MS No.: gmd-2014-175 revision

Review, 19 June 2015

Title: An observation-constrained multi-physics WRF ensemble for simulating European mega-heatwaves

Authors: A.I. Stegehuis, R. Vautard, P. Ciais, A.J. Teuling, D.G. Miralles, and M. Wild

Recommendation: [Minor Revision]

GENERAL COMMENTS:

We would like to thank the reviewer for all the constructive comments for improving the manuscript.

The revised manuscript is substantially improved compared to the original version both in clarity as in readability. Thanks to the amendment in the title (WRF instead of RCM) readers know beforehand which family of model formulations is the focus of the paper.

I still find it regrettable that variations in land-surface parameterization are not included in the investigation. Yet, though it came as a surprise when reading the author’s reply, I can accept their explanation that apparently none of the alternative land-surface schemes within WRF is capable of providing realistic results. Which makes me wonder, however, why the authors show so much confidence in announcing a future study dedicated to the use of multiple land-surface schemes.

In the last version of WRF a new land surface parameterization is included (Community Land Model Version 4 - CLM4), which would make it possible to perform such a future study.

There remain a couple of, mostly textual, issues to be resolved.

MAJOR POINTS:

1.) Abstract, 1st sentence. Actually, interchanging “often not” by “not often” does not take away my concern. In this sentence the authors imply a cause-consequence relation which isn’t there. Models are performing poorly in heat wave conditions , not because they have not been evaluated or calibrated with observations, but because the physical processes and their interactions controlling this type of atmospheric conditions are not adequately represented in the models. Evaluation in itself is an essential step, but will not prevent models from performing poorly, while calibration of climate models to certain conditions in the past might probably not be a preferable way to go in climate research.

I recommend to begin the abstract with the perception – which can only come from evaluation – that many climate type models have difficulties in properly reproducing heat wave conditions.

We thank the reviewer for the explanation, we agree that the lack of evaluation alone does not cause poor performance. We changed the sentence now into : 'Many climate models have difficulties in properly reproducing climate extremes, such as heat wave conditions'.

2.) In my opinion the abstract should explicitly mention that only one land-surface scheme has been considered in this study. The ideal place for doing that is the first part of the second sentence which I recommend to rephrase as “Here we use ...(WRF) regional climate model for a large number of configurations of different atmospheric physics schemes in combination with the NOAH land-surface scheme, with the goal ...”

This has been modified.

3.) The last sentence of the abstract is formulated in too general terms. Consider removing this line altogether because it does not highlight your findings. Usually an abstract focusses on what has been done, not on what has not been done. But if you still feel you have good grounds to retain this line, it should be sharpened and make explicit connection with the findings of this paper.

- First of all, the “varied physics schemes” is somewhat vague English, please use “configurations of different atmospheric physics schemes”. The question is, are these the five configuration found in the paper, or has the research to be carried out all over again.
- Secondly, instead of “varied land surface models” use “alternative land surface models”.
- Finally, the “sensitivity” of what could be included? And what are “land surface processes controlling soil moisture”. Please, clarify.

We agree with the referee that the sentences might not be necessary and removed it from the abstract.

4.) Line 55-57. I recommend to rewrite the 2nd part. “ ...underlying extreme temperatures during heat waves, because *it is difficult to separate model biases due to deficiencies in the model representation from* sensitivities to initial conditions.”

We changed the sentence following the referee's suggestion.

5.) Line 68-69: It is not clear to me what you want to say with this line. Different RCMs respond differently to the same large-scale forcings (re-analysis) because they utilize different “physics packages”. Is that what you mean with “internal variability”? Or are you referring to different “internal degrees of freedom” which allow RCMs with the same physics package to still choose different solutions. Also, it is not clear to me whether “its” is meant to refer to “atmospheric flow” or to “events”. In the latter case it should be “their”.

With 'its' we refer to the 'events, and so we have changed it into 'their'. For the internal variability we meant the internal degrees of freedom, which is now added to the text to make it more clear.

6.) Line 91-96: I still miss the point of the term “democracy-driven”, to me it sounds like a misinterpretation of the approach followed in assembling the 2nd type of ensemble. Please, leave it out.

Instead rephrase line 92-94 “ ... in coordinated experiments (see ...and references therein) *or* combinations of parameterizations...” as “ ... in coordinated experiments (see ...and references therein), *or by arbitrarily configured* combinations of parameterizations ...”. (Mind the interpunction following the “)”.) Then continue with “In the latter ensemble, ...”

The term 'democracy driven' is left out of the manuscript, and lines 92-94 are rephrased following the suggestions of the referee.

7.) The sensitivity tests you performed to quantify the response to changes in the initialization of soil moisture nicely illustrate that the feed back from soil moisture on temperature is potentially large, especially during episodes of blocked circulation with flow coming from the continent.

Regarding the experimental set up I wonder if 20% is with respect to the absolute amount (which it seems to be), or with respect to the range between field capacity and wilting point. The former would indeed be rather radical, because the perturbation of 20% would potentially bring soil moisture outside the physically realistic range. An alternative would have been to initialize soil moisture either at field capacity (“wet” run) or wilting point (“dry” run). Please, mention more specifically how soil moisture is modified in the sensitivity runs.

The soil moisture is modified with respect to the absolute amount. This is added to the manuscript in line 347 : 'In order to mimic radically different land surface processes, sensitivity tests in which the initial absolute amount of soil moisture was artificially increased and decreased by 20% all along the soil column have been conducted.'

Line 344-345: rephrase “a sensitivity test where initial soil moisture was artificially increased and decreased ... was conducted ...” as “sensitivity tests in which initial soil moisture was artificially increased or decreased ... have been conducted ...”

This has now been rephrased and we added the 'absolute amount ' regarding the previous remark of the referee.

Figure 8 is rather fuzzy and requires upgrading.

We used figures with the .PNG format in the text document, but the figures are also provided in .PDF format, which has a better quality. We suppose that the .PDF format will be used in the final manuscript.

Regarding the figure caption of Figure 8 I suggest to rephrase the last two sentences as “*Difference between the perturbed simulations (red indicates 20% reduction of initial soil moisture, blue 20% enhancement) performed with the five highest ranked configurations compared to their corresponding ‘control’ simulations. The darkest lines refer to the simulation conducted with to the best ranked configuration (1), while descending colour shade agrees with descending ranking (1-5).*”

Thank you for the suggestion. Indeed this makes the caption more clear. It has been rephrased.

8.) Concerning the new Figure 1, the two different shades of blue (light blue = cyan, and blue) are impossible to distinguish. It seems to me only one shade of ‘blue’ was used. Please, use ‘cyan’ for NOAH and ‘blue’ for RUC. (Or if that doesn’t provide enough contrast replace ‘cyan’ by e.g. ‘magenta’ or ‘red’)

We only used one shade in the original figure ; NOAH and RUC are distinguished by the type of line (dots versus solid line). In order to make the figure more clear we have now used two colours to separate the two land-surface schemes.

MINOR POINTS:

1.) Entire manuscript: Consider using “configuration” instead of “combination” where appropriate. The former is probably more adequate wording in this context than the latter.

Rephrased where appropriate.

2.) Line 40: “... problematic ...” “ ...impacting ...”

Rephrased.

3.) Line 53: “ ... new soil ...” “ ... new land surface hydrology ...”

Rephrased.

4.) Line 55: “ ... easily ...” - ” ... straightforwardly ...” (Nothing is easy, so not easily is a trivial phrase.)

Rephrased.

5.) Line 97-98: “ ... because of interacting physical processes and their biases.” Physical processes themselves have no biases, it is the way we describe them. Please rephrase like, for instance, “... because of shortcomings in the representation of physical processes and their interactions”.

Rephrased.

6.) Line 133: For the sake of completeness, expand “Boundary conditions come from ERA-Interim including sea surface temperatures ...”

This has been added.

7.) Line 157: “The NOAH scheme seemed more stable in the tests ...”. I suggest to replace “seemed more stable” by “appeared more realistic and robust” providing more proper wording.

Rephrased.

8.) Line 172: “..., and were not considered.”   “..., and have therefore not been considered.”

Rephrased.

9.) Line 187: Rephrase “ ... is the primary impacted variable and observations are reliable” as “ ... is the primary impacting variable, while corresponding observations are reliable”.

Rephrased.

10.) Line 176: Remove “validate and” , and only retain “we do not use them to rank ...”

'Validate and' has been removed.

11.) Line 191: Rephrase “The 1 K threshold is arbitrary but is used to avoid ...” as “The 1 K threshold was arbitrarily chosen and is used to avoid ...”.

Rephrased.

12.) Line 203-204: Rephrase “ ... which is shown to be able to impact ...” as “ ... which was shown to potentially impact ...”.

Rephrased.

13.) Line 224: “... during heat waves years, ...”   “... during heat wave years, ...”

Replaced.

14.) Line 292: “...to considerable overestimate ...”   “...to considerablyoverestimate ...”.

Replaced.

15.) Line 313: Suggest to rephrase “..., we found a large spread between the different physics for the simulations for ...” as “..., the multi-physics ensemble contained a large spread in ...”

Rephrased.

16.) Line 314: “..., three variables ...”   “..., the three variables ...”

Rephrased.

17.) Line 320: “ ... probable ...”   “ ... possible ...” (probable is too speculative)

Replaced.

18.) Line 320: “ ... largely ...” “... considerably ...” (largely is too strong, substantially is also possible).

Replaced.

19.)

a. Panels in figures 3, 5 and 7 have different size. Please, align.

b. Also for Fig 3a the text along the horizontal axis is not visible.

We think both a and b are due to the use of Word (the .PNG versions of the figures were used) and the conversion to a pdf file. The original figures (in .PDF) have the same format and size.

c. Figure 3d) is mentioned in the caption but not indicated in the Figure

This has been adapted.

d. Labels in Figure 4 are still (d), (e), (f), but should be (a), (b), (c).

This has been adapted (also for figure 2 and 6).

e. Figure 4c does not have labels along the vertical axis.

This has been adapted.

