The ExtremeX global climate model experiment: Investigating thermodynamic and dynamic processes contributing to weather and climate extremes

By Kathrin Wehrli, Fei Luo, Mathias Hauser, Hideo Shiogama, Daisuke Tokuda, Hyungjun Kim, Dim Coumou, Wilhelm May, Philippe Le Sager, Frank Selten, Olivia Martius, Robert Vautard, and Sonia I. Seneviratne

Reply to reviewer 1 (Paul Dirmeyer)

The modeling experiments described here shed light on the various roles of land versus atmosphere in extremes, going a step or two beyond what was done in the 1990s and 2000s in the "Koster style" studies of those days. It is interesting, adds to our scientific knowledge of climate variability, and should be published after revision. I do not wish to remain anonymous - Paul Dirmeyer

We thank Paul Dirmeyer for his thoughtful and positive evaluation of the manuscript and the helpful suggestions. In the following we will give answers to the comments in blue.

General comments:

1. Realizing this may be difficult without redesigning and rerunning the simulations, but I long to see a bit more separation in the various drivers, e.g., in the atmospheric component, could the roles of dynamics (circulation) versus physics (radiation, clouds, precipitation) be separated? At the land surface, could drivers acting through the energy balance terms versus water balance be quantified separately? Others have delved more into the process level (e.g., https://doi.org/10.1029/2012GL053703), and having models in hand for sensitivity studies enables many possibilities. Likely for "future work", but I wanted to bring up this question.
2. I greatly appreciate the message of the paper regarding the role of compensating errors and tuning. There remains among many in the model development community a strange hope that "fixing" one component of an Earth system model (e.g., upgrading the LSM) will somehow solve other problems. But often it just serves to expose those problems even more as the balance of errors has been disturbed. This paper also shines a bit more light on this issue.
3. Mainly in §5.1 but also conclusions: The conventional wisdom is that persistent anomalies in the atmospheric general circulation (which may have various causes themselves) establish conditions for heat waves and/or droughts, and then land-atmosphere feedbacks can exacerbate or prolong them. Is there any way to diagnose (confront or confirm) this idea from these experiments? Can the role of climate change on this evolutionary sequence be investigated here? These analyses are co-temporal and do not seem to account for the evolution over time of heat wave events, although you do consider persistence. It seems the two "approaches" (A) and (B) get at this somewhat (e.g. L343-344) but it is somewhat elusive.
4. There are a couple of recent papers that are quite germane to ExtremeX, particularly the notion that heat waves have a mix of land and atmosphere (which may ultimately be traced to remote ocean) drivers: https://doi.org/10.1029/2020AV000283, https://doi.org/10.1002/asl.948.

Thank you for the thoughts and ideas for the manuscript. We also appreciate the reference for the two additional papers in the fourth comment. We will mention these relevant and very recent results in our introduction. Regarding a further separation of drivers this is certainly a very relevant question and we agree that it would be great to have future studies going in this direction. We agree with the reviewer that this would require a new experiment design and the additional simulations would be pout of scope for the present work. Regarding a separation of the processes at the land surface we could think of experiments similar to those in Teng et al. (2019), where heating anomalies (from a dry simulation) are imposed in the atmospheric model.

As mentioned in the third comment the experiments are co-temporal. Hence, the ExtremeX setup is likely not ideal to confront or confirm whether circulation anomalies establish conditions for heatwaves or droughts and then land-surface feedbacks kick in by prolonging the events. Studies like Teng et al. (2019) and Martius et al. (2021) have shown that soil moisture anomalies can excite atmospheric circulation anomalies impacting the weather in other regions of the globe. Having experiments where the constraining of the soil moisture (or atmosphere) is confined to a certain period of time (and region) helps to isolate the processes and reduces other interactions. However, the soil moisture anomalies applied in the two studies mentioned would have to be created first which would likely be due to circulation anomalies. Going more deeply into this question would likely require dedicated case studies. Further, we think that the influence of climate change on the development of heatwaves/ droughts may be better investigated in fully coupled model simulations, potentially with a large ensemble to capture a sufficient number of events.

Reference:

Teng, H. Y., G. Branstator, A. B. Tawfik, and P. Callaghan, 2019: Circumglobal response to prescribed soil moisture over North America. J. Climate, 32, 4525–4545, https://doi.org/10.1175/JCLI-D-18-0823.1.

