A past, present and future perspective on the European summer vapour pressure deficit

Daniel F. Balting^{1*}, Simon Michel², Viorica Nagavciuc^{1,3}, Gerhard Helle⁴, Mandy Freund⁵, Gerhard H. Schleser⁶, David N. Steger⁷, Gerrit Lohmann^{1,8} and Monica Ionita^{1,9}

⁵ ¹Alfred-Wegener-Institute, Bremerhaven, 27570, Germany

² Institute for Marine and Atmospheric research Utrecht (IMAU), Department of Physics, Utrecht University, Utrecht, Netherlands

³ Faculty of Forestry, Ștefan cel Mare University, Suceava, 720229, Romania

⁴German Research Centre for Geosciences, Potsdam, 14473, Germany

⁵ Climate and Energy College, University of Melbourne, Melbourne, VIC 3010, Australia
 ⁶ Institute of Bio- and Geosciences IBG-3, Forschungszentrum Jülich, Jülich 52428, Germany
 ⁷ Physics Department, University of Bremen, Bremen, 28359, Germany
 ⁸ Deutsches Archäologisches Institut, Berlin, 14195, Germany

⁹ Emil Racovita Institute of Speleology, Romanian Academy, Cluj-Napoca, 400006, Romania

15

Correspondence to: Viorica Nagavciuc (viorica.nagavciuc@awi.de)

Supplementary Information

20

30 1 Random Forest Reconstruction

We use a Random Forest approach (simplified illustration is shown in Fig. S1) to reconstruct VPD at each grid point. RF is a machine learning method, which was introduced by Breiman (2001). Since RF encompass a large variety of regression and classification methods, we use the random forest regression known as random-input random forests (Breiman, 2001). This approach is used among others within the publication of Michel et al. (2020) and is performed using a model evaluation and

35 optimization framework by adding hold-out sampling and cross validation techniques to the initial RF algorithm. In this section, we first detail the general workflow with which a reconstructed timeseries is evaluated within the reconstruction framework. In a second step, we give details on how a RF model is built and optimized for a given pair of VPD and proxy data.

1.1 General workflow for the reconstruction and evaluation

- We here consider as predictands the observed fields of the climate quantity which we want to reconstruct. In our case, we
 independently reconstruct each of the observation time series from the VPD fields over the period 1900-1994 as predictand Y (applied to each grid point of the initial VPD data). The predictors which are needed to reconstruct Y are the δ¹⁸O_{cel} time series from our δ¹⁸O_{cel} network (denoted X₁,...,X₂₆). No proxy selection is made prior to the RF models fittings, since we use a relatively modest number of predictors (26) well geographically constrained to the area of interest (Europe). Nevertheless, it is worth noting that for the RF-based reconstruction of a given grid point, the RF method naturally attribute largest weights to δ¹⁸O_{cel} time series with the closest variations with the target VPD time series.
- To evaluate the reconstruction of each grid point, the common period of observational data and proxy data is randomly divided (i.e., a bootstrap sample) R=30 times into two distinct groups (i.e., hold-out sampling: Michel et al., 2020). RF models will then be set up using the first groups of each pair of bootstrap splits (called training or calibration periods), each covering 75% of the data. The remaining 25% of the data for each bootstrap split (called the validation periods) will allow evaluating
- 50 the quality of the RF models. The ratio between the validation and the training window is a compromise to have enough years to train the model as well as to calculate meaningful validation statistics. A too small time window for the validation period would make robust statements impossible, as a too short training period may result in the model not being built well due to missing or overlooked data. In other studies, the size of the training period varies between two-thirds and three-quarters of the total test period.
- 55 To quantify the RF model efficiency, the actual values of VPD are compared with the reconstructed VPD over the testing periods, that are given by RF models trained over the corresponding training periods from the same bootstrap split. In our study, we use the Nash–Sutcliffe model efficiency coefficient (CE; Nash and Sutcliffe, 1970) which is primary used in hydrological modelling and is an adapted version of the common R². The CE is defined as:

60
$$CE = 1 - \frac{\sum_{t=1}^{T} (Q_m^t - Q_o^t)^2}{\sum_{t=1}^{T} (Q_o^t - \overline{Q_o})^2}$$

where $\overline{Q_o}$ is the average of the observed VPD, Q_o^t is the RF modelled VPD and Q_o^t is the observed VPD at time t. It represents the following cases which are presented in the publication (Nash and Sutcliffe, 1970). If CE=1, the RF model fits perfectly which is characterized by an error variance equal to zero; if CE=0, it indicates that the RF model has the same

65 predictive skill as the mean of the time-series in terms of the sum of the squared error; if CE<0, it indicates that the observed mean is a better predictor than the model. Therefore, a CE=1 suggest a model with more predictive skill it is useful and meaningful to use. In our study, the hypothesis, if CE is significantly greater than 0 at the 95% confidence level, is tested with a one sided t-test.

