Capability of the variogram to quantify the spatial patterns of surface fluxes and soil moisture simulated by land surface models

Article

Accepted Version

Garrigues, S., Verhoef, A., Blyth, E., Wright, A., Balan-Sarojini, B., Robinson, E. L., Dadson, S., Boone, A., Boussetta, S. and Balsamo, G. (2021) Capability of the variogram to quantify the spatial patterns of surface fluxes and soil moisture simulated by land surface models. Progress in Physical Geography, 45 (2). pp. 279-293. ISSN 1477-0296 doi: https://doi.org/10.1177/0309133320986147 Available at http://centaur.reading.ac.uk/95949/

Progress in
Physical Geography

Capability of the variogram to quantify the spatial patterns of surface fluxes and soil moisture simulated by land surface models.

Journal:	*Progress in Physical Geography*
Manuscript ID	PPG-20-145.R1
Manuscript Type:	Progress Report
Keywords:	Variogram analysis, Land surface model, Surface flux, spatial pattern, spatial variability, spatial structure, spatial structure of accuracy
Abstract:	Up to now, relatively little effort has been dedicated to the quantitative assessment of the differences in spatial patterns of model outputs. In this paper, we employed a variogram-based methodology to quantify the differences in the spatial patterns of root-zone soil moisture, net radiation, and latent and sensible heat fluxes simulated by three land surface models (SURFEX/ISBA, JULES and CHTESSEL) over three European geographic domains, namely United-Kingdom, France and Spain. The model output spatial patterns were quantified through two metrics derived from the variogram: i) the variogram sill which quantifies the degree of spatial variability of the data and ii) the variogram integral range which represents the spatial length scale of the data. The higher seasonal variation of the spatial variability of sensible and latent heat fluxes over France and Spain compared to UK is related to a more frequent occurrence of a soil-moisture limited evapotranspiration regime during summer dry spells in the South of France and Spain. The small differences in spatial variability of net radiation between models indicates that the spatial patterns of net radiation are mostly driven by the climate forcing data set. However, the models exhibit larger differences in latent and sensible heat flux spatial variabilities, which are related to their differences in i) soil and vegetation ancillary datasets and ii) physical process representation. The highest discrepancies in spatial patterns between models are observed for soil moisture which is mainly related to the type of soil hydraulic function implemented in the models. This work demonstrates the capability of the variogram to enhance our understanding of the spatiotemporal structure of the uncertainties in land surface model outputs. We therefore strongly encourage the implementation of the variogram metrics in model intercomparison exercises.

Abstract

Up to now, relatively little effort has been dedicated to the quantitative assessment of the differences in spatial patterns of model outputs. In this paper, we employed a variogram-based methodology to quantify the differences in the spatial patterns of root-zone soil moisture, net radiation, and latent and sensible heat fluxes simulated by three land surface models (SURFEX/ISBA, JULES and CHTESSEL) over three European geographic domains, namely United-Kingdom, France and Spain. The model output spatial patterns were quantified through two metrics derived from the variogram: i) the variogram sill which quantifies the degree of spatial variability of the data and ii) the variogram integral range which represents the spatial length scale of the data.

The higher seasonal variation of the spatial variability of sensible and latent heat fluxes over France and Spain compared to UK is related to a more frequent occurrence of a soil-moisture limited evapotranspiration regime during summer dry spells in the South of France and Spain. The small differences in spatial variability of net radiation between models indicates that the spatial patterns of net radiation are mostly driven by the climate forcing data set. However, the models exhibit larger differences in latent and sensible heat flux spatial variabilities, which are related to their differences in i) soil and vegetation ancillary datasets and ii) physical process representation. The highest discrepancies in spatial patterns between models are observed for soil moisture which is mainly related to the type of soil hydraulic function implemented in the models.

This work demonstrates the capability of the variogram to enhance our understanding of the spatiotemporal structure of the uncertainties in land surface model outputs. We therefore strongly encourage the implementation of the variogram metrics in model intercomparison exercises.

Keywords: spatial variability, spatial structure, spatial pattern, variogram analysis, land surface model, surface fluxes, soil moisture.

1 Introduction

Accurate simulation of surface fluxes, such as evapotranspiration (ET), is essential to reduce the uncertainties in predictions of the long-term evolution of the terrestrial water cycle and the occurrence of hydro-meteorological hazards (Seneviratne et al. 2010; Vilasa et al., 2017). The different ways in which the control of soil moisture on ET dynamics is modelled has been identified as a major source of discrepancies among land surface model (LSM) predictions (Seneviratne et al. 2010). Various experiments have been designed to compare multiple LSMs and evaluate them against observations at both regional (Henderson-Sellers et al. 1993; Boone et al. 2009) and global scales (Dirmeyer, 2011, Koster et al., 2006). Important efforts have been dedicated to developing tools to evaluate and compare model performances at various spatial and temporal scales using ground data sets (e.g. FLUXNET) and satellite observations (Kumar et al., 2012; Luo et al., 2012; Best et al., 2015; Eyring et al., 2016).

Intercomparison of LSMs generally rely on first-order statistical metrics such as accuracy metrics (RMSE, bias), density function overlap with observations, model against observation scatterplots or Taylor diagrams, to quantify the discrepancies between models. Some process-based metrics have been developed to investigate a specific feature of the model such as the dry-down dynamics (Ukkola et al. 2016; Harris et al., 2017) or the bi-modal distribution of soil moisture (Vilasa et al., 2017).

While most studies have investigated the uncertainties in the temporal dynamics of surface fluxes, using the metrics mentioned above, far fewer efforts have been dedicated to quantify the uncertainties in their spatial dynamics (Harris et al., 2017). The evaluation of the differences in spatial patterns simulated by land surface models has been limited to qualitative assessment of spatial maps of key model output fluxes or variables, that provides little insight into the spatiotemporal structures of the model output uncertainties.

The analysis of the spatial structures displayed by geospatial data sets brings insight on the nature and the scale of spatial variation of Earth surface processes. Garrigues et al., 2007 employed stochastic models to characterize the nature of the spatial structures displayed by satellite imagery and to identify the underlying processes structuring distinct types of landscape. Besides, Garrigues et al., 2008 showed that the seasonal evolution of cropland spatial structures observed in satellite imagery can be related to vegetation dynamic

processes and agricultural practices. Tarnavsky et al.,2008 employed variogram analysis to assess the differences in spatial information content across multiple satellite-based vegetation maps and to identify the spatial resolution at which the differences in spatial information content between satellite data sets is reduced. Similarly, Pontius et al, 2008 showed that the analysis of the variations of the spatial structures of maps as a function of spatial resolutions brings insights on the differences in spatial information content between geospatial data sets. Pontius and Malanson, 2005 used this approach to assess the differences in accuracy between distinct land use models.

First-order statistics are not sufficient to reliably characterize the spatial structures of model outputs since they do not account for spatial correlation (Julesz, 1962). Second-order spatial statistic metrics are required to resolve the differences in the spatial patterns of geospatial data. A review of second-order spatial metrics applied to remote sensing imagery is given in Garrigues et al. (2006). These authors highlighted that the variogram is a robust tool to measure the spatial patterns of uncertainties in satellite observations. They exploited the variogram to quantify: (i) the spatial variability of the data over the area of interest and (ii) the spatial structures of the data, i.e. patches of similar outputs that repeat themselves independently within the studied region at a characteristic spatial length scale. Furthermore, the implementation of variogram is more straightforward than other metrics, such as wavelet decomposition (Csillag and Kabos, 2002).

