Response to Paul Dirmeyer

The manuscript is much improved. But I am now finding that I am unsure about some aspects of the procedures (see below). I think this can be clarified rather easily with some improvements to the description of the methodology. Also, I think the material on climate trends can be removed – it is not contributing to the main goals of the paper. Otherwise, specific comments are given below.

A1: We thank Paul Dirmeyer for his additional feedback on our study.

General comments:
1. I think the part on climate trends (Sections 2.7, 3.3 and Figure 5) can be dropped. With only 20 years of data considered, the likelihood of separating signal from "noise" (interannual to decadal variability) is slim. The method chosen, comparing statistics calculated from two adjacent decades, which necessarily have first moments separated by only 10 years, unsurprisingly finds nothing. There are other approaches such as those that focus on trend detection in time series that would have been more efficacious, but probably still would not produce significant results with fewer than 30-40 years of data. Furthermore, the reasons given for what are insignificant trends are themselves only speculative. Unless the authors feel compelled to include climate change because it was the topic of the funding grant, for instance, I think the paper would be better without this distracting section.
A2: We agree that the part on climate trends may be less robust than our other main results. For this reason, we decided to move the trend results into the appendix (former Figure 5 is now Figure A9 and former section 2.7 is now the description of Figure A9) such that in the main text we only shortly summarize the main findings of the trend analysis in lines 297-301:

"Furthermore, we study potential changes in the relevance of the considered drivers of hot extremes between the periods 2001-2010 and 2011-2020 (Fig. A9). We find only small changes. While these could be related to natural decadal variability, we find that radiation and EVI are becoming slightly more relevant at both considered time scales, which may be also related to global greening (Zhu et al., 2016, Chen et al., 2019) and global brightening (Wild, 2009). At the same time, the changes are significant only in relatively small fractions (<10% for most drivers) of the study area."

2. As the description of the methodology in section 2.3 has been clarified, I realize there is an important detail that was not mentioned. Figure 1 suggests that a separate set of analogues is determined independently for each variable (geopotential, EF, etc.). So, over what spatial domains are analogues determined? I understand that goodness-of-fit is determined by RMSE (Euclidian distance), I think this must be between 2-D fields, so is the domain always global, or is it some radius centered on each grid cell in question? It seems that the analogues should be regional in nature. Or am I completely misunderstanding the procedure? This is causing me to doubt my understanding of Figure A1 as well – is each map a composite of the difference at each grid cell for each grid cell's spatial analogue? But then Euclidian distance

should always be positive, and we have negative values too, so no I am unsure of the calculation behind Figure A1.

A3: Thank you for pointing this out. Actually, analogues are determined independently for each variable in each grid cell. So, the analogues are from the same spatial domain (same grid cell).

Euclidian distance should result with positive values; however in our case, we only subtract the analogue values from the observed value, which can result in positive or negative differences.

We have removed the one dimensional Euclidian distance term in line 137, and have added a sentence to clarify that analogues come from the same grid cell where the hot extreme was identified in lines 142-143:

"The analogue analysis is done separately for each grid cell, i.e. analogues always come from the same grid cell where the hot extreme was detected."

Specific comments:

1. Abstract: "Analogues" are the tool of choice for this research, but the term is not mentioned in the abstract. Something about the methodology should be mentioned in the abstract.

A4: The term was actually mentioned in the abstract, but not explained. We have now expanded this part of the abstract a bit in lines 15-16:

"Hot extremes are identified at daily and weekly time scales through the highest absolute temperature, and the relevance of the considered drivers is determined with an analogue-based approach. Thereby, temperature anomalies are analyzed from situations with driver values similar to that of the hot extreme."

2. L128: Missing period.

A5: Done

3. Fig A1: This shows percentage differences, but that doesn't work well when, for some variables, the mean is much larger than the variance (e.g., 500hPa geopotential height). Wouldn't it make more sense to normalize the differences, e.g., locally by the overall (across 20 years) standard deviation of the field at each point?

A6: We agree that normalizing would solve this issue. The differences are now normalized by the local (20-year) standard deviation. We have updated Figure A1 with the figure below.

