6th International Conference on Energy and Environment: bringing together Engineering and Economics
Guimarães, Portugal
6-7 June, 2024

A Markov-chain model for assessing heatwaves and droughts in Iberian Peninsula

Ebenezer Takyi[1], Pedro G. Lind[1,2,3,4], Ana Russo[5,*]

1 Department of Computer Science, OsloMet – Oslo Metropolitan University, Pilestredet 52, Oslo, N-0130, Norway
2 AI Lab – OsloMet Artificial Intelligence Lab, Pilestredet 52, Oslo, N-0166, Norway
3 NordSTAR – Nordic Center for Sustainable and Trustworthy AI Research, Pilestredet 52, Oslo, N-0166, Norway
4 Simula Research Laboratory, Numerical Analysis and Scientific Computing, , Oslo, N-0164, Norway
5 University of Lisbon, Faculty of Sciences, IDL, 1749-016 Lisboa, Portugal
* Corresponding author: acrusso@fc.ul.pt, University of Lisbon 1749-016 Lisboa, Portugal

KEYWORDS

Markov-chains, Compound events, Extreme events

ABSTRACT

The Iberian Peninsula subregion is known for the increasing frequency and intensity of heatwaves and drought conditions, but a comprehensive understanding of the statistical dependencies between these events is still underexplored. This undermines effective climate adaptation and mitigation strategies. We analyse ERA5 reanalysis daily mean temperature and precipitation data to investigate the statistical dependencies between heatwaves and drought conditions for the period 1950–2022 in the Iberian Peninsula. Through statistical analysis of precipitation and temperature data into binary states classifying the occurrence of heatwaves and droughts, we construct Markov-chain models describing the temporal sequence of these extreme events throughout the Iberian Peninsula. Moreover, using the binary states we introduce a similarity measure between pairs of geographic locations, based in the Hamming distance, which enables us to visualize a graph showing the most similar regions in Iberian Peninsula with respect to extreme climate events. We show that this statistical similarity changes significantly in time.

INTRODUCTION

Heatwaves and droughts, two types of extreme events with very different onsets, are becoming more frequent and with rising intensities due to climate change (Bell et al., 2018). They affect many human activities and the environment (Ferral et al., 2017). These two types of events often combine into extreme compound events (Leonard et al., 2014), which can cause superior impacts than isolated events (Zscheischler et al., 2017; Bevacqua et al., 2021). For example, the 2010 compound event in Russia, which combined a massive heatwave and drought (Schumacher et al., 2019), had widespread crop yield losses throughout Russia, northeastern Ukraine and northwestern Kazakhstan (Loboda et al., 2017). Therefore, it is vital to understand the relationship between them.

Among the panoply of multivariate approaches applied to assess the impacts of multiple climate hazards, the use Markov chain (MC) models have become widely used in climate analysis, especially in droughts and heatwaves (Hao et al., 2018). The multi-phase MC model was used by (Lennartsson et al., 2008) to build a model that is capable of predicting the precipitation occurrence at a particular grid station in Sweden. The MC has also been utilised in several studies that have investigated the predictability of El Nino-Southern Oscillation (ENSO) and its connections to the solar and atmospheric indices (Iqbal & Siddiqi, 2022; Kirov & Georgieva, 2002; Navarra et al., 2013). Subsequently, we further analysed the co-concurrency of extreme events on each grid point pair as similarly applied to explore drought spatial patterns across Europe (Giaquinto et al., 2023). Nevertheless, in the context of assessing the propensity of extreme climate events, MCs are not a typical choice.

In this paper we will process temperature and precipitation data into binary states, labelling occurrence or absence of extreme events with respect to those two climate variables, respectively heatwaves and droughts. The sequence of binary states in time at particular geographic locations will then be used to derive a MC describing the transition

between the binary states. We used the ERA5 reanalysis dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) daily data with a fine spatial resolution of 0.25 degrees from 1950 to 2022 (Hersbach et al., 2020).