An observation-constrained multi-physics WRF ensemble for simulating European mega heat waves

Annemiek I. Stegehuis[1], Robert Vautard[1], Philippe Ciais[1], Adriaan J. Teuling[2], Diego G. Miralles[3,4] and Martin Wild[5]

[1] LSCE/IPSL, Laboratoire CEA/CNRS/UVSQ, Gif-sur-Yvette, France

[2] Hydrology and Quantitative Water Management Group, Wageningen University, The Netherlands

[3] Department of Earth Sciences, VU University Amsterdam, Amsterdam, The Netherlands

[4] Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium

[5] ETH Zurich, Zurich, Switzerland

Corresponding author : A.I. Stegehuis, LSCE/IPSL, Laboratoire CEA/CNRS/UVSQ, 91191 Gif-sur-Yvette CEDEX, France

Abstract

~~C~~Many climate models ~~are not often evaluated or calibrated against observations of past climate extremes, resulting in poor performance during for instance~~have difficulties in properly reproducing climate extremes, such as heat wave conditions. Here we use the Weather Research and Forecasting (WRF) regional climate model with a large combination of different atmospheric physics schemes in combination with the NOAH land-surface scheme, with the goal of detecting the most sensitive physics and identifying those that appear most suitable for simulating the heat wave events of 2003 in Western Europe and 2010 in Russia. 55 out of 216 simulations combining different atmospheric physical schemes have a temperature bias smaller than 1 degree during the heat wave episodes, the majority of simulations showing a cold bias of on average 2-3°C. Conversely, precipitation is mostly overestimated prior to heat waves, and short wave radiation is slightly overestimated. Convection is found to be the most sensitive atmospheric physical process impacting simulated heat wave temperature, across four different convection schemes in the simulation ensemble. Based on these comparisons, we design a reduced ensemble of five well performing and diverse scheme ~~combinations~~configurations, which may be used in the future to perform heat wave analysis and to investigate the impact of climate change in summer in Europe. ~~Future studies could include the sensitivity to land surface processes controlling soil moisture, through the use of varied land surface models together with varied physics schemes.~~

1. Introduction

An increasing number of simulations and studies project a higher frequency of several types of extreme weather events in the future (e.g. Schär et al., 2004; Meehl et al., 2004; Della-Marta et al., 2007; Beniston et al., 2007; Kuglistsch et al., 2010; Fischer and Schär, 2010; Seneviratne et al., 2012; Orlowsky and Seneviratne, 2012). Since summer heat waves are among the most ~~problematic~~ impacting of such phenomena - threatening society and ecosystems - climate models used for future projections must provide accurate simulations of these phenomena, or at least their uncertainties should be documented. Even if climate models have been evaluated using observed weather in past decades, it is unclear whether they will be able to simulate extreme heat waves in future climates that may not have analogues in the historical record. At least, models should be able to reproduce the conditions measured during recent extreme heat wave cases, some of them having been shown to be unprecedented when considering the climate over the past five or six centuries (Chuine et al., 2004; Luterbacher et al., 2010; García-Herrera et al., 2010; Barriopedro et al., 2011; Tingley and Huybers, 2013).

Given the importance of forecasting summer heat waves well in advance, many studies have analyzed their predictability, which remains poor in seasonal forecasts. For instance the 2003 European heat wave was not simulated realistically (neither timing nor intensity) by the operational European Centre for Medium-Range Weather Forecasts (ECMWF) system, but improvements were clear with the use of a new ~~soil~~land surface hydrology, convection and radiation schemes (e.g. Weisheimer et al., 2011; Dole et al. 2011; Koster et al. 2010; van den Hurk et al. 2012). However seasonal forecasting experiments do not ~~easily~~ straightforwardly allow the assessment of model physical processes underlying extreme temperatures during heat waves because it is difficult to separate model biases due to deficiencies in the model representation from ~~model biases are mixed with~~ sensitivity to initial conditions. These may inhibit the effect of the representation of physical processes in reproducing the exact atmospheric circulation when starting simulations at the beginning of the season.

From a statistical perspective, extreme temperatures have been found to be reasonably well represented in global simulations of the current climate (IPCC, 2013), as well as in regional simulations (Nikulin et al., 2010). In recent regional modeling evaluation experiments, using an ensemble of state-of-the-art regional models guided by re-analysis at the boundaries of a

European domain, summer extreme seasonal temperatures were shown to be simulated with biases in the range of a few degrees (Vautard et al., 2013). Individual mega heat waves (2003 in Western Europe, 2010 in Russia) were reproduced by most models. However, it was difficult to infer whether these models could also simulate associated processes leading to the extreme heat waves. The exact same events with similar atmospheric flow and ~~its~~ their persistence could not be reproduced due to internal variability (internal degrees of freedom) of the models.

A comprehensive assessment of simulations of recent mega heat waves has only been the object of a limited number of such studies. Process-oriented studies of high extreme temperatures over Europe have focused on land-atmosphere feedbacks (e.g. Seneviratne et al., 2006 and 2010; Fischer et al., 2007; Teuling et al., 2009; Stegehuis et al., 2013; Miralles et al., 2014) because, beyond atmospheric synoptic circulation, these feedbacks are known to play an important role in summer heat waves. However, the sensitivity of simulated heat wave conditions to physical processes in models has not yet been explored in a systematic way. This could be important because error compensation among processes that involve land-atmosphere interactions, radiation and clouds may cause high temperatures for the wrong reasons (Lenderink et al., 2007).

The goal of the present study is threefold. First we examine the ability of a regional climate model, the Weather Research and Forecast (WRF, Skamarock et al., 2008), to simulate recent European mega heat waves, with a number of different model configurations. Analysis of these experiments then allows understanding which physical parameterizations are prone to reproduce the build-up of extreme temperatures, and thus the need for carefully constraining them in order to simulate these events properly. Finally, using observational constraints of temperature, precipitation and radiation, we select a reduced ensemble of WRF configurations that best simulates European heat waves, with different sets of physical schemes combinations.

This constrained multi-physics ensemble aims therefore at spanning a range of possible physical parameterizations in extreme heat wave cases while keeping simulations close to observations.

Our multi-physics regional ensemble approach contrasts with the classical multi-model ensembles that are constructed by the availability of model simulations in coordinated experiments (see e.g. Déqué et al., 2007 and references therein), or by arbitrarily configured combinations of parameterizations selected by different groups using the same model system (García-Díez et al., 2014). In the latter ~~“democracy-driven”~~ ensemble, the lack of overall design strategy may lead the uncertainty estimation to be biased and the models to be farther from observations. In addition, the real cause of model spread is difficult to understand because of shortcomings in the representation of ~~interacting~~ physical processes and their ~~biases~~interactions. Regional perturbed-physics or multi-physics ensembles could help understand and constrain uncertainties more effectively, but so far they have been seldom explored. García-Díez et al. (2014) showed that even a small multi-physics ensemble confronted to several climate variable observations can help diagnose mean biases of a RCM. Bellprat et al. (2012) showed that a well-constrained perturbed physics ensemble may encompass the observations. Their perturbed physics ensemble was designed by varying the values of a number of free parameters, and selecting only the configurations that were closest to the observations; however, the number of combinations of different physical parameterization schemes was limited to a total of eight different configurations.

The WRF model offers several parameterization schemes for most physical processes, and is thus suitable for a multi-physics approach. In fact, a WRF multi-physics approach has been used in several studies (e.g. García-Díez et al., 2011; Evans et al., 2012; Awan et al., 2011; Mooney et al., 2013), also with its predecessor MM5, but not specifically to simulate extreme heat waves.

Here we run an ensemble of 216 ~~combinations~~ configurations of WRF physical parameterizations, and compare each simulation with a set of observations of relevant variables in order to select a reduced set of 5 configurations ~~combinations~~ that best represent European summer mega heat waves. The evaluation is made over the extreme 2003 and 2010 events. The ensemble is also evaluated for a more regular summer (2007) in order to test the model configurations under non-heat wave conditions.

2. Methods

Simulations and general model setup

We use the WRF version 3.3.1 and simulate the three summers (2003, 2007, 2010) using an ensemble of physics scheme combinations. We first test the time necessary to initialize the soil moisture on a limited number of cases. Soil conditions are initialized using the ERA-Interim (Dee et al., 2011) soil moisture and temperatures; thereafter soil moisture and air temperature are calculated as prognostic variables by WRF. For the August 2003 case, we find that temperatures differ by less than 0.5°C among one another when starting experiments before May 1st. Thus in the current study, each simulation is run from the beginning of May to the end of August for the years 2003, 2007 and 2010. The regional domain considered is the EURO-CORDEX domain (Jacob et al., 2014; Vautard et al., 2013) and the low-resolution setup of 50 km x 50 km (~0.44 degree on a rotated lat-lon grid) is used – note that Vautard et al. (2013) recently concluded that a higher spatial resolution did not provide a substantial improvement in heat wave simulations. We use a vertical resolution with 32 levels for WRF. Boundary conditions come from ERA-Interim including sea surface temperatures, ~~(as well as~~ initial snow cover, and soil moisture and temperature). In order to focus on physical processes in the boundary layer and the soil-atmosphere interface, and to avoid chaotic evolution of large-scale atmospheric circulation, we constrain the model wind fields with ERA-Interim reanalyses above Model Level #15 (about 3000m), similar to previous studies (Vautard et al., 2014), using grid nudging, with a relaxation coefficient of 5.10^{-5} s^{-1}, corresponding to a relaxation time about equivalent to the input frequency (every six hours) (Omrani et al., 2013). Temperature and water vapor were not constrained, to let feedbacks fully develop.

Physics schemes

We test 216 combinations of physics schemes. We consider different physics of the planetary boundary layer and surface layer (PBL; 6 schemes), microphysics (MP; 3 schemes), radiation (RA; 3 schemes) and of convection (CU; 4 schemes). For each type of scheme, a few options were selected among the ensemble of possibilities offered in WRF. The selection was made to avoid variants of the same scheme, and to maximize the difference of temperature and precipitation outputs in preliminary experiments. At the time of study and model development stage, different land-surface schemes were available in WRF: 5-layer Thermal Diffusion Scheme (Dudhia, 1996), NOAH (Tewari et al., 2004), Rapid Update Cycle (RUC) (Benjamin et al., 2004) and Pleim-Xiu (Gilliam & Pleim, 2010). We decided however to only use one, the NOAH land surface scheme, in order to focus our study on atmospheric processes while limiting the number of simulations, and because the NOAH scheme is the most widely used in WRF applications. This was also motivated by the poor performance and extreme sensitivity of the RUC land surface scheme for the land latent and sensible heat flux as compared with local observations in 2003. It simulates strong latent heat fluxes in the beginning of the season and an extreme drying at the end, while sensible heat flux is overestimated. The NOAH scheme ~~seemed more stable~~appeared more realistic and robust in the tests that were done for capturing both latent and sensible heat fluxes during the 2003 heat wave at selected flux tower sites in Western Europe (Figure 1). Furthermore the Pleim-Xiu scheme is especially recommended for retrospective air quality simulations, and is developed with a specific surface layer scheme as coupled configuration (Gilliam & Pleim, 2010). The last possible option is the 5-layer thermal diffusion scheme (Dudhia, 1996) which predicts ground and soil temperatures but no soil moisture, and is therefore also not suitable for our study. Table 1 describes the physical schemes that were combined to simulate the weather over the three summer seasons.

Observational data

In order to evaluate the ensemble and to rank and select its best performing simulations we use gridded observed daily temperature and precipitation from E-OBS with a 0.25 degree resolution (version 7.0) (Haylock et al., 2008). Bilinear interpolation is used to regrid E-OBS data and the model output to the same grid. Furthermore we use station data of monthly global radiation from the Global Energy Balance Archive (GEBA) network (Wild et al., 2009). For France 2003 the data of 21 stations were available, for 2007 this number was 20. Observations over Russia were too scarce, and ~~were~~have therefore not been considered. Model data are interpolated to these stations using the nearest neighbor method. In addition, in order to check land-atmosphere fluxes and the partitioning of net radiation into sensible and latent heat fluxes, we use the satellite observation-driven estimates of daily latent heat fluxes from GLEAM (Miralles et al., 2011). Since the latter is not a direct measurement we do not use them to ~~validate and~~ rank the model configurations. Furthermore latent- and sensible heat flux measurements are used from three FLUXNET sites from the Carbo-Extreme database (Neustift/Stubai – Austria (Wohlfahrt et al., 2010); Tharandt-Anchor station – Germany (Grünwald & Bernhofer, 2007); and Soroe-LilleBogeskov – Denmark (Pilegaard et al., 2009)), for the evaluation of the land surface schemes.