In this paper, the variogram analysis methodology is applied to surface energy balance fluxes and root-zone soil moisture content simulated by three LSMs, at the regional scale, over three European domains associated with contrasted climate. The models are implemented using their standard configuration along with their standard soil and vegetation parameters. This progress report aims to demonstrate the capability of the variogram analysis to quantify the differences in the spatial patterns of surface fluxes and soil moisture, as predicted by distinct land surface models. The results of the variogram analysis are used to discuss the drivers of the differences in spatial patterns of model outputs.

2 Models and simulation design

2.1 Model description

Three LSMs are used in this work:

- SURFEX/ISBA (the Interactions between Soil, Biosphere, and Atmosphere; Noilhan and Planton, 1989) v8.0 developed at Meteo France;
- JULES (the Joint UK Land Environment Simulator; Best et al., 2011) configured as GL6 and based on the code version 4.4, developed by the Met Office;
- CHTESSEL, (most recent version of the Tiled ECMWF Scheme for Surface Exchanges Over Land; Basalmo et al. (2009)) employed at the European Centre for Medium-Range Weather Forecasts (ECMWF).

Table 1 provides the main characteristics of these models. These models are driven by a climatology of Leaf Area Index and vegetation height. Their spatial integration is achieved using their default soil and vegetation ancillary data sets and land cover map. All models calculate a single source energy balance and use a multi-layer diffusion scheme to simulate water and heat transfer in the soil. While SURFEX and JULES represent the functional coupling between the stomatal conductance (g_s) and the net assimilation of CO_2 (photosynthesis, A), CHTESSEL relies on an empirical parametric formulation of the stomatal conductance.

2.2 Experiment design

The simulations were conducted from 1994 to 2012 over a 0.5° grid over Europe and at a 3-hourly time step, using the WFDEI (WATCH Forcing Data methodology applied to ERA-Interim data, Weedon et al., 2014) atmospheric reanalysis. Two distinct simulations were conducted for JULES using distinct soil hydraulic parametrizations: one is based on the van Genuchten, 1980 scheme ($JULES_{VG}$) and one relies on the Clapp and Hornberger, 1978 scheme ($JULES_{CH}$). We used the 1979-1993 period as a spin up period and the model's outputs were analyzed from 1994 to 2012.

The spatial patterns of the model outputs were evaluated over three European geographic domains,

namely United-Kingdom, France and Spain, which were selected to represent contrasted climate, soil and vegetation characteristics (Table 2). The variogram analysis is applied to the latent heat (LE), sensible heat (H), net radiation (RN) fluxes, and the root-zone liquid soil moisture content (θ) variables. Three-hourly model outputs are first averaged over 10-days. Next, the 10-day values are averaged over the entire simulation period (1994-2012), to represent a mean climatological annual timeseries, i.e. thirty-six 10-day values. The variograms of the mean climatological values of each flux and variable are then computed.

3 Geostatistical methodology to quantify the spatial patterns of model outputs

The variogram analysis consists of two steps: i) computing the empirical variogram over the domain of interest and ii) fitting a theoretical variogram model to the empirical variogram to retrieve the spatial variability and the size of the spatial structures of the data.

3.1 Empirical variogram

The empirical variogram $\gamma^e(h)$ (Eq. 1) measures the dissimilarity between the model outputs $z(x_i)$ and $z(x_j)$ taken at two distinct grid cells x_i and x_j separated by a distance h (Eq.(1)). To provide statistically meaningful variogram values, the pairs of grid cells at a similar distance h are grouped into bins of separation distance. N(h) is the number of grid cell pairs within each bin. The variogram is computed as the average of the squared differences of the model values ($z(x_i)$ and $z(x_j)$) of all pairs of grid cells that fall within each distance bin. To properly sample the range of separation distance h between grid cells, the bin size is set to one grid cell. The variogram is computed to a maximum distance equal to half the domain size to ensure that there are enough grid cell pairs within in each distance bin (Garrigues et al. 2006).

$$\gamma^e(h) = \frac{1}{2N(h)} \sum_{x_i, x_j \| x_i - x_j \| = h} (z(x_i) - z(x_j))^2 \quad (1)$$

3.2 Variogram modelling

We assume that the model outputs ($z(x_i)$ and $z(x_j)$) are realizations of a random function Z(x). Under the second-order stationarity hypothesis (Chiles and Delfiner, 2012), which ensures the existence and the stationarity of the first and second moments of Z(x), the theoretical variogram γ(h) is defined by:

$$\gamma(h) = 0.5Var[Z(x+h) - Z(x)] = \sigma^2 - C(h) \quad (2)$$

where C(h) and σ^2 are the covariance and the variance of Z(x), respectively. γ(h) is a function starting from 0 for h = 0 and converging to the sill, σ^2 , as h tends to infinity. The range, r, of the theoretical variogram is the distance at which it reaches the sill. Data separated by a distance larger than the range are spatially uncorrelated.

The linear model of regionalization, which models the variogram as a linear combination of elementary conditionally definite negative functions $g_k(h)$, is defined by:

$$\gamma(h) = \sigma^2 \sum_{i=1}^{k} (b_k g_k(h)) \quad (3)$$

where b_k is the weight associated to each function $g_k(h)$. In an exploratory approach, distinct combinations of distinct g_k functions (spherical, exponential, Gaussian) were tested and it was found that a single exponential model was providing the best fit to our data.

Eq. (3) then becomes:

$$\gamma(h) = \sigma^2 \left[1 - exp\left(\frac{-3h}{r}\right)\right] \quad (4)$$

The parameters σ^2 and r of the exponential variogram model are estimated by least-square fitting to the experimental variogram.

The integral range derived from the variogram (A, Eq. 5) is used to quantify the spatial structures of the data. A is an area moment that represents the mean surface of the spatial structure of the variable over a given spatial domain (Chiles and Delfiner, 2012).

$$A = \frac{1}{\sigma^2}\int_{h \in \mathbb{R}^2}(\sigma^2 - \gamma(h)) \quad (5)$$

The value of A for an exponential variogram model can be analytically derived from Eq. 4 and is directly proportional to the variogram range:

$$A = \frac{2\pi r^2}{9} \quad (6)$$

3.3 Variogram modelling assumptions

The modelling of the variogram relies on the second order stationarity hypothesis which implies that the data does not exhibit any spatial trend within the geographic domain and allows to use the variogram range to quantify the length scale of the data. As demonstrated by Garrigues et al., (2006), the integral range can be used as a posteriori yardstick to judge if the size of the geographic domain is large enough to measure properly the length scale of the data with the variogram. We verified that the integral range of most of the modelled variograms are small enough compared to the surface of the geographic domain that justifies the validity of the second order stationarity hypothesis.

We chose to not include any nugget component (discontinuity at the origin) to represent unresolved sub-grid variability. We checked that including a nugget term in Eq.3 did not improve the quality of the variogram fit and does not change the estimation of the length scale of the data. This work is focused on the characterization of large-scale spatial structures and the characterization of sub-grid heterogeneity, which would require data at a finer spatial resolution, is beyond the scope of this work.