DATA AND METHODS

The sequence of binary states in time at particular geographic locations will then be used to derive a MC describing the transition between the binary states. We used the ERA5 reanalysis dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) daily data with a fine spatial resolution of 0.25 degrees from 1950 to 2022 (Hersbach et al., 2020). The study area is within the geographic coordinates of 36°N to 44°N latitude and -10°E to 4°E longitude. The ERA5 dataset provides daily precipitation and temperature data, which are the basis for calculating SPI and extreme temperature exceedances . Our methods are based on the statistical analysis of time-series of precipitation and temperature, as well as methods to derive the MC of binary states classifying the occurrence or absence of heatwaves and droughts. The mapping from continuous variables of temperature and precipitation into the binary states of heatwaves and droughts is based on the standard definition of extreme (high) temperatures which are heatwaves (above the 90th percentile), and extreme (low) precipitation which are droughts.

Processing precipitation data and performing drought analysis

The Standardized Precipitation Index (SPI) is one of the most commonly used indices for identifying and quantifying drought conditions due to its low data requirement and its ability to analyse the various aspects of drought based on varying time-scales. SPI plays a pivotal role in the identification and characterization of drought or wet events, enabling the determination of their duration, intensity, and magnitude. Renowned for its simplicity, the SPI, introduced by McKee et al. (1993), relies solely on precipitation data and has garnered widespread adoption, endorsed by the World Meteorological Organization (WMO, 2012), albeit not without considerations (Vicente-Serrano et al., 2010). The SPI's intrinsic multiscalar nature facilitates temporal and spatial comparability, thereby enabling the monitoring of drought response across diverse ecosystems. SPI targets only precipitation anomalies, being a robust tool for monitoring drought (WMO, 2012).

The calculation of SPI over different time scales (1, 2, 3, 6, 12, and 24 months) allows for the monitoring of drought conditions from shorter periods to longer periods. The 12-month period offers a view of precipitation over an entire year, reflecting long-term trends, which is key for detecting and managing hydrological droughts. We adopted this time scale for this study because SPI-12 captures long-term trends in precipitation, providing a broader picture of drought conditions. It is less influenced by short-term variability and can indicate more persistent anomalies in precipitation. Drought indices are classified into various categories, providing a standardized method to assess and communicate drought severity based on the SPI values as follows: 2.0 or more: (extremely wet), 1.5 to 2.0: (very wet), 1.0 to 1.5: (moderately wet), 0.0 to 1.0: (mild wet), -1.0 to 0: (near normal), -1.5 to - 1.0: (moderately dry), -2.0 to -1.5 (severely wet), -2.0 or less: (extremely dry).

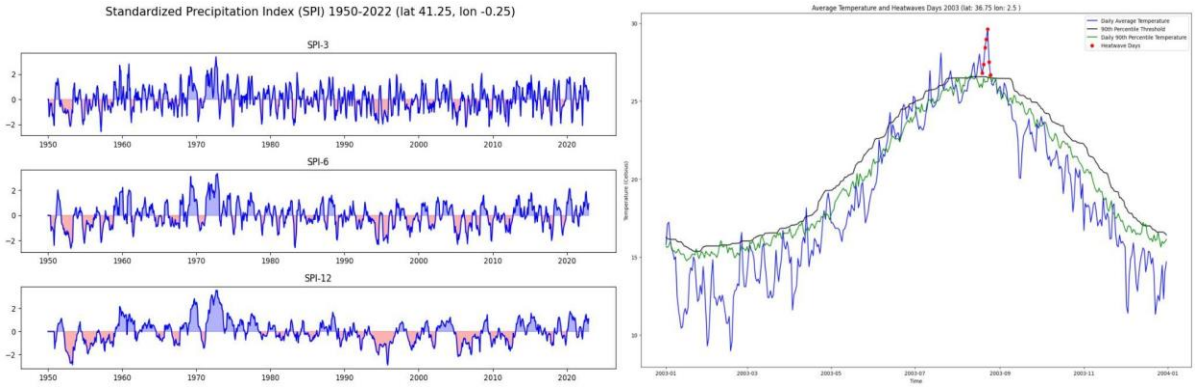

Figure 1: (Left) SPI data series for 3-, 6- and 12-months timescales for a given location (Blue line). The pink color indicates less than median precipitation in a series of SPI data series; (Right) Example of heatwave periods from temperature data series (red dots).