Evaluation and ranking of model simulations

For ranking, we set up several measures of model skill, based on the differences between observed and simulated spatial averages over two domains: France for 2003 and 2007 (5W–

5E & 44N–50N), and one in Russia for 2007 and 2010 (25E–60E & 50N–60N) (Fig. 2). A first scheme selection is made based on the skill to reproduce air temperature dynamics, since this is the primary impacted variable, while corresponding~~ and~~ observations are reliable. Because we are interested in heat waves, we select only those simulations that are within a 1 K regional average difference between simulated and observed temperature, for heat wave periods; these periods are defined as August 1st-15th for France (in 2003), and July 1st till August 15th for Russia (in 2010). The 1 K threshold ~~is~~was arbitrarily chosen ~~but~~and is used to avoid processing a large number of simulations that have unrealistic temperatures. Only 55 of the 216 simulations meet this criterion and are further considered. Then, the ranking of the retained simulations is done based on: (i) the daily temperature difference between simulations and observations during the heat wave periods (as above for 2003 and 2010), and during the period 1st-31st August for the normal year 2007, (ii) the root mean square error of monthly precipitation and radiation for the months July, June and August. The GEBA data set only contains scarce radiation observations over Russia, and therefore we could not consider this region for ranking models against incoming shortwave radiation. As a final step, an overall ranking is proposed by averaging the ranks obtained from the three variables (temperature, precipitation and radiation). From this final ranking, and in order to select an elite of multi-physics combinations, we selected the top-5 highest-ranked configurations. Note that observational uncertainty is not considered in this study, which ~~i~~was shown to ~~be able to~~potentially impact model ranking over Spain (Gomez-Navarro et al., 2012).

3. Results

3.1. Large systematic errors found during heat wave periods

Figure 3 shows the large temperature range spanned by the 216 ensemble members for the spatial average over the heat wave areas. The min-max range between ensemble members is up to 5°C during heat wave periods (Figure 3). Locally at 50 km resolution, the difference between the warmest and the coldest simulation during a heat wave is larger, reaching more than 10°C in 2003 (Figure 3d). In 2007, when summer temperatures were not extreme, the range is about twice as small. Only a few simulations match the observed high temperatures (Figure 3a-c). In Fig. 3a, we select two extreme configurations (blue and red lines), based on daily mean temperature over France during the 2003 heat wave. Interestingly, they are extreme in all regions and years, indicating that each ~~combination~~ configuration tends to induce a rather large systematic bias. This bias however, is different for the 'warm' and the 'cold' configuration. It seems not to be due to a misrepresentation of the diurnal cycle, since they remain when analyzing time series of maximum and minimum daily temperatures independently (see supplementary Figures 1a-f). However, minimum temperatures show a less consistent bias than maximum daily temperatures. A systematic temperature underestimation by WRF simulations over Europe has also been found in other multi-physics ensemble studies over Europe (e.g. Awan et al., 2011; García-Díez et al., 2011, 2014).

For monthly precipitation we obtain a large range of simulated values, with most configurations overestimating monthly summer rainfall (JJA) during heat wave~~s~~ years, and to a lesser extent during the wetter 2007 season (Fig. 4a-c). This is in line with the findings reported by Warrach-Sagi et al. (2013) and Awan et al. (2011), and with the overestimation of precipitation by many EURO-CORDEX models shown by Kotlarski et al. (2014). The two selected extreme ~~combinations~~ configurations (based on temperature, as explained above) are reproducing precipitation overall without a major bias. This suggests that the temperature bias in these two extreme simulations is not explicitly caused by a misrepresentation of the atmospheric water supply from precipitation. However soil moisture (the soil moisture over the whole column) does show a strong relation to temperature biases in model simulations. Figure 5a-d shows soil moisture at the end of July versus temperature in August 2003 for each model configuration. Configurations with low soil moisture level are associated with higher temperatures and vice versa, confirming the role of land-atmosphere feedbacks during heat waves, already pointed out by previous studies. This indicates that the evapotranspiration from spring to summer depleting soil moisture can be a critical process during summer for the development of heat waves, and that this process is not simply related to summer precipitation.

For solar radiation, the mean differences between our simulations over France 2003 and 2007 reaches approximately 100 Wm^{-2} (Fig. 6a,b). Observations for France (black dots) are found below the median value of the simulations so a slight overestimation of the ensemble is obtained. The first (warmest) extreme configuration (red dot) is associated with an overestimated radiation of 10-50 Wm^{-2} while the other (coldest, blue dot) extreme configuration exhibits an underestimated radiation by about the same amount. Since the warmest simulation agrees better with temperature observations than the coldest simulation, one may therefore suspect that it contains a cooling mechanism that partly compensates for the overestimated solar radiation.

3.2. Sensitivity of temperatures to physical parameterizations and sources of spread

In order to identify the physics schemes to which the development of heat waves is most sensitive, we examine how resulting temperatures are clustered as a function of the scheme used. We find that the spread between all simulations – both in terms of temperature and soil moisture – is mostly due to the differences in convection scheme (clustering of dots with the same color in Fig. 5a). For instance the Tiedtke scheme (blue dots) systematically leads to higher temperatures and lower soil moisture, while the Kain-Fritsch scheme (green dots) leads to wetter soils and lower temperatures, inhibiting heat waves. Microphysics and radiation schemes are also contributing to the spread of simulated temperature and soil moisture values

(Fig. 5b-c), although their effect is less marked than for convection. Heat wave temperatures and soil moisture seem to be least sensitive to the planetary boundary layer and surface layer physics schemes. The sensitivity of the convection scheme in WRF has already been mentioned in previous studies (Jankov et al., 2005; Awan et al., 2011;; Vautard et al., 2013; García-Díez et al., 2014). Note that the soil moisture simulated in early August 2003 is better correlated with preceding radiation than with precipitation (compare Supplementary Figures 2 and 3), indicating that the way clouds, and particularly convective clouds, affect radiation prior to the onset of heat waves is a major driver of the spread for the development of heat waves, higher radiation leading to drier soils and higher temperatures during heat waves.

3.3. A constrained reduced ensemble of best simulations

Focusing only on the 55 selected simulations that differ less than 1°C from the observations during the heat waves, we apply the ranking method introduced in Section 2 based on temperature, precipitation and radiation model-observation comparison metrics. The 5 highest ranked simulations are given in Table 2 and are actually the numbers 1-5 in Supplementary Table 1. Figure 7a confirms the ranking by showing that these simulations also perform well in terms of temperature, during the months prior to the heat wave. The same is furthermore found for the years 2007 in France (Supp. Fig. 5) and 2010 in Russia (Supp. Fig. 4), and also for other regions such as the Iberian Peninsula and Scandinavia (Supp. Fig. 6a,d). The selected simulations however performed less well for precipitation over France in 2003 (Fig. 7b), but do not show a large overestimation of precipitation either. Precipitation over Russia for the 5 highest-ranked simulations does show good performance (Supp. Fig. 4b), as well as for other European regions (Supp. Fig. 6). The mean radiation of the ensemble of the five best simulations is closer to the GEBA observations than in the case of the original ensemble (Fig. 7c).

Nonetheless, the better match of the reduced ensemble of the five highest-ranked simulations to the observations of temperature, precipitation and radiation is to a very large degree unsurprising: the selection was based on the fit to observations. However, it is still satisfactory to see that some simulations are capable of matching all three variables. Conversely, we also compare simulations against another key variable that was not used for evaluating and ranking simulations, namely the latent heat flux (Figure 7d). Albeit somehow reduced compared to the full-ensemble spread, the spread of the five best simulations for the latent heat flux remains large over the whole period, on average between 50 and 120 Wm^{-2} (observed values are around 75 Wm^{-2}). However, during the 2003 heat wave over France three of the five best simulations exhibit a close resemblance to the latent heat observations (approximately 5-10 Wm^{-2}) (Fig. 7d). The two simulations that are found to considerably~~e~~ overestimate latent heat flux by approximately 30-40 Wm^{-2} (as compared to GLEAM) are those that use a different convection scheme than the Tiedtke scheme. The overestimation of latent heat fluxes in these schemes is however not generalized for other regions and years (Suppl. Fig. 4c, 5d, 6c,f-h), for which the latent heat flux was fairly well simulated within the range of uncertainty of GLEAM.

A cross-comparison for the years 2003 and 2010, that is, using only the 2010 heat wave to select schemes and verify the performance of the selected schemes over 2003 and vice versa, yields some promising results. Table 3 shows the average ranking of the best (5, 10, 15, 20 and 25) simulations. When only using one heat wave to select the best configurations, they all lie in the top-ranked half, and even higher in the ranking in the case of the 2010 heat wave over Russia being used to select the best configurations. This suggests that the selection based upon one heat wave in one region should also provide better simulations for other heat waves or heat waves in other areas, i.e. that the bias of a member of the WRF ensemble is not local, but at least regional at the scale of Western Europe.

4. Concluding remarks

In this study we designed and analyzed a large multi-physics ensemble with the WRF model. It is made of all possible combinations of a set of different atmospheric physics parameterization schemes. They were evaluated for their ability to simulate the European heat waves of 2003 and 2010 using the regional climate model WRF based on temperature, precipitation and shortwave radiation. Even though the simulations were constrained by grid nudging, the multi-physics ensemble contained ~~we found~~ a large spread ~~between the different physics for the simulations for~~ in temperature, precipitation and incoming shortwave radiation, the three variables we used to create an overall configuration ranking. Most simulations systematically underestimate temperature and overestimate precipitation during heat waves, a model pattern that was already found in previous studies dealing with much smaller ensembles (e.g. Awan et al., 2011; García-Díez et al., 2011; Warrach-Sagi et al., 2013). The spread among ensemble members is amplified during the two extreme heat waves of study. Since we only considered a single land surface scheme, it is ~~probable~~ possible that the ensemble spread would ~~largely~~ considerably increase when incorporating the uncertainty associated with modeling land surface processes. Nevertheless, considering only atmospheric processes, the magnitude of the spread still reaches 5°C during the peak of the heat waves.

We also showed that among atmospheric process parameterizations, the choice of a convection scheme appears to dominate the ensemble spread. We found indications that the large differences between convection schemes seem to occur mostly through radiation, and therefore the way convective clouds affect the surface energy and water budget prior to and during heat waves. Changes in incoming radiation cause changes in evapotranspiration and therefore soil moisture, which may subsequently feed back on air temperature.

From this ensemble, we selected a small sub-ensemble with the five best ~~combinations~~

configurations of atmospheric physics schemes based on the fit to observations. These ~~combinations~~ configurations capture well the temperature dynamics during the mega heat waves of France and Russia, and they perform better than other ~~combinations~~ configurations in other regions of Europe. In addition, they are consistent with independent latent heat flux data used for cross-validation. This indicates that the constraints set for the selection reduce the uncertainty across the whole European continent and points towards the creation of an optimized ensemble of WRF configurations specific for heat waves, with reduced error compensations. A sub-ensemble that outperforms a larger ensemble was also found by Herrera et al. (2010). The sub-ensemble based on mean precipitation showed better results for extreme precipitation as well.

However a limitation of this study is the use of only one land-surface scheme; the five selected WRF configurations may actually all be affected by systematic errors of the NOAH land surface scheme. The importance of the selected land surface scheme is further confirmed by the larger spread of the “best” ensemble for latent heat (in Wm-2) than for shortwave radiation. In order to mimic radically different land surface processes, ~~a~~ sensitivity tests ~~where~~ in which the initial absolute amount of soil moisture was artificially increased and decreased by 20% all along the soil column ~~was~~ have been conducted. Results confirm the sensitivity of the temperature simulations to soil moisture, a variable partly controlled by the land surface scheme (Figure 8). The full answer to this question is left for a future study in which different atmospheric schemes and surface schemes will be jointly permuted.

