimilation technique; and (c) other methods used. This is followed by a results section, wherein we discuss how much the data assimilation technique affects the output variables including both hydrology-related variables and vegetation-growth-related variables. Before concluding, we have a second results section wherein we evaluate the land surface temperatures of the JULES model after data assimilation. In Section 3, we include a list of acronym definitions.

5 JULES

JULES, the Joint UK Land Environment Simulator (Best *et al.*, 2011; Clark *et al.*, 2011; Walters *et al.*, 2017; Wiltshire *et al.*, 2020), is the land surface component of the UK Met Office numerical weather prediction system and the UK climate models, including the Hadley Centre models and UKESM. Its primary purpose is to model the exchange of mass (in particular, carbon and water) and energy (radiation, heat and momentum) between the land and the atmosphere. However, it represents a large number of land surface processes and is increasingly being used to address other questions in the environmental sciences, in particular for climate impact studies. One such example is its contribution to the annual Global Carbon Project assessments (Friedlingstein *et al.*, 2025).

JULES can be run coupled with an atmospheric model (a so-called “online” simulation) or forced with meteorological data (“offline”). In Task 1.2 of EXPECT we have run JULES in offline mode.

We have used the UK Met Office Science Repository Service (MOSRS) for managing version control of the JULES code that we have used. At the UK MOSRS, our JULES code is archived as a ‘Rose/Cylc suite’ (Oliver *et al.*, 2019) with the suite identification u-dq237. Here, ‘Rose’ is a configuration and metadata tool that manages how jobs are configured and that stores model parameters in a structured way, and ‘Cylc’ is a workflow engine that controls when jobs are run on a supercomputer and that manages dependencies. The word ‘suite’ is primarily used to refer to a collection of related tasks (model runs, preprocessing, and postprocessing).

This suite, u-dq237, uses most of the same JULES parameter settings as the standard JULES Global Land GL7.0 suite (Walters *et al.*, 2019; Wiltshire *et al.*, 2020), which is identified as u-bx723. The exception is that for the entire modelled time period (1979-2019), we use constant 2015 values for atmospheric CO₂, which were approximately 400 ppm, instead of the measured CO₂ values which depend on the year. This was done in order to study more clearly the data assimilation without the confounding

effect of rising atmospheric CO₂ and its concomitant CO₂ fertilization of the vegetation.

This suite, u-dq237, has been further adapted in two ways in our work:

- a) In suite u-dx313, the data-assimilation window was extended from a single year to multiple years (2014-2018); and
- b) In suite u-du244, instead of using JULES to simulate all 50 ensemble members for 1980-2019, only the ensemble averages of both the prior and posterior soil parameters are used in the JULES simulations.

5.1 JULES forcing data

The forcing data used in the EXPECT project (in suite u-dq237) is Watch Forcing Data ECMWF ReAnalysis5 (WFDE5). This differs from standard ECMWF ReAnalysis5 (ERA5) forcing data mainly in its bias correction (Cucchi *et al.*, 2020). We do not use the first year of the data (1979) due to a meta-data problem, but we use the remainder of the currently available WFDE5 data (1980-2019). The WFDE5 data is hourly, unlike WFDEI data, which is only 3-hourly. The WFDE5 forcing data that we use here also differs from that used (CRU-NCEP-v7) (Viovy, 2018) in the standard GL7.0 suite (u-bx723). We used WFDE5 rather than CRU-NCEP-v7 as the forcing data mainly because of our familiarity with WFDEI forcing data.

5.2 JULES ancillary files

The WFDE5 version of the non-soil ancillary files or the prescribed data files (top model parameters, plant-functional-type vegetation fractions, land/sea mask, MODIS canopy heights, and MODIS leaf area indices) that are needed by JULES were specifically created for this EXPECT project using the ANTS software.

The WFDE5 version of the soil ancillary file that is needed by JULES was created for each of the 50 ensemble members using custom Python code, which is run each time the suite u-dq237 is run. As in Pinnington *et al.* (2021), we use the Tóth *et al.* (2015) pedotransfer functions (PTFs) to compute the physical soil parameters from the basic soil parameters provided by SoilGrids (Hengl *et al.*, 2017). The 15 unperturbed prior parameters of the Tóth PTF are taken from Tóth *et al.* (2015); see Table 1. Perturbed prior parameters are generated by sampling from a normal distribution centred on each unperturbed parameter, with a standard deviation equal to 10% of its value.

We do change one unperturbed prior parameter used by Pinnington *et al.* (2021): the 15th parameter, ϕ_0 , is changed from 0.10380 to 0.010380 (see Table 1). This revision matches the original Tóth value for the ϕ_0 parameter, and it allows for more reasonable estimates of the dependency of the saturated hydraulic conductivity (K_{sat}) upon the cation exchange capacity (CEC) via the equation: $\log_{10}(K_{sat}) = \phi_0 \text{ CEC} + \text{other terms}$. The need for this revision was noticed in preliminary JULES runs for the EXPECT

project by observing unusual values of K_{sat} in the western Russia region of our European domain.

Parameter	Prior Value	Description of Parameter
ϕ_a	0.63052	Constant term for the PTF of saturated soil moisture
ϕ_b	0.10262	Negative of the coefficient of squared bulk density for the PTF of saturated soil moisture
ϕ_c	0.0003335	Coefficient of clay content for the PTF of saturated soil moisture
ϕ_d	1.16518	Negative of the constant term for the PTF of $\log_{10}(\alpha)$, where α is the scaling factor for bubbling pressure
ϕ_e	0.16063	Negative of the coefficient of squared bulk density for the PTF of $\log_{10}(\alpha)$, where α is the scaling factor for bubbling pressure
ϕ_f	0.008372	Negative of the coefficient of clay content for the PTF of $\log_{10}(\alpha)$, where α is the scaling factor for bubbling pressure
ϕ_g	0.01300	Negative of the coefficient of silt content for the PTF of $\log_{10}(\alpha)$, where α is the scaling factor for bubbling pressure
ϕ_h	0.25929	Negative of the constant term for the PTF of $\log_{10}(N - 1)$, where N is the pore-size distribution index
ϕ_i	0.10590	Negative of the coefficient of squared bulk density for the PTF of $\log_{10}(N - 1)$, where N is the pore-size distribution index

ϕ_j	0.009004	Negative of the coefficient of clay content for the PTF of $\log_{10}(N - 1)$, where N is the pore-size distribution index
ϕ_k	0.001223	Negative of the coefficient of silt content for the PTF of $\log_{10}(N - 1)$, where N is the pore-size distribution index
ϕ_l	0.40220	Constant term for the PTF of $\log_{10}(K_{sat})$, where K_{sat} is the saturated hydraulic conductivity
ϕ_m	0.02329	Negative of the coefficient of clay content for the PTF of $\log_{10}(K_{sat})$, where K_{sat} is the saturated hydraulic conductivity
ϕ_n	0.01265	Negative of the coefficient of silt content for the PTF of $\log_{10}(K_{sat})$, where K_{sat} is the saturated hydraulic conductivity
ϕ_o	0.010380	Negative of the coefficient of cation exchange capacity (CEC) for the PTF of $\log_{10}(K_{sat})$, where K_{sat} is the saturated hydraulic conductivity; Note the value is 10 times smaller than in Pinnington <i>et al.</i> (2021)

Table 1 Table of the 15 parameters that we use for data assimilation with the Tóth *et al.* (2015) pedotransfer function (PTF).

The soil hydraulic model is based upon the van Genuchten (1980) equations. The values are as in Pinnington *et al.* (2021) except for ϕ_o , which is 10 times smaller. See Pinnington *et al.* (2021) for more details.

6 Data Assimilation

To tune JULES to observations of soil moisture, we use an approach which calibrates the soil hydraulic parameters in the model, which are provide as ancillary data on a per grid-cell basis. This is done by modifying parameters to the pedotransfer function that translates between data on soil texture and other properties to the soil hydraulic parameters needed by JULES. The new parameters for the pedotransfer function are assumed to be valid across the spatial domain of the data assimilation (Europe, in this case) but because the input data on soil texture vary spatially, the resulting soil hydraulic parameters also vary spatially.

The technique we use is derived from the standpoint of 4DVAR, which is commonly used in operational weather forecasting. The core of the standard 4DVAR algorithm is to minimize the cost function

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(h(x) - y)^T R^{-1}(h(x) - y), \quad 1$$

where x is the vector of unknown quantities to be determined (in our case, the pedotransfer function parameter), x_b and the covariance matrix B represent the first and second order moments of the prior distribution of the parameters, y is a vector containing the observations (soil moisture) and R is a covariance matrix representing the uncertainty in those observations. The non-linear operator $h(x)$ translates the unknowns into observation space and is provided by the JULES model in the case we describe here.

However, efficient minimisation of equation 1 typically requires a differentiated version of the non-linear operator $h(x)$ to allow the computation of the gradient of $J(x)$. This can be problematic for large, complex models like JULES because it is time consuming to generate the differentiated code. Consequently, we have adopted an approach that approximates the gradient information using ensembles runs of the operator.

We start by defining a perturbation matrix X'_b which is populated by an ensemble of random deviations from x_b drawn from a normal distribution such that

$$B \approx X'_b X'^T_b, \quad 2$$

and define a control variable transform

$$x - x_b = X'_b w, \quad 3$$

where w is a vector of weights with length equal to the number of ensemble members. In effect, the perturbation matrix translates between ensemble space and parameter space. We then perform an ensemble of JULES runs to make predictions of soil moisture for each of the perturbed parameter sets from which we can define a matrix of perturbations of the model predicted observations, HX'_b , such that

$$\hat{y} = HX'_b w + h(x_b), \quad 4$$

where $\hat{y}$ is the vector of predicted soil moisture given the weighting vector. From here, we can redefine Equation 1 in ensemble space, resulting in

$$J(w) = \frac{1}{2}w^T w + \frac{1}{2}(HX'_b w + h(x_b) - y)^T R^{-1}(HX'_b w + h(x_b) - y). \quad 5$$

Equation 5 can then be minimised to find the values for vector w that provides the best fit to the observations, and the optimal parameter set can be found by transforming back into parameter space. We can employ efficient gradient descent techniques for this purpose as we can now write the gradient of the cost function as

$$\nabla J(w) = w + X_b'^T H^T R^{-1} (H X_b' w + h(x_b) - y), \quad 6$$

thus avoiding the need for an explicitly differentiated version of JULES.

The method outlined above has been described by Pinnington *et al.* (2021) and shown to improve predictions of JULES soil moisture compared to independent ground-based observations from the COSMOS-UK network. The work of Pinnington *et al.* (2021) has also been used to assimilate SMAP soil moisture for 2017 (Reichle *et al.*, 2022; Chan *et al.*, 2016) into the JULES model throughout continental Africa (Maidment *et al.*, 2026). Two key advances in the EXPECT project are (a) a substantial increase in the spatial scale (the Pinnington study only considered a portion of the South East of England) and (b) the fact that we have used a much longer time series of data (the ESA CCI soil moisture data set: Dorigo *et al.*, 2015; Dorigo *et al.*, 2017) for the assimilation (2014-2018) and evaluation (1980-2019). Pinnington *et al.* (2021) also provide a mechanism for generating a posterior ensemble via which we can quantify the uncertainty in model prediction following data assimilation. The equations are not reproduced here, but they underlie the results in, for example, Figure 1 which shows the prior and posterior spread in the retrieved hydraulic parameters.

6.1 Data assimilation within the Rose/Cylc suite

Within the aforementioned Rose/Cylc suite, u-dq237, the data assimilation equations described in this section are implemented as Python code that is run after the 50-member 'prior' ensemble of FORTRAN-based JULES simulations are run. Each prior-ensemble member has the 15 prior soil parameters perturbed to be within a 10% gaussian spread of the unperturbed prior values. For each ensemble member, a new set of 15 posterior soil parameters is generated by the DA Python code, which is subsequently run as a 'posterior' ensemble of FORTRAN-based JULES simulations.

7 Other Methods

7.1 Automated plot generation

The Rose/Cylc suite that runs both the JULES and the DA also has Python code within it that generates .PNG files during postprocessing for the requested JULES output variables, perhaps averaged over each year. There is a .TXT file associated with each yearly plot for each variable that gives the vmin,vmax plotting limits and the spatial average and spatial RMS (computed over the whole European region). These .PNG and .TXT files are all downloaded to a MacBook Pro laptop, where another custom

Python program automatically sorts the plots into subplots on the correct, specified pages of a PowerPoint presentation. Plot annotations can be prescribed either programmatically in the Python code or from a template PowerPoint presentation. This semi-automated plot generation from the JASMIN supercomputer to the MacBook laptop into a PowerPoint file allows many different output variables for different years to be inspected in an easy-to-use way.

7.2 Calculation of the analysis increment and the RMSE reduction

In order to determine the change of a variable after the DA, we calculate the 'analysis increment', that is: the change of the variable after analysis, which in simple, generic terms (using the same notation as in the DA section above) is given by:

$$\delta x = x_a - x_b, \quad 7$$

where x_a is the value of the vector variable x after analysis, x_b is the value of the vector variable x before analysis (a.k.a., the background value), and δx is the analysis increment. This is similar to our equation 3. Privé and Errico (2013) give the analysis increment in terms of a Kalman matrix in their equation 1.