Processing temperature data and performing heatwave analysis

For heatwave detection, one calculates a daily threshold at the 90th-percentile temperatures for each calendar day considering the climatological temperature averages. A dynamic threshold for heatwave detection is computed using a moving centered window of 31-days applied to the climatological temperatures. This guarantees that climatological averages and anomalies are appropriately reflected in the 90th percentile threshold.

Further refinement is achieved by calculating a rolling sum of days flagged as exceeding the 90th percentile threshold: the first day of a heatwave is the 5th consecutive day above the 90th percentile, indicating a continuous sequence of high temperatures typical of a heatwave. If there is a drop below the 90th percentile between a heatwave day and an abnormal temperature, the drop day is still counted as part of a continuous heatwave period. This aligns with the definition of heatwaves, as events persist for multiple days. Figure 1 (right) illustrates the occurrence of a heatwave in a series of temperature measurements.

Binary states of extreme climate events, Markov-states and Hamming distance

Flagging each day of temperature and precipitation time series as belonging or not to a heatwave and drought respectively, we derive an auxiliary time series of binary states (S_H, S_D). State S_H, $S_D = 0$, indicates the absence of a heatwave or drought respectively, while a value of 1 indicates its respective occurrence (Figure 2).

By tracking simultaneously the derived series of binary states for both heatwaves and droughts, we then calculate the frequency of all possible transitions between the four possible states, $(S_H, S_D) = [(0,0); (0,1); (1,0); (1,1)]$. These frequencies are mapped into conditional probabilities

$$P((S_H(t+\Delta t), S_D(t+\Delta t))|(S_H(t),\ S_D(t))) = \frac{P((S_H(t+1), S_D(t+1)); (S_H(t), S_D(t)))}{P((S_H(t), S_H(t)))} \quad (1)$$

where Δt is the time within which the transition occurs. Here we consider $\Delta t = 1,2,3,4,5$ and 6 months. The set of conditional probabilities defines a MC. Figure 3 illustrates the Markov-chain (Left) and frequency of binary-states transitions (Top-Right) for a specific time delay ($\Delta t = 1$ month). We derive one Markov-chain as described above for each grid in the mesh covering this region (Figure 3, Bottom-Right).

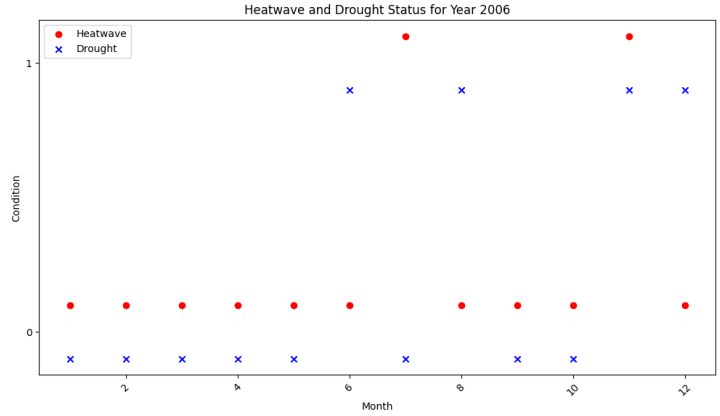

Figure 2: Illustrative example for the year 2006. Red dots indicate heatwave binary states and blue crosses indicate drought binary states. A low value (around 0) indicates absence of heatwave/drought, while high value (around 1) indicates the respective occurrence..