3.4 Spatial pattern analysis

The spatial patterns of the variables of interest are characterized by the square root of the variogram sill (i.e., σ), which quantifies the degree of spatial variability of the data, and by the variogram integral range (A), which quantifies the mean surface of the spatial structures (i.e patches) displayed by the data. The unit of σ is that of the variable (W m^{-2} for the surface fluxes and m^3 m^{-3} for soil moisture). The unit of A is that of the surface of a grid cell (0.25 deg^2).

The analysis involves quantifying the seasonal evolution of A and σ and assess the differences across geographic domains, variables and models. We use the Root-Mean-Square-Difference (RMSD) to quantify the differences in A and σ between model experiments.

Variograms were also computed for two key surface properties:

- the maximum water content available for the plant transpiration (MaxAWC) which is defined as the difference between the soil moisture at field capacity and the soil moisture at wilting point; this is a key driver of ET dynamics (Garrigues et al. 2015). MaxAWC is constant in time so one variogram is computed for each model and each geographic domain.
- the monthly Leaf Area Index (LAI) which represent the vegetation seasonal cycle. For each LSM, a variogram of LAI was computed over each geographic domain at a monthly timestep. In this report, we showed the results for France to illustrate the relationship between the spatial patterns of LAI and that of surface fluxes.

4 Drivers of the spatial patterns of surface fluxes and soil moisture

Impact of climate

Figure 1 shows that the seasonal variations of A and σ of net radiation are low compared to those calculated for latent and sensible heat fluxes. Net radiation is strongly driven by the climate and its spatial pattern does not have a strong seasonal signal.

Impact of surface characteristics

The lower values of A generally obtained over France indicate smaller size of spatial structures compared to UK and Spain. This is related to the smaller spatial structures observed for the soil hydraulic properties (Figure 5) and for LAI (Figure 6) over France.

Impact of vegetation dynamics

The spatial variability of sensible and latent heat fluxes exhibits strong seasonal variations and is the highest in late spring and early summer (Figure 2 and 3). This is related to the development of vegetation in spring and summer which results in a higher spatial variability of LAI as shown by Figure 7. The contrast between the response of vegetated and non-vegetated surface to the climate forcing is more important in summer which increases the spatial variability of surface fluxes. Also, evapotranspiration is more frequently limited by soil moisture during spring and summer; this contributes to the larger spatial variability in latent heat fluxes.

Impact of soil moisture stress

The larger seasonal variations of σ for sensible and latent heat fluxes observed over France and Spain compared to UK (Figure 2 and 3) is most likely related to a more frequent occurrence of the soil-moisture limited regime in the South of France and in Spain. Soil moisture stress generates large spatial contrasts in the energy partitioning between latent and sensible heat which explains the increase of the spatial variability of surface fluxes during dry spells.

5 Drivers of the differences in spatial patterns between models

Tables 3 and 4 indicate that the differences in A and σ are lower for the comparison between SURFEX and JULES compared to the comparison of these models with CHTESSEL which has the highest RMSD. The largest discrepancies between models are obtained for sensible and latent heat fluxes and root-zone soil moisture. We investigate below the main drivers of these discrepancies.

Impact of forcing data set

Figure 1 shows small differences in in A and σ of net radiation across models. This is supported by the higher RMSD of spatial variability between models, for H and LE, compared to those calculated for RN (Table 4). The main explanation is that the climate forcing data set, which contains the main drivers of net radiation, i.e., shortwave and longwave incoming radiation, is identical across experiments. Conversely latent and sensible heat fluxes exhibit larger differences in spatial patterns between experiments which arise from the differences in i) the soil and vegetation data sets implemented in each model and ii) physical process representation such as stomatal conductance parametrization.

Impact of soil hydraulic parametrization

The largest differences in σ between models are observed for the root-zone soil moisture. For France and Spain, we identified two combinations of models with similar σ values (Figure 4, Table 4):

- $JULES_{VG}$ and CHTESSEL with the highest σ values,
- $JULES_{CH}$ and SURFEX with the lowest σ values.

Conversely for UK, $JULES_{CH}$, $JULES_{VG}$ and SURFEX provide similar σ values of soil moisture while CHTESSEL shows much higher values than the rest of the models.

The simulation of surface fluxes is hugely dependent on the soil hydraulic parametrization and soil property map (Garrigues et al., 2015) which can exhibit different spatial patterns across models. SURFEX and $JULES_{CH}$ are based on the Clapp and Hornberger, 1978 model while $JULES_{VG}$ and CHTESSEL rely on the van Genuchten, 1980 model. This observation holds over France and Spain; this suggests that the impact of the type of soil hydraulic parametrization on the surface flux spatial patterns is stronger than the type of soil map used to infer the soil properties over these domains. Conversely, over the UK, the substantial differences in spatial variability between CHTESSEL and the rest of the model experiments are probably related to differences in the spatial distribution of soil

properties. This is confirmed by the variograms computed on the soil hydraulic properties (MaxAWC, Figure 5) which exhibit large discrepancies in the spatial variability of MaxAWC between models over UK.

Impact of LAI

The LAI value of each grid cell is the result of the LAI data set and the land cover map implemented in each model. While the models rely on the same satellite LAI data set, they use different land cover classifications (see Table 1) which may generate distinct spatial distributions of grid cell LAI values. This is illustrated over France where the models exhibit large discrepancies in both the LAI spatial structures (Figure 6) and the time course of the LAI spatial variability (Figure 7). Figure 7 shows that over France, SURFEX and JULES display more pronounced seasonal variations of LAI spatial variability than CHTESSEL which is concurrently observed for the latent heat flux in Figure 3. Besides, JULES shows the highest values of LAI spatial variability during the summer which concurrently results in the highest spatial variability of latent heat flux during the summer. The shift in the seasonal variation of LAI spatial variability of SURFEX and JULES observed over France (Figure 7) is also observed for the latent heat flux (Figure 3). These examples clearly show that the LAI spatial variability is a key driver of the spatial variability of the simulated surface fluxes.

6 Conclusions

This work shows that variogram modeling is a powerful tool to quantify the differences in land surface model outputs through two components:

- the degree of spatial variability quantified by the variogram sill,

- the surface of the spatial structure of the data represented by the variogram integral range.

Variogram analysis was applied to surface fluxes and soil moisture simulated by three land surface models over three European domains with contrasted climates (UK, France and Spain).

The main outcomes are

-The models show larger differences in the spatial variabilities of the turbulent heat fluxes, which are more prominently driven by i) soil and vegetation ancillary data and ii) physical process representation, compared to the net radiation spatial variability which is mainly driven by the radiative climate forcing.

- The seasonal evolution of the spatial variability of LAI and that of latent heat flux show large similarities. The increase in spatial variability of sensible and latent heat fluxes in spring and summer is related to the enhancement of the spatial variability of LAI which creates a larger spatial contrast in the surface response to the climate forcing. The higher seasonal variations of spatial variability in turbulent heat fluxes observed over France and Spain are most likely related to a more frequent occurrence of soil-moisture limited ET regime during summer dry spells in Spain and South of France.

- The highest discrepancies in spatial patterns between models are observed for soil moisture. Over France and Spain, the spatial patterns of soil moisture are primarily related to by the type of soil hydraulic function *(*van Genuchten (1980) versus Clapp, R. and G. Hornberger (1978)*)* while over UK the differences between models are related mainly to differences in soil maps that determine the parameters of the soil hydraulic functions.