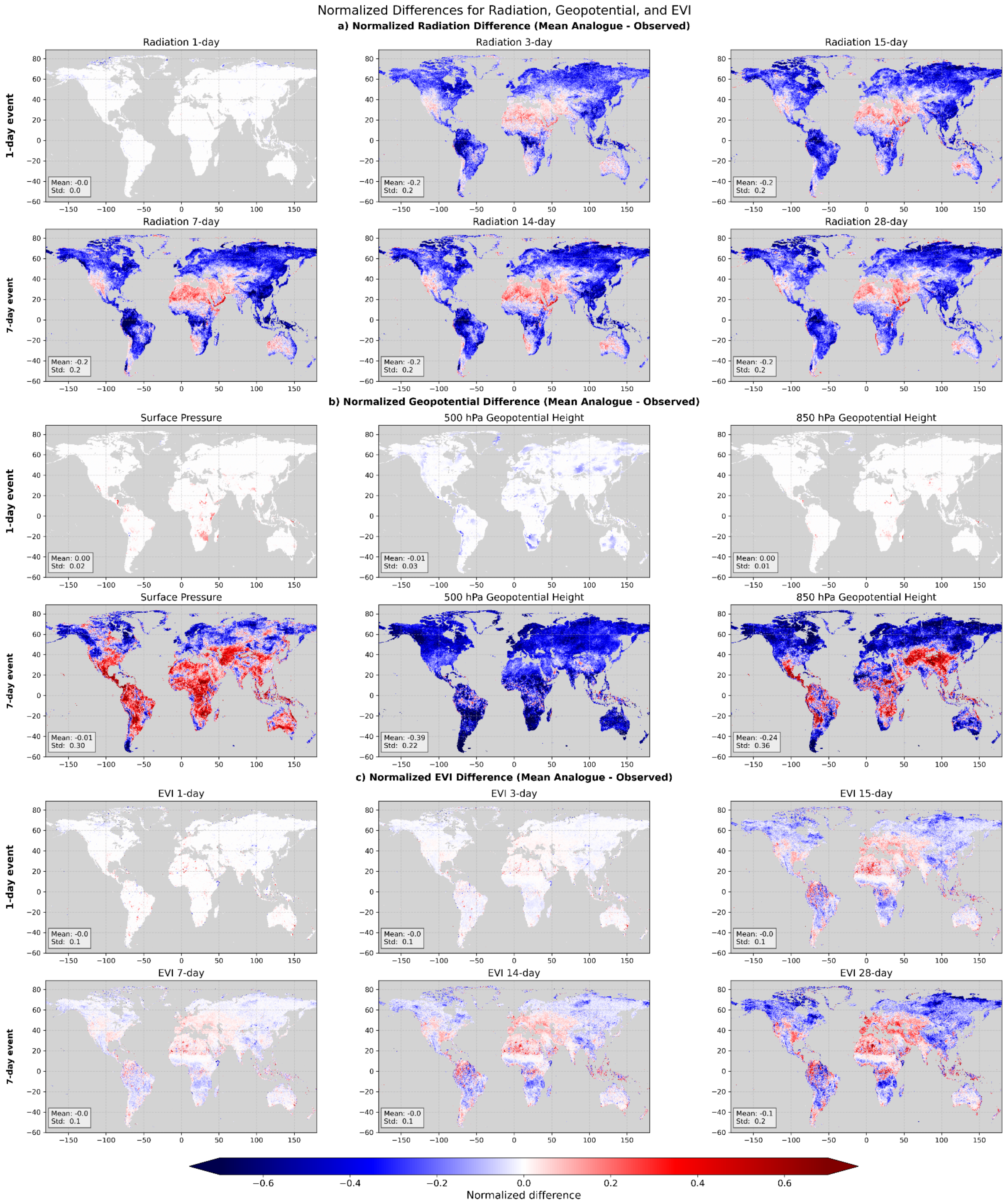


Figure A1. Spatial patterns of the normalized difference of geopotential height, radiation and EVI between the mean values of the five analogue periods and the values of the observed hot event divided by the local (20-year) standard deviation. Note that this figure shows the average differences across the three hottest periods. Mean and standard deviation values are denoted in the bottom left corner of each map