In a less generic notation, for example for the scalar NPP variable, the analysis increment (AI) (in absolute terms) for a given year y is:

$$AI_{abs}(NPP(y)) = NPP_{ave,posterior}(y) - NPP_{ave,prior}(y), \quad 8$$

where $NPP_{ave,prior}(y)$ is the annual-averaged NPP prior to the DA, and $NPP_{ave,posterior}(y)$ is the annual-averaged NPP after the DA. If desired, this can be normalized and converted to a percentage relative change as:

$$AI_{rel}(NPP(y)) = \frac{100 AI_{abs}(NPP(y))}{NPP_{ave,prior}(y)} \quad 9$$

The analysis increment for a given year can be compared to the analysis increment during the data assimilation period ranging from year y_1 to year y_2 . This is called the 'hindcast increment' (HI):

$$HI_{abs}(NPP(y)) = AI_{abs}(NPP(y)) - AI_{abs,DA}(NPP(y_1:y_2)), \quad 10$$

where the analysis increment during the data assimilation period is:

$$AI_{abs,DA}(NPP(y_1:y_2)) = NPP_{ave,posterior}(y_1:y_2) - NPP_{ave,prior}(y_1:y_2). \quad 11$$

The average prior NPP during the data assimilation period is $NPP_{ave,prior}(y_1:y_2)$, and similarly for the average posterior NPP.

The Root Mean Square Error (RMSE) of soil moisture (SM) is calculated as:

$$\text{RMSE}(\text{SM}(y)) = \sqrt{\frac{1}{N_d} \sum_{d=1}^{N_d} (\text{SM}_{\text{model}}(d, y) - \text{SM}_{\text{obs}}(d, y))^2}, \quad 12$$

where d is the day number, and N_d is the number of days (i.e., in a year).

The RMSE reduction of SM is computed in a similar way (but with an opposite sign) that the analysis increment is computed:

$$\text{RMSE reduction}(\text{SM}(y)) = \text{RMSE}_{\text{prior}}(\text{SM}(y)) - \text{RMSE}_{\text{posterior}}(\text{SM}(y)). \quad 13$$

The change in RMSE reduction of SM relative to the DA period is then calculated similarly to the hindcast increment:

$$\begin{aligned} &\text{change in RMSE reduction}(\text{SM}(y)) = \\ &\text{RMSE reduction}(\text{SM}(y)) - \text{RMSE reduction}(\text{SM}(y_1:y_2)). \end{aligned} \quad 14$$

The analysis increment, the hindcast increment, the RMSE, the RMSE reduction, etc. can all be computed as gridded variables over a selected region.

7.3 Calculation of Weeks of Max Wetting and Max Drying

The Wetting for a given week, w , is computed by taking the positive weekly differences of soil moisture, SM, in the topsoil layer in a grid cell for a given year y :

$$\text{Wetting}(w, y) = \max(\text{SM}(w, y) - \text{SM}(w - 1, y), 0). \quad 15$$

Likewise, the Drying is computed by taking the negative weekly differences of soil moisture in the topsoil layer in a grid cell:

$$\text{Drying}(w, y) = \min(\text{SM}(w, y) - \text{SM}(w - 1, y), 0). \quad 16$$

Note that the Drying values that we use are non-positive.

The Week of Max Wetting (WMW) for a given year is

$$\text{Week of Max Wetting}(y) = \arg \max_w [\text{Wetting}(w, y)], \quad 17$$

where the $\arg \max[\]$ function finds the week of the maximum wetting.

The Week of Max Drying (WMD) for a given year is

$$\text{Week of Max Drying}(y) = \arg \min_w [\text{Drying}(w, y)], \quad 18$$

where the $\arg \min[\]$ function finds the week of the maximum absolute drying, where drying is negative.

Likewise, the Wetting During Week of Max Wetting (WWMW) for a given year is

$$\text{Wetting during Week of Max Wetting}(y) = \max_w [\text{Wetting}(w, y)], \quad 19$$

where the $\max[\]$ function finds the maximum value of $\text{Wetting}(w, y)$.

The Drying during Week of Max Drying (DWMD) for a given year is

$$\text{Drying during Week of Max Drying (y)} = \min_w [\text{Drying}(w, y)], \quad 20$$

where the $\min[\]$ function finds the maximum absolute drying, where drying is negative.

7.4 Calculation of SSI hydrological drought index

After the DA, the modelled soil moisture should match the observed soil moisture better, and therefore, the modelled hydrological drought index could be considered to be an improved estimate after DA. Here, we compute the annual-averaged standardized soil-moisture index (SSI) as the hydrological drought index. The SSI was computed using daily soil moisture JULES data aggregated over a 30-day moving window. Only the top layer (0.1 meters) of the JULES soil-moisture was used here. For each grid cell, the 30-day soil-moisture series was standardized using an empirical cumulative distribution function and transformed to standard normal space. The resulting SSI-30 fields were subsequently averaged to annual means for analysis.

8 Results

For demonstration purposes, for many of the variables shown in this results section, we focus on the ensemble-mean behaviour for a single year (2003), which is a year when much of central Europe experienced a hydrological drought. We do have JULES data available for this work from 1980-2019, using the WFDE5 driving data. We also focus our analysis on the runs where the DA is only applied for latitudes north of Gibraltar. This means that the modelled soil moisture south of Gibraltar (i.e., in arid regions like the Sahara) will not affect the DA.

8.1 Change of soil PTF parameters after data assimilation

After the DA, the posterior distributions of the Tóth pedotransfer function (PTF) parameters change from the prior distributions (see Figure 1 and Table 1), in order to minimize the cost function in equation 1 over the European region (for latitudes north of Gibraltar) for the chosen data assimilation period of 2014-2018. The mean values of the posterior PTF parameters are usually shifted from the mean values of the prior PTF parameters, and the widths of the posterior PTF parameter distributions (over ensemble members) are usually narrower than the widths of the prior PTF parameter distributions. For example, the ϕ_e parameter indicates the coefficient of p^2 in the PTF for $\log_{10}(\alpha)$, or in other words, $\log_{10}(\alpha) = \phi_e p^2 + \text{other terms}$. The mean value of the ϕ_e parameter increases from 0.16063 to 0.337, and the standard deviation of the ϕ_e parameter distribution goes down considerably, from 0.0161 (10%) to 0.00395 (1.17%). The higher values of ϕ_e imply a higher sensitivity of $\log_{10}(\alpha)$ to p^2 . Since α represents the inverse air-

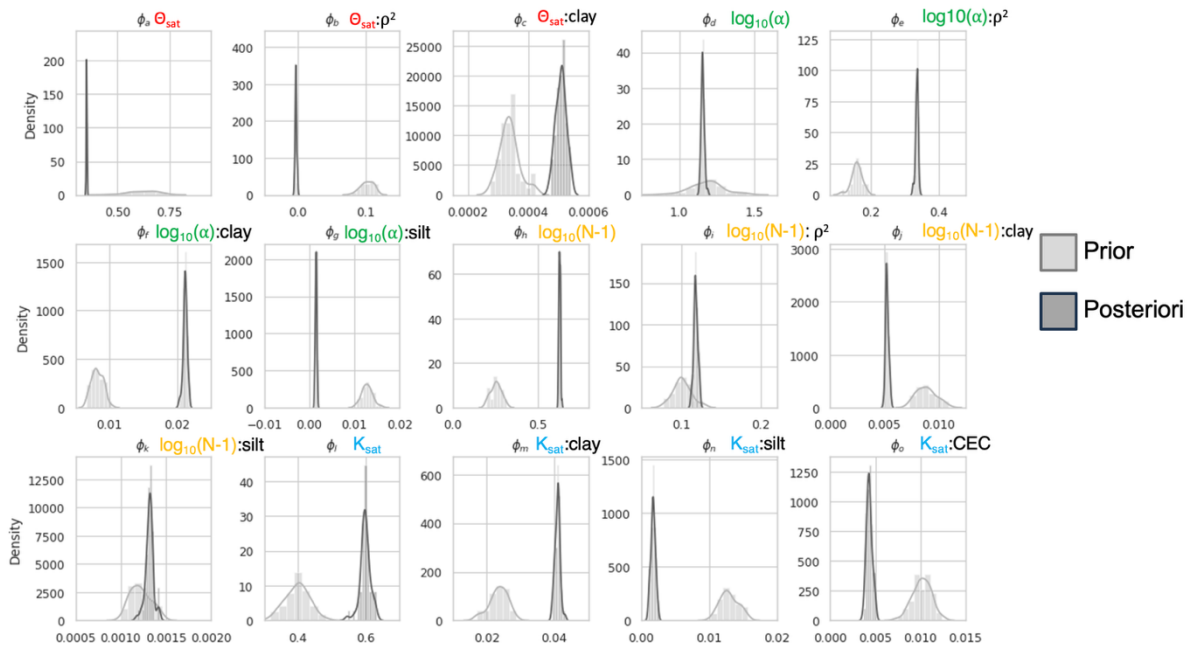


Figure 1: Data assimilation gives posteriori estimates of Tóth Pedotransfer-Function (PTF) soil parameters. These Tóth PTF parameter values ($\phi_{\alpha-o}$) are constant over the whole European region for each ensemble member. These parameter values are used together with SoilGrids grid-cell estimates of the (top) soil basic properties (clay, silt, sand, CEC, ρ , & pH) in order to estimate the soil physical properties (θ_{sat} , $\log_{10}(\alpha)$, $\log_{10}(N-1)$, K_{sat}) for each grid cell. The distributions shown here for the prior Tóth PTF parameter values indicate the assumed fractional standard deviation of 10% over the 50 ensemble members, the distributions shown here are for the posterior Tóth PTF parameter values indicate the computed variation after data assimilation for the 50 ensemble members over the European region (north of Gibraltar) for the data assimilation period of 2014-2018.

entry suction, this means that areas within Europe with higher bulk density (compacted soils, dense soils) will now show larger increases in α , meaning they drain more readily in the model after DA. Likewise, areas with lower bulk density will show larger decreases in α , meaning stronger water retention after DA. Furthermore, the lower standard deviation of the ϕ_e parameter suggests that the DA of the ESA CCI satellite soil moisture is constraining the parameter more tightly than assumed in the prior.

8.2 Reduction of soil moisture RMSE after data assimilation

We calculate the reduction of soil moisture RMSE after DA for each year and for each grid cell in Europe. The DA was done previously with an ensemble of 50 members for either the period 2014-2017 or the period 2014-2018. The latter period adds an extra year which happens to be a drought year (2018) for central Europe.

In Figure 2, the Prior RMSE for soil moisture for the year 2003 is shown in a) and the Posterior RMSE for soil moisture for the same year is shown in b). Note that the Posterior RMSE is considerably reduced (it is bluer) in significant areas. This reduction is shown

more clearly by the preponderance of red colours in c), where the RMSE reduction (the difference between a) and b)) is displayed.

Figure 2d is the change in RMSE reduction (see equation 14) for 2003, relative to the DA period of 2014-2018. This sub-figure suggests that for 2003, in Western and Central Europe (excluding the Nordics), that the RMSE reduction is rather similar to the RMSE reduction during the DA period of 2014-2018. However, in Western Russia and parts of the Nordics, the change in RMSE reduction is significantly higher (more RMSE reduction) in 2003 as compared to the DA period of 2014-2018.

In Figure 3, we observe that the European average of soil moisture RMSE reduction is > 0 for all years (1980-2019). We also observe that the European average of RMSE reduction is about 0.007 for the years 2014-2016 and is about 0.011 for year 2017 (all 4 years are within either of the assimilation window periods). These values are higher than the single assimilation-year values (not shown here). By inspecting the red curves in Figure 3, we see that an addition of a fifth year to the Assimilation Window Period causes a small increase in RMSE reduction for all years. This is evidence that the DA is reducing the RMSE even when there is a possible disturbance to the values of the soil parameters caused by the drought in 2018 in Central Europe.

