Supporting Information for "Guidelines for studying diverse types of compound weather and climate events"

Emanuele Bevacqua[1,2], Carlo De Michele[3], Colin Manning[4], Anaïs Couasnon[5], Andreia F. S. Ribeiro[6,7], Alexandre M. Ramos[7], Edoardo Vignotto[8], Ana Bastos[9], Suzana Blesić[10], Fabrizio Durante[11], John Hillier[12], Sérgio C. Oliveira[13], Joaquim G. Pinto[14], Elisa Ragno[15], Pauline Rivoire[16], Kate Saunders[17], Karin van der Wiel[18], Wenyan Wu[19], Tianyi Zhang[20], and Jakob Zscheischler[1,21,22]

[1]Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

[2]Department of Meteorology, University of Reading, Reading, United Kingdom

[3]Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy

[4]School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, United Kingdom

[5]Institute for Environmental Studies, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

[6]Institute for Atmospheric and Climate Science, ETH Zurich, Universitätstrasse 16, Zurich 8092, Switzerland

[7]Instituto Dom Luiz (IDL), Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

[8]Research Center for Statistics, University of Geneva, 24 rue du Général-Dufour, Geneva 1211

[9]Max Planck Institute for Biogeochemistry, Dept. of Biogeochemical Integration, 07745 Jena, Germany

[10]Institute for Medical Research, University of Belgrade and Center for Participatory Science, Belgrade, Serbia

[11]Department of Economic Sciences, University of Salento, Lecce, Italy.

[12]Geography, Loughborough University, Epinal Way, Loughborough, LE11 3TU, UK

[13]Centre for Geographical Studies, Institute of Geography and Spatial Planning, Universidade de Lisboa, Lisboa, Portugal

[14]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany

[15]Delft University of Technology, Faculty of Civil Engineering and Geosciences, 2628 CN, Delft, Netherlands

[16]Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland

[17]School of Mathematical Sciences, Queensland University of Technology, Gardens's Point, Brisbane, Australia, 4000

[18]Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

[19]Department of Infrastructure Engineering, Faculty of Engineering and Information Technology, The University of Melbourne, Australia

[20]State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

[21]Climate and Environmental Physics, University of Bern, Bern, Switzerland

[22]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

Contents of this file

Text S1. Additional information on the tools employed for the preconditioned event analysis

Classification of Crop and Forest Grid Cells. The categorisation of grid cells into forest or crop grid cells (Figure 2a) is done using the European Space Agency Land Cover dataset (Santoro et al., 2017). This provides a land classification at a 300m resolution for the period 1992-2018, which we regrid to a 0.5° lat-lon grid using the LC-CCI User Tool. A grid cell is classified as a crop grid cell if rainfed croplands are the majority class, while it is classified as a forest grid cell if the majority class is either evergreen, deciduous broadleaf, needle leaved, or mixed forests. To exclude human influence due to land-cover change, we keep only grid cells where the sum of the absolute changes over the 27 year period is below 5%. We also exclude irrigated croplands. The locations of the resulting crop and forest grid cells is shown in Figure 2a.

Application of Logistic Regression. Using a logistic regression model, we estimate the probability of extremely low vegetation activity in summer, defined as occurrences of LAI in summer (LAI_{JJA}) below its 5^{th} percentile. In this case, the impact Y is a binary time series (1: $LAI_{JJA} < 5^{th}$ percentile; 0: $LAI_{JJA} > 5^{th}$ percentile), while the predictors $X_1, X_2, ..., X_n$ are continuous variables. The probability of a low anomaly is estimated as:

$$P[Y = 1] = \frac{1}{1 + exp(b_0 + b_1\ X_1 + ... + b_n\ X_n)}$$

where b_0, b_1, ... , b_n are the regression coefficients. The logistic regression model is fitted using the glm function from the stats package in R (R Core Team, 2020). We fit two models for both cropland and forest grid cells separately, a base model that includes summer precipitation and temperature as predictors and a second model which includes the two summer predictors along with spring radiation. After an iterative process, this was found to be the most parsimonious model with good performance. Similar performance is achieved using spring precipitation instead of spring radiation, though we choose to use spring radiation (see discussion in the main text). Finally, no improvement is achieved when adding other variables (i.e summer radiation or spring temperature). Before fitting, all predictors are checked for multi-collinearity using the variation inflation factor (VIF). Some variables have moderate correlation but it is small enough that it will not adversely affect the results.

Text S2. Additional information on the tools employed for the temporally compounding event analysis

We looked for clustering of precipitation events over multiple windows before the landslide of interest. The amplitude of the shortest window ([-4,0] days) is in line with the 3 days thresholds used to disentangle nearby daily precipitation amounts arising from individual weather systems (i.e., used for removing high-frequency clustered events). Due to the multiple windows, multiple tests are involved when assessing whether the number of precipitation events is above the 95^{th} percentile of the Binomial distribution (where some tests may be dependent due to the overlapping of time windows), hence we considered the Fuzzy Benjamini-Hochberg correction (Kulinskaya & Lewin, 2009).

Given that we look for precipitation clustering up to 90 days before the landslide, which occurred during the November-March period in each hydrological year, we searched for clustering backward until August when considering landslides occurred in November. Hence, the parameter p (equal to 0.056) of the Binomial distribution was estimated based on data during August-March, specifically as N_{tot}/L, where N_{tot} is the total number of selected precipitation events and L is the length of the August-March time series. Note that in further analysis, a dependency of the parameter p from time could be considered, such as to account for seasonality effects. The goodness of fit of the Binomial distribution was tested (significance level of 0.05) for all temporal windows using the Chi-square goodness-of-fit test.

We quantified the association between precipitation clustering and landslide occurrences as the fraction of landslide events preceded by precipitation clustering. To assess whether this association is significant, we compared the observed fraction with that 95^{th} percentile of the fraction obtained in 5000 synthetic datasets that assume no association between clustered precipitation events and landslides. In these synthetic datasets, the landslides occurrence is shuffled in time. In particular, we shuffled the year of occurrence of the landslide events but left that day and month of occurrence as in the original dataset to preserve the temporal structure of landslide occurrences in the synthetic dataset (Witt et al., 2010; Bevacqua et al., 2019). The resulting increase of the confidence interval width with the temporal window in Figure 7b is in line with the tendency of precipitation events to cluster more at higher temporal windows.

References

Bevacqua, E., Maraun, D., Vousdoukas, M. I., Voukouvalas, E., Vrac, M., Mentaschi, L., & Widmann, M. (2019). Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change. *Science Advances*, *5*(9), eaaw5531.

Kulinskaya, E., & Lewin, A. (2009). On fuzzy familywise error rate and false discovery rate procedures for discrete distributions. *Biometrika*, *96*(1), 201–211.

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Santoro, M., Kirches, G., Wevers, J., Boettcher, M., Brockmann, C., Lamarche, C., & Defourny, P. (2017). Land Cover CCI: Product User Guide Version 2.0. *Climate Change Initiative Belgium: Belgium, Leuven.*

Witt, A., Malamud, B. D., Rossi, M., Guzzetti, F., & Peruccacci, S. (2010). Temporal correlations and clustering of landslides. *Earth Surface Processes and Landforms*, *35*(10), 1138–1156.