Martius, O., Wehrli, K., & Rohrer, M. (2021). Local and Remote Atmospheric Responses to Soil Moisture Anomalies in Australia, *Journal of Climate*, *34*(22), 9115-9131. Retrieved Dec 6, 2021, from https://journals.ametsoc.org/view/journals/clim/34/22/JCLI-D-21-0130.1.xml

Specific comments:

L75: Technical point: an ensemble of one is not an ensemble. It is just a single run.

This will be corrected.

Fig 1: It would be more clear to replot with the X-axis in a time dimension, e.g., label it as the e-folding (relaxation) time scale.

We think that for the manuscript it is more illustrative to keep the plot with the nudging intensity on the x-axis. However, we will add the formula used to compute the relaxation term to clear things up.

L111: Change "allows to isolate" to "allows isolation of".

This will be corrected.

L124-125: Which models nudged and which replace soil moisture states? And for those that nudged, what was the relaxation time scale?

Thank you for the question. It turned out that it was a misunderstanding among the modeling groups that in MIROC a soil moisture nudging was used. In fact, all models replace, hence prescribe, the simulated soil moisture. We will correct this throughout the manuscript, which will also simplify the terminology used.

L131: I think there was more than one version (combination of inputs) for the LandFlux-Eval data set for ET - which was used?

The mean from the merged ET synthesis product was used (hence their diagnostic, reanalysis and land surface model-based data sets). We will add this information to the description of the reference data sets.

§2.4: This would benefit from a schematic. Could you reproduce or recreate a figure based on Fig 1 of Wehrli et al. 2019? It would be very helpful. And doesn't differences in the results from approaches (A) and (B) shed light on the nonlinearities in the responses (evidence of feedbacks)?

A simplified figure based on the one in the 2019 paper will be added. Indeed, differences in the results following the two approaches show nonlinearities in the responses. We will add a sentence to mention this. The results from the two approaches were found to be qualitatively similar therefore we will not explore the differences.

L285-288: To this list should be added "unrepresented processes" in models, particularly those unresolved due to grid scale: non-hydrostatic atmospheric processes in coarse resolution models, unresolved mesoscale circulations, sub-grid surface heterogeneity.

We thank the reviewer for this suggestion and we will amend the list with processes unrepresented in models.

L288-289: Atmospheric modelers in particular are fixated on 500 hPa geopotential height errors as a metric of circulation fidelity.

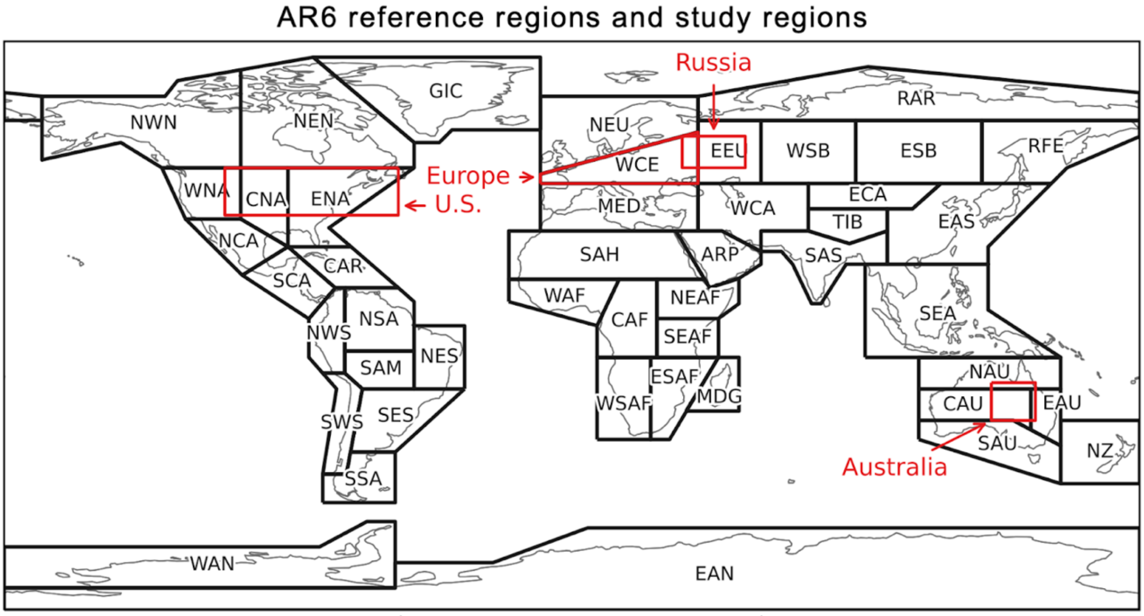
The predecessor papers Wehrli et al., 2018 and 2019 looked into the 500 hPa geopotential height for nudged CESM simulations. We will look into this also for the other models and describe the results for all three models in the manuscript.