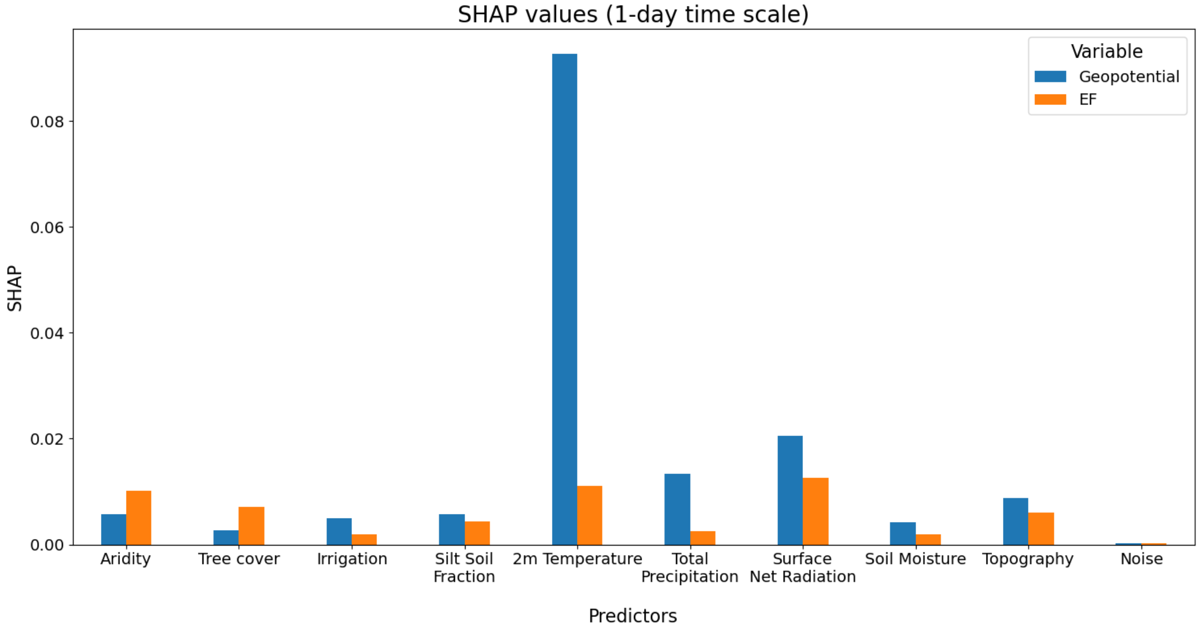
4. L177: change “cells” to “cell”

A7: Done

5. L192: I am not familiar with SHAP. Is there some threshold for significance that can be determined? What SHAP value would noise produce? Looking at Fig A3, I have no intuition of what values are important.

A8: SHAP values are a way to interpret the relative importance of each predictor for each target variable. In our case, aridity, tree cover, surface net radiation, soil moisture, and topography are among the predictors used, while geopotential height and EF serve as target variables. Since SHAP values only indicate the relative importance of each predictor, there is no significance threshold. Therefore, the importance of a SHAP value is best understood by comparing it to the values of other predictors within the same model.

We additionally plotted the same figure including a noise variable (see Figure R below), which was generated from a random normal distribution. This figure demonstrates that the noise predictor makes almost no contribution to explaining geopotential height and EF compared to the other predictors.