This report highlights the capability of the variogram to measure the spatial patterns of land surface model outputs, to monitor the seasonal changes in spatial patterns and to quantify the differences in spatial patterns across variables, geographic domains and models. The integral range is a powerful metric to verify if the model simulations resolve similar spatial features. The variogram sill allows the detection of differences in the spatial variability across models that can be related to differences in the spatial distribution of the surface parameters used by each model. We therefore strongly encourage the implementation of the variogram metrics in model intercomparison exercises to

enhance our understanding of the spatiotemporal structure of the uncertainties associated with land surface models. Besides, the variogram analysis proposed in this work can be applied to satellite observations of surface fluxes and surface characteristics to evaluate the accuracy of the spatial patterns of surface fluxes simulated by land surface models.

A key assumption of the variogram modelling approach, used in this work, is the second-order stationarity which implies that the data does not exhibit any spatial trend and allows to use the variogram to quantify the length scale of the data. The integral range, which represents the average surface of the data spatial structure, can be used to verify as posteriori the validity of this assumption. The size of the geographic domain should be adjusted to have an integral range much smaller than the surface of the geographic domain used to compute the variogram. Besides, in this work we chose to model the variogram using a single exponential model to demonstrate the capability of the variogram metrics to characterize the model outputs spatial patterns. A first improvement could be to include additional elementary functions in the linear model of regionalization (Eq. 2) to describe multiple length scales in the data. While this can be relevant for small spatial domains, we did not find any improvements for the large domains investigated in this work. Another possible avenue for improvement is to use alternative types of covariance models. Genton and Kleiber (2015) demonstrate the flexibility of the Matern covariance function to describe correlation at both large (scale parameter) and small distances (smoothness parameter). In addition, in this work, we modeled the spatial patterns for each 10-day time step and then we observed the temporal evolution of the variogram metrics. An alternative approach could be to implement a space-time version of the linear model of regionalization to describe the spatiotemporal structure of the model outputs (De Iaco et al., 2013). Variograms can be modelled within a multivariate framework to investigate the co-variability between variables. This could be applied in further research to analyze the spatial dynamics of ET-soil moisture relationships, and to identify the discrepancies in the representation of the spatial dynamics of drought across models.

315

320

References

Abramowitz, G. (2012) Towards a public, standardized, diagnostic benchmarking system for land surface models. *Geosci. Model Dev.* 5: 819-827. https://doi.org/10.5194/gmd-5-819-2012, 2012.

Balsamo, G, P. Viterbo, A. C. M. Beljaars, B. J. J. M. van den Hurk, M. Hirschi, A. K. Betts, and K. Scipal, (2009) A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the ECMWF-IFS. *J. Hydrometeor.* 10: 623–643.

Balsamo, G, P. S. Boussetta, E. Dutra, A. C. M. Beljaars, P. Viterbo and B. J. J. M. Van de Hurk, (2011) Evolution of land surface processes in the IFS. ECMWF Newsletter, No. 127, ECMWF, Reading, United Kingdom, 17–22

Best, M.J., and co-authors (2011) The Joint UK Land Environment Simulator (JULES), Model description – Part 1: Energy and water fluxes. *Geosci. Model Dev. Discuss*. 4: 595–640. doi:10.5194/gmdd-4-595-2011.

Best, M.J., G. Abramowitz, H.R. Johnson, A.J. Pitman, G. Balsamo, A. Boone, M. Cuntz, B. Decharme, P.A. Dirmeyer, J. Dong, M. Ek, Z. Guo, V. Haverd, B.J. van den Hurk, G.S. Nearing, B. Pak, C. Peters-Lidard, J.A. Santanello, L. Stevens, and N. Vuichard (2015) The Plumbing of Land Surface Models: Benchmarking Model Performance. *J. Hydrometeor.* 16**:**1425–1442. https://doi.org/10.1175/JHM-D-14-0158.1

Boone, A. and Etchevers, P, (2001) An intercomparison of three snow schemes of varying complexity coupled to the same land-surface model: Local scale evaluation at an Alpine site. *J. Hydrometeorol.* 2:374–394.

Boone, A., Habets, F., Noilhan, J., Clark, D., Dirmeyer, P., Fox, S., Gusev, Y., Haddeland, I., Koster, R., Lohmann, D., Mahanama, S., Mitchell, K., Nasonova, O., Niu, G.-Y., Pitman, A.,Polcher, J., Shmakin, A. B., Tanaka, K., van den Hurk, B., Verant, S., Verseghy, D., Viterbo, P., and Yang, Z.-L. (2004) The Rhone-Aggregation Land Surface Scheme Intercomparison Project: an overview. *J. Climate.* 17: 187–208.

Boussetta, S., Balsamo G., Beljaars A., Agustí-Panareda A., Calvet J.-C., Jacobs C., van den Hurk B., Viterbo P., Lafont S., Dutra E., Jarlan L., Balzarolo M., Papale D., and van der Werf, G., 2013: Natural land carbon dioxide exchanges in the ECMWF Integrated Forecasting System: Implementation and offline validation. *J. Geophys. Res.* 118: 5923–5946. doi: 10.1002/jgrd.50488

Brooks, R. H. and A. T. Corey, (1964) Properties of porous media affecting fluid flow. *J. Irrig. Drain. Am. Soc. Civil Eng.* 2: 61–88, 1966.

Carrer, D., Roujean, J.-L., Lafont, S., Calvet, J.-C., Boone, A., Decharme, B., Delire, C., Gastellu-Etchegorry, J.-P. (2013) A canopy radiative transfer scheme with explicit FAPAR for the interactive vegetation model ISBA-A-gs: Impact on carbon fluxes. *J. Geophys. Res. Biogeosciences.* 118:888–903. doi:10.1002/jgrg.20070

Calvet, J-C., J. Noilhan, J.-L. Roujean, P. Bessemoulin, M. Cabelguenne, A. Olioso, and J.-P. Wigneron (1998) An interactive vegetation SVAT model tested against data from six contrasting sites. *Agr. Forest Meteorol. (***92):**73-95.

Calvet, J.-C., V. Rivalland, C. Picon-Cochard, and J.-M. Guehl, (2004) Modelling forest transpiration and CO_2 fluxes – response to soil moisture stress. *Agr. Forest Meteorol.* (**124**(3–4)):

143–156, doi:10.1016/j.agrformet.2004.01.007.

Calvet, J.-C., Lafont, S., Cloppet, E., Souverain, F., Badeau, V. and C. Le Bas, (2012) Use of agricultural statistics to verify the interannual variability in land surface models: a case study over France with ISBA-A-gs. *Geosci. Model Dev*. **5:**37–54, doi:10.5194/gmd-5-37-2012.

Clapp, R. and G. Hornberger, (1978) Empirical equations for some soil hydraulic properties. Water Resour. Res. 14: 601–604.

Chiles J, Delfiner P (2012) Geostatistics : Modeling spatial uncertainty, 2nd edn. Wiley, New York

Csillag, F., & Kabos, S. (2002) Wavelets, boundaries and the analysis of landscape pattern. *Ecoscience*. 9: 177–190.

Decharme, B., A. Boone, C. Delire, and J. Noilhan (2011) Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions. *J. Geophys. Res*. **116:** D20126, doi:10.1029/2011JD016002.

Decharme, B., E. Martin, and S. Faroux (2013): Reconciling soil thermal and hydrological lower boundary conditions in land surface models. *J. Geophys. Res. Atmos*. **118:**7819–7834, doi:*10.1002/jgrd.50631*.