Figs 5, 6 and associated text in §5.1: "Midwest" as a region name does not sit well in the global context, as it is a subregion of the U.S. In the other three cases, "Russia", "Europe" and "Australia" do not designate those entire areas, but a portion within each. Thus, "Midwest" should be replaced with "U.S."

The name of the region will be changed according to the reviewer's suggestion. For the events analysed in section 5.1 we will add that the U.S. heatwave was "also known as the Midwest heatwave" since the names of the events were chosen to match with existing literature.

For Russia (line 315) and the U.S. (line 336), how do these areas overlap or intersect the AR6 designated areas? Neither Fig A2 nor any of the other map plots in this manuscript show latitudes and longitudes, so it is difficult to compare by eye.

Figure A2 from the manuscript will be replaced by Figure R1 shown here. We will also update the figures to show longitude and latitudes where possible. Some figures already show rather small maps and we will make sure to not lose information or readability of the figures.



Abbrev.	Name	Abbrev.	Name	Abbrev.	Name
GIC	Greenland/Iceland	NEU	N.Europe	RFE	Russian-Far-East
NWN	N.W.North-America	WCE	West&Central-Europe	WCA	W.C.Asia
NEN	N.E.North-America	EEU	E.Europe	ECA	E.C.Asia
WNA	W.North-America	MED	Mediterranean	TIB	Tibetan-Plateau
CNA	C.North-America	SAH	Sahara	EAS	E.Asia
ENA	E.North-America	WAF	Western-Africa	ARP	Arabian-Peninsula
NCA	N.Central-America	CAF	Central-Africa	SAS	S.Asia
SCA	S.Central-America	NEAF	N.Eastern-Africa	SEA	S.E.Asia
CAR	Caribbean	SEAF	S.Eastern-Africa	NAU	N.Australia
NWS	N.W.South-America	WSAF	W.Southern-Africa	CAU	C.Australia
NSA	N.South-America	ESAF	E.Southern-Africa	EAU	E.Australia
NES	N.E.South-America	MDG	Madagascar	SAU	S.Australia
SAM	South-American-Monsoon	RAR	Russian-Arctic	NZ	New-Zealand
SWS	S.W.South-America	WSB	W.Siberia	EAN	E.Antarctica
SES	S.E.South-America	ESB	E.Siberia	WAN	W.Antarctica
SSA	S.South-America				

Figure R1: Reference regions of the IPCC AR6 as defined in Iturbide et al. (2020). Red outlines show the study regions considered in Section 5.1.

L353: You discuss results from MIROC, but what about the other two models?

The individual rations for MIROC were mentioned to highlight that both approaches lead to the conclusion that SM dominates over the atmospheric circulation contribution. However, we understand that it is confusing why the individual ratios for the other models are not mentioned and we will revise this paragraph.

Fig 7: There seems to a growing proportion of contribution from soil moisture as the anomaly periods grow longer (which would be reasonable, as locally soil moisture represents a slower manifold, a redder spectrum than tropospheric variables). It appears this could be easily quantified. Showing the area-weighted average of the metric in the figure (e.g., the SM-dominant percentage, averaged over unmasked areas only) in each panel would show a growing value with warm spell duration in each model, showing the growing relative importance of the land surface states for long-duration events (which would get at the "conventional wisdom" point above, to some degree).

We thank the reviewer for this suggestion and we will update the figure and description accordingly.

