



*Supplement of*

## **Detection of compound and seesaw hydrometeorological extremes in New Zealand: A copula-based approach**

**Morgan J. Bennet et al.**

*Correspondence to:* Morgan J. Bennet ([morgan.bennet@postgrad.otago.ac.nz](mailto:morgan.bennet@postgrad.otago.ac.nz))

The copyright of individual parts of the supplement might differ from the article licence.

**S1. Detailed vine copula construction, including mathematical notation:**

As described by Sklar (1959), the joint multivariate distribution ( $F$ ) can be described as:

$$F(x) = C\{F_1(x_1), \dots, F_d(x_d)\} \quad (\text{S1})$$

where  $x_1, \dots, x_d$  denotes a setting of  $d$  relevant variables,  $F_1, \dots, F_d$  the marginal distributions and  $C$  is a copula (itself a  $d$ -dimensional distribution function on  $[0, 1]_d$  with uniform margins) (Erhardt and Czado, 2018).

Following the methodology of Erhardt and Czado (2018), the copula data that are obtained from the marginal models corresponding to  $d$  different drought variables can be described as:

$$u := (u_1, \dots, u_d) \quad (\text{S2})$$

where  $u_j = (u_j)$ ,  $j = 1, \dots, d$  and  $u_j$  are the copula data corresponding to variable  $j$ .

As noted by Aas *et al.* (2009), vine copulas can be used to model multivariate data acting on two variables at a time, offering a way in which to construct higher-dimension copulas. For a three variable example, the vine copula density  $c$  is given as:

$$c(u_1, u_2, u_3; \theta) = c_{1,2}(u_1, u_2; \theta_{1,2})c_{1,3}(u_1, u_3; \theta_{1,3})c_{2,3;1}\{h_{2|1}(u_2, u_1; \theta_{1,2}), h_{3|1}(u_3, u_1; \theta_{1,2}); \theta_{2,3;1}\} \quad (\text{S3})$$

where  $c_{1,2}$ ,  $c_{1,3}$  and  $c_{2,3;1}$  are the pair copula densities corresponding to the copulas  $C_{1,2}$ ,  $C_{1,3}$  and  $C_{2,3;1}$ .

The  $h$ -functions involved are defined as  $h_{b|a}(u_b, u_a; \theta) := C_{b|a}(u_b|u_a; \theta)$ , where  $C_{b|a}$  denotes the conditional distribution function of  $U_b$  given  $U_a$  (Erhardt and Czado, 2018).

Having fitted the copula, the resultant copula data were transformed into independent data in the  $[0, 1]$  space using the Rosenblatt transformation (Rosenblatt, 1952). The Rosenblatt transform  $v := (v_1, \dots, v_d)$  on the basis of the selected vine copula  $C$  for the data  $u = (u_1, \dots, u_d)$  is defined as:

$$v_d := C_{d|1, \dots, d-1}(u_d|u_1, \dots, u_{d-1}) \quad (\text{S4})$$

where  $C_{j|1, \dots, j-1}$  is the conditional cumulative distribution function for variable  $j$  given the variables  $1, \dots, j-1$  for all  $j = 2, \dots, d$ . In the Canonical Vine (C-Vine) context the order of the variables is determined by the selected order of root variables (Erhardt and Czado, 2018).

The elements from this multivariate probability integral transformation were transformed to the standard normal distribution  $p$ , using the inverse of the cumulative distribution function of a standard normal distribution:

$$p_j = \Phi^{-1}(v_j) \quad (\text{S5})$$

24 Finally, these uniform elements were then aggregated and standardised to the final Standardised  
 25 Multivariate Index (SMI):

$$SMI = \frac{1}{\sqrt{d}} \sum_{j=1}^d p_j \quad (\text{S6})$$

26 Limits of -3/3 were also imposed (Stagge *et al.*, 2015). Index construction was performed using R  
 27 software (R Core Team, 2017), and the following packages:

28 **Standardised Indices:**

29 SCI Package: Stagge *et al.* (2015)

30 fitdistrplus: Delignette-Mueller and Dutang (2015)

31 **Copula:**

32 copula: Hofert *et al.* (2020)

33 **Vine Copula:**

34 VineCopula: Nagler *et al.* (2022)

35 rvinecopulib: Nagler and Vatter (2021)

## **S2. Method Description:**

The following provides a more specific description of the methodological procedure used in the construction of the Standardised Multivariate Index (SMI) and Standardised Bivariate Index (SBI), and aligns with the flowchart illustrated in Fig. S1.

### **Overview**

Maximum temperature, precipitation and soil moisture were selected as the variables from which an extreme event index was constructed, using the European ReAnalysis 5th Generation Land Component (ERA5-Land) dataset (Muñoz-Sabater *et al.*, 2021).

### **Standardised Indices**

The extreme event index was constructed using copulas, after first constructing individual standardised indices: the Standardised Precipitation Index (SPI) (McKee *et al.*, 1993) (precipitation), the Standardised Temperature Index (STI) (Zscheischler *et al.*, 2014) (maximum temperature) and the Standardised Soil Moisture Index (SSMI) (Xu *et al.*, 2018) (soil moisture). Gamma, Normal and Beta distributions were used to fit the SPI, STI and SSMI respectively across a 30-day accumulation period.