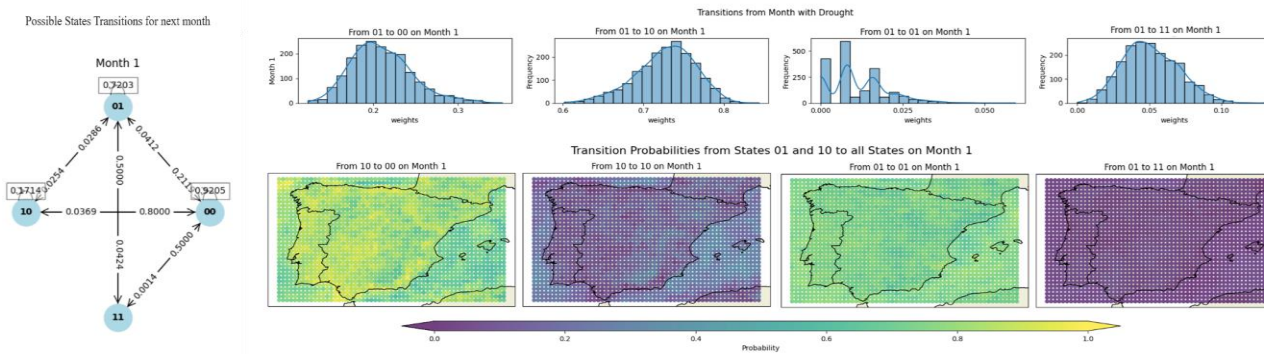

Figure 3: (Left) Markov chain probability of changing to another state Heatwaves and SPI-12 (Top-Right) Histograms of Possible transitions from one State to another. (Bottom-Right) Transition Probabilities from States 01 and 10 to Other States on Month 1.

Together with the binary states labelling the occurrence of heatwaves (H) and droughts (D), *(S_H, S_D)* which are used to derive one MC for each location in Iberian Peninsula, we also consider the Hamming distance (Mackay, 2003) between both the heatwave and drought series of binary states. The Hamming distance is defined for each pair of locations, l_1 and l_2, and it is given by: $H_{total}(l_1, l_2) = \sqrt{H_d^2(l_1, l_2) + H_h^2(l_1, l_2)}$

where $H_d(S_1^{(d)}, S_2^{(d)}) = \frac{1}{N}\sum_{i=1}^{N} |s_1^{(d)}(i) - s_2^{(d)}(i)|$ and $H_h(S_1^{(h)}, S_2^{(h)}) = \frac{1}{N}\sum_{i=1}^{N} |s_2^{(h)}(i) - s_2^{(h)}(i)|$, with $s_2^{(d)}(i)$ and $s_2^{(d)}(i)$ the binary state (0 or 1) with respect to heatwave ("h") or drought ("d") at location *l1* in day *i*, and similarly for location *l2*. In other words, the Hamming distance measures the proportion of differing bits between sequences, providing a quantitative measure of sequence dissimilarity.

RESULTS

The 90th percentile temperatures differ across Iberia, with the southern regions showing a higher average temperature than the northern region. The Pyrenees Mountains between Spain and France show a much lower average 90th percentile temperature than the remaining areas. Regarding precipitation, the northern coastal regions of Portugal and Spain, as well as the Pyrenees Mountains, show higher annual values (not shown).
By computing the Hamming distance between each pair of geographic locations (longitude and latitude), one can draw a weighted graph of similarities, where nodes represent locations and edges represent the similarity, with a thickness proportional to the inverse of the Hamming distance. Figure 4 collects the main results of this manuscript showing, from left to right, three consecutive periods, 1951-1975, 1976-2000 and 2001-2022. The last column shows the graphs for the total time period 1951-2022.

The distribution of the similarity values, *1/Htotal*, follows approximately a power-law (not shown). For this reason, plotting the last quartile of highest similarities leads still to a too dense graph, difficult to visualize. Therefore, we opted by showing in Figure 4 only the 1% highest similarities (first row), 0.1% (second row) and 0.01% (third row).

Four points are worth highlighting. Firstly, for all time periods, the highest similarities occur typically between close – or adjacent – geographic locations. Second, through time the "clusters" of high similarities have covered a larger and larger region of the peninsula. This is particularly true for the last periods, from 1976-2000 to 2001-2022. Third, through time the value of the similarities for the highest percentiles has decreased. Lastly, the areas where the highest similarities occur do not seem to be connected with the occurrence of higher temperatures or lower precipitation regions.