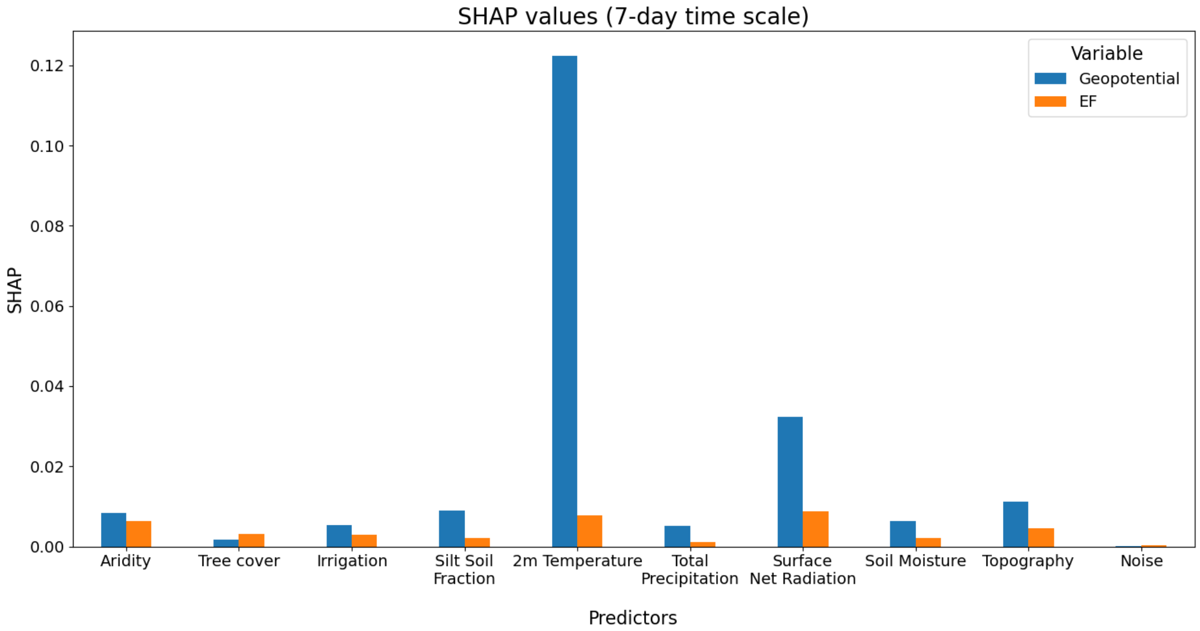


Figure R. Relative importance (Shapley Additive Explanations, SHAP values) of multiple factors to explain the spatial patterns of geopotential height and EF as main drivers for 1-day and 7-day hot extremes. Additionally, a noise predictor is added, which is an artificial data with a normal distribution.

6. L 291: Change "In a next step, we are determining…" to "In the next step, we determine…"
A9: Done

7. L296: Change "are grouping" to "group".
A10: Done

8. Table S1: I believe this should be in reference to Figure A4, not A2. However…
A11: Done

9. Figure A4: The colors are all so similar that it is impossible to tell much about spatial distributions beyond the purple vs brown. Only the percentages next to the colorbars are informative, and that information is in Table A1. With so many gradations (20 categories), it is not possible to produce readable high-resolution maps. If this information is vital to portray in map form, the authors need to consider a different method – it may require 3 to 6 maps for each time scale, with one or two variables (among six sets listed) per map. Otherwise, maybe only the table is necessary.
A12: We appreciate the reviewer's feedback. To improve clarity and readability, we have revised Figure A4 by reducing the complexity of the maps. Specifically, we plotted each variable group (e.g. geopotential height at different pressure levels) resulting with 6 maps for each time scale. This adjustment enhances the distinction between different variable groups and makes the spatial distribution patterns easier to interpret visually.