### **Pre-processing Work**

Here, the pre-treatment guidelines proposed by Tootoonchi *et al.* (2022) are addressed (Fig. S1). The first step, an initial exploratory analysis, was performed in Bennet *et al.* (2023). Data ties, predominately within precipitation data (i.e. zero precipitation), are treated via the use of the Weibull plotting position for zero precipitation values (Stagge *et al.*, 2015) when constructing the SPI. Autocorrection is addressed by the approach of Kao and Govindaraju (2010) via the constructing of daily subsets in the time series, resulting in 365 individual time series for each index. Stationarity is not addressed, with the baseline period (1961-1990) used for the standardised indices generally considered free from non-stationarity (World Meteorological Organisation, 2017), and thus copula parameters are able to be stabilised.

### **Vine Copula Structure**

A Canonical Vine (C-Vine) is a sub-class of R-Vine (itself a nested set of trees used for pair copula construction as building blocks; Erhardt and Czado (2018) and Bedford and Cooke (2002)), having a

star like structure where a centre node is linked to all remaining nodes (Wu *et al.*, 2021). This star like structure allows the order of variables to be set to the order of importance, with the variable of highest importance being set as the root variable (Erhardt and Czado, 2018). In the current work, this C-Vine structure was selected, while testing (not shown; utilizing the algorithm of Dissmann *et al.* (2013)) revealed the conditional/root variable precipitation as the chosen optimal variable.

## **Parameter Estimation**

A semi-parametric approach for the estimation of copula parameters is common in hydroclimatic studies (Pham *et al.*, 2016; Tootoonchi *et al.*, 2022), and is herein employed. The semi-parametric approach first establishes the individual marginal distributions based on their rank behaviour and by forming pseudo observations, before copula parameters are estimated by maximizing the pseudo likelihood function (Kim *et al.*, 2007; Tootoonchi *et al.*, 2022). Copula selection was from a family set made up of Clayton, Frank and Joe (no rotations), selected via Akaike Information Criterion (AIC). This process was performed for each calendar day, for each grid cell (Figure 1; a-c). The selection of only three copula families was made to reduce the overall spatial and temporal complexity associated with fitting highly unique copula families on the fine scales of this study ( $0.10^\circ \times 0.10^\circ$  and daily resolution), while a range of families was still required to capture tail behaviour. Additional results for the optimal copula selection for each copula pairing are shown in the supplementary material (Table S1). Further details on the selected copula families and their unique characteristics are contained in the works of Joe (2014), Tootoonchi *et al.* (2022), and Wu *et al.* (2021).

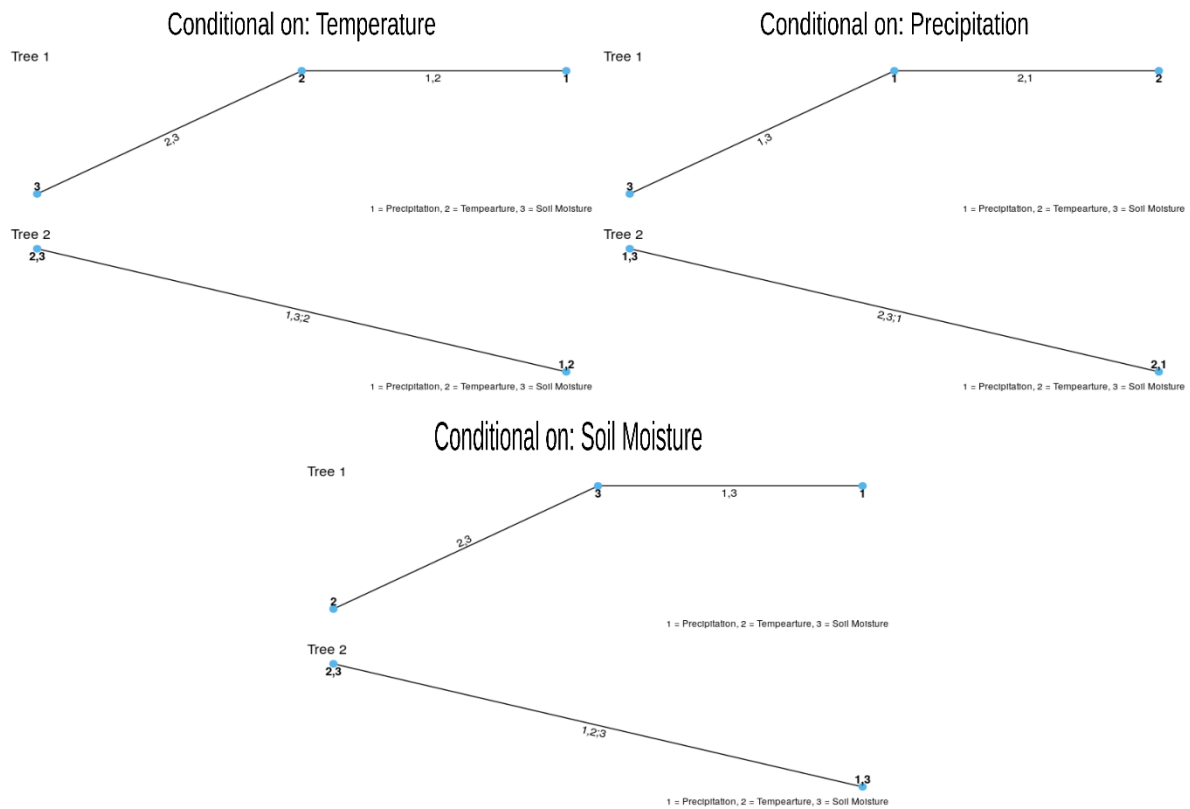
The automatic fitting and model selection contained within the “rvinecopulib” and “VineCopula” packages of Nagler and Vatter (2021) and Nagler *et al.* (2022) provides for parameter estimation of semi-parametric models (i.e. after forming pseudo observations) using either maximum likelihood or the inversion of Kendall’s  $\tau$ . Both parameter estimation methods were trialled by creating two copula models. Each model automatically selected the optimal copula families at each tree edge, for each calendar day and each grid cell (and for each accumulation period), using the “RVineCopSelect” function of Nagler *et al.* (2022), with the parameter estimation method being the only difference across the two models (either maximum likelihood or the inversion of Kendall’s  $\tau$ ). Parameter estimation using maximum likelihood was the best performing vine copula model, with 95.24% of grid cells having the best fit (log-likelihood) compared to the copula set.

