Exploring the joint probability of precipitation and soil moisture over Europe using copulas

Carmelo Cammalleri[1,2], Carlo De Michele[1], Andrea Toreti[2]

[1]Dipartimento di Ingegneria Civile e Ambientale (DICA), Politecnico di Milano, Milan, 20133, Italy.
[2]European Commission, Joint Research Centre (JRC), Ispra, 21027, Italy.

Correspondence to: Carmelo Cammalleri (carmelo.cammalleri@polimi.it)

Abstract. The joint probability of precipitation and soil moisture is here investigated over Europe with the goal to extrapolate meaningful insights on the potential joint use of these variables for the detection of agricultural droughts within a multivariate probabilistic modeling framework. The use of copulas is explored, being the framework often used in hydrological studies for the analysis of bivariate distributions. The analysis is performed for the period 1996-2020 on the empirical frequencies derived from ERA5 precipitation and LISFLOOD soil moisture datasets, both available as part of the Copernicus European Drought Observatory. The results show an overall good correlation between the two standardized series (Kendall's τ = 0.42±0.1), but also clear spatial patterns in the tail-dependence derived with both non-parametric and parametric approaches. About half of the domain shows symmetric tail-dependence, well reproduced by the Student-t copula; whereas the rest of the domain is almost equally split between low and high tail-dependences (both modeled with the Gumbel family of copulas). These spatial patterns are reasonably reproduced by a random forest classifier, suggesting that this outcome is not driven by chance. This study stresses how a joint use of standardized precipitation and soil moisture for agriculture drought characterization may be beneficial in areas with strong low tail-dependence, and how this behavior should be carefully considered in multivariate drought studies.

1. Introduction

Agricultural drought, defined as a condition of unusually high precipitation shortages and/or soil water deficits causing adverse effects on crop yields and production (Panu and Sharma, 2002), is probably the most recognized of the four main drought types or phases (Wilhite and Glantz, 1985). This is mainly due to the more direct and easier to understand impacts compared to the other types of droughts (Mishra and Singh, 2010). The scientific literature on agricultural drought provides a large variety of indices (WMO and GWP, 2016), with the aim of reproducing the temporal dynamics of crop water deficit through a combination of climatic observations, hydrological modeling, and remote sensing data (Zargar et al., 2011).

The difficulty in capturing the multi-facet nature of agricultural drought events across the world with a single approach (Sivakumar et al., 2011) is confirmed by the absence of consensus in the scientific literature on the most reliable agricultural drought index. Despite the large range of available indices, some common characteristics can be identified, such as the focus on some proxy variables of plant water availability – through soil moisture (Dutra et al., 2008), actual evapotranspiration (Anderson et al., 2011) or basic meteorological information (Vicente-Serrano et al., 2010) – and the need to account for deviations from long-term conditions (i.e., use of standardized anomalies).

Meteorological drought indicators computed on appropriate aggregation time scales (McKee et al., 1993; Vicente-Serrano et al., 2010) have demonstrated a good capability to represent agricultural drought conditions in several case studies (e.g., Bachmair et al., 2018; Mohammed et al., 2022; Tian et al., 2018). They have been successfully integrated in a number of operational drought monitoring systems, thanks to their minimal input data requirement and ease of use. Among those indices, the Standardized Precipitation Index (SPI, McKee et al., 1993) computed on short-to-medium aggregation periods (i.e., SPI-3 and -6) is often adopted as a suitable proxy variable for agricultural droughts (WMO, 2012).

As highlighted by Sheffield and Wood (2007), simplified indices for drought monitoring, such as the Palmer Drought Severity index (PDSI; Palmer, 1965) or the previously mentioned meteorological indicators, have been slowly integrated with indices directly based on modeled soil moisture data. This transition is fostered by the increasing availability worldwide of process-based hydrological models. Soil moisture percentile, or similarly standardized quantities, are often used for this scope (Mo and Lettenmeier, 2013; Xia et al., 2014). The ever-growing records of remote

54 sensing-based estimates of soil moisture are becoming an additional data source to support the

55 development of dedicated soil moisture-based drought indices (Cammalleri et al., 2017; Carrão et

56 al., 2016).