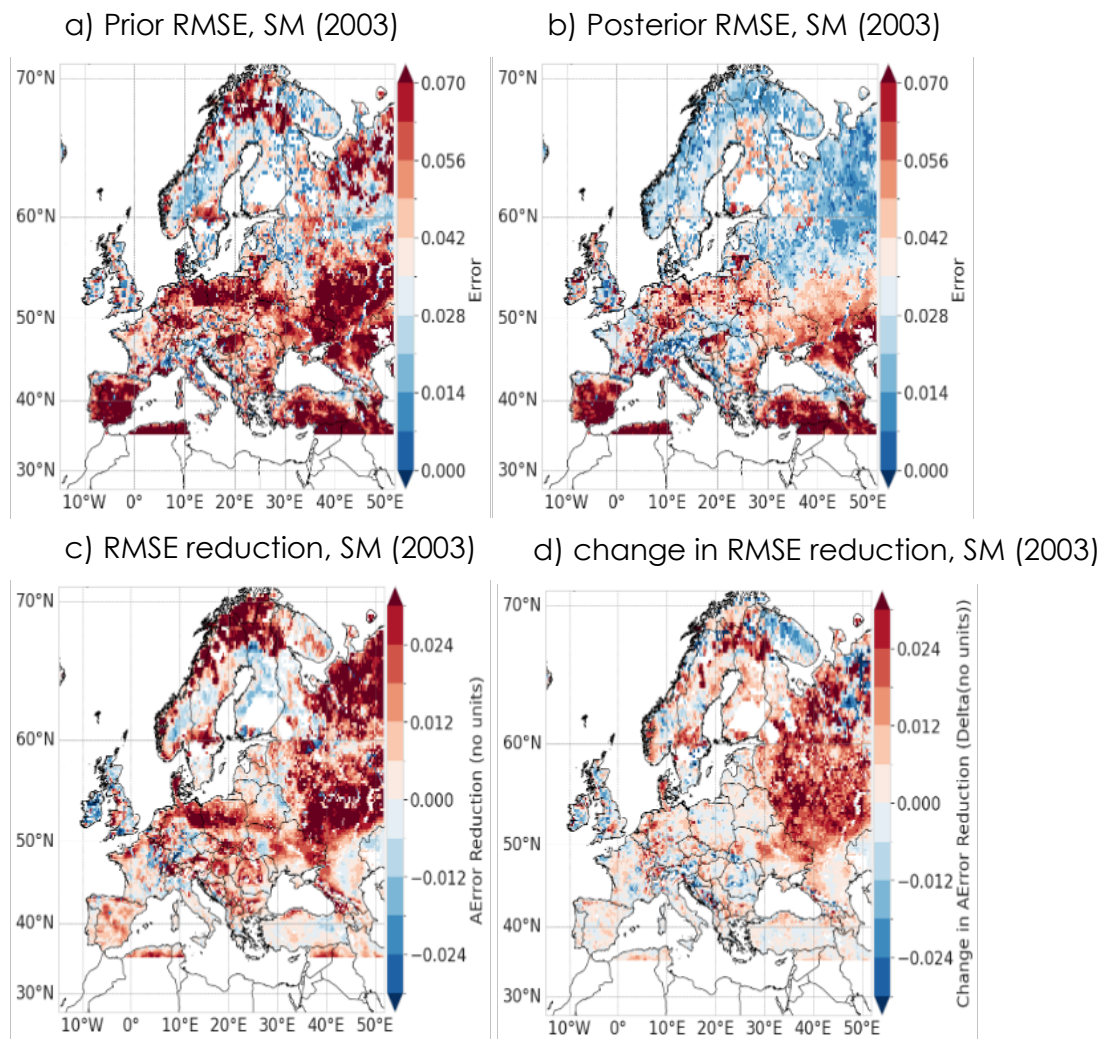


Figure 2: The Root Mean Square Error (RMSE) of soil moisture (SM) for 2003 (units = None, plot limits of ± 0.07), before (a) and after (b) DA.

The RMSE reduction for the year 2003 is shown in (c) with plot limits of ± 0.03 , and the change in RMSE reduction for the year 2003 relative to the DA period of 2014-2018 is shown in (d) with the same plot limits. During the year 2003, there was a drought in central Europe.

RMSEreduction: European ave. $\pm$ StdDev

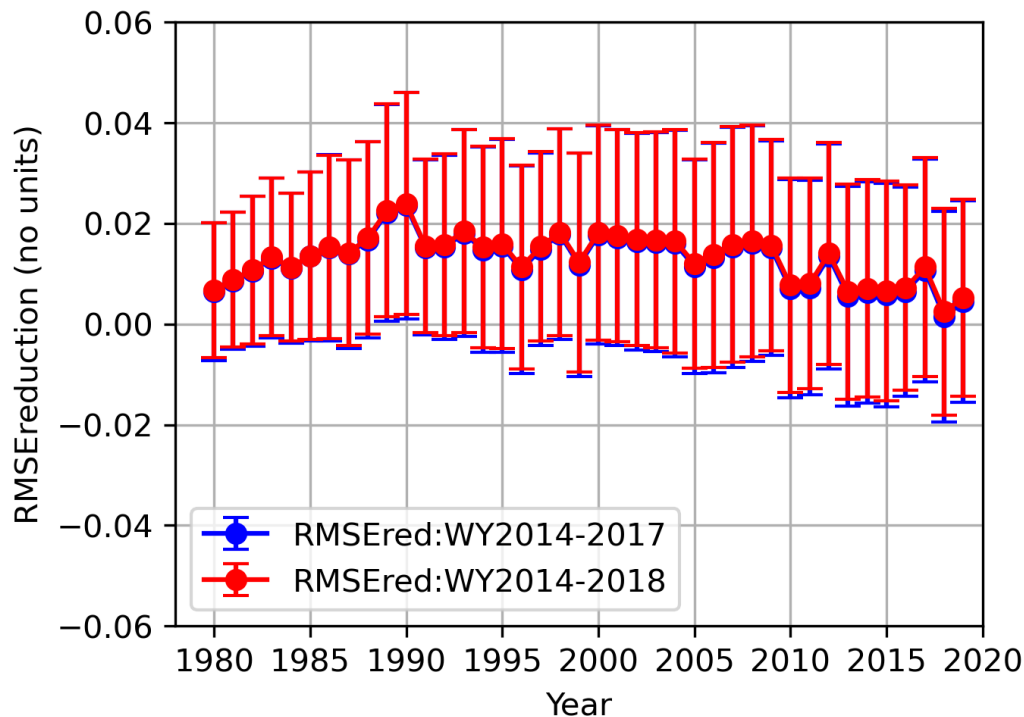


Figure 3: For the years 1980-2019, the European average ($\pm$ the spatial standard deviation) of the reduction of soil moisture RMSE after the data assimilation is plotted. The DA was performed previously over a pre-defined time period. A spinup was done first by running the model for 4 years of 1980. The DA is restricted to latitudes north of Gibraltar. The blue curves show the results for when the DA period covers the window years (WY) 2014-2017, and the red curves show the results for when the DA period covers WY 2014-2018.

8.3 Re-estimation of soil moisture after data assimilation

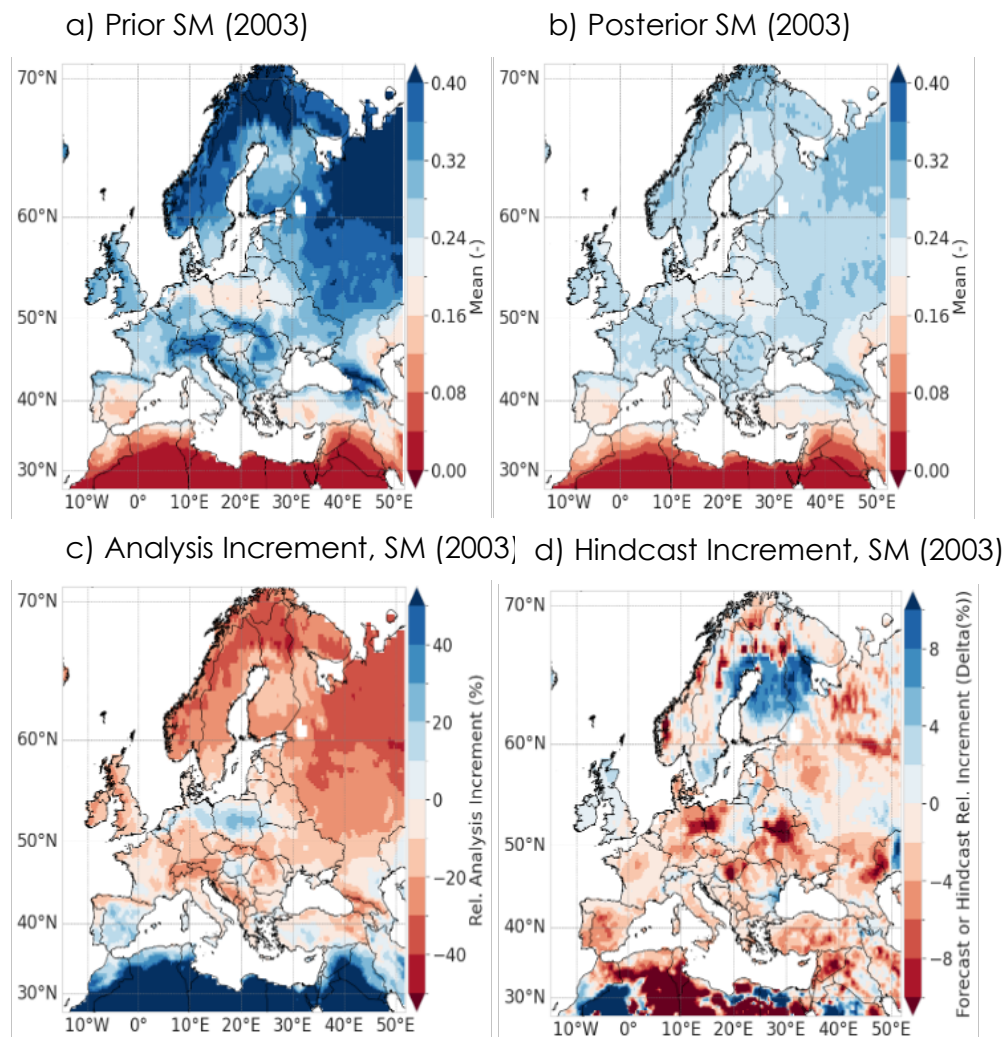


Figure 4: The ensemble mean of soil moisture (SM) for 2003 (units = None, plot limits of 0.0-0.40), before (a) and after (b) DA.

The analysis increment of SM for the year 2003 is shown in (c) with plot limits of $\pm 50\%$, and the hindcast increment for the year 2003 relative to the DA period of 2014-2018 is shown in (d) with plot limits of $\pm 10\%$. During the year 2003, there was a drought in central Europe.

We also estimate the ensemble-average of the soil moisture for each year for each grid cell in Europe, both before DA and afterwards. In Figure 4, the Prior ensemble-average of soil moisture for the year 2003 is shown in a) and the Posterior ensemble-average of soil moisture for the same year is shown in b). Note that the Posterior SM is considerably reduced in most areas besides the Sahara, where the DA is not currently enabled. This reduction (drying) is also observed in subfigure c), where the analysis increment is plotted. For the area where DA is enabled (north of Gibraltar), the spatial average of the analysis increment for 2003 of SM is $-24.5\% \pm 9.4\%$. The drying is significant. Relative to the DA period of 2014-2018, as observed in subfigure d), there is accentuated drying (of order 5-10%) in eastern Germany, western Poland, Hungary,

NW Ukraine, and southern Belarus; and similarly, there are enhanced wetter soils in Finland.

8.4 Re-estimation of SSI hydrological drought index

Using the SSI as the hydrological drought index, as described in section 7, we can estimate the annual-averaged SSI before and after the DA. An example is shown in Figure 5, which is for the year 2003. The year 2003 is a central European drought year. In Figure 5a, before the DA, there is an apparent drought across much of Europe, with annual-averaged SSI values ranging from -1.5 up to -0.5 in many grid cells (the value of -1 is a 1-sigma below-average year for soil-moisture). In Figure 5b, after the DA, the reddish-coloured area has qualitatively become slightly redder than before the DA, meaning that the hydrological drought is estimated to be slightly stronger than previously estimated. The difference between the values from before the DA to after the DA is shown in Figure 5c as the absolute analysis increment