### 93    **Copula Families**

94    The Frank copula enables the capture of symmetry and negative dependence in the dataset, while  
95    Clayton and Joe copula can capture the lower and upper tail dependence, respectively. In this case,  
96    lower tail dependence indicates a stronger relationship between low values (i.e. low soil moisture  
97    and high temperature) and higher tail dependence refers to high values (high soil moisture and  
98    precipitation).

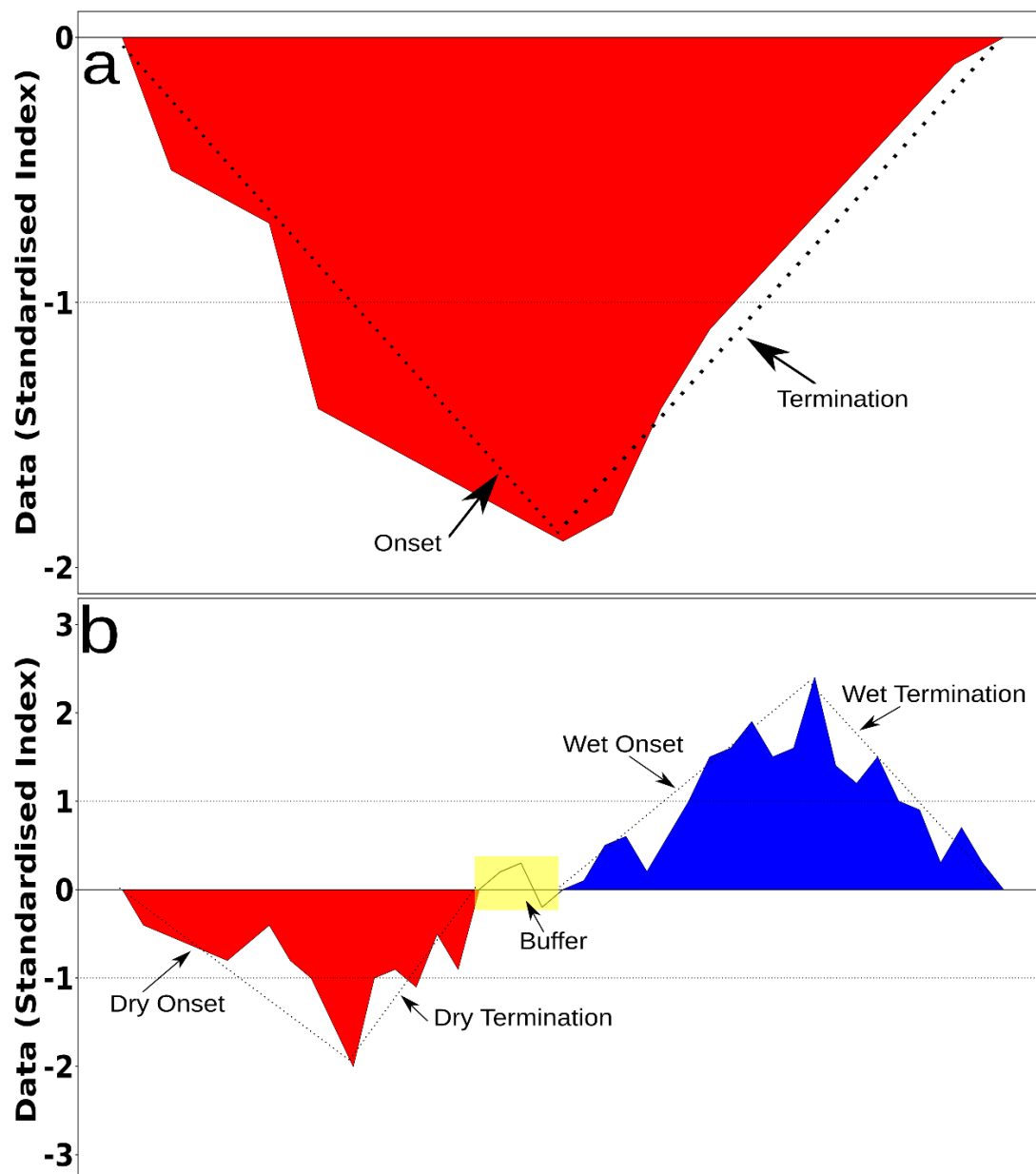
### 99    **Goodness of Fit**

100    To minimise computational load, goodness of fit tests were performed on the vine copula model  
101    fitted to the 15th day of each month, before aggregation of the test results. Testing was performed  
102    using Cramér-von Mises (CvM) and Kolmogorov-Smirnov (K-S) statistics. Results indicated strong  
103    agreement in the goodness of fit tests, with an overall good fitment of the copula model.