57   In the context of agricultural drought, an overall good agreement between SPI and soil

58 moisture indices has been demonstrated over a large range of agricultural practices, crop types and

59 climatic conditions. Halwatura et al. (2017) showed how SPI-3 represents a good approximation

60 of modeled soil moisture over three different climatic regions in eastern Australia. Sims et al.

61 (2002) found high correlation between short-term precipitation deficit and soil moisture variations

62 in North Carolina, while Ji and Peters (2003) highlighted the high correlation between SPI-3 and

63 vegetation growth over croplands and grasslands in the U.S. Great Plains. Wang et al. (2015)

64 observed a good matching between soil moisture dynamics and SPI at the scale of 1-3 months

65 when testing various indices over China. In Europe, Manning et al. (2018) highlighted how

66 precipitation is the main driver of soil moisture droughts for a set of both dry and wet sites.

67   In spite of the above-mentioned consistencies, the outcome of any drought analysis is

68 inevitably affected by the index selected to characterize drought conditions over a certain study

69 region, as also highlighted by Quiring and Papakryiakou (2003) in testing different indices over

70 the Canadian prairies. These authors suggest that a variety of drought indices should always be

71 tested to determine the most appropriate one for a given application. It follows that the synergy

72 between multiple indices can be exploited by the use of multivariate indicators (Hao and Singh,

73 2015), a family of approaches that encompasses a variety of merging strategies, including

74 combined cascading indices (Cammalleri et al., 2021a; Rembold et al., 2019), composite and

75 integrated approaches (Brown et al., 2008; Svoboda et al., 2002), and joint probability functions

76 (Bateni et al., 2018; Hao and AghaKouchak, 2013; Kanthavel et al., 2022).

77   The latter category, in particular, aims at capturing the complex statistical dependence

78 among different drought-related variables (Hao and Singh, 2015), and it has seen a growing

79 relevance in many hydrological applications thanks to the introduction of copula functions and

80 their ability to model a wide range of dependence structures (Nelsen, 2006; Salvadori et al., 2007;

81 Joe, 2015). In the field of drought indices, the approach proposed by Kao and Govindaraju (2010)

82 for the computation of the Joint Deficit Index (JDI) has been applied to a variety of drought-related

83 quantities over different regions, often including precipitation and soil moisture (i.e., Dash et al.,

84 2019; Kwon et al., 2019).

85   A key feature in using joint probability is the possibility to characterize the so-called tail-
86   dependence (TD), namely the asymptotical dependence of the extremes (Frahm et al., 2005). While
87   TD has received large attention in the scientific literature of hydrological extremes (e.g.,
88   Aghakouchak et al., 2010; Poulin et al., 2007; Serinaldi, 2008), its use is largely unexploited in
89   studies focusing on combined drought indices.

90   Studies on the marginal distribution of either precipitation or soil moisture usually adopt
91   the Gamma distribution for precipitation and the Beta distribution for soil moisture. The use of the
92   Gamma family for the implementation of the SPI at different accumulation periods has become a
93   standard practice in many applications (e.g., Mo and Lyon, 2015; Yuan and Wood, 2013). While
94   other distributions have also proven to be reliable, such as the exponentiated Weibull (Pieper et
95   al., 2020) and the Person Type III (Ribeiro and Pires, 2016), fitting the Gamma is still the most
96   adopted approach. Over Europe, Stagge et al. (2015) demonstrated how the Gamma outperformed
97   the other tested distributions across all accumulation periods and regions.