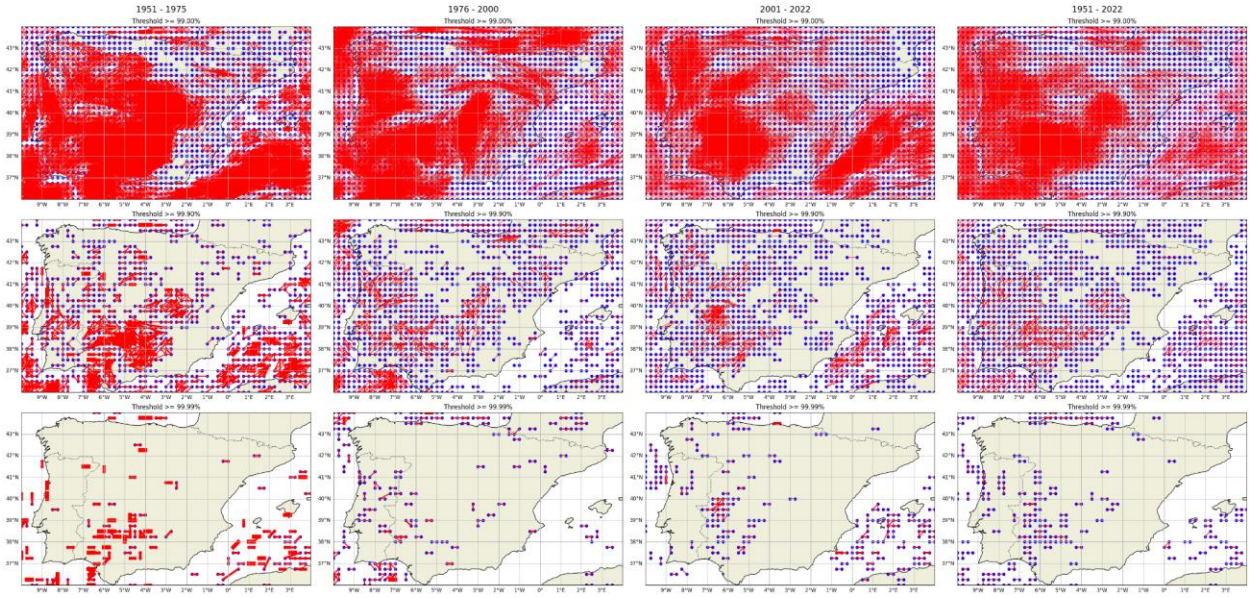

Figure 4: Assessing the similarity with respect to extreme climate events, between different regions in Iberian Peninsula. (Col. 1, Left) Graph with the cumulative similarities between 1951 and 1975, (Col. 2) between 1976 and 2000, (Col. 3) between 2001 and 2022, and (Col. 4) the total time period 1951-2022. (First row, Top) Graphs showing the 1% highest similarities, (2nd row) the 0.1% highest similarities and (3rd row) the 0.01% highest similarities. The similarities are measured by the inverse of the corresponding Hamming distance between each two geographic locations. Blue dots represent grid points, whereas red edges represent the similarity "strength" between pairs of points.

DISCUSSION AND CONCLUSIONS

The results described previously show evidence that high similarities have become weaker through time and, simultaneously have spread through larger regions. On one hand, this might indicate that coping with possible occurrences of extreme events can become more difficult, since similarities between different geographic regions seems to weaken since the last decades. On the other hand, the more homogeneous distribution of highest similarities throughout the entire Peninsula, shows that more and more the occurrence of an extreme event, when observed, tends to spread out through larger regions.

All in all, we have introduced and developed a framework to analyze the ubiquity of transitions from and to extreme events related with temperature and precipitation. The framework includes two parts: one, described through a Markov chain model, related with the probability of state transition at individual locations, and another, based in Hamming distance, which measures how similarly these transitions occur in two distinct locations. A possible next step in this approach is to include similarities incorporating a non-zero time-lag between the pair of locations. In this way, high similarities between different locations, for a particular time lag, could open the possibility to uncover possible causal connections between different regions.

ACKNOWLEDGEMENTS

The authors acknowledge support from AIClimate@EU project (EEA Bilateral Initiative 2014-2021, Portugal-Norway) and Nordic Center for Sustainable and Trustworthy AI Research (Oslo Metropolitan University, Norway). A.R. also aknowledges funding from the Portuguese Fundação para a Ciência e a Tecnologia (FCT) I.P./MCTES through national funds (PIDDAC) – Instituto Dom Luiz (https://doi.org/10.54499/UIDP/50019/2020 and https://doi.org/10.54499/LA/P/0068/2020), and projects DHEFEUS (https://doi.org/10.54499/2022.09185.PTDC). A.R. was supported by FCT under the grant agreement https://doi.org/10.54499/2022.01167.CEECIND/CP1722/CT0006.