Although our ensemble is trained on only summer conditions, our results have several implications for climate modeling. First, the constrained WRF ensemble may be used in future studies of climate change; each of the five members may exhibit a different sensitivity to future climate change conditions, leading to a constrained exploration of the uncertainty. Then it is important to notice that our study pinpoints the need to carefully design or adjust the convection scheme for a proper representation of the summer climate during heat waves. This is particularly important in order to evaluate the impacts of climate change on ecosystems, health, carbon cycle, water and cooling capacity of thermal energy plants, since heat waves in the mid latitudes are expected to be of the most impacting phenomena in a human altered climate. Therefore, impact studies can be designed based on the selected configurations.

Acknowledgments

AIS acknowledges CEA for funding as well as of the GHG-Europe FP7 project. AJT acknowledges financial support from The Netherlands Organisation for Scientific Research through Veni grant 016.111.002. P.C. acknowledges support of the ERC-SYG project P-IMBALANCE. The authors acknowledge K. Pilegaard, A. Ibrom, C. Bernhofer, G. Wohlfahrt and CarboEurope for sharing FLUXNET data. We would like to thank the reviewers for their useful comments and suggestions for improving the manuscript.

References

Awan, N. K., H. Truhetz & A. Gobiet (2011) Parameterization-induced error characteristics of MM5 and WRF operated in climate mode over the Alpine region: an ensemble-based analysis. *Journal of Climate*, 24, 3107-3123, doi:10.1175/2011JCLI3674.1.

Barriopedro, D., E. M. Fischer, J. Luterbacher, R. Trigo & R. Garcia-Herrera (2011) The hot summer of 2010: redrawing the temperature record map of Europe. *Science*, 332, 220-224, doi: 10.1126/science.1201224.

Beljaars, A.C.M. (1994) The parameterization of surface fluxes in large-scale models under free convection. *Quart. J. Roy. Meteor. Soc.*, 121, 255–270.

Bellprat, O., S. Kotlarski, D. Luthi & C. Schär (2012) Exploring perturbed physics ensembles in a regional climate model. *Journal of Climate*, 25, 4582-4599, doi: 10.1175/JCLI-D-11-00275.1.

Beniston, M., D. B. Stephenson, O. B. Christensen, C. A. T. Ferro, C. Frei, S. Goyette, K. Halsnaes, T. Holt, K. Jylha, B. Koffi, J. Palutikof, R. Scholl, T. Semmler & K. Woth (2007) Future extreme events in European climate: an exploration of regional climate model projections. *Clim. Change*, 81, 71-95 , doi: 10.1007/s10584-006-9226-z.

Benjamin, S. G., G. A. Grell, J. M. Brown & T. G. Smirnova (2004) Mesoscale weather prediction with RUC hybrid isentropic-terrain-following coordinate model. *Mon. Wea. Rev*, 132, 473-494.

Chou, M.–D. & M. J. Suarez (1999) A solar radiation parameterization for atmospheric studies. *NASA Tech. Memo 104606* 40, Greenbelt, Maryland.

Chuine, I., P. Yiou, N. Viovy, B. Seguin, V. Daux & E. L. Ladurie (2004) Historical phenology: grape ripening as a past climate indicator. *Nature*, 432, 289-290, doi: 10.1038/432289a.

Collins, W. D., P. J. Rasch, B. A. Boville, J. J. Hack, J. R. McCaa, D. L. Williamson, J. T. Kiehl, B. Brieglib, C. Bitz, S.-J. Lin, M. Zhang & Y. Dai (2004) Description of the NCAR Community Atmosphere Model (CAM 3.0). NCAR Tech. Note NCAR/TN–464+STR. 214 pp.

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

Della-Marta, P.M., M. R. Haylock, J. Luterbacher & H. Wanner (2007) Doubled length of western European summer heat waves since 1880. *J. Geophys. Res.*, 112, D15103, doi:10.1029/2007JD008510.

Déqué, M, D. P. Rowell, D. Luthi, F. Giorgi, J. H. Christensen, B. Rockel, D. Jacob, E. Kjellstrom, M. de Castro & B. van den Hurk (2007) An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Climatic Change*, 81, 53-70, doi:10.1007/s10584-006-9228-x.

Dole, R., M. Hoerling, J. Perlwitz, J. Eischeid, P. Pegion, T. Zhang, X.-W. Quan, T. Y. Xu & D. Murray (2011) Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.*, 38, L06702, doi:10.1029/2010GL046582.

Dudhia, J. (1996) A multi-layer soil temperature model for MM5. Sixth Annual PSU/NCAR Mesoscale Model Users’ Workshop. Boulder CO, July 1996, 49-50.

Evans, J. P., M. Ekstrom & F. Ji (2012) Evaluating the performance of a WRF physics ensemble over South-East Australia. *Clim. Dyn.*, 39, 1241-1258, doi:10.1007/s00382-011-1244-5.

Fischer, E. M., S. I. Seneviratne, D. Luthi & C. Schär (2007) Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophy. Res. Lett.*, 34, L06707, doi:10.1029/2006GL029068.

Fischer, E. M. & C. Schär (2010) Consistent geographical patterns of changes in high-impact European heat waves. *Nat. Geosci.*, 3, 398-403.

García-Díez, M., J. Fernández, L. Fita & C. Yague (2011) Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. *Q. J. R. Meteorol. Soc.*, 139, 501-514, doi:10.1002/qj.1976.

García-Díez, M., J. Fernández & R. Vautard (2014) An RCM multi-physics ensemble over Europe : Multi-variable evaluation to avoid error compensation. *Clim. Dyn.*, submitted.

García-herrera, R., J. Diaz, R. M. Trigo, J. Luterbacher & E. M. Fischer (2010) A review of the European summer heat wave of 2003. *Critical Reviews in Environmental Science and Technology*, 40, 267-306, doi: 10.1080/10643380802238137.

Gomez-Navarro, J. J., J. P. Montávez, S. Jerez, P. Jimenez-Guerrero & E. Zorita (2012) What is the role of the observational dataset in the evaluation and scoring of climate models? Geophys. Res. Lett. 39:L24701.

Grell, G. A. & D. Devenyi (2013) A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, 29, 1693, doi:10.1029/2002GL015311.

Grünwald, T. & C. Bernhofer (2007) A decade of carbon, water and energy flux measurements of an old spruce forest at the Anchor Station Tharandt. *Tellus,* 59B, 387–396.

Han, J., & H. Pan (2011) Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Wea. Forecasting*, 26, 520–533.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones & M. New (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *J. Geophys. Res.*, 113, D20119, doi: 10.1029/2008JD010201.

Herrera, S., L. Fita, J. Fernández & J. M. Gutierrez (2010) "Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel" J. Geophys. Res. 115:D21117.

Hong, S.–Y., & J.–O. J. Lim (2006a) The WRF single–moment 6–class microphysics scheme (WSM6). J. Korean Meteor. Soc., 42, 129–151.

Hong, S.-Y., Y. Noh, J. Dudhia (2006b) A new vertical diffusion package with an explicit treatment of entrainment processes. Mon. Wea. Rev., 134, 2318–2341.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough & W. D. Collins (2008) Radiative forcing by long–lived greenhouse gases: Calculations with the AER radiative transfer models. J. Geophys. Res., 113, D13103, doi:10.1029/2008JD009944.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, doi:10.1017/CBO9781107415324.

Jacob, D., J. Petersen, B. Eggert, A. Alias, O. B. Christensen, L. M. Bouwer, A. Braun, A. Colette, M. Deque, G. Georgievski, E. Georgopoulou, A. Gobiet, L. Menut, G. Nikulin, A. Haensler, N. Hempelmann, C. Jones, K. Keuler, S. Kovats, N. Kroner, S. Kotlarski, A. Kriegsmann, E. Martin, E. Van Meijgaard, C. Moseley, S. Pfeifer, S. Preuschmann, C. Radermacher, K. Radtke, D. Rechid, M. Rounsevell, P. Samuelsson, S. Somot, J. F. Soussana, C. Teichmann, R. Valentini, R. Vautard, B. Weber & P. Yiou (2014) EURO-CORDEX: new-high-resolution climate change projections for European impact research. *Regional Environmental Change*,14, 563-578.

Janjic, Z. I. (1994) The Step–Mountain Eta Coordinate Model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945.

Janjic, Z. I., (2002) Nonsingular implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model. *NCEP Office Note No. 437*, 61 pp., National Centers for Environmental Prediction, College Park, MD.

Jankov, I., W. A. Gallus, M. Segal, B. Shaw & S. E. Koch (2005) The impact of different

WRF model physical parameterizations and their interactions on warm season WCS rainfall. *Weather and Forecasting*, 20, 1048-1060, doi:10.1175/WAF888.1.

Kain, J. S. (2004) The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.*, 43, 170–181.

Koster, R. D., S. P. P. Mahanama, T. J. Yamada, G. Balsamo, A. A. Berg, M. Boisserie, P. A. Dirmeyer, F. J. Doblas-Reyes, G. Drewitt, C. T. Gordon, Z. Guo, J. H. Jeong, D. M. Lawrence, W. S. Lee, Z. Li, L. Luo, S. Malyshev, W. J. Merryfield, S. I. Seneviratne, T. Stanelle, B. J. J. M. Van den Hurk, F. Vitart & E. F. Wood (2010) Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.*, 37, L02402, doi: 10.1029/2009GL041677.

Kotlarski, S., K. Keuler, O. B. Christensen, A. Colette, M. Déqué, A. Gobiet, K. Goergen, D. Jacob, D. Lüthi, E. van Meijgaard, G. Nikulin, C. Schär, C. Teichmann, R. Vautard, K. Warrach-Sagi & V. Wulfmeyer (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geosci. Model Dev.*, 7, 1297-1333, doi:10.5194/gmd-7-1297-2014, 2014.

Kuglitsch, F. G., T. Toreti, E. Xoplaki, P. M. Della-marta, C. S. Zerefos, M. Turkes & J. Luterbacher (2010) Heat wave changes in the eastern Mediterranean since 1960. *Geophys. Res. Lett.*, 37, L04802, doi: 10.1029/2009GL041841.

Lenderink, G., A. van Ulden, B. van den Hurk, E. van Meijgaard, 2007, Summertime inter-annual temperature variability in an ensemble of regional model simulations: analysis of the surface energy budget. *Climatic Change*, 81:233–247

Luterbacher, J., S. J. Koenig, J. Franke, G. Van der Schrier, E. Zorita, A. Moberg, J. Jacobeit, P. M. Della-marta, M. Kuttel, E. Xoplaki, D. Wheeler, T. Rutishauser, M. Stossel, H. Wanner, R. Brazdil, P Dobrovolny, D. Camuffo, C. Bertolin, A. Van Engelen, F. J.

Gonzalez-Rouco, R. Wilson, C. Pfister, D. Limanowka, O. Nordli, L. Leijonhufvud, J. Soderberg, R. Allan, M. Barriendos, R. Glaser, D. Riemann, Z. Hao & C. S. Zerefos (2010) Circulation dynamics and its influence on European and Mediterranean January-April climate over the past half millennium: results and insights from instrumental data, documentary evidence and coupled climate models. *Climate change*, 101, 201-234, doi: 10.1007/s10584-009-9782-0.