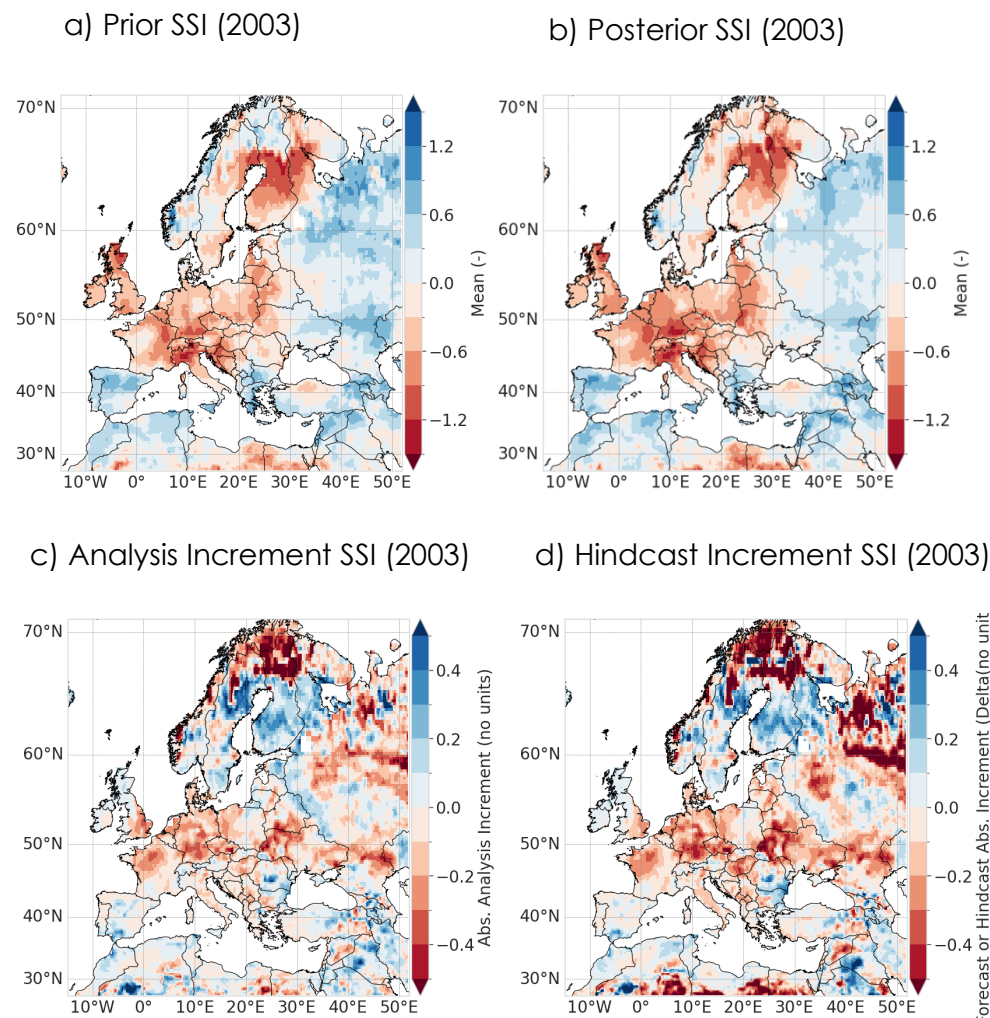
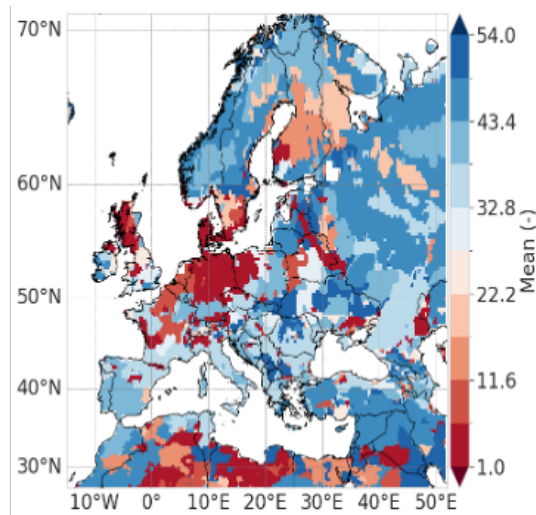


Figure 5: The annual standardized soil-moisture index (SSI) for the Central Europe drought year 2003 (units = standard deviations), before (a) and after (b) DA. In (c), the absolute analysis increment of SSI is shown for 2003. In (d), the absolute hindcast increment of SSI is shown for 2003.

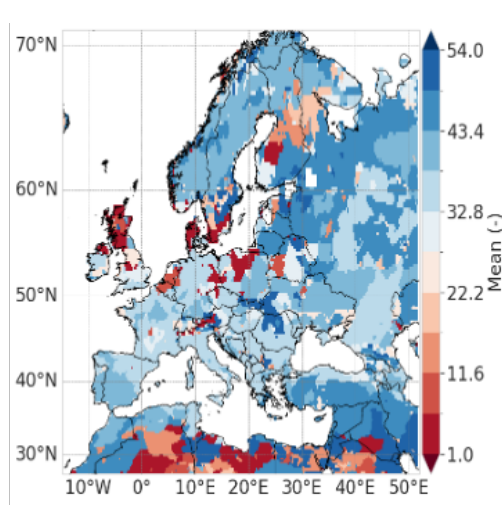
to be between -0.3 and -0.2 for much of Central Europe, indicating 0.2-0.3 lower SSI after the DA. We don't see such a clear, distinctive difference in the annual-averaged SSI for the other years (whether they be drought years or not), but there are differences, which can be evaluated in a more thorough analysis. One possible next step is to repeat the analysis but focusing on a seasonal analysis of the SSI.

8.5 Re-estimation of the Week of Maximum Wetting

a) Prior ensemble mean 2003 WMW



b) Posterior ensemble mean 2003 WMW



c) Timeseries of dSMCL (depth<=0.10 m) for 2003 averaged over northern Germany

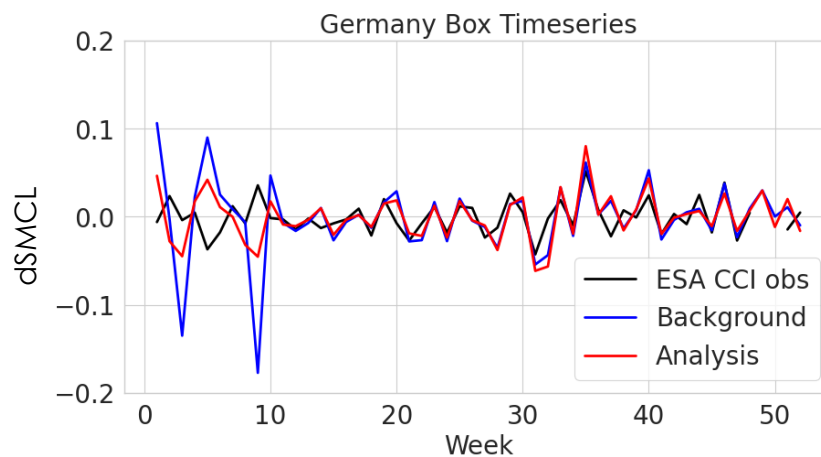


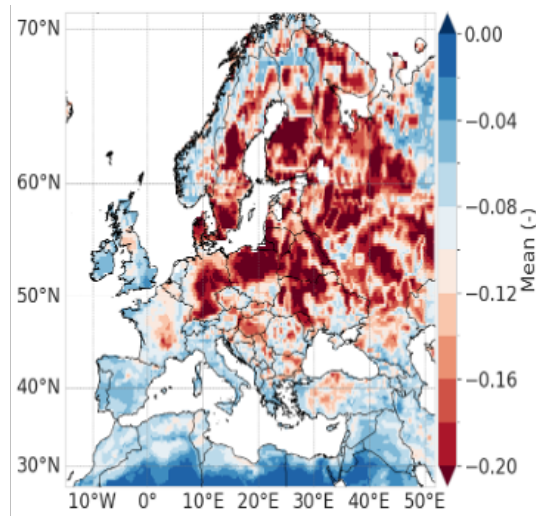
Figure 6: For the year 2003: (a-b) "Week of max wetting" (WMW), (a) map of prior ensemble mean WMW, (b) map of posterior ensemble mean WMW, and (c) timeseries of weekly change of soil moisture (dSMCL: no units) averaged over depths ≤ 0.10 m (for background and analysis) and over northern Germany (50-53°N, 8-13°E) with ESA CCI observations in black, prior ensemble mean in blue, and posterior ensemble mean in red.

In Figure 6, we show the effects of using the DA system upon the “Week of Maximum Wetting” (WMW) variable (see Equation 17) that we compute for the prior and posterior ensemble means. The maps of the WMW are shown for the year 2003 in Figure 6a-b and are computed from the weekly change in soil-moisture (dSMCL) timeseries for each grid-cell (not spatially-averaged over grid-cells), similar to the spatially-averaged (over northern Germany) timeseries shown in Figure 6c. The map of the ensemble-mean of the prior WMW (before the DA) is shown in Figure 6a, and the map of the ensemble-mean of the posterior WMW (before the DA) is shown in Figure 6b. Note that for the prior WMW (Figure 6a), there are many locations in Europe with WMW < 10, whereas for the posterior WMW (Figure 6b), most of these same locations have WMW between 30-40. This can be understood by inspection of the dSMCL timeseries plot in Figure 6c, where for northern Germany, the large positive spikes in weeks 1 and 5 are reduced by the DA, which are closer to the averages of the ESA CCI observations of dSMCL. This has the effects of changing the WMW to values later in the year and/or reducing the Wetting during the WMW to more reasonable values. The reduction of extreme values in dSMCL is one of the expectations of the DA system.

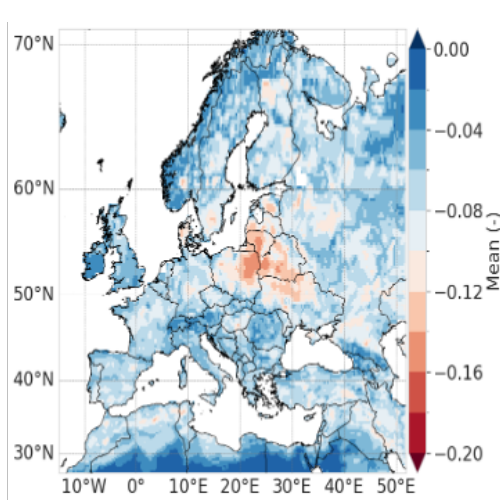
8.6 Re-estimation of Drying during the Week of Maximum Drying

In Figure 7, we show the effects of using the DA system upon the “Drying during the Week of Maximum Drying” (DWMD) variable (see equation 20) that we compute for the prior and posterior ensemble means. The maps of the DWMD are shown for the year 2003 (which was a drought year for central Europe) in Figure 7a-b and are computed from the weekly change in soil-moisture (dSMCL) timeseries for each grid-cell (not spatially-averaged over grid-cells), similar to the spatially-averaged (over northern Germany) timeseries shown in Figure 7c. The map of the ensemble-mean of the prior DWMD (before the DA) is shown in Figure 7a, and the map of the ensemble-mean of the posterior DWMD (before the DA) is shown in Figure 7b. Note that for the prior DWMD (Figure 7a), there are many locations in Europe with DWMD < -0.16, whereas for the posterior DWMD (Figure 7b), most of these locations have DWMD > -0.10, with a residual localized spot of high (negative) DWMD centred in eastern Poland. This can be understood by inspection of the dSMCL timeseries plot in Figure 7c, where for northern Germany, the large negative spikes in weeks 3 and 9 are reduced by the DA, which are closer to the averages of the ESA CCI observations of dSMCL. This has the effects of changing the Week of Maximum Drying (WMD) and/or reducing the DWMD to more reasonable values. The reduction of extreme values in dSMCL is one of the expectations of the DA system.

a) Prior ensemble mean 2003 DWMD



b) Posterior ensemble mean 2003 DWMD



c) Timeseries of dSMCL (depth<=0.10 m) for 2003 averaged over northern Germany

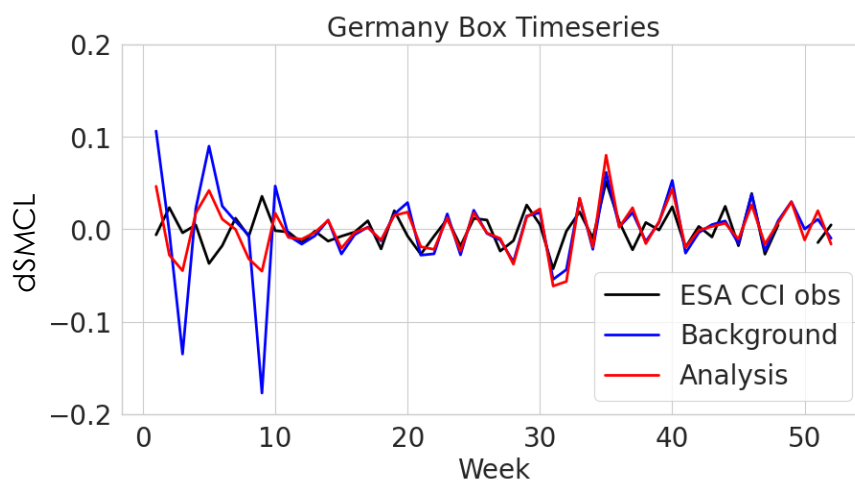


Figure 7: For the year 2003, which is a drought year in Central Europe: (a-b) "Drying during week of max drying" (DWMD), (a) map of prior ensemble mean DWMD, (b) map of posterior ensemble mean DWMD, and (c) (same as Figure 6c) timeseries of weekly change of soil moisture (dSMCL: no units) averaged over depths ≤ 0.10 m (for background and analysis) and over northern Germany (50-53°N, 8-13°E) with ESA CCI observations in black, prior ensemble mean in blue, and posterior ensemble mean in red.

8.7 Re-estimation of the net primary production (NPP)

We use the annual-averaged net primary production (NPP), as defined by $NPP = GPP - RA$, where GPP is the gross primary production and RA is the autotrophic respiration. The NPP statistic measures the rate that the vegetation is retaining assimilated carbon as biomass, after accounting for the autotrophic respiration of CO_2 . In Figure 8, we first estimate the analysis increment (difference of NPP before and

after DA), and then we estimate the hindcast increment for each year relative to the mean over the data assimilation window of 2014-2018.

The drought year in central Europe of 2003 has the NPP hindcast increment being negative for central Europe, which is perhaps the effect of the drought. The negative values of the NPP hindcast increment in eastern Europe and western Russia in 1999 and 2002 might suggest agricultural droughts in those regions for those years.