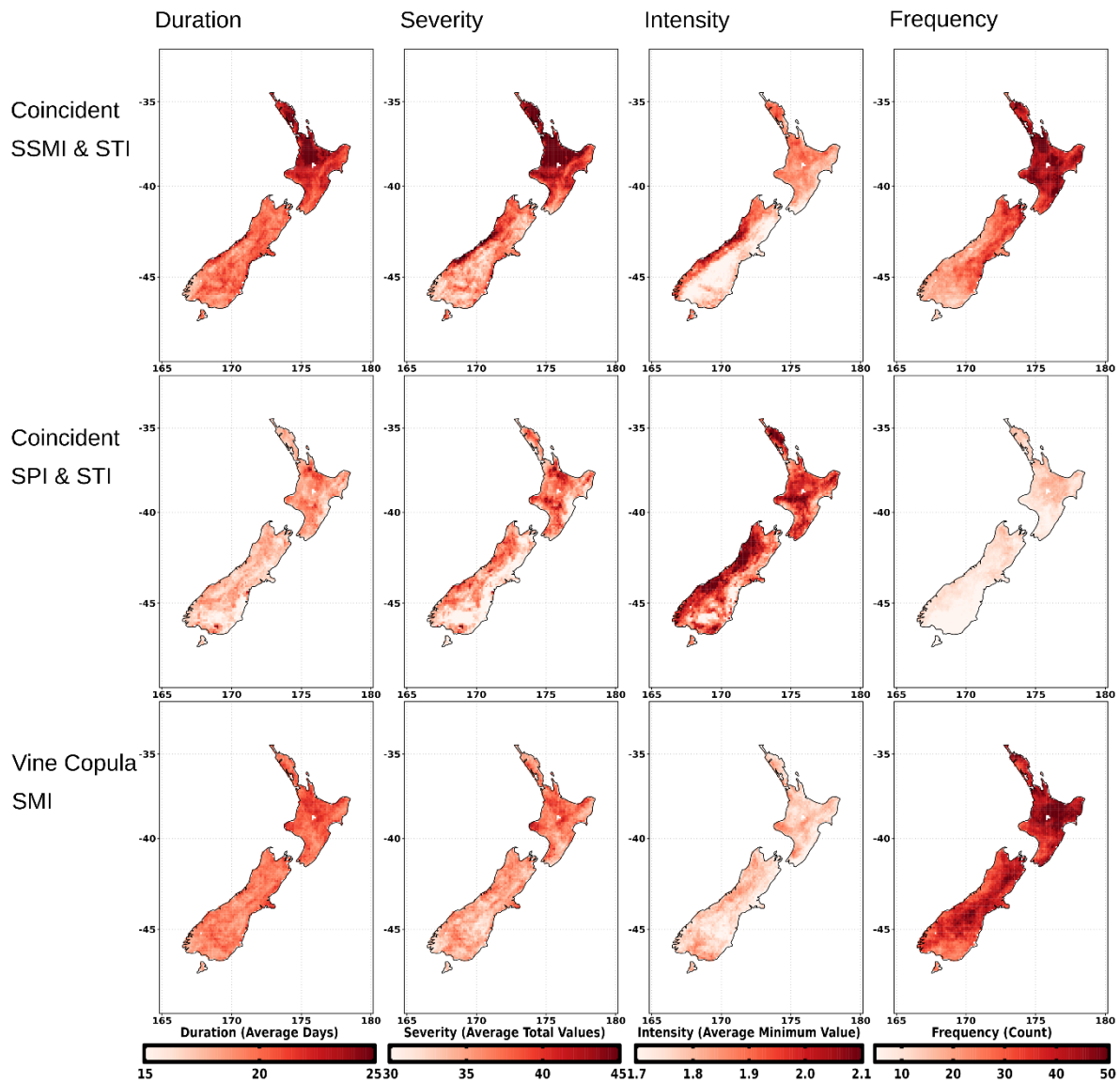
104

105 **Fig. S1** Tree structure of optimal variables orders detected using the Dissmann *et al.* (2013)  
 106 algorithm, applied across the entirety of New Zealand. Headings refer to the root/conditional  
 107 variable: temperature (top left), precipitation (top right) and soil moisture (bottom).