Meehl, G.A. & C. Tebaldi (2003) More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, 305, 994-997, doi:10.1126/science.1098704.

Miralles, D. G., T. R. H. Holmes, R. A. M. De Jeu, J. H. Gash, A. G. C. A Meesters & A. J. Dolman (2011) Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.*, 15, 453-469, doi:10.5194/hess-15-453-2011.

Miralles, D. G., A. J. Teuling, C. C. van Heerwaarden & J. Vilà-Guerau de Arellano (2014) Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature Geosci*., 7, 345-349, doi:10.1038/ngeo2141.

Mooney, P. A., F. J. Mulligan & R. Fealy (2013) "Evaluation of the Sensitivity of the Weather Research and Forecasting Model to Parameterization Schemes for Regional Climates of Europe over the Period 1990–95", J. Clim. 26:1002-1017.

Morrison, H., G. Thompson & V. Tatarskii (2009) Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One– and Two–Moment Schemes. *Mon. Wea. Rev.*, **137**, 991–1007.

Nakanishi, M. & H. Niino (2006) An improved Mellor–Yamada level 3 model: its numerical stability and application to a regional prediction of advecting fog. *Bound. Layer Meteor.*, 119, 397–407.

Nakanishi, M. & H. Niino (2009) Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912.

Nikulin, G., E. Kjellstrom, U. Hansson, G. Strandberg and A. Ullerstig (2010) Evaluation and future projections of temperature, precipitation and wind extremes over Europe in an ensemble of regional climate simulations. *Tellus A*, 63, 41-55.

Omrani, H., P. Dobrinski, P & T. Dubos (2013) Optimal nudging strategies in regional climate modelling: investigation in a Big-Brother experiment over the European and Mediterranean regions. *Climate Dynamics*, 41, 2451-2470.

Orlowsky, B. & S. I. Seneviratne (2012) Global changes in extreme events: regional and seasonal dimension. *Climatic Change*, 110, 669-696, doi:10.1007/s10584-011-0122-9.

Pilegaard, K., A. Ibrom, M. S. Courtney, P. Hummerlshøj & N. O. Jensen (2011) Increasing net CO_2 uptake by a Danish beech forest during the period from 1996 to 2009. *Agricultural and Forest Meteorology*, 151, 934-946.

Pleim, J. E. (2007) A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part I: Model Description and Testing. *J. Appl. Meteor. Climatol.*, 46, 1383–1395.

Schär, C., P. L. Vidale, D. Luthi, C. Frei, C. Haberli, M. A. Liniger & C. Appenzeller (2004) The role of increasing temperature variability in European summer heatwaves. *Nature*, 427**,** 332-336, doi:10.1038/nature02300.

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

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

Seneviratne, S. I., N. Nicholls, D. Easterling, C. M. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, and X. Zhang, 2012: Changes in climate extremes and their impacts on the natural physical environment. In: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* [Field, C. B., V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, and P. M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 109-230.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang & J. G. Powers (2008) A description of the Advanced Research WRF version 3. NCAR Tech. Note 1–125, http://nldr.library.ucar.edu/collections/TECH-NOTE-000-000-000-855.

Stegehuis, A., R. Vautard, P. Ciais, A. J. Teuling, M. Jung & P. Yiou (2013) Summer temperatures in Europe and land heat fluxes in observation-based data and regional climate model simulations. *Clim. Dyn.*, 41, 455-477, doi:10.1007/s00382-012-1559-x.

Sukoriansky, S., B. Galperin, & V. Perov (2005) Application of a new spectral model of stratified turbulence to the atmospheric boundary layer over sea ice. *Bound.–Layer Meteor.*, 117, 231–257.

Teuling, A. J., M. Hirschi, A. Ohmura, M. Wild, M. Reichstein, P. Ciais, N. Buchmann, C. Ammann, L. Montagnani, A. D. Richardson, G. Wohlfahrt & S. I. Seneviratne (2009) A regional perspective on trends in continental evaporation. *Geophys. Res. Lett.*, 36, L02404, doi: 10.1029/2008GL036584.

Tewari, M., F. Chen, W. Wang, J. Dudhia, M. A. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel & R. H. Cuenca (2004) Implementation and verification of the unified NOAH land surface model in the WRF model. *20th conference on weather analysis and forecasting/16th conference on numerical weather prediction*, pp. 11–15. Seattle, WA, American Meteorological Society.

Thompson, G., P. R. Field, R. M. Rasmussen & W. D. Hall (2008) Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. *Mon. Wea. Rev.*, 136, 5095–5115.

Tiedtke, M., (1989) A comprehensive mass flux scheme for cumulus parameterization in large–scale models. *Mon. Wea. Rev.*, 117, 1779–1800.

Tingley, M. P. & P. Huybers (2013) Recent temperature extremes at high northern latitudes unprecedented in the past 600 years. *Nature*, 496, 201-205, doi: 10.1038/nature11969.

Van den Hurk, B., F. Doblas-Reyes, G. Balsamo, R. D. Koster, S. I. Seneviratne & H. Camargo (2012) Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Clim. Dyn.*, 38, 349-362, doi: 10.1007/s00382-010-0956-2.

Vautard, R., A. Gobiet, D. Jacob, M. Belda, A. Colette, M. Deque, J. Fernandez, M. García-Díez, K. Goergen, I. Guttler, T. Halenka, T. Karacostas, E. Katragkou, K. Keuler, S. Kotlarski, S. Mayer, E. Van Meijgaard, G. Nikulin, M. Patarcic, J. Scinocca, S. Sobolowski, M. Suklitsch, C. Teichmann, K. Warrach-Sagi, V. Wulfmeyer & P. Yiou (2013) The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Clim. Dyn.*, 41, 2555-2575, doi:10.1007/s00382-013-1714-z.

Vautard, R., F. Thias, I. Tobin, F.-M. Breon, J.-G. Devezeaux de Lavergne, A. Colette, P. Yiou & P. M. Ruti (2014) Regional climate model simulations indicate limited climatic impacts by operational and planned European wind farms. *Nature Communications*, 3196, doi:10.1038/ncomms4196.

Warrach-Sagi, K., T. Schwitalla, V. Wulfmeyer & H. S. Bauer (2013) Evaluation of a climate simulation in Europe based on the WRF-NOAH model system: precipitation in

Germany. *Clim. Dyn.*, 41, 755-774, doi:10.1007/s00382-013-1727-7.

Weisheimer, A., F. J. Doblas-Reyes, T. Jung & T. N. Palmer (2011) On the predictability of the extreme summer 2003 over Europe. *Geophys. Res. Lett.*, 38, L05704, doi: 10.1029/2010GL046455.

Wild, M., B. Trussel, A. Ohmura, C. N. Long, G. Konig-Langlo, E. G. Dutton & A. Tsvetkov (2009) Global dimming and brightening: An update beyond 2000. *J. Geophys. Res.*, 114, D00D13, doi: 10.1029/2008JD011382.

Wohlfahrt, G., S. Pilloni, L. Hörtnagl & A. Hamerle (2010) Estimating carbon dioxide fluxes from temperature mountain grasslands using broad-band vegetation indices. *Biogeosciences*, 7, 683-694.

Zhang, C., Y. Wang & K. Hamilton (2011) Improved representation of boundary layer clouds over the southeast pacific in ARW–WRF using a modified Tiedtke cumulus parameterization scheme. *Mon. Wea. Rev.*, 139, 3489–3513.

Table and figure captions

Table 1. Physics schemes used in this study (with references). All possible permutations are made, yielding a total of 216 simulations. The numbers in the table refer to the number the schemes have in the Weather Research and Forecasting (WRF) model.

Table 2. The five best performing ~~combinations~~ configurations of physics in ranked from the first to the fifth best.

Table 3. Cross-comparison between France 2003 and Russia 2010. The (5, 10, 15, 20 and 25) best simulations, when only using one heat wave to select the best configurations and vice versa, are taken and compared with their ranking for the other heat wave. If there would be no correlation between the two years, the average ranking would lay approximately at half of the total number of simulations for both years that lay within a first selection of 1K (column 8). In bold the rankings that are lower than this number. Because observations of radiation are lacking over Russia, we tested France with and without including radiation in the ranking.

Figure 1. Time series of daily land heat fluxes in 2003 from May to the end of August on three different FLUXNET sites, with latent heat flux (LH) on the first row, sensible heat flux (SH) on the second row, and evaporative fraction (EF – latent heat flux divided by the sum of latent and sensible heat flux) on the last row (DOY is day of year). The three columns represent three sites, with Neustift/Stubai (Austria – ATneu 47N, 11E) in the first column, Tharandt (Germany – DETha, 51N, 4E) in the second, and Soroe-LilleBogeskov (Denmark – DKsor, 66N, 11E) in the third column. Vegetation types on the three sites are respectively grassland (GRA), evergreen needleleaf forest (ENF), and deciduous broadleaf forest (DBF).

In grey all 216 simulations with the NOAH scheme. Observational data is shown in black (FLUXNET). The ~~solid light blue~~ green line is one configuration with NOAH, while the ~~blue dots~~blue line represents the same configuration but with RUC instead of NOAH.

Figure 2. Domains used in this study: France, Iberian Peninsula, Russia and Scandinavia.

Figure 3. Time series of daily mean temperature over France in 2003 (a) and 2007 (b) and Russia in 2010 (c). Every simulation is shown in gray and observations of E-OBS in black. The blue and red lines are the coldest and the warmest simulations over France during the heat wave. These lines have the same set of physics in all the figures (3, 4, 5). Figure d shows the simulated temperature min-max range during the heatwave of 2003 (1-15 August). The range is calculated as the difference between the warmest and the coldest simulation during the heat wave period between the 216 members of the ensemble.

Figure 4. Monthly precipitation over France in 2003 (a) and 2007 (b) and Russia 2010 (c). The boxplots show the extremes, 25th, 50th, and 75th percentiles. The blue and red dots are the coldest and the warmest simulations over France during the heat wave (as in figure 3).

Figure 5. Scatter plot of soil moisture content at July 31, and temperature in August. Every point is one simulation. Different colors and symbols represent different physics for convection (CU) (a), microphysics (MP) (b), radiation (RA) (c) and planetary boundary layer-surface (PBL-SF) (d).

Figure 6. Monthly radiation over France in 2003 (a) and 2007 (b); no radiation data being available in Russia for 2010. The boxplots show the extremes, 25th, 50th, and 75th percentiles.

The blue and red dots are the coldest and the warmest simulations over France during the heat wave (as in figure 3).

Figure 7. Daily time series of temperature (a) and latent heat flux (c); monthly time series of precipitation (b) and incoming shortwave radiation (d). Observations are shown in black, and the five best performing runs in colors. Gray lines indicate other simulations. All figures are a spatial average over France during summer 2003.

Figure 8. Sensitivity test of the initialization of soil moisture. Difference between the perturbed ~~‘control’~~ simulations (red indicates 20% reduction of initial soil moisture, blue 20% enhancement) performed with ~~and the perturbed ones (minus (red) and plus (blue) 20% initial soil moisture) of~~ the five highest ranked configurations compared to their corresponding ‘control’ simulations. The darkest lines refer to ~~are~~ the ~~best~~ simulation~~s~~ conducted with the best ranked configuration (1), ~~and~~while ~~–~~descending colour shade agrees with descending ranking (1-5).