The final gridded reconstruction is obtained by reapplying the RF model over the whole period 1900-1994 for a given win-

70 dow (nest), and its final score is obtained as the averaged CE scores over the 30 training/testing splits. In our study, a given grid point will be considered as robustly reconstructed if its CE scores over the 30 training/testing splits are significantly greater than 0 at the 95% confidence, based on a one sided Student t-test.

1.2 Regression tree and random forest

85

For all RF models either based on training samples or complete timeseries over the historical period, we here denote Y the 75 VPD time series for a given grid point, $X = (X_j)_{1 \le j \le p}$ are the p proxy records over the same period than Y (either a training period or the full 1900-1994 period), and $X' = (X'_j)_{1 \le j \le p}$ are the same p proxy records, but for the period over which Y is reconstructed (either a testing period or a period prior to 1900).

The RF is an aggregate of the outputs given by different regression trees (Fig. S1), which are based on randomly drawn subsample of the initial p predictors. To build a single regression tree the X and Y data are gathered at the root of the tree as an initial node. This node is then out into two shild nodes. Any out can be written as:

80 initial node. This node is then cut into two child nodes. Any cut can be written as:

$$\{X^j \le d\} \cup \{X^j \ge d\}$$

where $j = \{1, ..., p\}$ is the proxy (variable) index and $d \in \mathbb{R}$ is the value of the cut. As a consequence, all observations with a jth variable greater than d are placed in one child node and all observations with a jth variable lesser than d are placed in another child node. The optimal (j,d) pair is determined by minimizing the sum of the intra-nodes variance of Y within the two the child nodes (Michel et al., 2020). Therefore, (j,d) is found by minimizing the following convex problem:

$$(j,d) = \arg \min_{\substack{d \in \mathbb{R} \\ 1 \le j \le p}} \sum_{i:X_j^i < d} (Y_i - \frac{\sum_{i:X_j^i < d} Y_i}{\#(\{i:X_j^i < d\})})^2 + \sum_{i:X_j^i > d} (Y_i - \frac{\sum_{i:X_j^i > d} Y_i}{\#(\{i:X_j^i > d\})})^2$$

Where p is the number of proxy records used which varies with the inferior boundary of the nested time frame (section 2.1), X_i^i denotes the ith value of the jth proxy, and # is the cardinal operator. We then recursively apply the same procedure to the

next child nodes with the same variables until one node contains less than s=5 observations. Based on this method, a decision

- 90 or regression tree is created. Hence, the RF model is obtained by building a number of N such trees, but each are based on m randomly picked variables from the p initial ones, with m < p. The proxy records values over the reconstruction period of interest (*i.e.*, X²) are then used to browse each of the RF trees according to the cuts previously made during the model fitting. The RF reconstruction is then given as the average of the outputs given by the N trees.
- It is worth noting that there are inherent parameters to the RF method (i.e., control parameters). These parameters are the number of randomly used proxies in each tree (*m*), the number of trees (*N*) and the amount of data required in a node to stop the construction of regression trees (*s*). The optimization of three parameters is very consuming in terms of time and energy, but a very large advantage of the RF method is that it only needs one to be effectively optimized (Breiman, 2001; Oshiro et al., 2012). Indeed, for a large suite of simulated datasets, Oshiro et al. (2012) have shown that using more than *N*=128 trees have a very low likelihood to gain in robustness, which they then recommended to use whatever the source of modelled data.
- In the same way, Breiman (2001) has shown that s=5 always give very satisfying results, while the gain of robustness by its optimisation does not worth the time it takes to. Therefore, we here further optimise *m*, for each RF model built in this study. To estimate the optimal m for each of the RF models, we use the K-fold cross-validation approach (Stone, 1974; Geisser, 1975) with K=5 similarly to Michel et al. (2020). It consists in building RF models for every possible value of *m* (0<*m*<*p*, *m* ∈ N) over 5 further training/testing samples determined with each of the K folds being used as the testing sample (see
- 105 Michel et al. 2020 for more details on the cross validation procedure). The resulting optimised RF model is then applied to the entire data of *X* and *Y*.