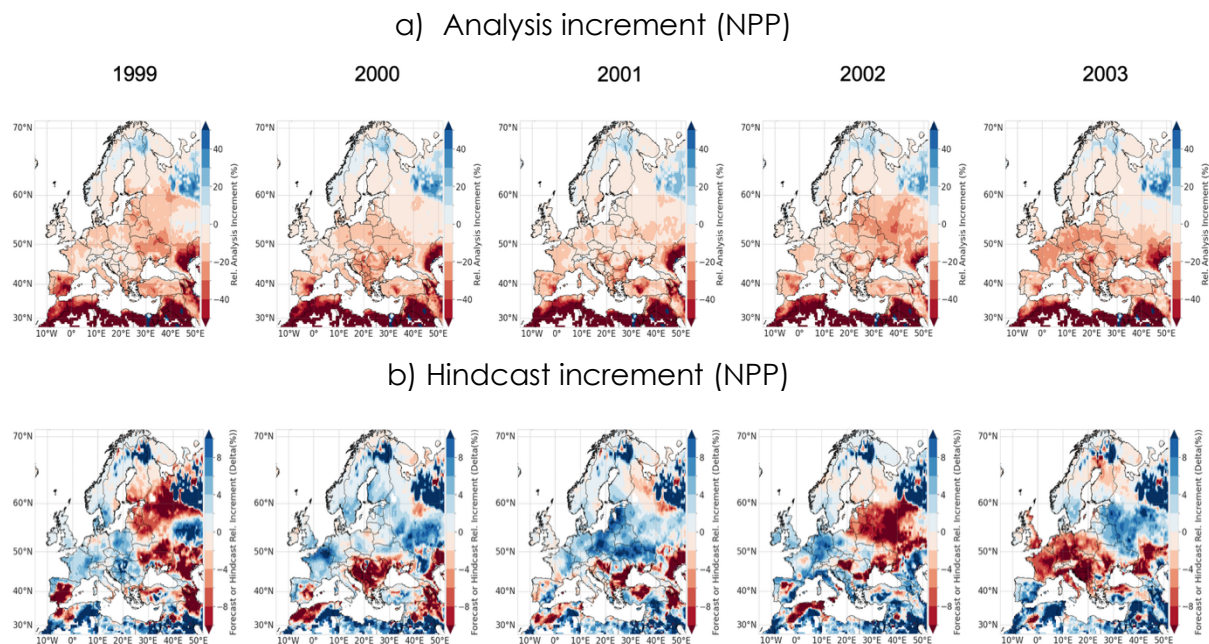


Figure 8: Analysis increments (a) and Hindcast increments (b) of annual net primary production (NPP) for 1999-2003 (units = %).

The hindcast increments are calculated with respect to the mean of NPP over the data assimilation period (2014-2018). The analysis increments are shown in (a) with plot limits of $\pm 50\%$, whereas the hindcast increments are shown in (b) with plot limits of $\pm 10\%$. The NPP is defined as $GPP - RA$. During the year 2003, there was a drought in central Europe.

8.8 Re-estimation of the carbon use efficiency (CUE)

We use the annual-averaged carbon use efficiency (CUE), as defined by $CUE = NPP/GPP$, where NPP is the net primary production and GPP is the gross primary production. The CUE statistic measures how efficient the vegetation is at retaining assimilated carbon as biomass, after accounting for autotrophic respiration. Before and after the DA, as shown in Figure 9, for much of western Europe, the DA does not affect the CUE for 2003 that much. In western Europe, the annual-averaged CUE for 2003 is usually between 0.4-0.5, and the analysis increment of the CUE is usually relatively close to 0, with slight negative values ranging down to only -10%. However, in parts of western Russia, Finland, and northern Sweden, the CUE is often less than 0.3, and the analysis increment of the CUE fluctuates dramatically with both

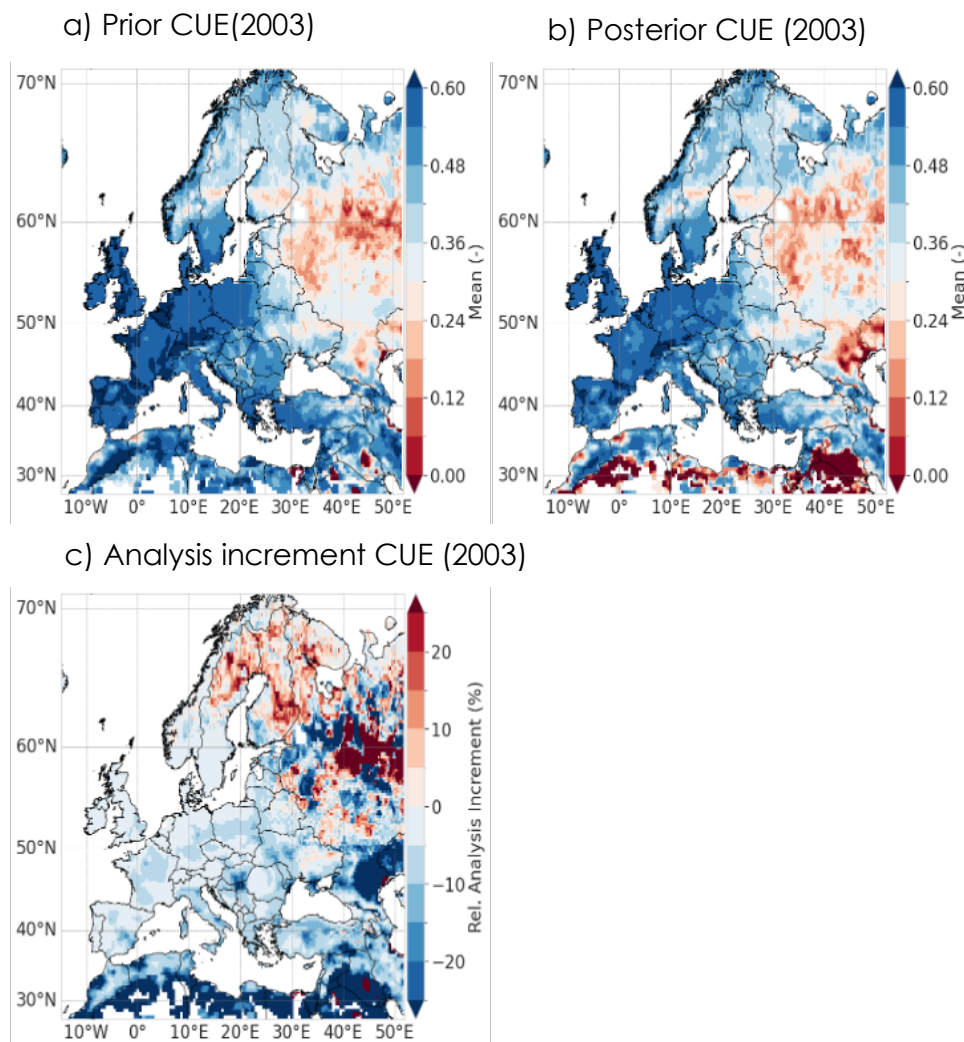


Figure 9: The annual carbon use efficiency (CUE) for 2003 (units = None), before (a) and after (b) DA. The analysis increment for the year 2003 is shown in (c) with plot limits of $\pm 25\%$. The CUE is defined as NPP/GPP . During the year 2003, there was a drought in central Europe.

positive and negative values sometimes exceeding $\pm 25\%$. We estimate from the global MODIS observational analysis of CUE from 2000-2003 by Zhang *et al.* (2009) (see their Figure 1 which is reprinted here as Figure 10) that there are regions of

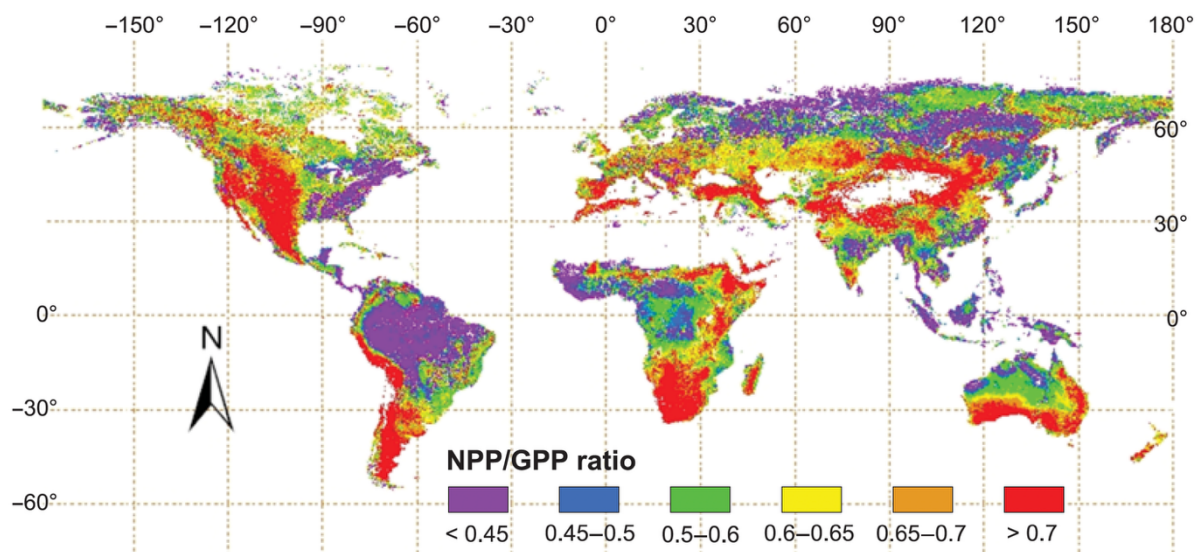


Figure 10: The global spatial pattern of average NPP/GPP (carbon use efficiency, CUE) from a MODIS analysis of 2000-2003 data. Reprinted from Zhang *et al.* (2009).

northeastern Europe with CUE < 0.45, whereas much of western Europe has values of CUE > 0.60. Zhang *et al.* (2009) conclude from their MODIS analysis of CUE that forested ecosystems have lower CUE than do herbaceous and shrub ecosystems. Therefore, the regional dichotomy in our modelled CUE is borne out by Zhang *et al.*'s CUE analysis of MODIS observations.

8.9 Re-estimation of the water use efficiency (WUE)

We use the annual-averaged water use efficiency (WUE), as defined by $WUE = GPP/ET$, where GPP is the gross primary production and ET is the evapotranspiration. The units of WUE are chosen to be gC/mmH₂O. The effects of dew and snow have not been accounted for. Here, the ET is computed from the latent heat flux without adding the terms for the sublimation from snow or for the condensation to dew. The WUE statistic measures how efficient the vegetation is at assimilating carbon per unit of water use. Before and after the DA, as shown in Figure 11, in much of western Europe (exceptions include western Spain, Hungary, and northern Scandinavia), the annual-averaged WUE for 2003 is usually between 0.9-1.5 gC/mmH₂O, and the analysis increment of the WUE is usually between $\pm 10\%$. However, in parts of western Russia, eastern Spain, Hungary, and northern Scandinavia, the WUE is often less than 0.75 gC/mmH₂O. The analysis increment of the WUE is normally between -15% and -5% for latitudes < 55°N.

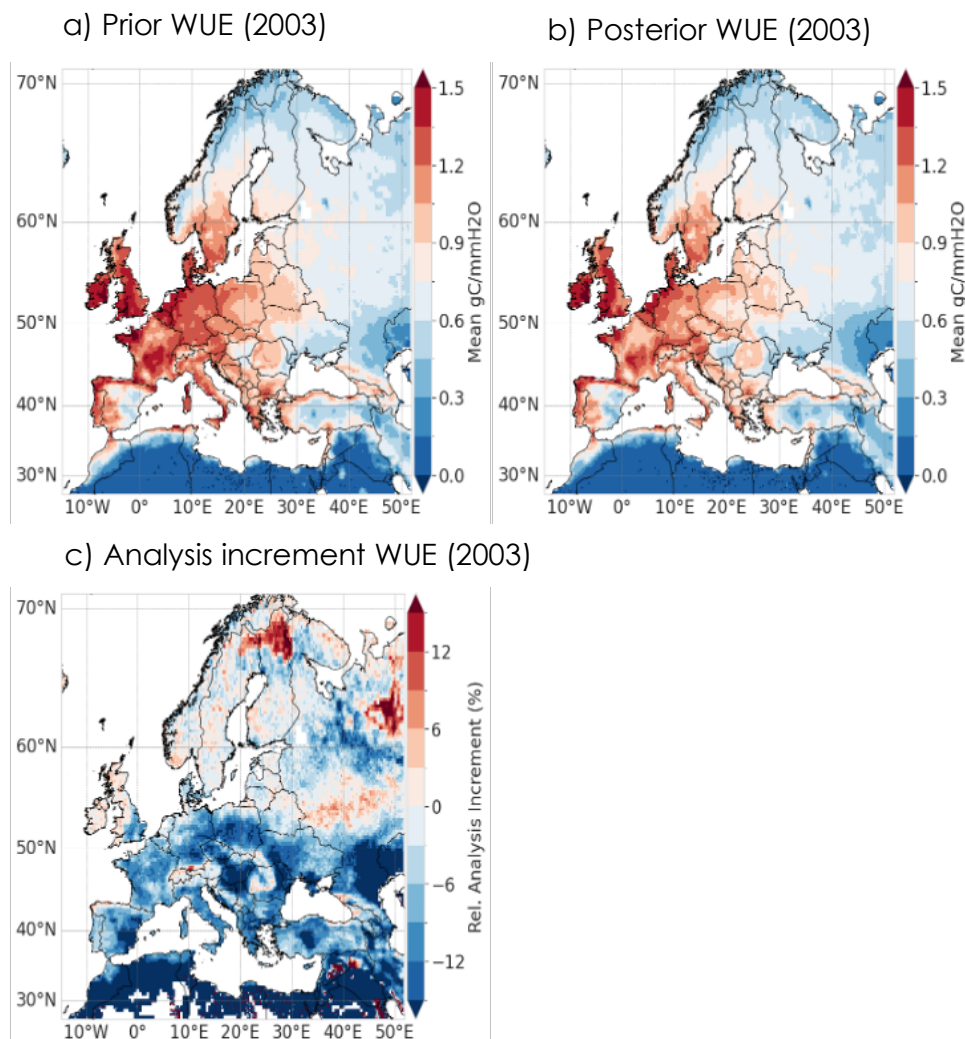


Figure 11: The annual water use efficiency (WUE) for 2003 (units = $\text{gC}/\text{mmH}_2\text{O}$), before (a) and after (b) DA, with plot limits of 0-1.5.