Decharme, B., E. Brun, A. Boone, C. Delire, P. Le Moigne and S. Morin, (2016) Impacts of snow and organic soils parameterization on northern Eurasian soil temperature profiles simulated by the ISBA land surface model. *The Cryosphere*, **10**: 853–877. doi:10.5194/tc-10-853-2016.

De Iaco , S., Myers, D. E., Palma, M. and Posa, D., (2013) Using simultaneous diagonalization to identify a space–time linear coregionalization model. *Math. Geosci*. 45: 69–86. MR3018750

Desborough, C. E.(1997). The Impact of Root Weighting on the Response of Transpiration to Moisture Stress in Land Surface Schemes, *Mon. Weather Rev*. 125:1920–1930.

Dirmeyer, P. A., (2011) A history and review of the Global Soil Wetness Project (GSWP). *J. Hydrometeor.*12:729–749, doi:10.1175/JHM-D-10-05010.1.

Dolman, A. J. and De Jeu, R. A. M., (2010) Evaporation in Focus. *Nat. Geosci.*,3:296, doi:10.1038/ngeo849.

Egea, G., Verhoef, A., Vidale, P. L. (2011). Towards an improved and more flexible representation of water stress in coupled photosynthesis stomatal conductance models. *Agr. Forest Meteorol*.151 :1370-1384.

Eyring, V., Righi, M., Lauer, A., Evaldsson, M., Wenzel, S., Jones, C., Anav, A., Andrews, O., Cionni, I.,Davin, E. L., Deser, C., Ehbrecht, C., Friedlingstein, P., Gleckler, P., Gottschaldt, K.-D., Hagemann, S.,Juckes, M., Kindermann, S., Krasting, J., Kunert, D., Levine, R., Loew, A., Mäkelä, J., Martin, G.,Mason, E., Phillips, A. S., Read, S., Rio, C., Roehrig, R., Senftleben, D., Sterl, A., Ulft, L. H., Walton, J.,Wang, S., Williams, K. D. (2016). ESMValTool (v1.0)—A community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development.* **9**(5): 1747–1802. https://doi.org/10.5194/gmd-9-1747-2016

Faroux, S., A. T. Kaptué Tchuenté, J.-L. Roujean, V. Masson, E. Martin, and P. Le Moigne (2013) Ecoclimap-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models. *Geosci. Model Dev.* **6:**563-582, doi:10.5194/gmd-6-563-2013.

ECMWF, (2011), IFS documentation Cy37r2, Part IV: Physical processes, Chapter 8, https://www.ecmwf.int/en/elibrary/9239-part-iv-physical-processes

FAO (2003). Digital soil map of the world (DSMW). Technical report, Food and Agriculture Organization of the United Nations, re-issued version.

Garrigues, S., Allard, D., Baret, F., & Weiss, M. (2006): Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sensing of Environment.* 103:81−96.

Garrigues, S., Allard, D., Baret, F., (2007). Using first- and second-order variograms for characterizing landscape spatial structures from remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*. **45** (6): 1823-1834. DOI: 10.1109/TGRS.2007.894572

Garrigues, S., Allard, D., Baret, F., (2008). Modeling temporal changes in surface spatial heterogeneity over an agricultural site. *Remote Sensing of Environment.* ***112*** (2):588-602. **DOI : 10.1016/j.rse.2007.05.014**

Garrigues, S., A. Olioso, D. Carrer, B. Decharme, J.-C. Calvet, E. Martin, S. Moulin, and O. Marloie (2015) Impact of climate, vegetation, soil and crop management variables on multi-year ISBA-A-gs simulations of evapotranspiration over a Mediterranean crop site. *Geosci. Model Dev.*, **8**:3033–3053. doi:10.5194/gmd-8-3033-2015.

Garrigues S., A.Boone, B. Decharme, A. Olioso , C. Albergel, J-C. Calvet , S. Moulin, S. Buis, E.. Martin (2018) Impacts of the soil water transfer parametrization on the simulation of evapotranspiration over a 14-year Mediterranean crop succession. *Journal of Hydrometeorology*, 19:3-25. **DOI:**JHM-D-17-0058.1

Genton, M.G and Kleiber, W., (2015) Cross-covariance functions for multivariate geostatistics, Statistical Science, 30: 147-163.

Goudriaan, J., H.H. van Laar, H. van Keulen, and W. Louwerse, (1985) Photosynthesis, CO2 and plant production. *Wheat Growth and Modelling*, W. Day, and R.K. Atkin, Eds, NATO ASI Series, Plenum Press, New York, Series A, **86**, 107–122.

Harris, P.P., S.S. Folwell, B. Gallego-Elvira, J. Rodríguez, S. Milton, and C.M. Taylor (2017) An Evaluation of Modeled Evaporation Regimes in Europe Using Observed Dry Spell Land Surface Temperature. *J. Hydrometeor.,* 18: 1453–1470, https://doi.org/10.1175/JHM-D-16-0227.1

Henderson-Sellers, A., Yang, Z.-L., & Dickinson, R. E. (1993). The Project for intercomparison of land-surface parameterization schemes. *Bulletin of the American Meteorological Society*. **74**(7):1335– 1349.

HWSD, 2009, Harmonized World Soil Database, Documentation, www.fao.org/docrep/018/aq361e/aq361e.pdf

Jarvis, P. G., (1976). The interpretation in the variations of leaf water potential and stomatal conductance found in canopies in the field. *Philos. Trans. R. Soc. London*, B. 273:593–610.

Julesz, B. (1962). Visual pattern discrimination. *Institute of Radio Engineers Transactions on Information Theory.* 8:84–92

Koster, R. D., Sud, Y. C., Guo, Z., Dirmeyer, P. A., Bonan, G., Oleson, K. W., et al. (2006) GLACE: The Global Land-Atmosphere Coupling Experiment. Part I: Overview. *Journal of Hydrometeorology*. 7(4): 590–610. https://doi.org/10.1175/JHM510.1

Kumar, S. V., Peters-Lidard, C. D., Santanello, J., Harrison, K., Liu, Y., and Shaw, M. (2012) Land surface Verification Toolkit (LVT) – a generalized framework for land surface model

evaluation. *Geosci. Model Dev*.5:869-886, https://doi.org/10.5194/gmd-5-869-2012.

Luo, Y. Q., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., Ciais, P.,Dalmonech, D., Fisher, J. B., Fisher, R., Friedlingstein, P., Hibbard, K., Hoffman, F., Huntzinger, D.,Jones, C. D., Koven, C., Lawrence, D., Li, D. J., Mahecha, M., Niu, S. L., Norby, R., Piao, S. L., Qi, X.,Peylin, P., Prentice, I. C., Riley, W., Reichstein, M., Schwalm, C., Wang, Y. P., Xia, J. Y., Zaehle, S.,Zhou, X. H. (2012) A framework for benchmarking land models. *Biogeosciences.* **9**(10):3857– 3874. https://doi.org/10.5194/bg-9-3857-2012

Masson, V., et al., (2013) The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geosci. Model Dev.* 6: 929-960.

Noilhan, J., and S. Planton (1989) A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev*.**117:**536-549.