L419-420: Is this true? The atmospheric nudging is very weak in the lower troposphere, and other studies have shown the effect of land surface anomalies on the atmosphere is largely constrained to the boundary layer (e.g., https://doi.org/10.1175/1525-7541(2001)002%3C0329:AEOTSO%3E2.0.CO;2) except over elevated terrain where heating anomalies from the land surface can get into the upper troposphere directly (https://doi.org/10.5194/gmd-14-4465-2021).

As the reviewer mentioned, the effect of land surface anomalies on the atmosphere in general is local and constrained to the boundary layer. We will rephrase the lines in the manuscript to make clear that the present study does not disagree with this statement. In the second study mentioned nudging of the horizontal wind was used to initialize the model before perturbing the land surface temperature. In the present study horizontal wind is nudged for every model time step during the whole simulation period. Indeed, we found that there is only negligible variability between ensemble members due to the setup of the atmospheric nudging. This is not only true for horizontal winds in the free atmosphere as shown for CESM in Wehrli et al. (2018) but also for land surface conditions. We illustrate this in Figure R2 by showing the daily maximum temperature anomaly compared to the 1982-2008 climatology for the Russian heatwave as in Figure 5a) but for the AF_SI experiment instead of AF_SF.

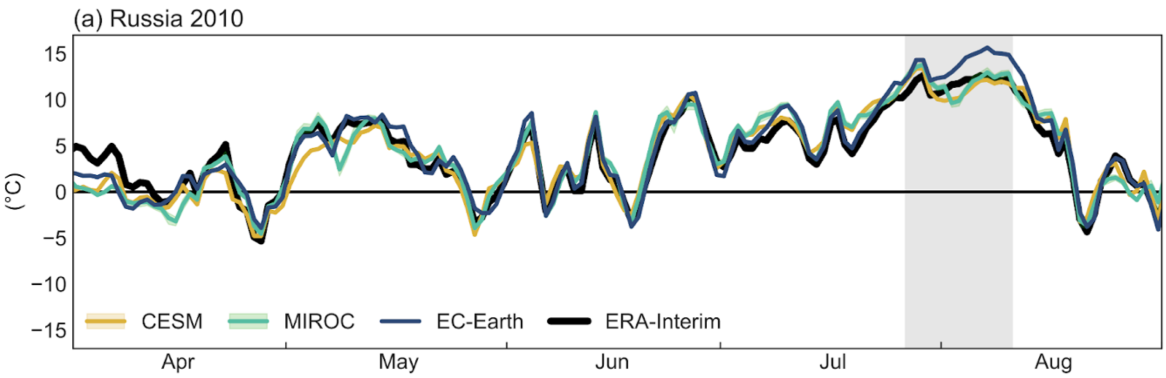


Figure R2: Daily maximum temperature anomaly compared to the 1982–2008 climatology for the nudging (AF_SI) experiment for the three models and for ERA-Interim (black line). The 15-day event period is highlighted in light grey. The shading shows the full ensemble spread and lines the ensemble mean (or single simulation for EC-Earth). The thick black line shows the values from ERA-Interim.

For CESM and MIROC five members for AF_SI were available and for EC-Earth only one. Figure R2 shows that the variability between the ensemble members is very small for both CESM and MIROC. For April to August (time period shown) daily standard deviation between ensemble members varies from 0.02°C to 0.19°C (0.07°C averaged over the whole time period) for CESM and 0.07°C to 0.47°C (0.22°C on average) for MIROC. The AF_SI experiment also captures the temporal evolution of TX anomaly similarly well as AF_SF. In Figure R3 the daily maximum temperature anomaly for all experiments is shown. For CESM (top) the AF_SI and AF_SF experiments barely differ while they do for MIROC (middle) and EC-Earth (bottom). This agrees with the findings from Figure 2 in the manuscript that AF_SI and AF_SF for CESM have a very similar climatology (and hence RMSE), which is not found for the other two models. This is due to the differences in how soil moisture was prescribed in the models and we will discuss this point in the manuscript.