REFERENCES

1. Bell, J. E., Brown, C. L., Conlon, K., Herring, S., Kunkel, K. E., Lawrimore, J., Luber, G., Schreck, C., Smith, A., & Uejio, C. (2018). Changes in extreme events and the potential impacts on human health. *Journal of the Air & Waste Management Association*, *68*(4), 265–287. https://doi.org/10.1080/10962247.2017.1401017

2. Bevacqua, E., Zappa, G., Lehner, F., & Zscheischler, J. (2022). Precipitation trends determine future occurrences of compound hot–dry events. *Nature Climate Change*, *12*(4), 350–355. https://doi.org/10.1038/s41558-022-01309-5

3. Ferral, A., Solis, V., Frery, A., Orueta, A., Bernasconi, I., Bresciano, J., & Scavuzzo, C. M. (2017). Spatio-temporal changes in water quality in an eutrophic lake with artificial aeration. *Journal of Water and Land Development*, *35*(1), 27–40. https://doi.org/10.1515/jwld-2017-0065

4. Giaquinto, D., Marzocchi, W., & Kurths, J. (2023). Exploring meteorological droughts' spatial patterns across Europe through complex network theory. *Nonlinear Processes in Geophysics*, *30*(2), 167–181. https://doi.org/10.5194/npg-30-167-2023

5. Hao, Z., Singh, V., & Hao, F. (2018). Compound Extremes in Hydroclimatology: A Review. *Water*, *10*(6), 718. https://doi.org/10.3390/w10060718

6. Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

7. Iqbal, A., & Siddiqi, T. A. (2022). Markovian descriptors based stochastic analysis of large-scale climate indices. *Stochastic Environmental Research and Risk Assessment*, *36*(4), 955–968. https://doi.org/10.1007/s00477-021-02108-8

8. Kirov, B., & Georgieva, K. (2002). Long-term variations and interrelations of ENSO, NAO and solar activity. *Physics and Chemistry of the Earth, Parts A/B/C*, *27*(6–8), 441–448. https://doi.org/10.1016/S1474-7065(02)00024-4

9. Lennartsson, J., Baxevani, A., & Chen, D. (2008). Modelling precipitation in Sweden using multiple step markov chains and a composite model. *Journal of Hydrology*, *363*(1–4), 42–59. https://doi.org/10.1016/j.jhydrol.2008.10.003

10. Leonard, M., Westra, S., Phatak, A., Lambert, M., Van Den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., & Stafford-Smith, M. (2014). A compound event framework for understanding extreme impacts. *WIREs Climate Change*, *5*(1), 113–128. https://doi.org/10.1002/wcc.252

11. Loboda, T., Krankina, O., Savin, I., Kurbanov, E., & Hall, J. (2017). Land Management and the Impact of the 2010 Extreme Drought Event on the Agricultural and Ecological Systems of European Russia. In G. Gutman & V. Radeloff (Eds.), *Land-Cover and Land-Use Changes in Eastern Europe after the Collapse of the Soviet Union in 1991* (pp. 173–192). Springer International Publishing. https://doi.org/10.1007/978-3-319-42638-9_8

12. MacKay, D. J. C. (2019). *Information theory, inference, and learning algorithms* (22nd printing). Cambridge University Press.

13. Navarra, A., Tribbia, J., & Conti, G. (2013). Atmosphere–Ocean Interactions at Strong Couplings in a Simple Model of El Niño. *Journal of Climate*, *26*(23), 9633–9654. https://doi.org/10.1175/JCLI-D-12-00763.1

14. Schumacher, D. L., Keune, J., van Heerwaarden, C. C., Vilà-Guerau de Arellano, J., Teuling, A. J., & Miralles, D. G. (2019). Amplification of mega-heatwaves through heat torrents fuelled by upwind drought. *Nature Geoscience*, *12*(9), 712–717. https://doi.org/10.1038/s41561-019-0431-6

15. Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

16. Zscheischler, J., & Seneviratne, S. I. (2017). Dependence of drivers affects risks associated with compound events. *Science Advances*, *3*(6), e1700263. https://doi.org/10.1126/sciadv.1700263