Table 1

Microphysics (MP)	PBL+Surface (PBL-SF)	Radiation (RA)	Convection (CU)	Soil
6) WRF-SM6 (Hong et al. 2006a)	1-1) Yonsei Uni-MM5 (Hong et al. 2006b; Beljaars, 1994)	3) CAM (Collins et al. 2004)	1) Kain-Fritsch (Kain 2004)	2) NOAH (Tewari et al. 2004)
8) New Thompson (Thompson et al. 2008)	2-2) MYJ-ETA (Janjic et al. 1994; Janjic, 2002)	4) RRTMG (Iacono et al. 2008)	3) Grell-Devenyi (Grell & Devenyi, 2012)	
10) Morrison DM (Morrison et al. 2009)	4-4) QNSE-QNSE (Sukoriansky et al. 2005)	5) Goddard (Chou & Suarez, 1999)	6) Tiedtke (Tiedtke 1989; Zhang et al. 2011)	
	5-2) MYNN-ETA (Nakanishi & Niino, 2006, 2009; Janjic, 2002)		14) New SAS (Han & Pan, 2011)	
	5-5) MYNN-MYNN (Nakanishi & Niino, 2006, 2009)			

	7-1) ACM2-MM5 (Pleim 2007; Beljaars, 1994)			

687

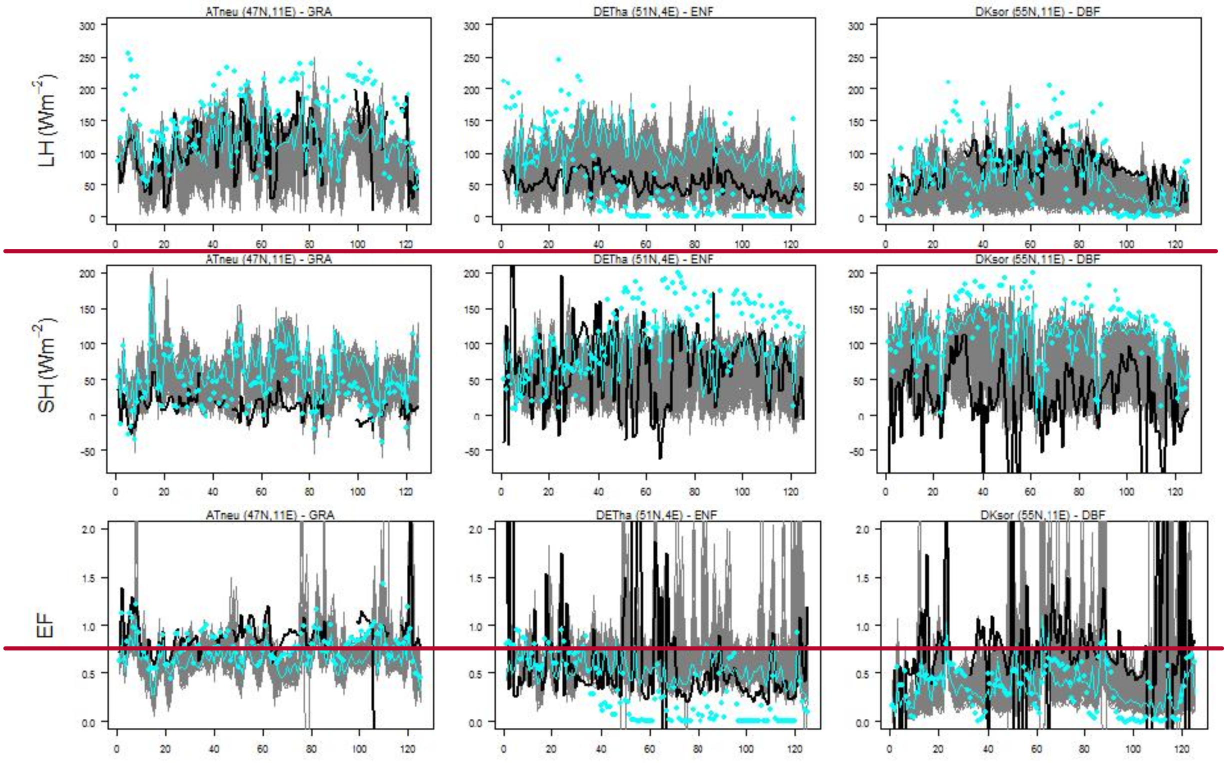
Table 2

Microphysics	PBL-Surface	Radiation	Convection	Soil	Rank
Morrison DM	Yonsei Uni-MM5	RRTMG	Tiedtke	NOAH	1
WRF-SM6	MYNN-MYNN	RRTMG	Grell-Devenyi	NOAH	2
WRF-SM6	ACM2-MM5	Goddard	Tiedtke	NOAH	3
New Thompson	MYNN-MYNN	RRTMG	New SAS	NOAH	4
New Thompson	ACM2-MM5	RRTMG	Tiedtke	NOAH	5

Table 3

		Average ranking of 5, 10, 15, 20 and 25 best simulations					
		5	10	15	20	25	Number of simulations within 1ºC
With radiation	Average rank Fr-Ru	**22.6**	**21.8**	**25.3**	**23.1**	**27.5**	104
With radiation	Average rank Ru-Fr	**15.75**	**15.2**	**14.7**	**13**	39.3	58
Without radiation	Average rank Fr-Ru	53	**37**	**28.4**	**27.6**	**25.5**	104
Without radiation	Average rank Ru-Fr	**20.25**	**16.8**	**18.1**	**17**	**19.9**	58

Figure 1


ATneu (47N,11E) - GRA
DETha (51N,4E) - ENF
DKsor (55N,11E) - DBF
LH (Wm^{-2})
SH (Wm^{-2})
EF
0
50
100
150
200
250
300
-50
0.0
0.5
1.0
1.5
2.0
0
20
40
60
80
100
120

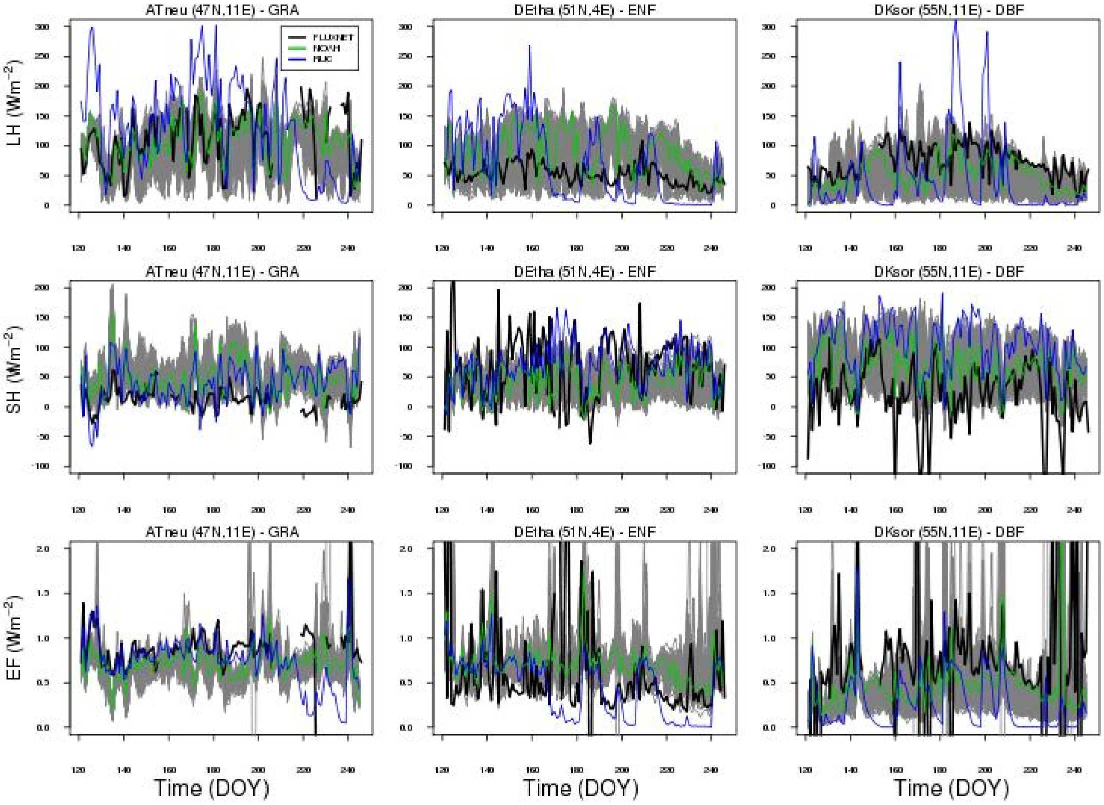

ATneu (47N,11E) - GRA
DEtha (51N,4E) - ENF
DKsor (55N,11E) - DBF
FLUXNET
NOAH
RUC
LH (Wm^{-2})
SH (Wm^{-2})
EF (Wm^{-2})
0
50
100
150
200
250
300
-100
-50
0.0
0.5
1.0
1.5
2.0
120
140
160
180
200
220
240
Time (DOY)