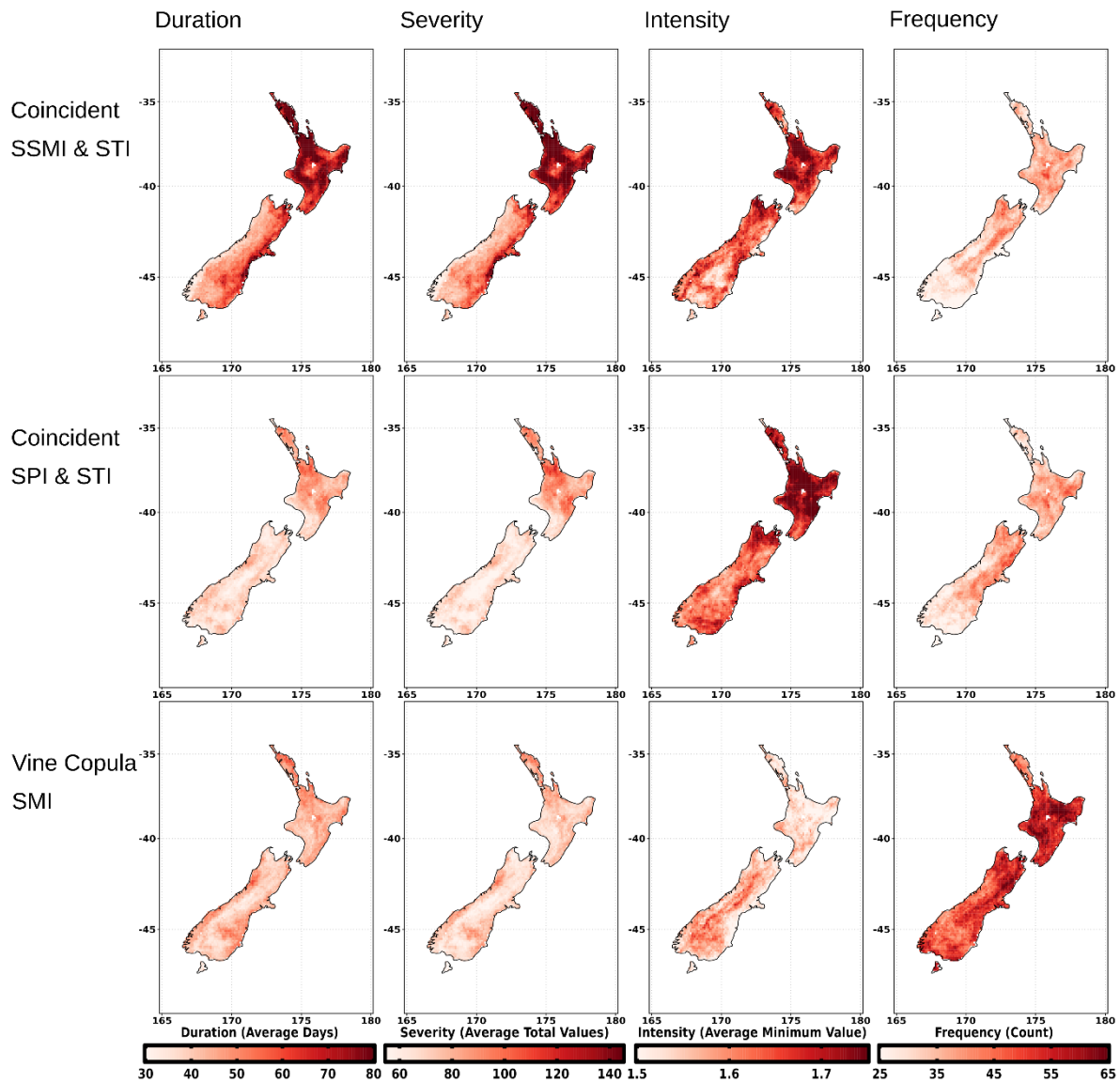


**Fig. S2** Example of compound (a) and seesaw (b) events, adapted from Rashid and Wahl (2022). In the present work, the buffer within a seesaw event refers to a 30-day period, to ensure adequate sample size of events. Event characteristics are noted in the figure; for seesaw events representing onset rates of dry and wet phases, as well as termination rates of dry and wet phases, while for compound events representing onset and termination rates.