98   A more limited number of applications based on soil moisture data are available in the
99   scientific literature compared to SPI. The use of the Beta distribution for soil moisture data has
100   been introduced as early as the late '70s, with the pioneer study of Ravelo and Decker (1979),
101   following the consideration that soil moisture is a double-bounded quantity, ranging between
102   residual and saturation. Sheffield et al. (2004) successfully applied this standardization for drought
103   analyses over the US, while the same distribution has been adopted by Cammalleri et al. (2016)
104   on modeled data over Europe. Most recently, the Beta distribution was also used to characterize
105   the frequency of global satellite soil moisture data (Sadri et al., 2020).

106   Conversely, no standard approaches have been identified for the application of copulas to
107   model the bivariate joint distribution of precipitation and soil moisture, mainly due to the large
108   variety of probabilistic structures than may be observed between these two quantities. Common
109   fitting strategies rely on the application of various copula families to identify the optimal for each
110   specific site (e.g., Hao and AghaKouchak, 2013), or are based on an a-priori selection of a copula
111   family following empirical evidence (e.g., Dixit and Jayakumar, 2021). Independently from the
112   selection strategy, the adopted copula implicitly assumes an underling TD behavior, which
113   influence on extreme detection should be properly accounted.

114   A comprehensive study on the joint probabilistic dynamics of precipitation and soil
115   moisture is currently lacking in the scientific literature of multivariate drought modeling. Hence,

116   the main goal of this study is to fill this gap, by investigating the mutual relationship between the

117   empirical frequencies of precipitation (cumulated over 3 months, as for SPI-3) and soil moisture

118   datasets as available over Europe as part of the European Drought Observatory of the Copernicus

119   Emergency Management Service (EDO, https://edo.jrc.ec.europa.eu).

120   A large set of copulas is tested for this purpose across the entire European domain, to

121   identify an optimal modeling of the dependence especially in proximity of the tails (given its major

122   role in extreme detection). The spatial distribution of the results is analyzed to infer evidence of

123   common patterns and behavior, which may support future operational applications based on

124   similar parametric approaches.

125

126   **2. Materials and Methods**

127   **2.1 Precipitation and soil moisture datasets**

128   The study focuses on Europe and makes use of the dataset of indicators available over the region

129   as part of EDO. Precipitation data accumulated over consecutive 3-month periods are used here,

130   as the quantity at the base of the SPI-3 index. Hourly total precipitation maps from the ECMWF

131   ERA5 global atmospheric reanalysis model (https://www.ecmwf.int/en/forecasts/dataset/ecmwf-

132   reanalysis-v5) are collected through the Copernicus Climate Change Service (C3S,

133   https://climate.copernicus.eu/) and cumulated at monthly updates (no missing values are present

134   in the reanalysis dataset). This dataset has proven to be quite reliable over Europe for drought

135   analyses (e.g., Cammalleri et al., 2021b; van der Wiel et al., 2022), as it is currently employed in

136   near-real time as part of the operational tools of EDO. Empirical frequencies of 3-month

137   precipitation are derived from the rainfall records, in order to obtain a non-parametric calculation

138   of the standardized anomaly, SPI-3, without the possible artifact introduced by the fitting of a

139   theoretical distribution (i.e., Gamma distribution) (see Soľáková et al., 2014). From here on, we

140   will refer to this dataset as standardized precipitation.

141   Soil moisture records over the entire European domain are derived from the simulations of

142   the LISFLOOD distributed hydrological rainfall–runoff model (de Roo et al., 2000). LISFLOOD

143   runs in near-real time as part of the European Flood Awareness System (Thielen et al., 2009), and

144   it provides daily soil moisture maps for the root zone at a spatial resolution of 5-km. Daily modeled

145   data are averaged at monthly scale and converted into a Soil Moisture Index (SMI) as in

146   Seneviratne et al. (2010). The model is calibrated and validated over an extensive network of river

147 discharge stations following the procedure described in Arnal et al. (2019), and it has been

148 successfully tested for drought analyses over Europe as part of EDO for the computation of the

149 Soil Moisture Anomaly (SMA) index (Cammalleri et al., 2015). Similar to precipitation, empirical

150 frequencies are computed from the monthly soil moisture data in order to obtain a non-parametric