Pontius, R G and J. Malanson, (2005). Comparison of the structure and accuracy of two land change models. *International Journal of Geographical Information Science.* 19(2): 243–265

Pontius, R,G., Thontteh,O., Chen, H., (2008). Components of information for multiple resolution comparison between maps that share a real variable. **Environ. Ecol. Stat**. 15:111–142, DOI 10.1007/s10651-007-0043-y

Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär (2006) Land–atmosphere coupling and climate change in Europe. *Nature.* 443:205–209, https://doi.org/10.1038/nature05095.

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth Sci. Rev.* 99 :125–161. doi:10.1016/j.earscirev.2010.02.004

Senevirtane, and Coauthors, (2013) Impact of soil moisture–climate feed-backs on CMIP5 projections: First results from the GLACE-CMIP5 experiment. *Geophys. Res. Lett.* 40:5212–5217,doi:10.1002/grl.50956.

Tarnavsky, E. V., Garrigues, S., Brown M. E., (2008). Multiscale Geostatistical Analysis of AVHRR, SPOT-VGT, and MODIS NDVI products. *Remote Sensing of Environment*. **112**(2): 535-549. DOI:10.1016/j.rse.2007.05.008

Ukkola, A. M., De Kauwe, M. G., Pitman, A. J., Best, M. J., Abramowitz, G., Haverd, V Decker, M., Haughton, N. (2016) Land surface models systematically overestimate the intensity, duration and magnitude of seasonal scale evaporative droughts. *Environmental Research Letters.* 11: http://dx.doi.org/10.1088/1748

van Genuchten, M. T. (1980), A closed-form equation for predicting the

hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* 44 :892–898

Verhoef, A. and Egea, G. (2014) Modeling plant transpiration under limited soil water: Comparison of different plant and soil hydraulic. *Agr. Forest Meteorol.*191 :22–32.

Vilasa, L., D. G. Miralles, R. A. M. de Jeu, and A. J. Dolman (2017), Global soil moisture bimodality in satellite observations and climate models, J. Geophys. Res. Atmos., 122, 4299–

4311, doi:10.1002/2016JD026099.

Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J. and Viterbo, P. (2014) The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research.* 50:7505–7514, doi:10.1002/2014WR015638.

Table 1: Characteristics of SURFEX, JULES and CH-TESSEL land surface models. A-gs denotes the functional coupling between the stomatal conductance (gs) and the net assimilation of CO_2 (A)

	SURFEX	JULES	CH-TESSEL
Energy balance	One-source	One-source	One-source
Canopy radiative transfer	Multi-layer (10) canopy	Multi-layer canopy	Big-leaf
Heat and water soil transfers	Multi-layer soil diffusion scheme (14 layers)	Multi-layer soil diffusion scheme (4 layers)	Multi-layer soil diffusion scheme (4 layers)
Soil retention model	Clapp, R. and G. Hornberger, 1978	Clapp, R. and G. Hornberger, 1978 or van Genuchten (1980)	van Genuchten (1980)
Root-zone depth	1 to 3 m depends on plant functional type and location	1.*00m*	2.*89m*
Field capacity pressure head	-3.3 m	-3.3 m	- 1.0 m
Wilting point pressure head	-150.0 m	-150.0 m	-150.0 m
Stomatal conductance	A-gs (Jacobs et al., 1996)	A-gs (Jacobs et al., 1996)	Jarvis approach (Jarvis, 1976)
Photosynthesis model	Goudriaan et al. (1985)	Collatz et al. (1991) and Collatz et al. (1992)	Goudriaan et al. (1985)
Soil water stress parameter	Available soil water content in the root-zone	Available soil water content in the root-zone	Available soil water content in the root-zone
Application of soil water stress	Mesophyll and stomatal conductance	Photosynthesis parameters	Jarvis approach
Leaf Area Index	MODIS climatology	MODIS climatology	MODIS climatology
Land cover map	ECOCLIMAP	IGBP	GLC2000
Soil map	HSWD	HSWD	FAO/ Digital Soil Map of the World (DSMW)
Plant functional types	needle leaf trees, evergreen broadleaf trees, deciduous broadleaf trees, C3 crops, C4 crops, C4 irrigated crops, herbaceous, tropical herbaceous, and wetlands	(broadleaf forest, needleleaf forest, C3 grass, C4 grass and shrub	PFTs (semi-desert, water and land mixtures, crops, short grass, evergreen needleleaf, deciduous needleleaf, deciduous broadleaf, deciduous needleleaf, tall grass, tundra, irrigated crops, bogs and marshes, evergreen shrubs,

			deciduous shrubs, mixed forest, interrupted forest).
References	Masson et al., 2012 Decharme et al., 2011 Garrigues et al., 2015	Best et al., 2011 Clark et al., 2011	Balsamo et al., 2009 ECMWF, 2011 Boussetta et al., 2013

Table 2: Geographic domains. ULC and LRC mean upper left corner and lower right corner, respectively.

Domain	ULC Latitude	ULC Longitude	LRC Latitude	LRC Longitude
France (FR)	50.25	-2.25	43.75	5.75
Spain (SP)	43.25	-9.25	36.25	2.25
United-Kingdom (UK)	58.75	-9.25	50.75	1.75

Table 3: Root-Mean-Square-Difference (RMSD) in integral range (A) of 10-day mean values of RN, H, LE and root-zone soil moisture between SURFEX, $JULES_{CH}$, $JULES_{VG}$ and CHTESSEL.

Experiments	RN			H			LE			Soil moisture		
$Model_x$ vs $Model_y$	UK	FR	SP	UK	FR	SP	UK	FR	SP	UK	FR	SP
SURFEX vs $JULES_{VG}$	11.91	5.66	15.23	21.33	8.53	12.09	27.20	17.12	5.79	8.92	20.87	11.30
SURFEX vs $JULES_{CH}$	11.98	5.82	15.43	16.78	8.34	9.40	27.72	16.66	7.14	22.10	15.72	11.78
SURFEX vs CHTESSEL	15.01	3.01	6.72	35.61	7.80	14.00	34.54	20.68	11.38	5.12	20.24	11.30
$JULES_{VG}$ vs $JULES_{CH}$	0.12	1.35	1.76	13.58	4.78	4.56	14.80	5.72	2.52	17.67	19.81	4.63
$JULES_{VG}$ vs CHTESSEL	10.72	6.74	9.04	33.60	9.41	7.06	27.04	13.67	8.57	10.28	26.34	0.00
$JULES_{CH}$ vs CHTESSEL	10.75	6.87	9.28	33.95	7.95	7.93	23.31	14.93	9.20	20.75	23.31	4.63
Mean values	10.08	4.90	9.58	25.81	7.80	9.17	25.77	14.80	7.43	14.14	21.05	7.27

Table 4: Root-Mean-Square-Difference (RMSD) of spatial variability (σ) of 10-day RN, H, LE and root-zone soil moisture between SURFEX, $JULES_{CH}$, $JULES_{VG}$ and CHTESSEL.