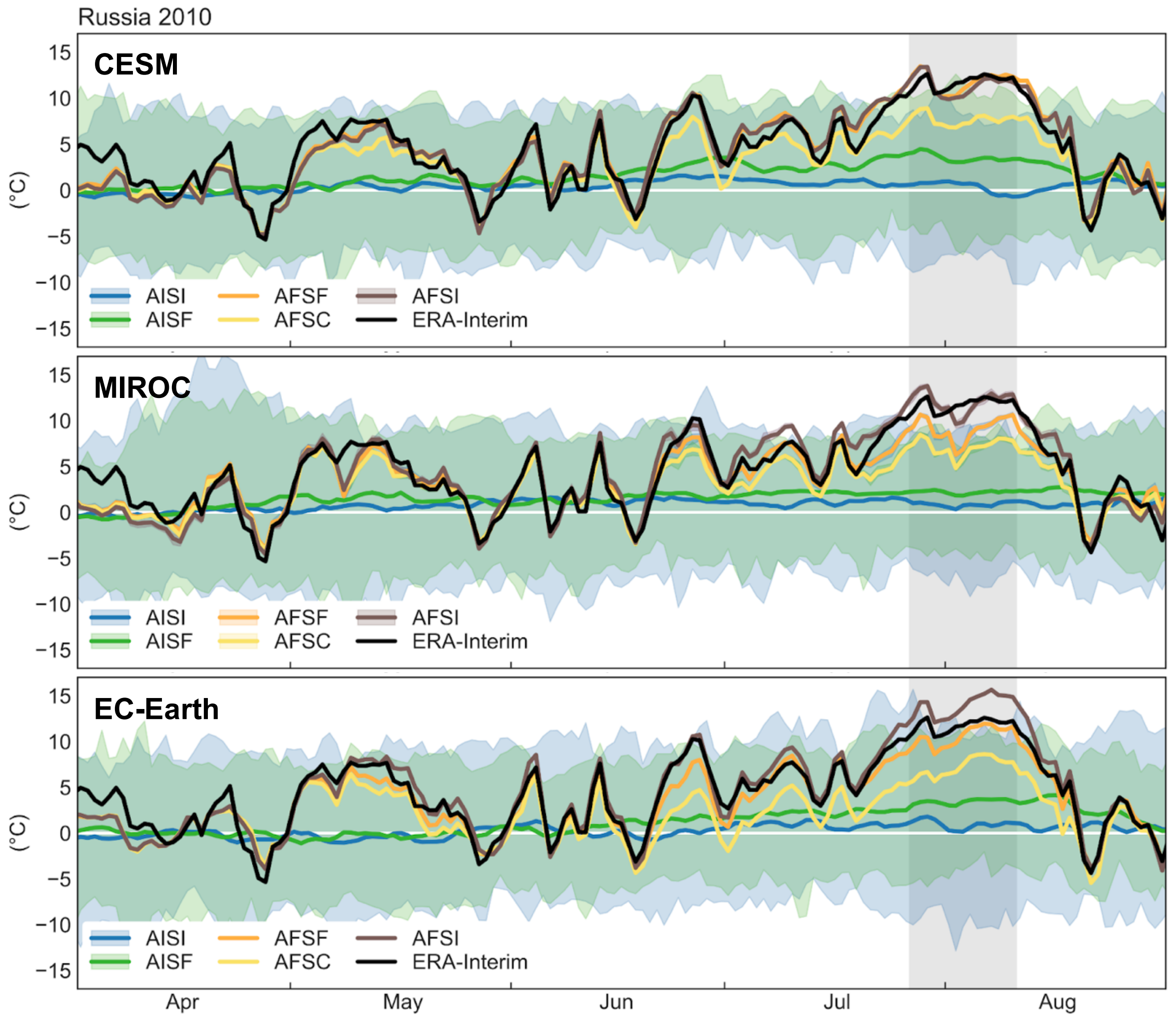


Figure R3: Daily maximum temperature anomaly compared to the 1982–2008 climatology for all experiments per model. Shown is the year 2010 and values are averaged for the region considered for the Russian heatwave. From top down: CESM, MIROC, EC-Earth. The shading shows the full ensemble spread and lines the ensemble mean (or single simulation). The black line shows the values from ERA-Interim.

All map figures: Since soil moisture as a climate driver has no meaning over (under) permanent ice, glacial areas like Greenland should be masked from the maps.

Yes, we agree and figures will be redone to mask out glaciated areas.

Code and data availability: This is not consistent with COPDESS / FAIR data standards to which EGU journals adhere. Public data and/or code repositories should be used and indicated with permanent hyperlinks.

We will make the relevant fields for the figures shown available with the revised manuscript.