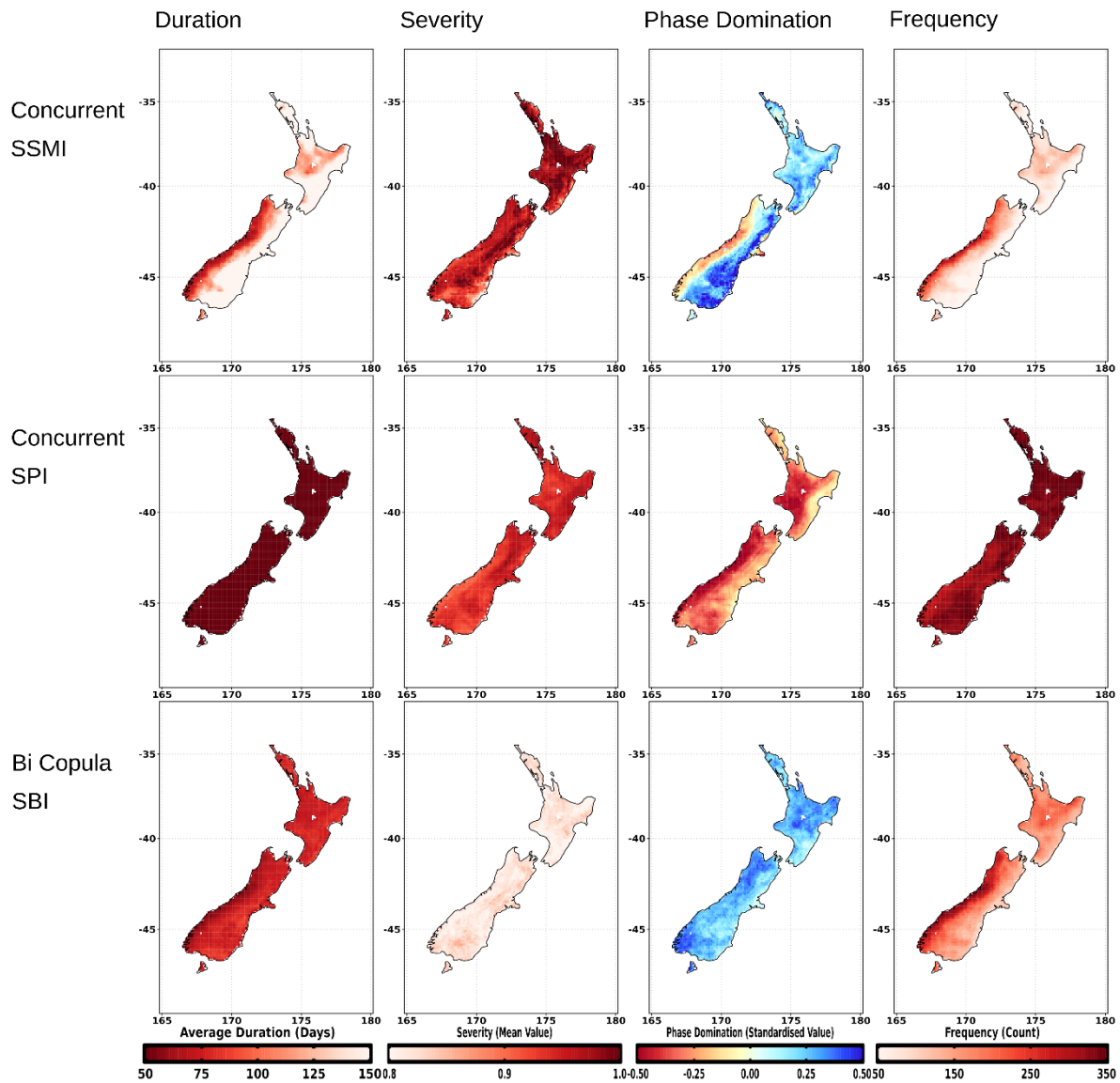




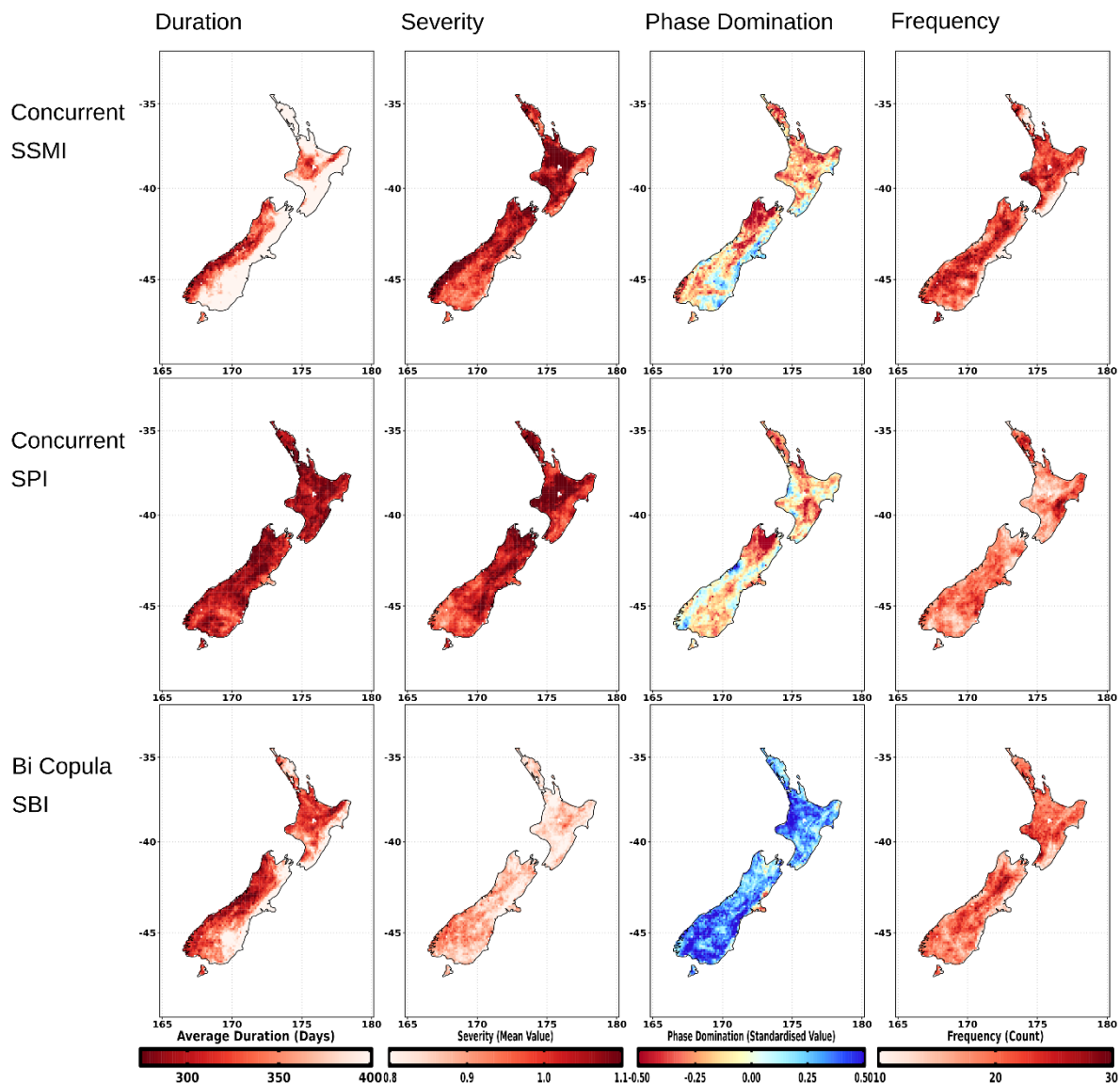
**Fig. S3** Sensitivity testing for 10-day accumulation: Average duration (a-c), severity (d-f), intensity (g-i) and frequency (j-l) of compound events on a grid cell basis for the period 1950-2021. Showing the coincident of -1 SPI and STI (a, d, g, j), coincident of -1 SSMI and STI (b, e, h, k) and co-occurrence of -1 SMI and STI (c, f, i, l). Note the differing scales.