Experiments	RN			H			LE			Soil moisture		
$Model_x$ vs $Model_y$	UK	FR	SP	UK	FR	SP	UK	FR	SP	UK	FR	SP
SURFEX vs $JULES_{VG}$	0.06	0.03	0.14	0.23	0.21	0.14	0.24	0.17	0.29	1.83	2.65	5.65
SURFEX vs $JULES_{CH}$	0.06	0.03	0.15	0.21	0.21	0.15	0.24	0.17	0.18	3.21	0.44	0.36
SURFEX vs CHTESSEL	0.13	0.09	0.09	0.26	0.25	0.19	0.24	0.25	0.21	7.28	2.10	4.63
$JULES_{VG}$ vs $JULES_{CH}$	0.01	0.01	0.01	0.04	0.05	0.13	0.03	0.07	0.19	1.72	2.89	5.54
$JULES_{VG}$ vs CHTESSEL	0.17	0.10	0.21	0.15	0.16	0.17	0.08	0.19	0.25	8.75	0.67	1.05
$JULES_{CH}$ vs CHTESSEL	0.17	0.10	0.21	0.13	0.13	0.08	0.08	0.13	0.10	10.38	2.33	4.52
Mean	0.1	0.06	0.135	0.17	0.17	0.14	0.15	0.16	0.20	5.53	1.85	3.63

Figure 1: Temporal evolution of the integral range (A) and the spatial variability (σ) of RN at 10-day time step over UK, France and Spain, for SURFEX (magenta), $JULES_{VG}$ (blue), $JULES_{CH}$ (green) and CHTESSEL (red) experiments.

Figure 2: Temporal evolution of the integral range (A) and the spatial variability (σ) of H at 10-day time step over UK, France and Spain, for SURFEX (magenta), $JULES_{VG}$ (blue), $JULES_{CH}$ (green) and CHTESSEL (red) experiments.

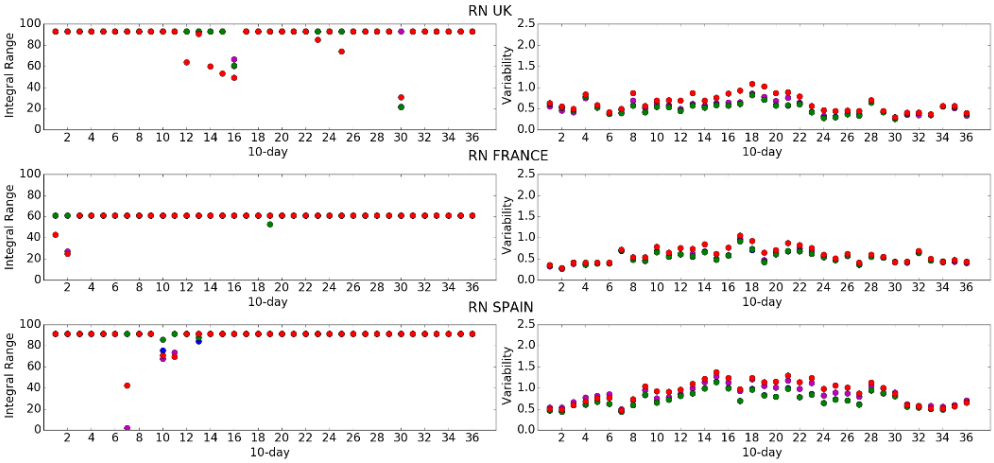
Figure 3: Temporal evolution of the integral range (A) and the spatial variability (σ) of LE at 10-day time step over UK, France and Spain, for SURFEX (magenta), $JULES_{VG}$ (blue), $JULES_{CH}$ (green) and CHTESSEL (red) experiments.

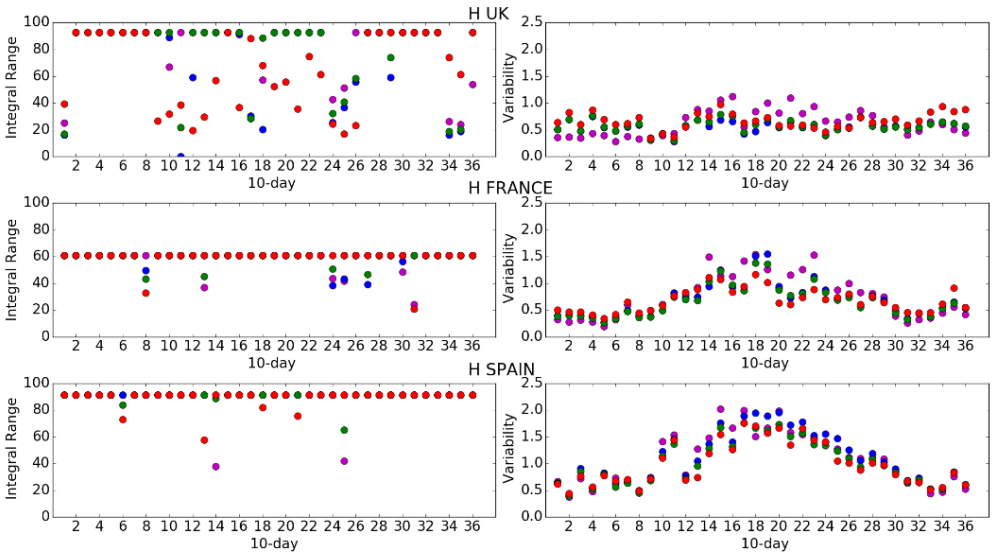
Figure 4: Temporal evolution of the integral range (A) and the spatial variability (σ) of the root-zone soil moisture at 10-day time step over UK, France and Spain, for SURFEX (magenta), $JULES_{VG}$ (blue), $JULES_{CH}$ (green) and CHTESSEL (red) experiments.

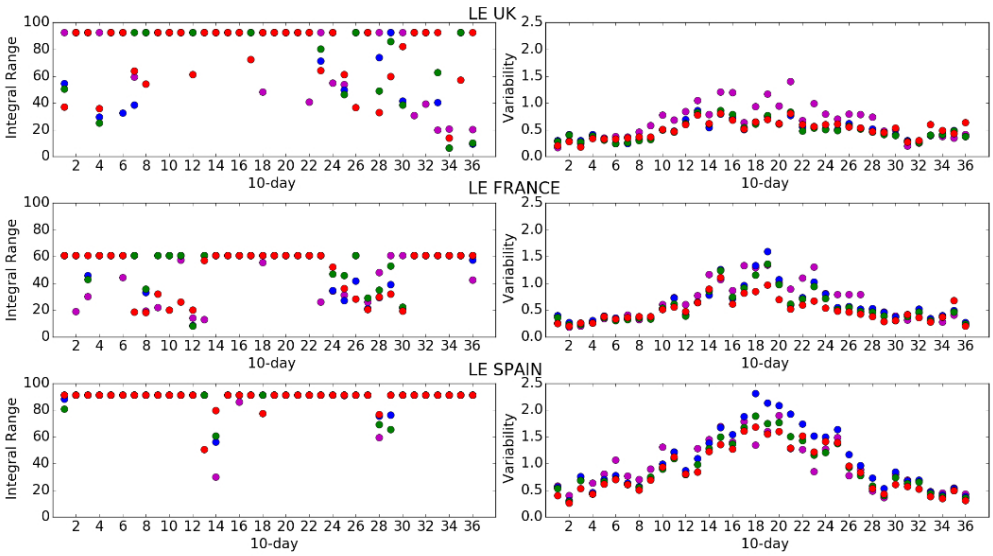
Figure 5: Comparison of the integral range (A) and the spatial variability (σ) of the maximum available water content (MaxAWC) across models and geographic domains. Models are represented by distinct colors: SURFEX in magenta, $JULES_{VG}$ in blue, $JULES_{CH}$ in green and CHTESSEL in red. Geographic domains are represented by distinct symbols: UK (triangle), France (circle), Spain (square).