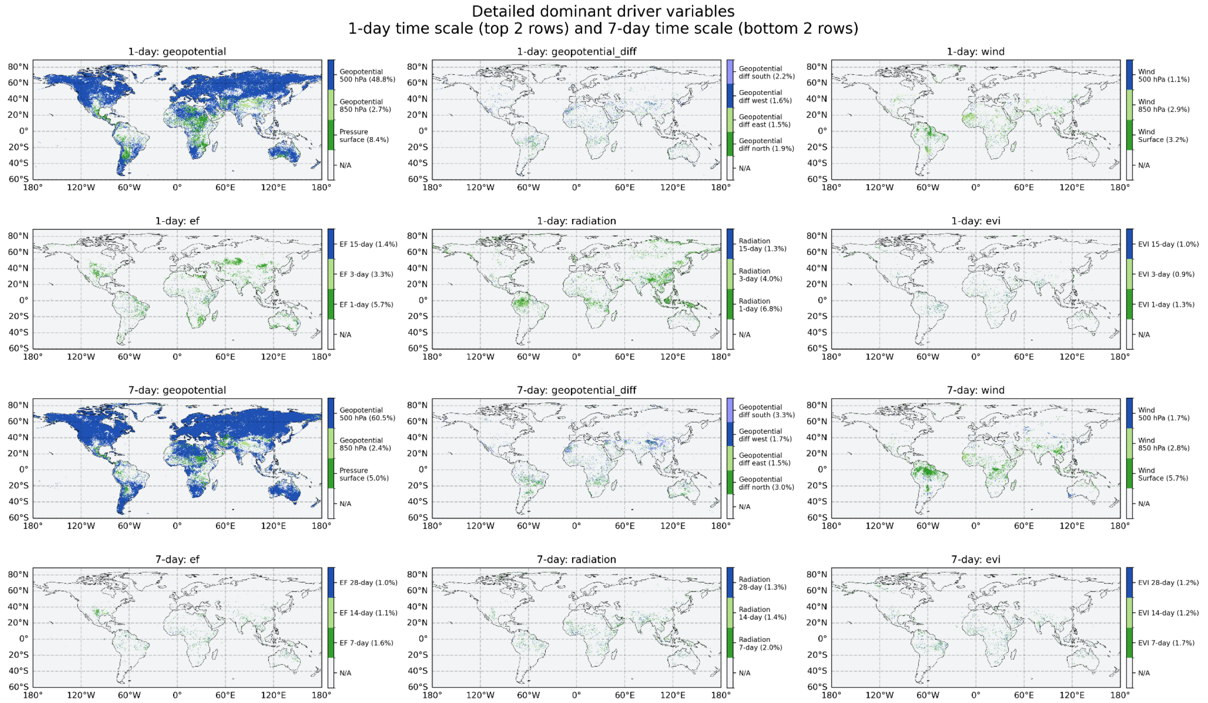
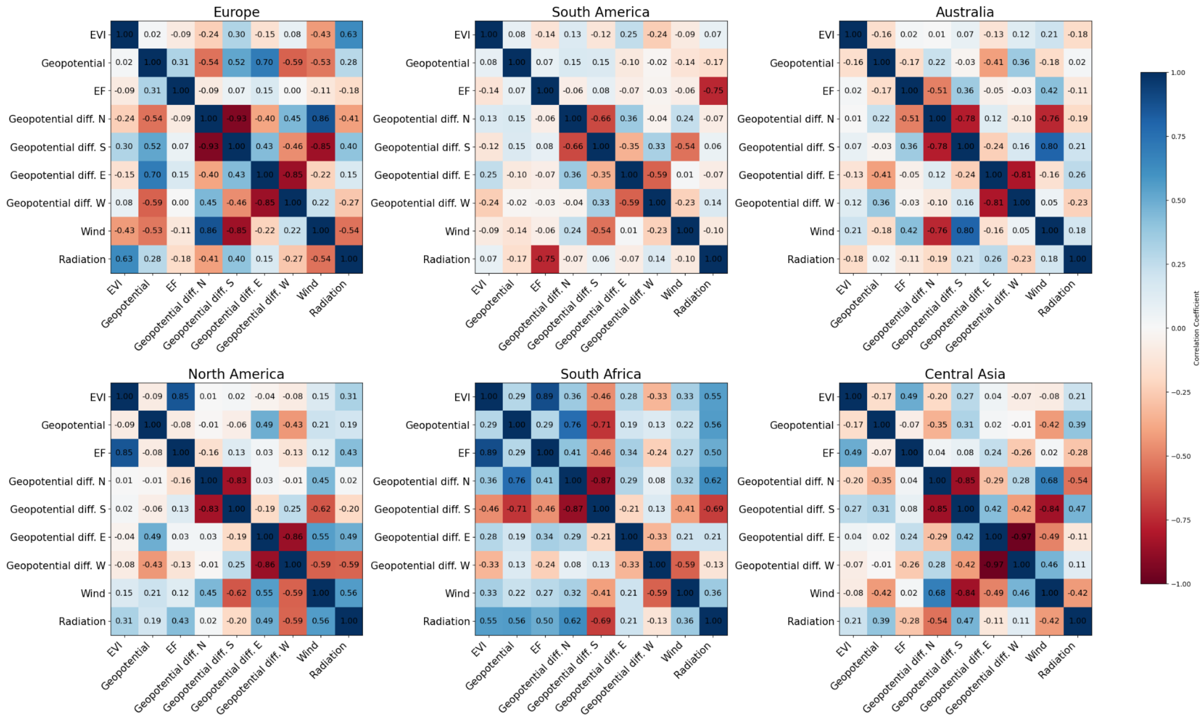


Figure A4. Detailed dominant driver variables identified for 1-day (top two rows) and 7-day (bottom two rows) time scales are shown for six variable groups: geopotential height, geopotential height difference, wind speed, EF, radiation, and EVI. Each colored grid cell indicates the dominant variable within the respective group. Grey grid cells (N/A) indicate areas where the dominant driver either belongs to a different variable group than the one currently plotted or has missing data. Percentages provided in parentheses on each colorbar indicate the area-weighted fraction of the total analyzed area where each variable is identified as the dominant driver. These percentages represent the area over which each variable is dominant when considering all variables collectively within each time scale separately. These results are summarized in Table A5.