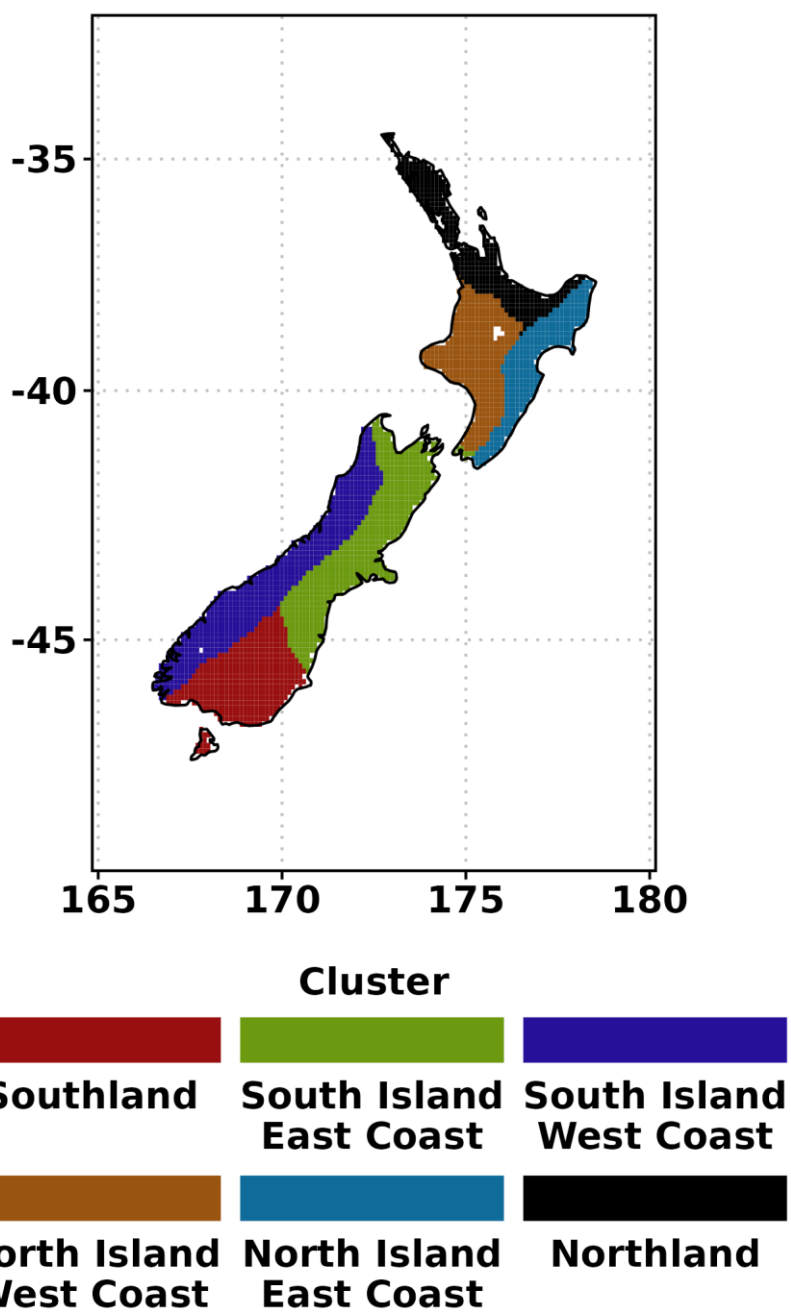
**Fig. S4** Sensitivity testing for 90-day accumulation: Average duration (a-c), severity (d-f), intensity (g-i) and frequency (j-l) of compound events on a grid cell basis for the period 1950-2021. Showing the coincident of -1 SPI and STI (a, d, g, j), coincident of -1 SSMI and STI (b, e, h, k) and co-occurrence of -1 SMI and STI (c, f, i, l). Note the differing scales.



**Fig. S5** Sensitivity testing for 10-day accumulation: Average duration (a-c), severity (d-f), phase domination (g-i) and frequency (j-l) of seesaw events on a grid cell basis for the period 1950-2021. Showing the concurrent classifications of SPI (a, d, g, j) and SSMI (b, e, h, k), as well as the bivariate SBI (c, f, i, l). Note the differing scales.



**Fig. S6** Sensitivity testing for 90-day accumulation: Average duration (a-c), severity (d-f), phase domination (g-i) and frequency (j-l) of seesaw events on a grid cell basis for the period 1950-2021. Showing the concurrent classifications of SPI (a, d, g, j) and SSMI (b, e, h, k), as well as the bivariate SBI (c, f, i, l). Note the differing scales.



**Fig. S7** New Zealand wide clustering performed on vine copula index (SMI) on a 1-day time step to remove any serial autocorrelation.

138 **Table S1** Optimal bivariate copula family for each season (DJF, MAM etc.), identified as the most common copula family (i.e. most days) for  
139 each season from either Frank, Clayton or Joe copulas, for all grid cells. Identification of optimal copula is performed using AIC selection  
140 criteria.

| Distribution | Precipitation-Soil Moisture Copula |        |        |        | Precipitation-Temperature Copula |         |        |         | Temperature-Soil Moisture Copula |        |        |        |
|--------------|------------------------------------|--------|--------|--------|----------------------------------|---------|--------|---------|----------------------------------|--------|--------|--------|
|              | Summer                             | Autumn | Winter | Spring | Summer                           | Autumn  | Winter | Spring  | Summer                           | Autumn | Winter | Spring |
| 30-Day       | Frank                              | Frank  | Frank  | Frank  | Clayton                          | Clayton | Frank  | Clayton | Clayton                          | Frank  | Frank  | Frank  |