The analysis increment for the year 2003 is shown in (c) with plot limits of $\pm 15\%$. The WUE is defined as GPP/ET . The effects of dew and snow have not been accounted for. During the year 2003, there was a drought in central Europe.

8.10 Comparing analysis increments of WUE, CUE, GPP, NPP, latent heat flux, and soil moisture

In Figure 12, we plot the analysis increments for 2003 of WUE, CUE, GPP, NPP, latent heat flux, and soil moisture. The analysis increment of the WUE is normally between -15% and -5% for latitudes $< 55^\circ\text{N}$. For most of continental Western Europe, the analysis increment of GPP is between -30% and -10%, and the analysis increment of the latent heat flux (LHF) is between -10% and -2%. Since for snow-free and dew-free surfaces, latent heat flux can be divided by a constant (the latent heat of vaporization) to transform it to evapotranspiration, we can take the ratio of the midpoints of the

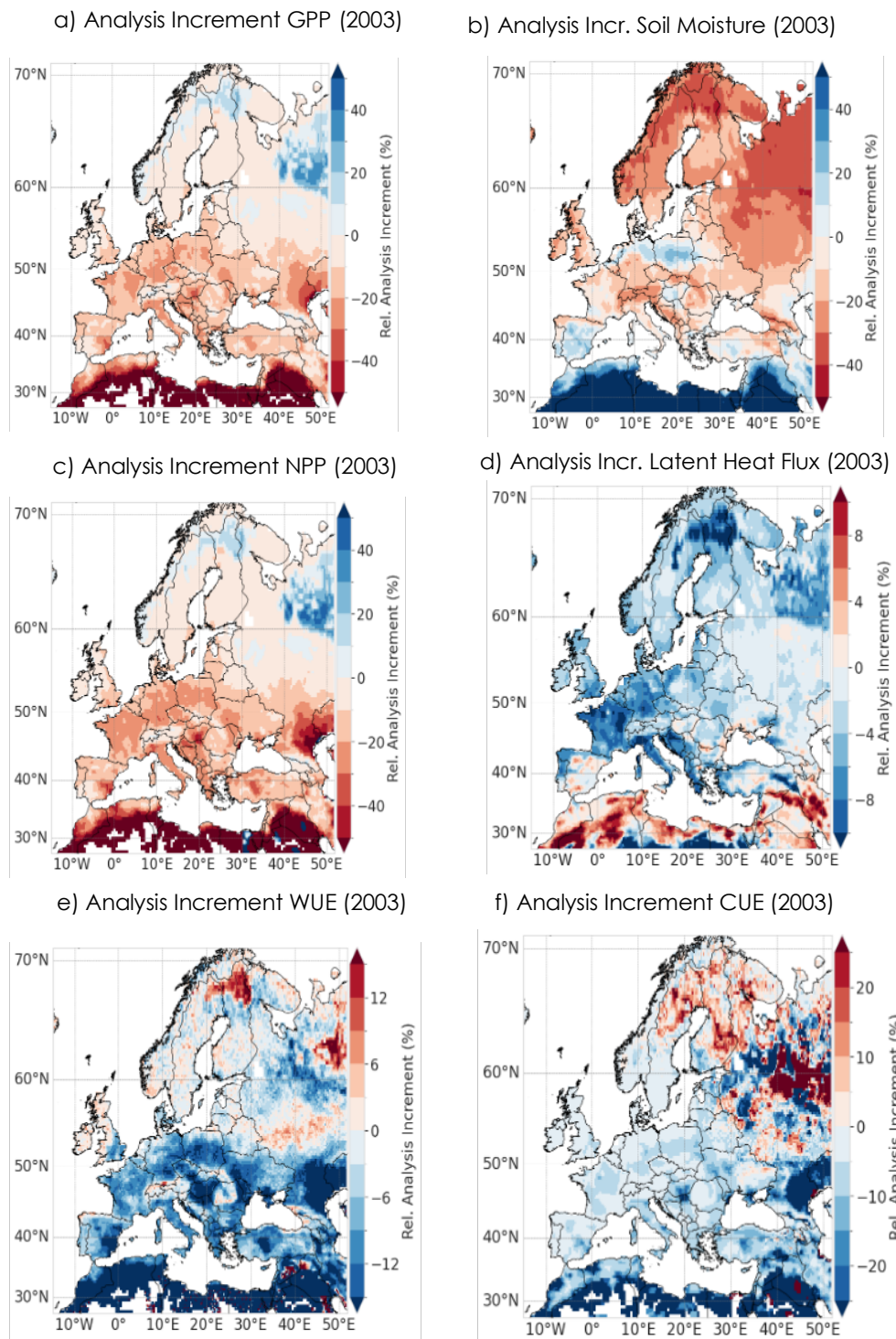


Figure 12: The ensemble-mean maps of the analysis increment in 2003 for a) GPP, b) Soil Moisture, c) NPP, d) Latent Heat Flux (can be used as evapotranspiration), e) WUE, and f) CUE.

analysis increments of GPP (-20%) and LHF (-6%): $(100\%-20\%)/100\%-6\% = 80\%/94\% = 0.80/0.94 = 0.85 = 85\% = (100\%-15\%)$. This back of the envelope estimate of the analysis increment for the ratio $WUE = GPP/ET$ thus gives -15%, which is consistent with the endpoint of the range plotted for WUE. Therefore, the larger factor here (at least for the year 2003 and for the cross-Europe spatial scale) for determining the analysis increment of WUE is the analysis increment of GPP, not the analysis increment of ET.

We also observe that the analysis increment of NPP has a similar range for most of continental Western Europe for 2003 as the analysis increment of GPP (between -30% and -10%). This is consistent with the analysis increment of CUE being about 0% for most of this region, since $CUE = NPP/GPP$.

Some outstanding questions remain:

- a) Why is the analysis increment of GPP positive instead of negative for northern Russia and northern Scandinavia?
- b) Why is the analysis increment of GPP near 0% further south in both Russia and Scandinavia?

We speculate the answer to both questions is related to the rather dry soils observed in the analysis increment of soil moisture ($< -40\%$, see Figure 12b or Figure 4c) for Scandinavia and Russia. The Prior SM for these regions (Figure 4a) is quite high (> 0.4), and the Posterior SM for these regions (Figure 4b) is in the normal range. It appears that the DA takes the high SM and puts it into the normal, less extreme range (as it should). Since JULES only uses soil-moisture drought stress functions and not water-logged stress functions, this change to less water-logged soils after DA should not have the effect of increasing the GPP. However, this was observed to be the case, so further investigation is warranted.

9 Land Surface Temperature evaluation

In this evaluation we focused on the 2004-2019 period (since this is the period with overlap with MSG LST data (2004-NRT)) and on the monthly means of the mean diurnal cycle of LST in clear sky conditions, computed independently for the LSA SAF LST (hereafter LSA), ERA5 reanalysis (Hersbach *et al.*, 2020) and the JULES long simulations PRIOR (before the data assimilation) and POST (after the data assimilation). For ERA5, PRIOR, and POST, clear-sky filtering was applied using the ERA5 total cloud cover threshold below 0.3. To characterise the diurnal cycle, which reflects the partitioning of available energy, we used a metric linked to its amplitude. For each month and grid point, we first computed the mean diurnal cycle from hourly data. We then calculated the difference between the monthly diurnal maximum and the monthly mean, which we refer to as the monthly diurnal cycle amplitude. In the following results, we also present results based on the diurnal maximum and minimum.

Figure 13 presents the mean differences in the diurnal, maximum, mean, and minimum LST for June–August 2004–2019 between the different model simulations, the reanalysis,

and the satellite product. June–August was selected because it shows the strongest signals, although the main conclusions also hold for March–April and September–November (not shown). ERA5 has a cold bias in the diurnal maximum and a warm bias in the diurnal minimum, leading to an underestimation of the diurnal amplitude. The JULES PRIOR simulation does not display such a consistent cold/warm bias across the domain; instead, the daily minimum shows a clear contrast, with a cold bias over Mediterranean regions and a warm bias over central and northern Europe. The mean differences between POST and PRIOR vary spatially for the daily maximum, while the daily minimum shows a systematic warming in POST. When analysing the differences in absolute errors, $|POST-LSA| - |PRIOR-LSA|$ positive values indicate an overall error increase in POST relative to PRIOR, whereas negative values indicate an error reduction, using LSA as reference. The results show a largely neutral impact for the daily mean, a patchy pattern of positive and negative values for the daily maximum, and a clear improvement in POST for the daily minimum over the Mediterranean region, with some deterioration over central and northern Europe. This latter pattern reflects the general warming of the daily minimum in POST compared with PRIOR.

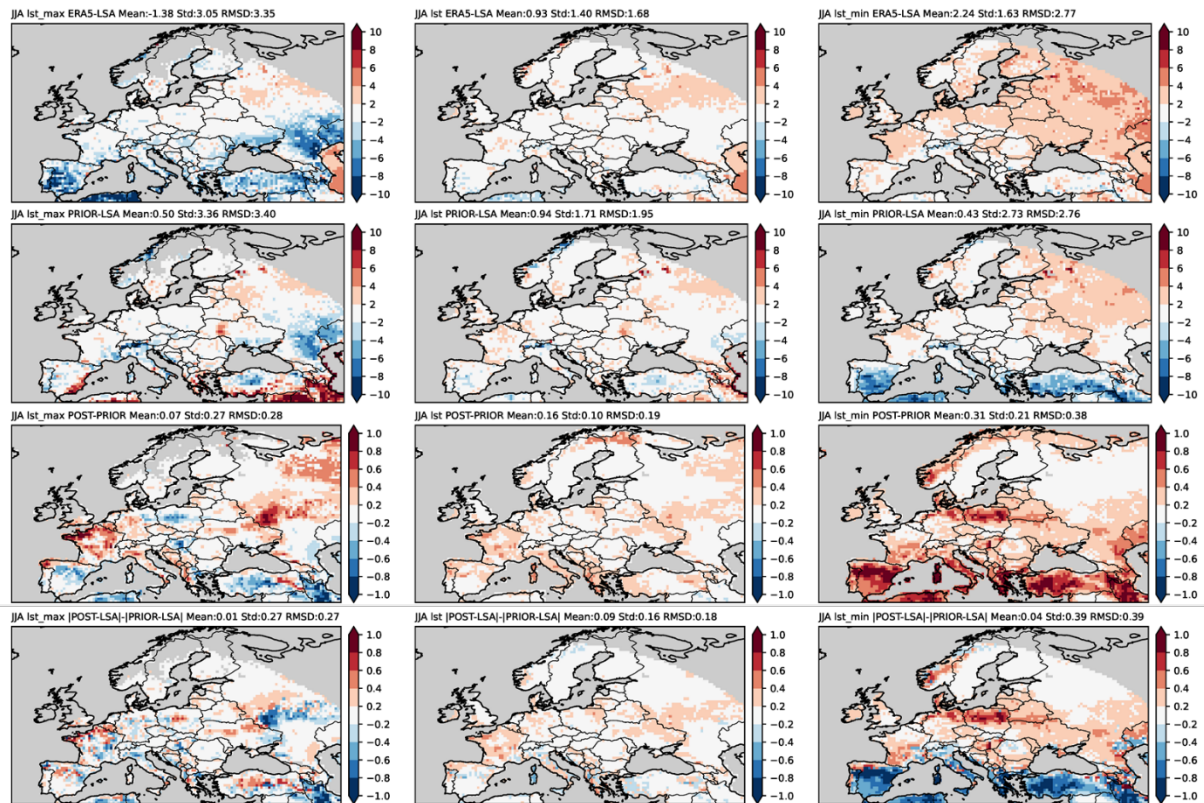


Figure 13: Mean differences of land surface temperature daily maximum (*lst_max*, left panels), daily mean (*lst*, middle panels) and daily minimum (right panels) for the months June–August (2004–2019) between: 1st row ERA5-LSA, 2nd row PRIOR-LSA, 3rd row POST-PRIOR, and 4th row $|POST-LSA| - |PRIOR-LSA|$. The results are only valid in clear-sky conditions, and areas masked as grey have less than 10% of valid data.