10. Figure A6: I do not understand how Wind and Geopotential difference can be strongly anticorrelated in Australia, but positively correlated elsewhere. Is there perhaps some factor in calculating geopotential difference that changes sign in the Southern Hemisphere (e.g., the Coriolis term) that creates this inconsistency? But then why is it not also present for South Africa and South America? I find this puzzling.

A13: We previously used only northward geopotential height differences in that figure, but we have now included all other geopotential height differences. In this updated figure, we observe that the negative correlation identified earlier might be related to a strong positive correlation in the opposite direction, influenced by local conditions (e.g., aridity and topography) specific to this site and not representative of other Southern Hemisphere locations. Furthermore, since the geopotential height differences are computed at 500 hPa while wind speed is measured at the surface, we suggest that a strong geopotential height gradient at 500 hPa does not necessarily translate to strong surface wind speeds. This discrepancy could be attributed to local land surface characteristics such as topography and vegetation cover.



Europe

	EVI	Geopotential	EF	Geopotential diff. N	Geopotential diff. S	Geopotential diff. E	Geopotential diff. W	Wind	Radiation
EVI	1.00	0.02	-0.09	-0.24	0.30	-0.15	0.08	-0.43	0.63
Geopotential	0.02	1.00	0.31	-0.54	0.52	0.70	-0.59	-0.53	0.28
EF	-0.09	0.31	1.00	-0.09	0.07	0.15	0.00	-0.11	-0.18
Geopotential diff. N	-0.24	-0.54	-0.09	1.00	-0.93	-0.40	0.45	0.86	-0.41
Geopotential diff. S	0.30	0.52	0.07	-0.93	1.00	0.43	-0.46	-0.85	0.40
Geopotential diff. E	-0.15	0.70	0.15	-0.40	0.43	1.00	-0.85	-0.22	0.15
Geopotential diff. W	0.08	-0.59	0.00	0.45	-0.46	-0.85	1.00	0.22	-0.27
Wind	-0.43	-0.53	-0.11	0.86	-0.85	-0.22	0.22	1.00	-0.54
Radiation	0.63	0.28	-0.18	-0.41	0.40	0.15	-0.27	-0.54	1.00

South America

	EVI	Geopotential	EF	Geopotential diff. N	Geopotential diff. S	Geopotential diff. E	Geopotential diff. W	Wind	Radiation
EVI	1.00	0.08	-0.14	0.13	-0.12	0.25	-0.24	-0.09	0.07
Geopotential	0.08	1.00	0.07	0.15	0.15	-0.10	-0.02	-0.14	-0.17
EF	-0.14	0.07	1.00	-0.06	0.08	-0.07	-0.03	-0.06	-0.75
Geopotential diff. N	0.13	0.15	-0.06	1.00	-0.66	0.36	-0.04	0.24	-0.07
Geopotential diff. S	-0.12	0.15	0.08	-0.66	1.00	-0.35	0.33	-0.54	0.06
Geopotential diff. E	0.25	-0.10	-0.07	0.36	-0.35	1.00	-0.59	0.01	-0.07
Geopotential diff. W	-0.24	-0.02	-0.03	-0.04	0.33	-0.59	1.00	-0.23	0.14
Wind	-0.09	-0.14	-0.06	0.24	-0.54	0.01	-0.23	1.00	-0.10
Radiation	0.07	-0.17	-0.75	-0.07	0.06	-0.07	0.14	-0.10	1.00

Australia

	EVI	Geopotential	EF	Geopotential diff. N	Geopotential diff. S	Geopotential diff. E	Geopotential diff. W	Wind	Radiation
EVI	1.00	-0.16	0.02	0.01	0.07	-0.13	0.12	0.21	-0.18
Geopotential	-0.16	1.00	-0.17	0.22	-0.03	-0.41	0.36	-0.18	0.02
EF	0.02	-0.17	1.00	-0.51	0.36	-0.05	-0.03	0.42	-0.11
Geopotential diff. N	0.01	0.22	-0.51	1.00	-0.78	0.12	-0.10	-0.76	-0.19
Geopotential diff. S	0.07	-0.03	0.36	-0.78	1.00	-0.24	0.16	0.80	0.21
Geopotential diff. E	-0.13	-0.41	-0.05	0.12	-0.24	1.00	-0.81	-0.16	0.26
Geopotential diff. W	0.12	0.36	-0.03	-0.10	0.16	-0.81	1.00	0.05	-0.23
Wind	0.21	-0.18	0.42	-0.76	0.80	-0.16	0.05	1.00	0.18
Radiation	-0.18	0.02	-0.11	-0.19	0.21	0.26	-0.23	0.18	1.00