 $\,$ Figure S1: Simplified illustration of the RF approach for our VPD reconstruction.



Figure S2: Investigation of the different spectra of the rolling mean time series of the three reconstructed areas. A, North Europe, B, Central Europe and C, Mediterranean region. The white contour lines indicate significant frequencies (p < 0.01). All the plots are done with the *WaveletComp* package in R (Rosch and Schmidbauer, 2018).

Table S1: Description of the used climate model ensemble for the historical run and three future scenarios. In the entire analysis, the first realization (r1i1p1f1) has been used.

	Historical	SSP1-2.6	SSP2-4.5	SSP5-8.5	References
ACCESS-CM2	1	1	1	1	(Dix et al., 2019)
ACCESS-ESM1-5	1	1	1	1	(Ziehn et al., 2019)
EC-Earth3	1	1	1	1	(EC-Earth Consortium (EC-Earth), 2019a)
EC-Earth3-Veg	1	1	1	1	(EC-Earth Consortium (EC-Earth), 2019b)
EC-Earth3-Veg-LR	1	1	1	1	(EC-Earth Consortium (EC-Earth), 2020)
GFDL-ESM4	1	1	1	1	(Krasting et al., 2018)
INM-CM4-8	1	1	1	1	(Volodin et al., 2019a)
INM-CM5-0	1	1	1	1	(Volodin et al., 2019b)
MIROC6	1	1	1	1	(Tatebe and Watanabe, 2018)
MPI-ESM1-2-HR	1	1	1	1	(Jungclaus et al., 2019)
MPI-ESM1-2-LR	1	1	1	1	(Wieners et al., 2019)
MRI-ESM2-0	1	1	1	1	(Yukimoto et al., 2019)

125 References

Breiman, L.: Random Forests, Mach. Learn., 45, 5-32, https://doi.org/10.1023/A:1010933404324, 2001.

Dix, M., Bi, D., Dobrohotoff, P., Fiedler, R., Harman, I., Law, R., Mackallah, C., Marsland, S., O'Farrell, S., Rashid, H., Srbinovsky, J., Sullivan, A., Trenham, C., Vohralik, P., Watterson, I., Williams, G., Woodhouse, M., Bodman, R., Dias, F. B., Domingues, C., Hannah, N., Heerdegen, A., Savita, A., Wales, S., Allen, C., Druken, K., Evans, B., Richards, C., Ri-

130 dzwan, S. M., Roberts, D., Smillie, J., Snow, K., Ward, M., and Yang, R.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP (20200715), https://doi.org/10.22033/ESGF/CMIP6.2281, 2019.

EC-Earth Consortium (EC-Earth): EC-Earth-Consortium EC-Earth3 model output prepared for CMIP6 CMIP (20200715), https://doi.org/10.22033/ESGF/CMIP6.181, 2019a.

EC-Earth Consortium (EC-Earth): EC-Earth-Consortium EC-Earth3-Veg model output prepared for CMIP6 CMIP (20200715), https://doi.org/10.22033/ESGF/CMIP6.642, 2019b.

EC-Earth Consortium (EC-Earth): EC-Earth-Consortium EC-Earth3-Veg-LR model output prepared for CMIP6 CMIP (20210201), https://doi.org/10.22033/ESGF/CMIP6.643, 2020.

Geisser, S.: The Predictive Sample Reuse Method with Applications, J. Am. Stat. Assoc., 70, 320–328, https://doi.org/10.2307/2285815, 1975.