Figure 2

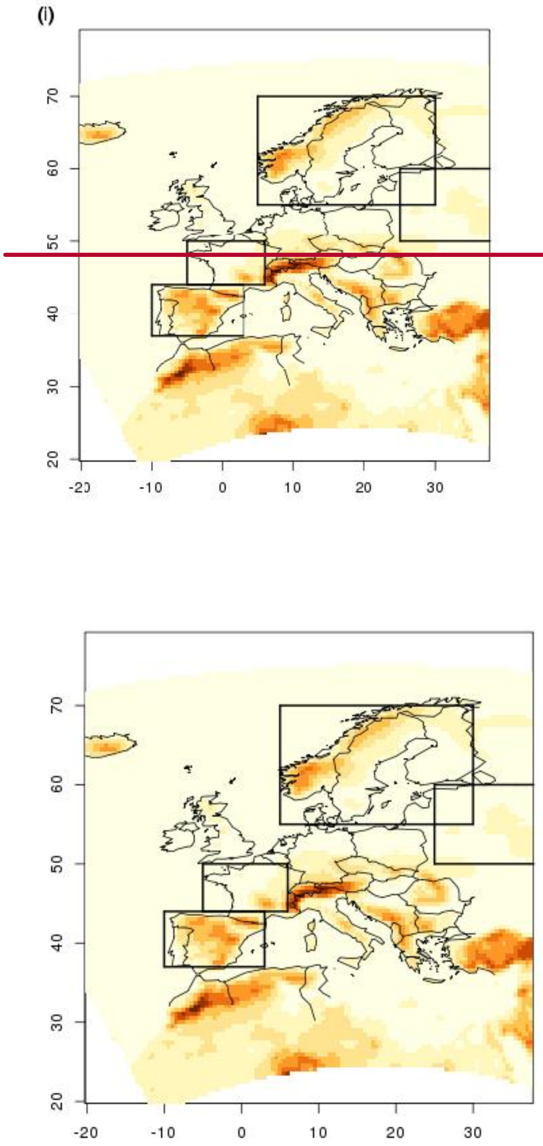


Figure 3

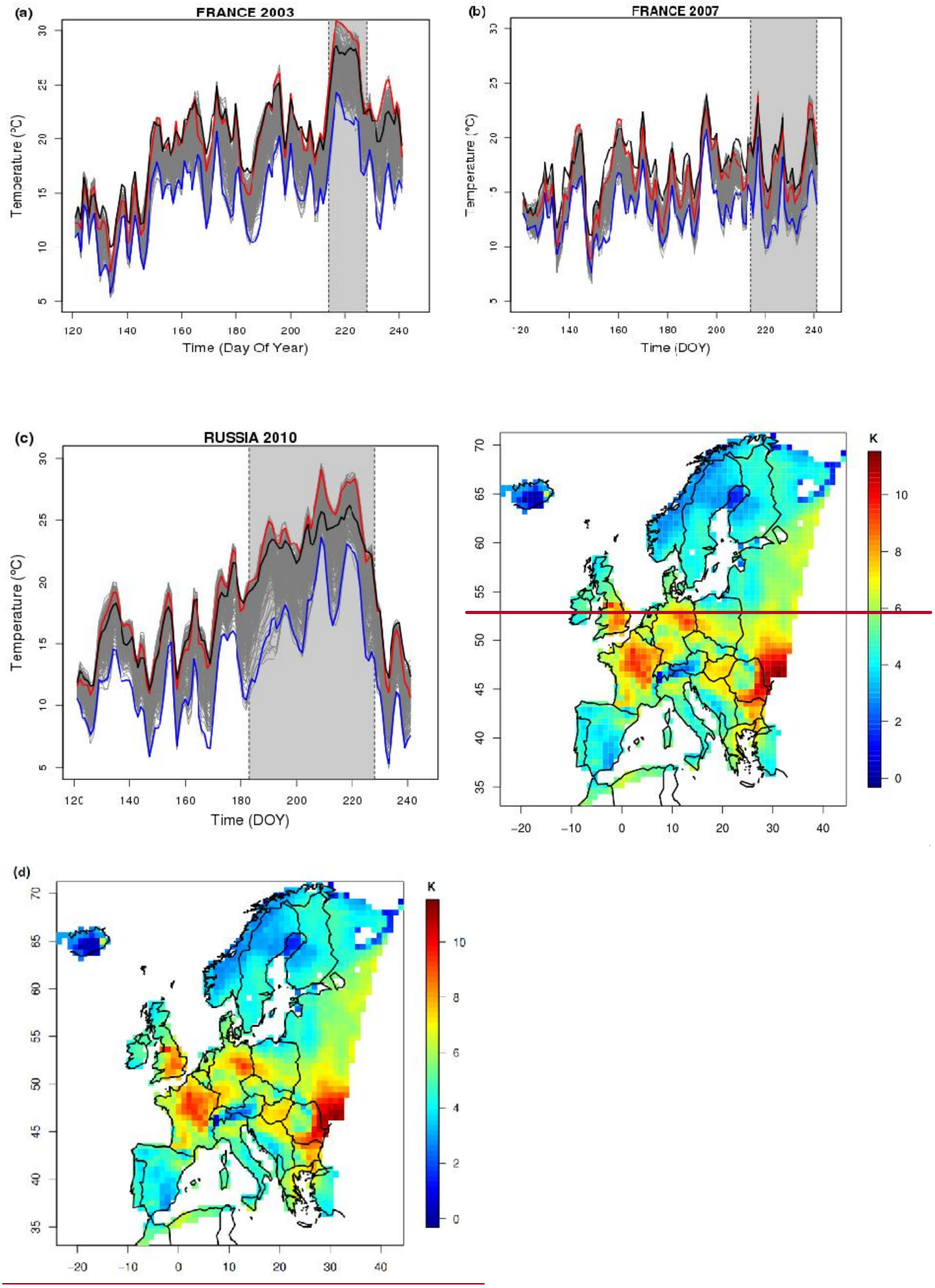
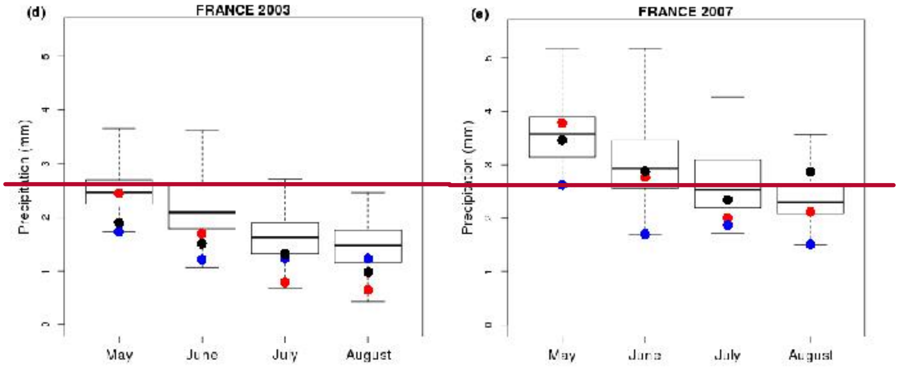
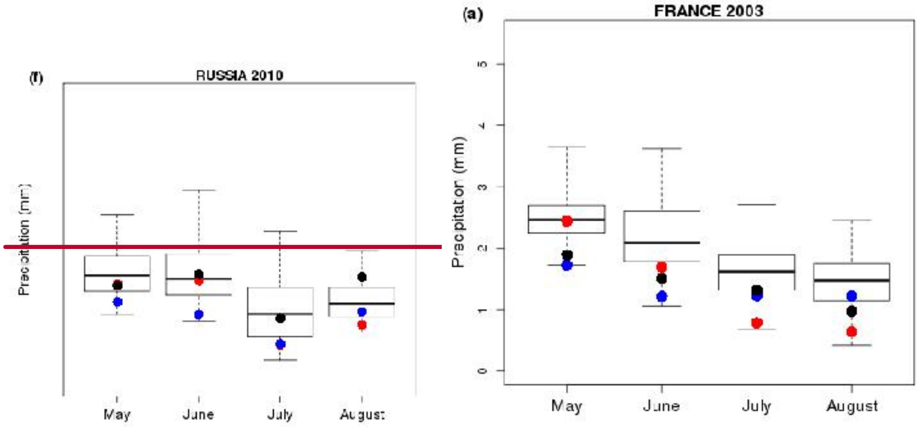

(a)
FRANCE 2003
Temperature (°C)
Time (Day Of Year)
(b)
FRANCE 2007
Temperature (°C)
Time (DOY)
(c)
RUSSIA 2010
Temperature (°C)
Time (DOY)
K
(d)
K

Figure 4a-c


(d)
FRANCE 2003
Precipitation (mm)
0
1
2
3
4
5
May
June
July
August
(e)
FRANCE 2007
Precipitation (mm)
0
1
2
3
4
5
May
June
July
August


(f)
RUSSIA 2010
Precipitation (mm)
May
June
July
August
(a)
FRANCE 2003
Precipitation (mm)
0
1
2
3
4
5
May
June
July
August

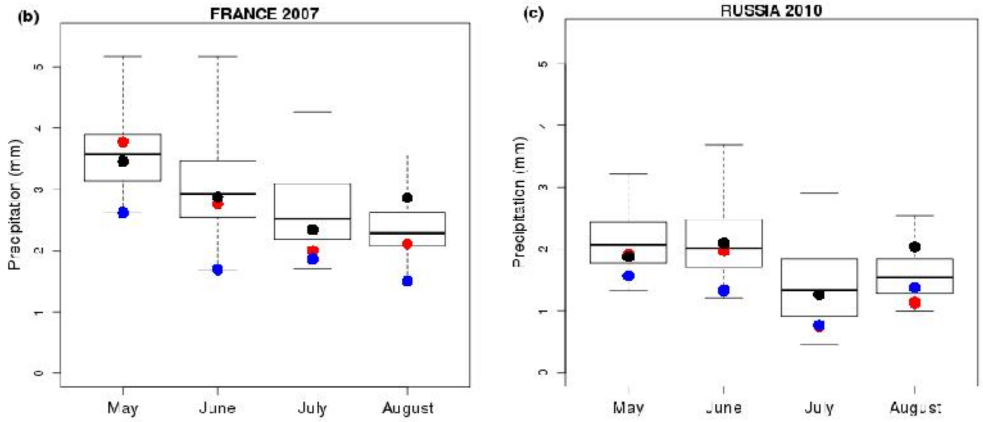
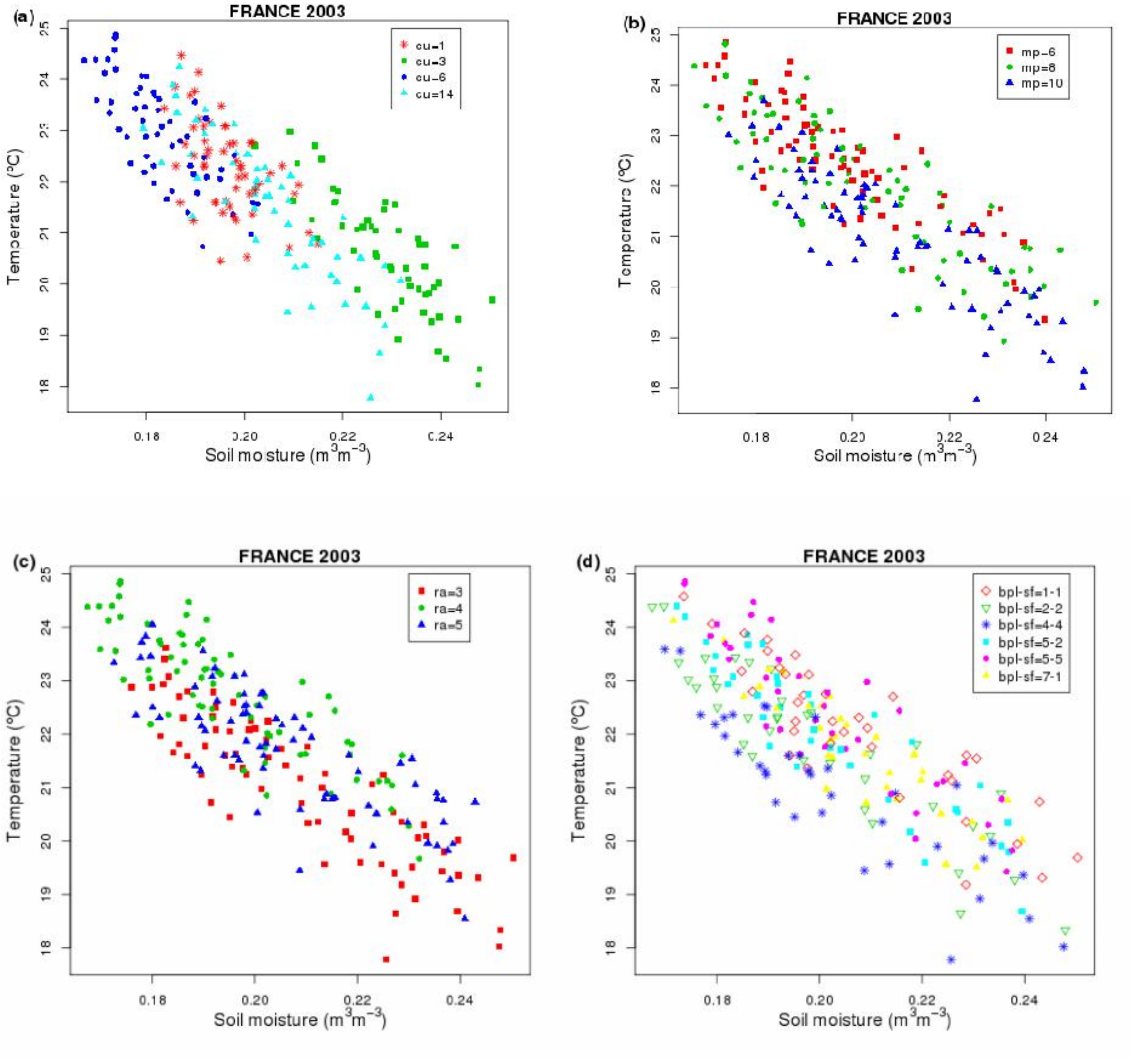
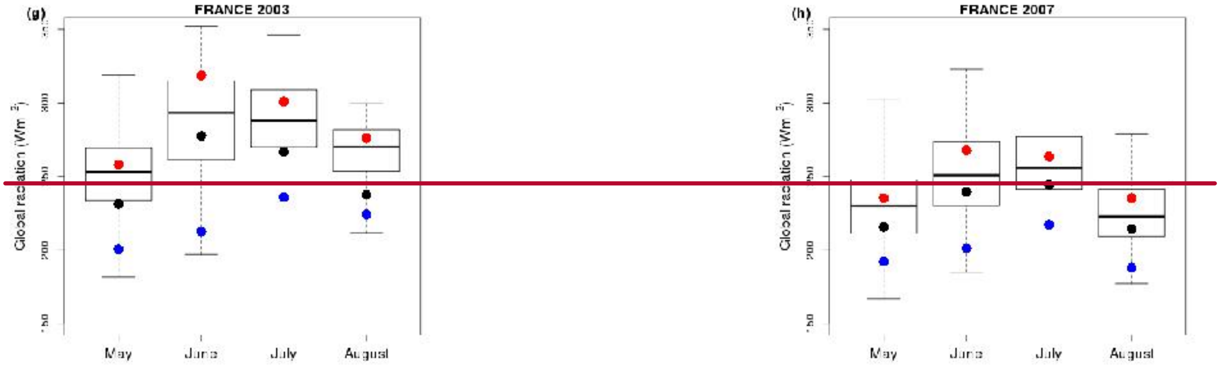