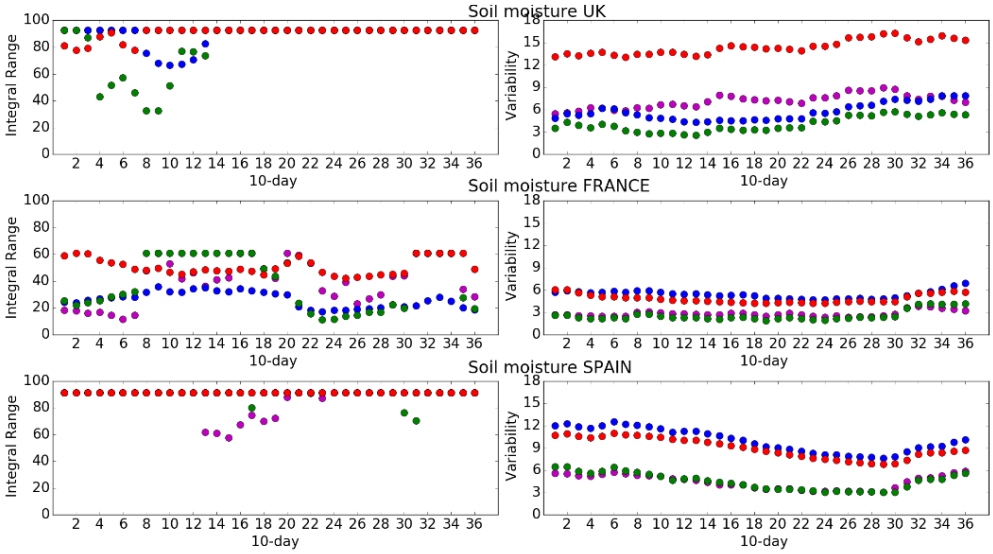
Figure 6: Seasonal evolution of the integral range (A) of LAI over France, for SURFEX (magenta), $JULES_{VG}$ (blue), and CHTESSEL (red) models.

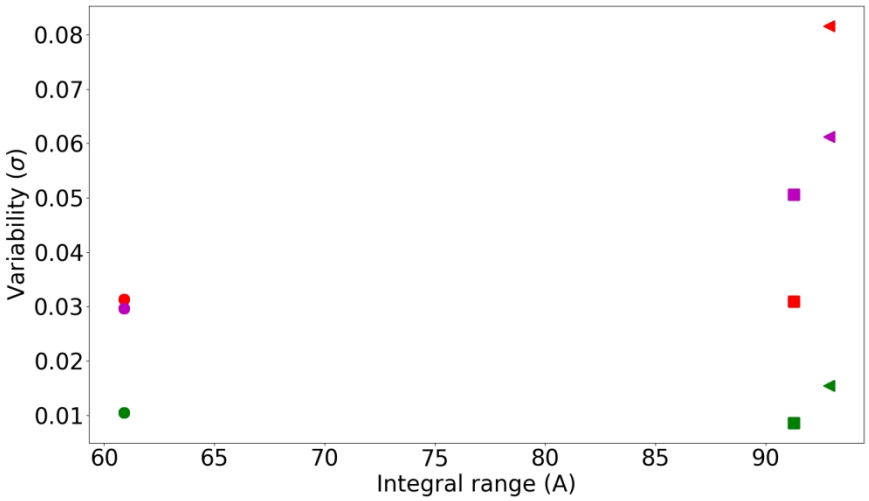
Figure 7: Seasonal evolution of the spatial variability (σ) of LAI over France, for SURFEX (magenta), $JULES_{VG}$ (blue), and CHTESSEL (red) models.

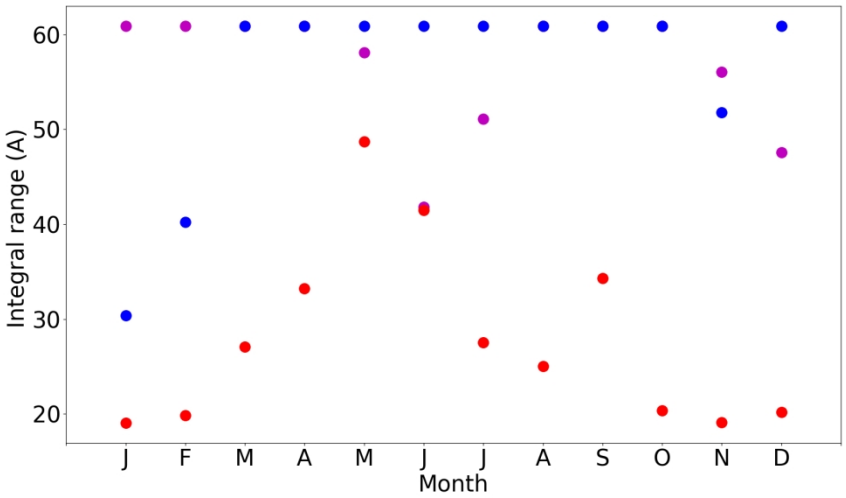

RN UK
Integral Range
Variability
10-day
RN FRANCE
Integral Range
Variability
10-day
RN SPAIN
Integral Range
Variability
10-day

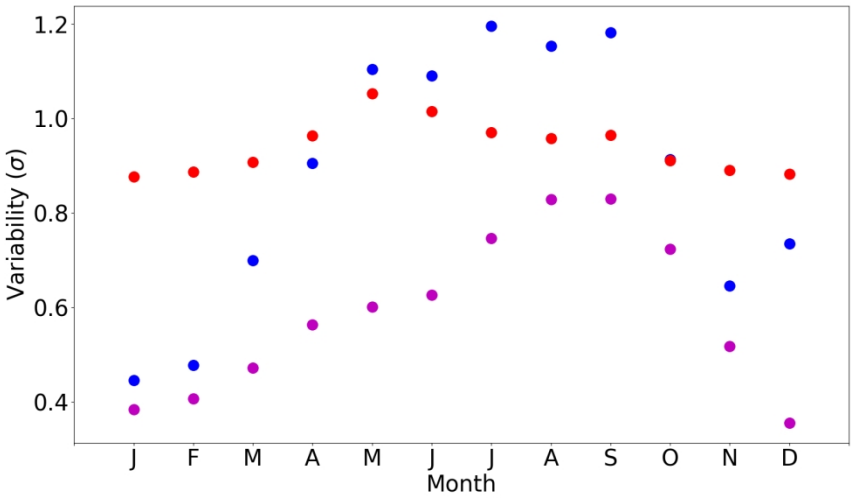

H UK
H FRANCE
H SPAIN
Integral Range
Variability
10-day


LE UK
Integral Range
Variability
10-day
LE FRANCE
Integral Range
Variability
10-day
LE SPAIN
Integral Range
Variability
10-day


Soil moisture UK
Integral Range
Variability
10-day
Soil moisture FRANCE
Integral Range
Variability
10-day
Soil moisture SPAIN
Integral Range
Variability
10-day


0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
Variability (σ)
60
65
70
75
80
85
90
Integral range (A)


Integral range (A)
60
50
40
30
20
J
F
M
A
M
J
J
A
S
O
N
D
Month


1.2
1.0
0.8
0.6
0.4
Variability (σ)
J
F
M
A
M
J
J
A
S
O
N
D
Month