- 140 Jungclaus, J., Bittner, M., Wieners, K.-H., Wachsmann, F., Schupfner, M., Legutke, S., Giorgetta, M., Reick, C., Gayler, V., Haak, H., de Vrese, P., Raddatz, T., Esch, M., Mauritsen, T., von Storch, J.-S., Behrens, J., Brovkin, V., Claussen, M., Crueger, T., Fast, I., Fiedler, S., Hagemann, S., Hohenegger, C., Jahns, T., Kloster, S., Kinne, S., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Müller, W., Nabel, J., Notz, D., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Rast, S., Schmidt, H., Schnur, R., Schulzweida, U., Six, K., Stevens, B., Voigt, A., and Roeck-
- 145 ner, E.: MPI-M MPIESM1.2-HR model output prepared for CMIP6 CMIP (20200715), https://doi.org/10.22033/ESGF/CMIP6.741, 2019.

Krasting, J. P., John, J. G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., Rand, K., Zadeh, N. T., Balaji, V., Durachta, J., Dupuis, C., Menzel, R., Robinson, T., Underwood, S., Vahlenkamp, H., Dunne, K. A., Gauthier, P. P., Ginoux, P., Griffies, S. M., Hallberg, R., Harrison, M., Hurlin, W., Malyshev, S., Naik, V., Paulot, F., Paynter, D. J., Ploshay, J.,

150 Schwarzkopf, D. M., Seman, C. J., Silvers, L., Wyman, B., Zeng, Y., Adcroft, A., Dunne, J. P., Dussin, R., Guo, H., He, J., Held, I. M., Horowitz, L. W., Lin, P., Milly, P. C. D., Shevliakova, E., Stock, C., Winton, M., Xie, Y., and Zhao, M.: NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 CMIP (20200715), https://doi.org/10.22033/ESGF/CMIP6.1407, 2018.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Oshiro, T. M., Perez, P. S., and Baranauskas, J. A.: How Many Trees in a Random Forest?, in: Machine Learning and Data Mining in Pattern Recognition, Berlin, Heidelberg, 154–168, https://doi.org/10.1007/978-3-642-31537-4_13, 2012.

Rosch, A. and Schmidbauer, H.: WaveletComp 1.1: A guided tour through the R package, 59, 2018.

Stone, M.: Cross-validation and multinomial prediction, Biometrika, 61, 509–515, https://doi.org/10.1093/biomet/61.3.509, 160 1974.

Tatebe, H. and Watanabe, M.: MIROC MIROC6 model output prepared for CMIP6 CMIP (20201102), https://doi.org/10.22033/ESGF/CMIP6.881, 2018.

Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., Gusev, A., Kostrykin, S., Iakovlev, N., Shestakova, A., and Emelina, S.: INM INM-CM4-8 model output prepared for CMIP6 CMIP (20200715), https://doi.org/10.22033/ESGF/CMIP6.1422, 2019a.

Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., Gusev, A., Kostrykin, S., Iakovlev, N., Shestakova, A., and Emelina, S.: INM INM-CM5-0 model output prepared for CMIP6 CMIP historical (20191113), https://doi.org/10.22033/ESGF/CMIP6.5070, 2019b.

- Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., Bittner, M., Legutke, S., Schupfner, M., Wachsmann, F.,
 Gayler, V., Haak, H., de Vrese, P., Raddatz, T., Mauritsen, T., von Storch, J.-S., Behrens, J., Brovkin, V., Claussen, M.,
 Crueger, T., Fast, I., Fiedler, S., Hagemann, S., Hohenegger, C., Jahns, T., Kloster, S., Kinne, S., Lasslop, G., Kornblueh, L.,
 Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Müller, W., Nabel, J., Notz, D., Peters, K., Pincus, R.,
 Pohlmann, H., Pongratz, J., Rast, S., Schmidt, H., Schnur, R., Schulzweida, U., Six, K., Stevens, B., Voigt, A., and Roeckner, E.: MPI-M MPIESM1.2-LR model output prepared for CMIP6 CMIP (20200715),
- 175 https://doi.org/10.22033/ESGF/CMIP6.742, 2019.

165

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yoshimura, H., Shindo, E., Mizuta, R., Ishii, M., Obata, A., and Adachi, Y.: MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP (20201217), https://doi.org/10.22033/ESGF/CMIP6.621, 2019.

 Ziehn, T., Chamberlain, M., Lenton, A., Law, R., Bodman, R., Dix, M., Wang, Y., Dobrohotoff, P., Srbinovsky, J., Stevens,
 L., Vohralik, P., Mackallah, C., Sullivan, A., O'Farrell, S., and Druken, K.: CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP (20200715), https://doi.org/10.22033/ESGF/CMIP6.2288, 2019.