(b)
FRANCE 2007
Precipitation (mm)
0
1
2
3
4
5
May
June
July
August
(c)
RUSSIA 2010
Precipitation (mm)
0
1
2
3
4
5
May
June
July
August

Figure 5a-d



(a)
FRANCE 2003
cu=1
cu=3
cu=6
cu=14
Temperature (ºC)
Soil moisture (m^{3}m^{-3})
(b)
FRANCE 2003
mp=6
mp=8
mp=10
(c)
FRANCE 2003
ra=3
ra=4
ra=5
(d)
FRANCE 2003
bpl-sf=1-1
bpl-sf=2-2
bpl-sf=4-4
bpl-sf=5-2
bpl-sf=5-5
bpl-sf=7-1
18
19
20
21
22
23
24
25
0.18
0.20
0.22
0.24

Figure 6a-b


(g)
FRANCE 2003
Global radiation (Wm⁻²)
May
June
July
August
(h)
FRANCE 2007
Global radiation (Wm⁻²)
May
June
July
August

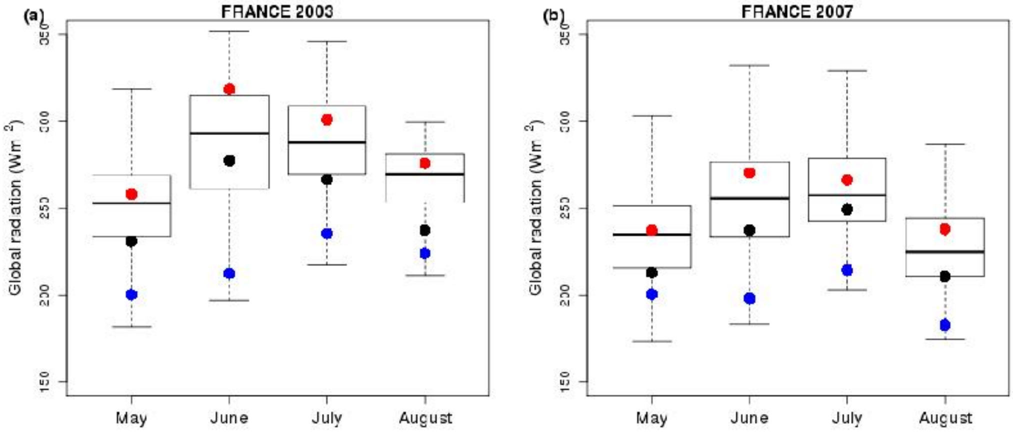

(a)
FRANCE 2003
Global radiation (Wm⁻²)
350
300
250
200
150
May
June
July
August
(b)
FRANCE 2007
Global radiation (Wm⁻²)
350
300
250
200
150
May
June
July
August

Figure 7a-d

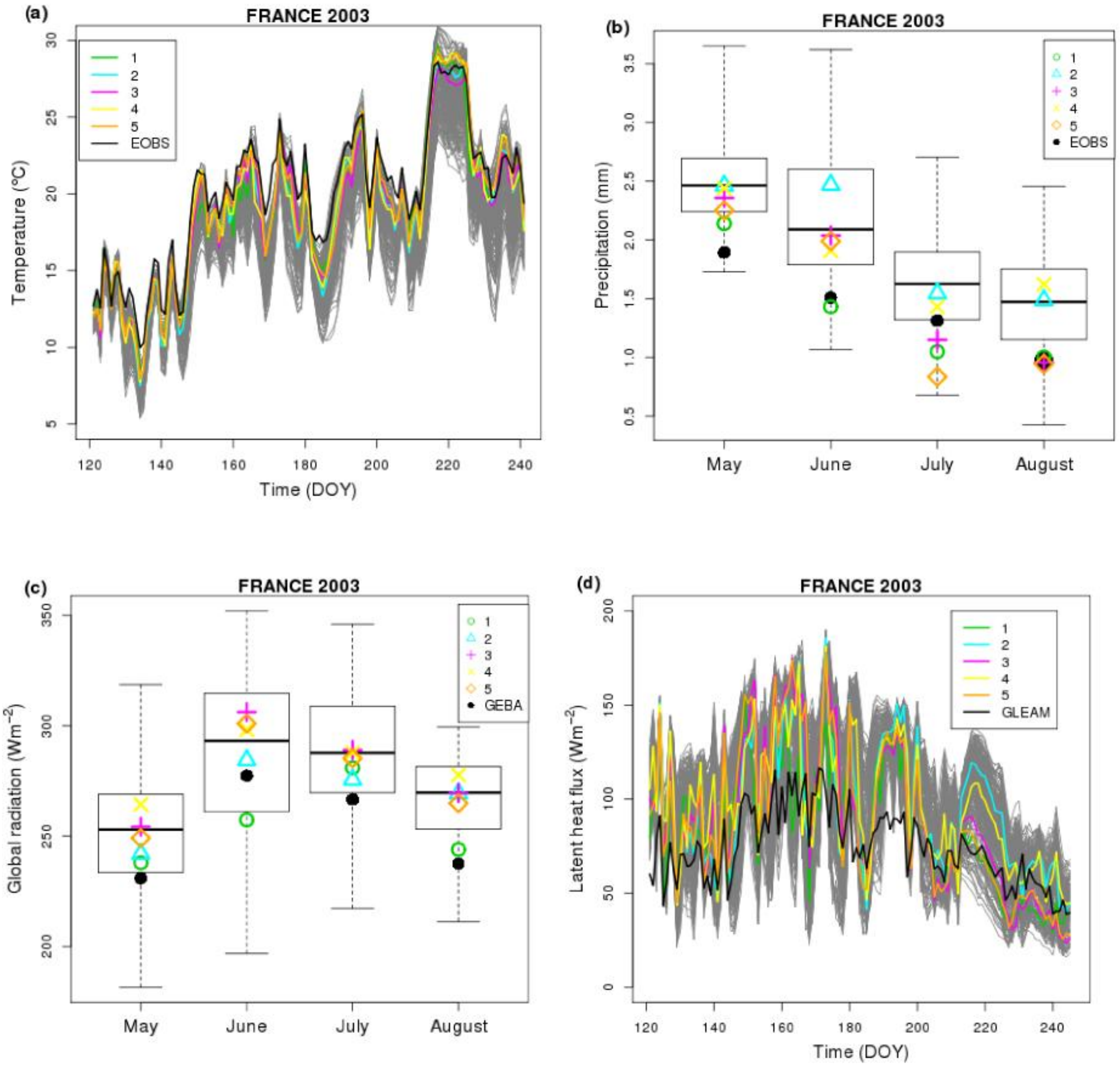
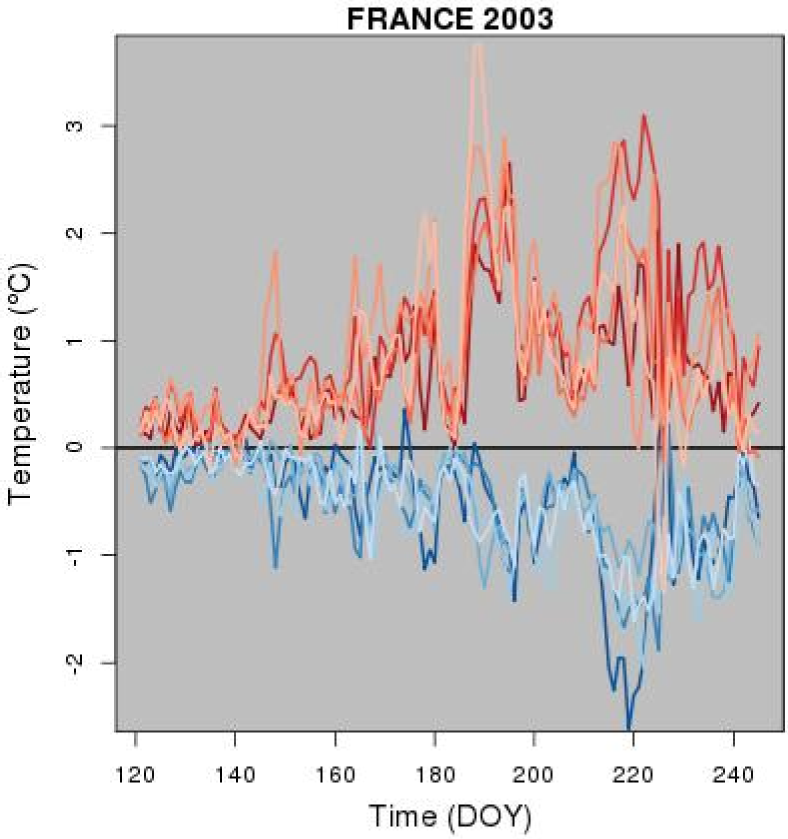

(a)
FRANCE 2003
Temperature (°C)
Time (DOY)
1
2
3
4
5
EOBS
5
10
15
20
25
30
120
140
160
180
200
220
240
(b)
FRANCE 2003
Precipitation (mm)
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1
2
3
4
5
EOBS
May
June
July
August
(c)
FRANCE 2003
Global radiation (Wm^{-2})
200
250
300
350
1
2
3
4
5
GEBA
May
June
July
August
(d)
FRANCE 2003
Latent heat flux (Wm^{-2})
Time (DOY)
0
50
100
150
200
1
2
3
4
5
GLEAM
120
140
160
180
200
220
240

Figure 8


FRANCE 2003
Temperature (°C)
3
2
1
0
-1
-2
120
140
160
180
200
220
240
Time (DOY)