North America

	EVI	Geopotential	EF	Geopotential diff. N	Geopotential diff. S	Geopotential diff. E	Geopotential diff. W	Wind	Radiation
EVI	1.00	-0.09	0.85	0.01	0.02	-0.04	-0.08	0.15	0.31
Geopotential	-0.09	1.00	-0.08	-0.01	-0.06	0.49	-0.43	0.21	0.19
EF	0.85	-0.08	1.00	-0.16	0.13	0.03	-0.13	0.12	0.43
Geopotential diff. N	0.01	-0.01	-0.16	1.00	-0.83	0.03	-0.01	0.45	0.02
Geopotential diff. S	0.02	-0.06	0.13	-0.83	1.00	-0.19	0.25	-0.62	-0.20
Geopotential diff. E	-0.04	0.49	0.03	0.03	-0.19	1.00	-0.86	0.55	0.49
Geopotential diff. W	-0.08	-0.43	-0.13	-0.01	0.25	-0.86	1.00	-0.59	-0.59
Wind	0.15	0.21	0.12	0.45	-0.62	0.55	-0.59	1.00	0.56
Radiation	0.31	0.19	0.43	0.02	-0.20	0.49	-0.59	0.56	1.00

South Africa

	EVI	Geopotential	EF	Geopotential diff. N	Geopotential diff. S	Geopotential diff. E	Geopotential diff. W	Wind	Radiation
EVI	1.00	0.29	0.89	0.36	-0.46	0.28	-0.33	0.33	0.55
Geopotential	0.29	1.00	0.29	0.76	-0.71	0.19	0.13	0.22	0.56
EF	0.89	0.29	1.00	0.41	-0.46	0.34	-0.24	0.27	0.50
Geopotential diff. N	0.36	0.76	0.41	1.00	-0.87	0.29	0.08	0.32	0.62
Geopotential diff. S	-0.46	-0.71	-0.46	-0.87	1.00	-0.21	0.13	-0.41	-0.69
Geopotential diff. E	0.28	0.19	0.34	0.29	-0.21	1.00	-0.33	0.21	0.21
Geopotential diff. W	-0.33	0.13	-0.24	0.08	0.13	-0.33	1.00	-0.59	-0.13
Wind	0.33	0.22	0.27	0.32	-0.41	0.21	-0.59	1.00	0.36
Radiation	0.55	0.56	0.50	0.62	-0.69	0.21	-0.13	0.36	1.00

Central Asia

	EVI	Geopotential	EF	Geopotential diff. N	Geopotential diff. S	Geopotential diff. E	Geopotential diff. W	Wind	Radiation
EVI	1.00	-0.17	0.49	-0.20	0.27	0.04	-0.07	-0.08	0.21
Geopotential	-0.17	1.00	-0.07	-0.35	0.31	0.02	-0.01	-0.42	0.39
EF	0.49	-0.07	1.00	0.04	0.08	0.24	-0.26	0.02	-0.28
Geopotential diff. N	-0.20	-0.35	0.04	1.00	-0.85	-0.29	0.28	0.68	-0.54
Geopotential diff. S	0.27	0.31	0.08	-0.85	1.00	0.42	-0.42	-0.84	0.47
Geopotential diff. E	0.04	0.02	0.24	-0.29	0.42	1.00	-0.97	-0.49	-0.11
Geopotential diff. W	-0.07	-0.01	-0.26	0.28	-0.42	-0.97	1.00	0.46	0.11
Wind	-0.08	-0.42	0.02	0.68	-0.84	-0.49	0.46	1.00	-0.42
Radiation	0.21	0.39	-0.28	-0.54	0.47	-0.11	0.11	-0.42	1.00

Figure A7 Spearman cross-correlation matrix for the six land-surface (EVI, EF, radiation) and atmospheric variables (geopotential, wind, geopotential difference (north, south, east, west)), averaged across three subregions within each region (See Section 2.5)