The interannual variability of LST in the simulations and in ERA5 was compared with LSA by first computing anomalies for each dataset independently (removing the mean annual cycle for 2004–2019) and then analysing the standard deviation of the monthly anomalies (Figure 14) and the temporal correlation between the model simulations and LSA (Figure 15). For these metrics, which focus on interannual variability, the differences between PRIOR and POST are negligible, and therefore only the PRIOR results are shown. The interannual variability of the LST anomalies is highest for the diurnal maximum, followed by the daily minimum, and lowest for the diurnal amplitude—consistently across LSA, ERA5, and PRIOR (Figure 14). We therefore focus on the diurnal amplitude as a potential proxy for how land surface conditions (e.g., available soil moisture, vegetation state) influence the surface energy budget and are subsequently reflected in LST. ERA5 shows a systematic underestimation of the interannual variability of the LST amplitude, whereas PRIOR is in closer agreement with LSA. Despite these differences in variability, the temporal correlations of the LST anomalies in ERA5 and PRIOR with LSA are similar, with a slight tendency for PRIOR to show lower agreement in the southern part of the domain—coinciding with the regions most affected by the data assimilation, which produced a warming of the diurnal minimum.

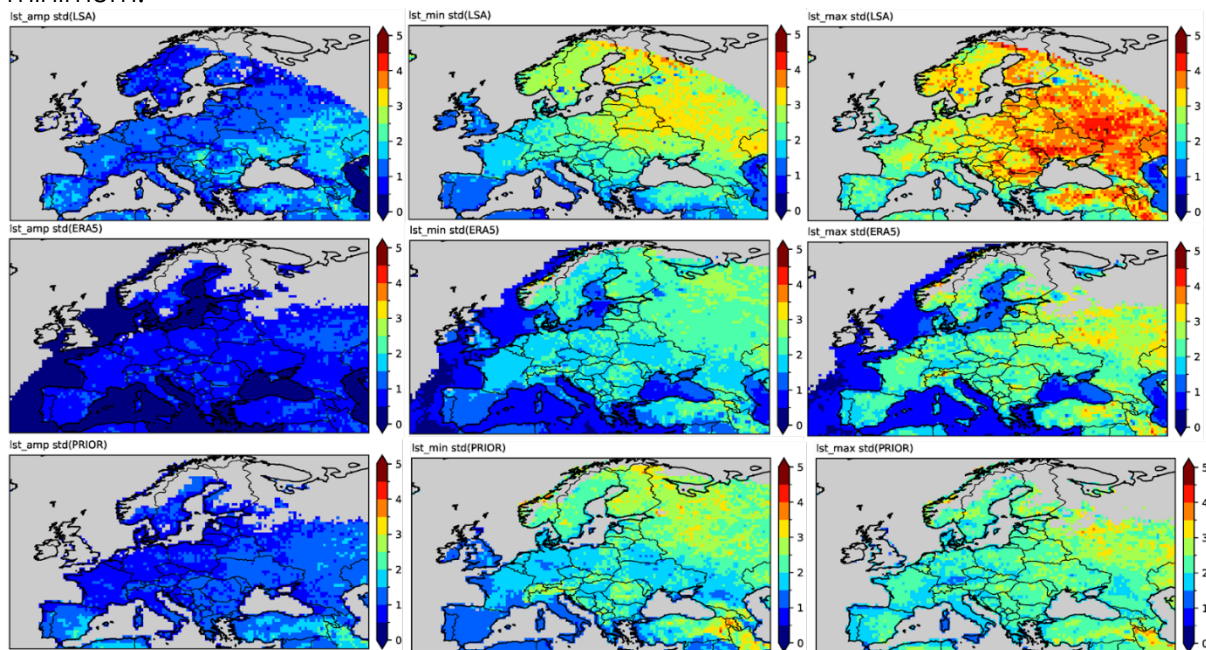


Figure 14: Standard deviation of monthly anomaly time series (removing the mean annual cycle, 2004–2019) of land surface temperature diurnal amplitude (left panel), daily minimum (middle panels) and daily maximum (right panels) in LSA product (top panels), ERA5 (centre panels) and PRIOR (bottom panels). Pixels are shown in grey when the sample size is smaller than 36 (minimum of 3 years)

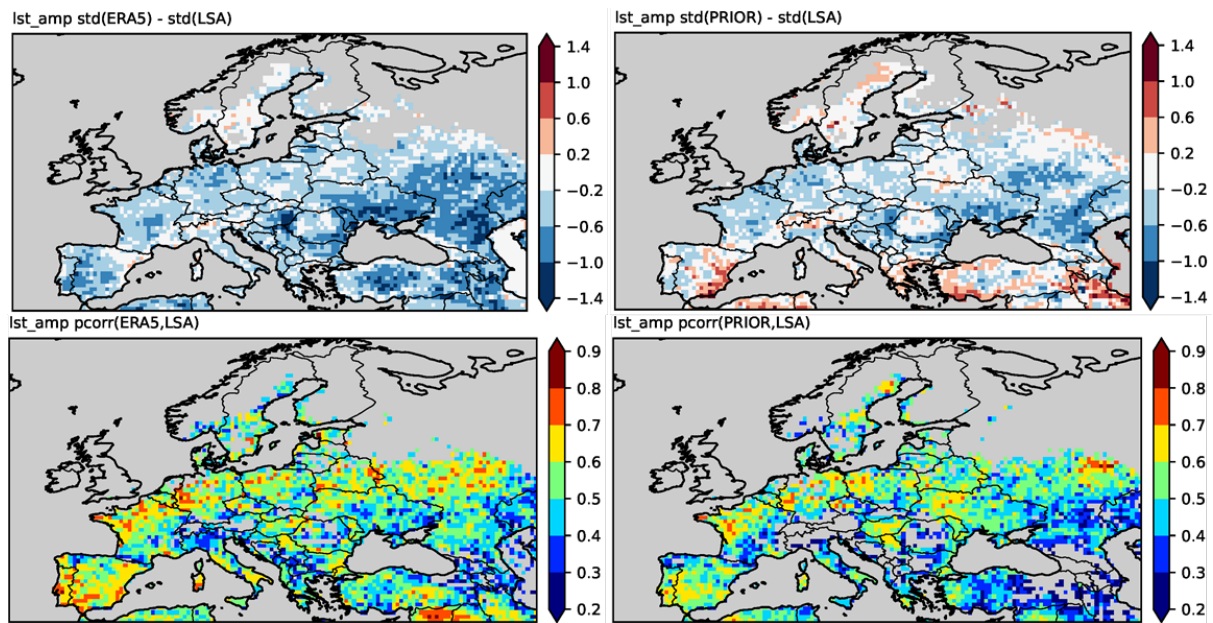


Figure 15: Top panels: Difference of standard deviation of monthly anomaly time series (removing the mean annual cycle, 2004-2019) of land surface temperature diurnal amplitude ERA5-LSA (top left) and PRIOR-LSA (top right). Bottom panels: Pearson correlation coefficient of the monthly anomaly time series for the land surface temperature amplitude between ERA5vsLSA (bottom left) and PRIORvsLSA (bottom right). Grey pixels in all panels are masked when the sample size is smaller than 36 (3 years) and in the bottom panels also when the correlations are not significant ($p > 0.01$).

10 Conclusions

We have implemented the Pinnington *et al.* (2021) data assimilation (DA) technique that was previously limited to SE England for the broader European region.

Results when running the DA on all of Europe show that the Tóth PTF soil parameters are adjusted significantly from their prior values, but not unreasonably so (see Figure 1). In order to accomplish this, we did need to revert Pinnington's 15th PTF parameter back to the original value defined by Tóth *et al.* (2015). The European spatial average of the soil-moisture RMSE (between the model and the satellite observations) is reduced after the DA (Figure 2, Figure 3), and this reduction of the spatial average RMSE stays positive and persists for all years (between 1980-2019) outside the DA window (of 2014-2018).

Further results, when focusing on the Central European drought year of 2003, demonstrate that after the DA:

- c) most areas of Europe have drier soils after the DA (see Figure 4, where the European average is a reduction of soil moisture by $-24.5\% \pm 9.4\%$);
- d) the annual standardized soil-moisture index is slightly more indicative of a drought in central Europe (Figure 5);
- e) the "Week of max wetting" is not early in the year for northern Germany, as some outlier weeks of the weekly change in soil moisture have been eliminated so they are not outliers (Figure 6);
- f) the "Drying during week of max drying" is not as extreme, as some outlier weeks (e.g., in northern Germany) have been tempered so they are not outliers (Figure 7);
- g) the Hindcast Increments of NPP are substantially negative for much of Central Europe in 2003, which may indicate the presence of an agricultural drought that year (Figure 8);
- h) the Hindcast Increments of NPP are also substantially negative over regions of Eastern Europe and Western Russia for the years of 1999 and 2002, which may indicate the presence of agricultural droughts for those two years in these regions (Figure 8);
- i) there is a spatial dichotomy of carbon use efficiency with higher values in Western Europe than in Western Russia (Figure 9);
- j) this spatial dichotomy of modelled CUE is also indicated in an observational MODIS analysis of CUE for 2000-2003 (Figure 10) by Zhang *et al.* (2009);
- k) the Analysis Increment of water use efficiency is between -15% and -5% for much of Europe for latitudes $< 55^{\circ}\text{N}$ for 2003, indicating that the prior model's WUE was adjusted downward moderately by the DA for that year (Figure 11); and
- l) the Analysis Increment of carbon use efficiency is roughly 0% for much of Western Europe. However, it varies widely for northern Russia and northern Scandinavia, being caused by the Analysis Increment of GPP crossing over to positive values in these regions (Figure 12).

Overall, this DA technique performed rather well over Europe, and the output data can be usefully analysed more thoroughly for the remainder of the EXPECT project (i.e. in upcoming work for WP5 and WP6) and beyond. This DA technique could also be reasonably and efficiently be extended to other regions of the world as well as to global models.

10.1 Conclusions from evaluation with LST

We evaluate over the period (2004–2019) using monthly mean diurnal cycles from LSA, ERA5, and JULES PRIOR/POST, with consistent clear-sky filtering and a focus on the diurnal amplitude as a potential proxy of how land surface conditions modulate the surface energy budget. We conclude:

- m) from the Mean differences plot (Figure 13) that: ERA5 underestimates the diurnal amplitude (cold daytime / warm nighttime bias), while JULES PRIOR exhibits region-dependent biases; POST warms the daily minimum, improving Mediterranean performance but degrading central/northern Europe;
- n) that the Interannual variability analysis (Figure 14, Figure 15) shows: the strongest variability in the diurnal maximum and the weakest in the diurnal amplitude; ERA5 underestimates amplitude variability, whereas PRIOR aligns more closely with LSA; and
- o) that the Temporal correlations with LSA are similar for ERA5 and PRIOR, with slightly lower agreement for PRIOR in southern regions affected by data assimilation impacts on nighttime temperatures.

10.2 Links to other parts of EXPECT

The data we have produced will be used by Task 3.1. The long time series of modelled soil moisture, tuned to satellite observations will help to identify land surface drivers of decadal changes in boreal summer atmospheric circulation. We also anticipate potential applications in Tasks 4.1, 4.2, and 5.2.

11 References

M.J. Best, M. Pryor, D.B. Clark, G.G. Rooney, R.L.H. Essery, C.B. Ménard, J.M. Edwards, M.A. Hendry, A. Porson, N. Gedney, L.M. Mercado, S. Sitch, E. Blyth, O. Boucher, P.M. Cox, C.S.B. Grimmond, and R.J. Harding. ["The Joint UK Land Environment Simulator \(JULES\), model description – Part 1: Energy and water fluxes"](https://doi.org/10.5194/gmd-4-677-2011), *Geoscientific Model Development*, 4, pp. 677-699 (2011).
<https://doi.org/10.5194/gmd-4-677-2011>

G. Chagnaud, C.M. Taylor, L.S. Jackson, C.E. Birch, J.H. Marsham, and C. Klein. ["Wet-bulb temperature extremes locally amplified by wet soils"](https://doi.org/10.5194/ggl-52-8-e2024-gl112467), *Geophysical Research Letters*, 52(8), e2024GL112467 (2025).