141

## References

- Aas, L., Czado, C., Frigessi, A. and Bakken, H.: Pair-copula constructions of multiple dependence, *Insurance: Math. Econ.*, 44, 182–198, doi:10.1016/j.insmatheco.2008.11.001, 2009.
- 145 Bedford, T. and Cooke, R.: Vines: A New Graphical Model for Dependent Random Variables, *Ann. Stat.*, 30, 1031–1068, doi:10.1214/aos/1023051046, 2002.
- Bennet, M., Kingston, D. and Cullen, N.: Extreme Compound and Seesaw Hydrometeorological Events in New Zealand: An Initial Assessment, *J. Geophysical Research*, 128, e2022JD038346, doi:10.1029/2022JD038346, 2023
- 150 Delignette-Mueller, M. and Dutang, C.: fitdistrplus: An R package for fitting distributions, *J. Stat. Softw.*, 64, 1–34, doi:10.18637/jss.v064.i04, 2015.
- Dissmann, J., Brechmann, E., Czado, C. and Kurowicka, D.: Selecting and estimating regular vine copulae and application to financial returns, *Comput. Stat. Data Anal.*, 59, 52–69, doi:10.1016/j.csda.2012.07.005, 2013.
- Erhardt, T. and Czado, C.: Standardized drought indices: a novel univariate and multivariate approach, *J. R. Stat. Soc. C*, 67, 643–664, doi:10.1111/rssc.12251, 2018.
- 155 Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J.: Copula: Multivariate dependence with copulas [R package version 1.0-1], <https://CRAN.R-project.org/package=copula>, 2020.
- Joe, H.: *Dependence Modeling with Copulas*, CRC Press, Boca Raton, FL, U.S.A., 462 pp., 2014.
- Kao, S. and Govindaraju, R.: A copula-based joint deficit index for droughts, *J. Hydrol.*, 380(1), 121–134, 2010.
- 160 Kim, G., Silvapulle, M. and Silvapulle, P.: Comparison of semiparametric and parametric methods for estimating copulas, *Comput. Stat. Data Anal.*, 51(6), 2836–2850, 2007.
- McKee, T., Doesken, N. and Kleist, J.: The relationship of drought frequency and duration to time scales, 8th Conf. Appl. Climatol., American Meteorological Society, Anaheim, CA, U.S.A., 17–22 January, 1993.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D., Piles, M., Rodríguez-Fernández, N., Zsoter, E., Buontempo, C. and Thépaut, J.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, *Earth Syst. Sci. Data*, 13(9), 4349–4383, 2021.
- 165 Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E., Graeler, B. and Erhardt, T.: VineCopula: Statistical Inference of Vine Copula [R package version 2.4.4], <https://github.com/tnagler/VineCopula>, 2022.
- 170 Nagler, T. and Vatter, T.: rvinecopulib: High Performance Algorithms for Vine Copula Modeling [R package version 0.5.5.1.1], <https://CRAN.R-project.org/package=rvinecopulib>, 2021.

- Pham, M., Vernieuwe, H., De Baets, B., Willems, P. and Verhoest, N.: Stochastic simulation of precipitation-consistent daily reference evapotranspiration using vine copulas, *Stoch. Environ. Res. Risk Assess.*, 30(8), 2197–2214, 2016.
- 175 R Core Team: R: A language and environment for statistical computing, R Found. Stat. Comput., Vienna, Austria, <https://www.R-project.org/>, 2017.
- Rashid, M. and Wahl, T.: Hydrologic risk from consecutive dry and wet extremes at the global scale, *Environ. Res. Commun.*, 4(7), 71001–71014, 2022.
- Rosenblatt, M.: Remarks on a Multivariate Transformation, *Ann. Math. Stat.*, 23(3), 470–472, 1952.
- 180 Sklar, M.: Fonctions de Répartition à n Dimensions et Leurs Marges, *Publ. Inst. Stat. Univ. Paris*, 8(1), 229–231, 1959.
- Stagge, J., Tallaksen, L., Gudmundsson, L., Van Loon, A. and Stahl, K.: Candidate distributions for climatological drought indices (SPI and SPEI), *Int. J. Climatol.*, 35(13), 4027–4040, 2015.
- Tootoonchi, F., Sadegh, M., Haerter, J., Rätty, O., Grabs, T. and Teutschbein, C.: Copulas for hydroclimatic analysis: A practice-oriented overview, *WIREs Water*, 9(2), 1–28, 2022.
- 185 World Meteorological Organisation: Guidelines on the Calculation of Climate Normals, World Meteorol. Org., Geneva, Switzerland, 29 pp., 2017.
- Wu, Z., Su, X., Singh, V., Feng, K. and Niu, J.: Agricultural Drought Prediction Based on Conditional Distributions of Vine Copulas, *Water Resour. Res.*, 57(8), 1–23, 2021.
- 190 Xu, Y., Wang, L., Ross, K., Liu, C. and Berry, K.: Standardized Soil Moisture Index for Drought Monitoring Based on Soil Moisture Active Passive Observations and 36 Years of North American Land Data Assimilation System Data: A Case Study in the Southeast United States, *Remote Sens.*, 10(3), 301, 2018.
- Zscheischler, J., Michalak, A., Schwalm, C., Mahecha, M., Huntzinger, D., Reichstein, M., Berthier, G., Ciais, P., Cook, R., El-Masri, B., Huang, M., Ito, A., Jain, A., King, A., Lei, H., Lu, C., Mao, J., Peng, S., Poulter, B., Ricciuto, D., Shi, X., Tao, B., Tian, H., Viovy, N., Wang, W., Wei, Y., Yang, J. and Zeng, N.: Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data, *Global Biogeochem. Cycles*, 28(6), 585–600, 2014.
- 195