<https://doi.org/10.1029/2024GL112467>

D.B. Clark, L.M. Mercado, S. Sitch, C.D. Jones, N. Gedney, M.J. Best, M. Pryor, G.G. Rooney, R.L.H. Essery, E. Blyth, O. Boucher, R.J. Harding, C. Huntingford, and P.M. Cox. [“The Joint UK Land Environment Simulator \(JULES\), Model Description – Part 2: Carbon Fluxes and Vegetation Dynamics”](#), *Geoscientific Model Development*, 4, 701-722 (2011).
<https://doi.org/10.5194/gmd-4-701-2011>

S.K. Chan, R. Bindlish, P.E. O'Neill, E. Njoku, T. Jackson, A. Colliander, F. Chen, M. Burgin, S. Dunbar, J. Piepmeier, S. Yueh, D. Entekhabi, M.H. Cosh, T. Caldwell, J. Walker, X. Wu, A. Berg, T. Rowlandson, A. Pacheco, H. McNairn, M. Thibeault, J. Martínez-Fernández, A. González-Zamora, M. Seyfried, D. Bosch, P. Starks, D. Goodrich, J. Prueger, M. Palecki, E.E. Small, and M. Zreda, J.-C. Calvet, W.T. Crow, and Y. Kerr. [“Assessment of the SMAP Passive Soil Moisture product”](#), *IEEE Transactions on Geoscience and Remote Sensing*, 54, pp. 4994-5007 (2016).
<https://doi.org/10.1109/TGRS.2016.2561938>

M. Cucchi, G.P. Weedon, A. Amici, N. Bellouin, S. Lange, H. Müller Schmied, H. Hersbach, and C. Buontempo. [“WFDE5: bias-adjusted ERA5 reanalysis data for impact studies”](#), *Earth System Science Data*, 12, pp. 2097–2120 (2020).
<https://essd.copernicus.org/articles/12/2097/2020/>

W.A. Dorigo, A. Gruber, R.A.M. De Jeu, W. Wagner, T. Stacke, A. Loew, C. Albergel, L. Brocca, D. Chung, R.M. Parinussa, and R. Kidd. [“Evaluation of the ESA CCI soil moisture product using ground-based observations”](#), *Remote Sensing of Environment*, 162, pp. 380-395 (2015).
<https://doi.org/10.1016/j.rse.2014.07.023>

W. Dorigo, W. Wagner, C. Albergel, F. Albrecht, G. Balsamo, L. Brocca, D. Chung, M. Ertl, M. Forkel, A. Gruber, E. Haas, P.D. Hamer, M. Hirschi, J. Ilkonen, R. de Jeu, R. Kidd, W. Lahoz, Y.Y. Liu, D. Miralles, T. Mistelbauer, N. Nicolai-Shaw, R. Parinussa, C. Pratola, C. Reimer, R. van der Schalie, S.I. Seneviratne, T. Smolander, and P. Lecomte, [“ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions”](#), *Remote Sensing of Environment*, 203, pp. 185-215 (2017).
<https://doi.org/10.1016/j.rse.2017.07.001>

S.C. Freitas, I.F. Trigo, J.M. Bioucas-Dias, and F.-M. Gottsche. [“Quantifying the uncertainty of land surface temperature retrievals from SEVIRI/Meteosat”](#), *IEEE Transactions on Geoscience and Remote Sensing*, 48(1), pp. 523-534 (2010).
<https://doi.org/10.1109/tgrs.2009.2027697>

P. Friedlingstein, M. O'Sullivan, M.W. Jones, R.M. Andrew, J. Hauck, P. Landschützer, C. Le Quéré, H. Li, I.T. Lujikx, A. Olsen, G.P. Peters, W. Peters, J. Pongratz, C. Schwingshackl, S. Sitch, J.G. Canadell, P. Ciais, R.B. Jackson, S.R. Alin, A. Arneth,

V. Arora, N.R. Bates, M. Becker, N. Bellouin, C.F. Berghoff, H.C. Bittig, L. Bopp, P. Cadule, K. Campbell, M.A. Chamberlain, N. Chandra, F. Chevallier, L.P. Chini, T. Colligan, J. Decayeux, L. Djeutchouang, X. Dou, C. Duran Rojas, K. Enyo, W. Evans, A. Fay, R.A. Feely, D.J. Ford, A. Foster, T. Gasser, M. Gehlen, T. Gkritzalis, G. Grassi, L. Gregor, N. Gruber, Ö. Gürses, Ian Harris, M. Hefner, J. Heinke, G.C. Hurtt, Y. Iida, T. Ilyina, A.R. Jacobson, A. Jain, T. Jarníková, A. Jersild, F. Jiang, Z. Jin, E. Kato, R.F. Keeling, K. Klein Goldewijk, J. Knauer, J. Ivar Korsbakken, S.K. Lauvset, N. Lefèvre, Z. Liu, J. Liu, L. Ma, S. Maksyutov, G. Marland, N. Mayot, P.C. McGuire, N. Metz, N.M. Monacchi, E.J. Morgan, S.-I. Nakaoka, C. Neill, Y. Niwa, T. Nützel, L. Olivier, T. Ono, P.I. Palmer, D. Pierrot, Z. Qin, L. Resplandy, A. Roobaert, T.M. Rosan, C. Rödenbeck, J. Schwinger, T.L. Smallman, S. Smith, R. Sospedra-Alfonso, T. Steinhoff, Q. Sun, A.J. Sutton, R. Séférian, S. Takao, H. Tatebe, H. Tian, B. Tilbrook, O. Torres, E. Tourigny, H. Tsujino, F. Tubiello, G. van der Werf, R. Wanninkhof, X. Wang, D. Yang, X. Yang, Z. Yu, W. Yuan, X. Yue, S. Zaehle, N. Zeng, and J. Zeng, *"ESSD, Earth System Science Data*, 17, pp. 965–1039 (2025). <https://essd.copernicus.org/preprints/essd-2024-519/>

L. Fróis, I.F. Trigo, W. Dorigo, P. de Rosnay, and E. Dutra. *"A satellite-based assessment of ERA5 soil moisture and land surface temperature interannual co-variability"*, *Environmental Research Letters*, 21(2), 024024 (2026). <https://doi.org/10.1088/1748-9326/ae3789>

M. Th. van Genuchten, *"A closed-form equation for predicting the hydraulic conductivity of unsaturated soils"*, *Soil Science Society of America Journal*, 44.5, pp. 892-898 (1980). <https://doi.org/10.2136/sssaj1980.03615995004400050002x>

T.R. Gurung and L. Chen. *"Understanding the influence of soil moisture on heatwave characteristics in the contiguous United States"*. *Environmental Research Letters*, 19(6), 064070 (2024). <https://doi.org/10.1088/1748-9326/ad4dbb>

T. Hengl, J. Mendes de Jesus, G. B. M. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagočić, W. Shangguan, M.N. Wright, X. Geng, B. Bauer-Marschallinger, M.A. Guevara, R. Vargas, R.A. MacMillan, N.H. Batjes, J.G.B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, and B. Kempen. *"SoilGrids250m: Global gridded soil information based on machine learning"*, *PLOS ONE*, 12, pp. 1-40 (2017). <https://doi.org/10.1371/journal.pone.0169748>

H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R.J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, J.-N. Thépaut. *"The ERA5 global reanalysis"*. *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp. 1999-

2049 (2020).

<https://doi.org/10.1002/qj.3803>

F. Johannsen, S. Ermida, J.P.A. Martins, I.F. Trigo, M. Nogueira, and E. Dutra. "[Cold bias of ERA5 summertime daily maximum land surface temperature over Iberian Peninsula](#)". *Remote Sensing*, 11(21), 2570 (2019).

<https://doi.org/10.3390/rs11212570>

R.I. Maidment, T. Quaife, E. Pinnington, E. Black, A. Ejigu, D. Kumar, and S. Thomas. "[A new, long-term root zone soil moisture dataset for operational agricultural drought monitoring over Africa](#)", *Scientific Data* (2026, in press).

<https://doi.org/10.1038/s41597-026-06585-w>

D. Maraun, R. Schiemann, A. Ossó, and M. Jury. "[Changes in event soil moisture-temperature coupling can intensify very extreme heat beyond expectations](#)". *Nature Communications*, 16, 734 (2025).

<https://doi.org/10.1038/s41467-025-56109-0>

M. Nogueira, S. Boussetta, G. Balsamo, C. Albergel, I.F. Trigo, F. Johannsen, D.G. Miralles, and E. Dutra. "[Upgrading land-cover and vegetation seasonality in the ECMWF coupled system: Verification with FLUXNET sites, Meteosat satellite land surface temperatures, and ERA5 atmospheric reanalysis](#)", *Journal of Geophysical Research. Atmospheres*, 126(15), e2020JD034163 (2021).

<https://doi.org/10.1029/2020JD034163>

H. Oliver, M. Shin, D. Matthews, O. Sanders, S. Bartholomew, A. Clark, B. Fitzpatrick, R. van Haren, R. Hut, and N. Dros. "[Workflow Automation for Cycling Systems The Cylc Workflow Engine](#)", *Computing in Science and Engineering*, 21, pp. 7-21 (2019)

<https://doi.org/10.1109/MCSE.2019.2906593>

E. Pinnington, J. Amezcua, E. Cooper, S. Dadson, R. Ellis, J. Peng, E. Robinson, R. Morrison, S. Osborne, and T. Quaife. "[Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of SMAP satellite data](#)", *Hydrology and Earth System Sciences*, 25(3), pp.1617-1641 (2021).

<https://doi.org/10.5194/hess-25-1617-2021>

N.C. Privé and R.M. Errico. "[The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment](#)", *Tellus A: Dynamic Meteorology and Oceanography*, 65, 21740 (2013).

<https://doi.org/10.3402/tellusa.v65i0.21740>

R. Reichle, G. De Lannoy, R.D. Koster, W.T. Crow, J.S. Kimball, Q. Liu, and M. Bechtold (2022). "[SMAP L4 global 3-hourly 9 km EASE-Grid surface and root zone soil moisture analysis update](#)", (SPL4\$MAU, Version 7). Boulder, Colorado USA. NASA National

Snow and Ice Data Center Distributed Active Archive Center.

<https://doi.org/10.5067/LWJ6TF5SZRG3>

S.I. Seneviratne, T. Corti, E.L. Davin, M. Hirschi, E.B. Jaeger, I. Lehner, B. Orlowsky, and A.J. Teuling. ["Investigating soil moisture–climate interactions in a changing climate: A review"](#), *Earth-Science Reviews*, 99(3-4), pp.125-161 (2010).

<https://doi.org/10.1016/j.earscirev.2010.02.004>

B. Tóth, M. Weynants, A. Nemes, A. Makó, G. Bilas, and G. Tóth. ["New generation of hydraulic pedotransfer functions for Europe"](#), *European Journal of Soil Science*, 66, pp. 226-238 (2015).

<https://doi.org/10.1111/ejss.12192>

I.F. Trigo, C.C. Dacamara, P. Viterbo, J.-L. Roujean, F. Olesen, C. Barroso, F. Camacho-de-Coca, D. Carrer, S.C. Freitas, J. García-Haro, B. Geiger, F. Gellens-Meulenberghs, N. Ghilain, J. Meliá, L. Pessanha, N. Siljamo, and A. Arboleda. ["The Satellite Application Facility for Land Surface Analysis"](#), *International Journal of Remote Sensing*, 32, pp. 2725-2744 (2011).

<https://doi.org/10.1080/01431161003743199>

K.P. Tripathy, A.K. Mishra, S.V. Kumar, and D.G. Miralles. ["Lagged soil moisture controls on the persistence of drought and heatwaves in the United States"](#). *Geophysical Research Letters*, 52(22), e2025GL115811 (2025).

<https://doi.org/10.1029/2025GL115811>

N. Viovy. ["CRUNCEP Version 7 - Atmospheric Forcing Data for the Community Land Model."](#) NSF National Center for Atmospheric Research. (2018).

<https://doi.org/10.5065/PZ8F-F017>.

D. Walters, A.J. Baran, I. Boutle, M. Brooks, P. Earnshaw, J. Edwards, K. Furtado, P. Hill, A. Lock, J. Manners, C. Morcrette, J. Mulcahy, C. Sanchez, C. Smith, R. Stratton, W. Tennant, L. Tomassini, K. Van Weverberg, S. Vosper, M. Willett, J. Browse, A. Bushell, K. Carslaw, M. Dalvi, R. Essery, N. Gedney, S. Hardiman, B. Johnson, C. Johnson, A. Jones, C. Jones, G. Mann, S. Milton, H. Rumbold, A. Sellar, M. Ujiie, M. Whittall, K. Williams, and M. Zerroukat. ["The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations"](#), *Geoscientific Model Development*, 12, pp. 1909-1963 (2019).

<https://doi.org/10.5194/gmd-12-1909-2019>

A.J. Wiltshire, M.C. Duran Rojas, J.M. Edwards, N. Gedney, A.B. Harper, A.J. Hartley, M.A. Hendry, E. Robertson, and K. Smout-Day. ["JULES-GL7: the Global Land configuration of the Joint UK Land Environment Simulator version 7.0 and 7.2"](#), *Geoscientific Model Development*, 13, pp. 483-505 (2020).

<https://doi.org/10.5194/gmd-13-483-2020>

Y. Zhang, M. Xu, H. Chen, and J. Adams. ["Global pattern of NPP to GPP ratio derived from MODIS data: effects of ecosystem type, geographical location and climate"](https://doi.org/10.1111/j.1466-8238.2008.00442.x), *Global Ecology and Biogeography*, 18, pp. 280-290 (2009)
<https://doi.org/10.1111/j.1466-8238.2008.00442.x>