151 calculation of the standardized anomaly, SMA, thus independent from a theoretical fitting (i.e.,

152 Beta distribution). We will refer to this dataset as standardized soil moisture from hereafter.

153   In this study, data collected for the most recent 25 years (1996-2020) are used as a common

154 period. This period is chosen to minimize the effects of non-stationarity in precipitation records

155 and to avoid the inclusion of early LISFLOOD records that are affected by a lower number of

156 ground meteorological stations in the forcing (Thieming et al., 2022). The time series of both

157 standardized precipitation and soil moisture at grid-cell scale are preliminarily tested for auto-

158 correlation using the partial auto-correlation function (PACF, Box and Jenkins, 1976). This

159 analysis returned positive and statistically significant (95% confidence interval) values only at lag

160 = 1, suggesting a substantial absence of auto-correlation beyond what is expected for time series

161 with smooth temporal dynamics such as 3-month cumulative precipitation and soil moisture.

162   The 300 maps (12 months × 25 years) for the two standardized datasets are then spatially

163 interpolated on a common Lambert azimuthal equal-area (LAEA) projection on a regular grid of

164 5-km using the nearest neighbor algorithm. This is done to preserve the high-resolution

165 information of the soil moisture and by considering the smooth spatial dynamics of precipitation

166 accumulated over 3 months.

167 **2.2 Copula families**

168 The introduction of copulas in multivariate probability modeling has provided to hydrologists a

169 flexible tool to reproduce the joint probability of multiple dependent variables characterized by a

170 variety of marginal distributions (De Michele and Salvadori, 2003; Salvadori and De Michele,

171 2004).

172   Limiting the focus on bivariate variables, the joint probability distribution, F, of two

173 random variables (X_1 and X_2) can be expressed, thanks to the Sklar's theorem, as:

174 $$F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$ (1)

175 where F_1 and F_2 are the marginal distribution of X_1 and X_2, respectively, and C is the copula

176 function (Salvadori et al., 2007).

177   A large variety of parametric formulations has been introduced in the literature to explicitly

178   link the marginal distributions to the joint probability, with some of the most common copula

179   families used in hydrology belonging to the Elliptical and Archimedean copulas (Chen and Guo,

180   2019). Two measures of dependence play a major role in parametric copula inference. The Kendall

181   rank correlation coefficient (τ) is commonly used as a non-parametric measure of overall ordinal

182   association, while the so-called Tail-Dependence (TD) coefficients (Salvadori et al., 2007) are

183   used to estimate the asymptotical degree of dependence in the upper and lower extremes (upper

184   tail-dependence, λ_U, and lower tail-dependence, λ_L, respectively). The estimation of TD non-

185   parametrically is not an easy task, as highlighted by Serinaldi et al. (2015), as it aims at assessing

186   an asymptotic behavior from a finite sample. Several formulations are proposed in the scientific

187   literature (see Frahm et al., 2005), and the method proposed by Schmidt and Stadtmueller (2006)

188   is here used to obtain non-parametric estimates of both TD coefficients.

189   In this study, the parametric bivariate probability of standardized precipitation and soil

190   moisture is assessed by using the R package "VineCopula" (Aas et al., 2009; Dissman et al., 2013).

191   The Akaike Information Criterion (AIC, Stoica and Selen, 2004) is used to select, for each spatial

192   grid cell, the best fitting copula among the wide range of families available in the package. The

193   main properties of some relevant copulas are reported in Table 1, as they will be useful to interpret

194   the successive results.

195   In particular, from the data in Table 1 it is important to highlight how the BB7 copula is a

196   combination of Joe and Clayton copulas, of which it inherits the tail-dependences, and how the

197   TD behavior of a copula can be inverted (i.e., the upper tail-dependence can become the lower and

198   *vice versa*) by simply considering the reciprocal marginals (commonly known as rotated forms,

199   identified by the suffix 180). Information from both non-parametric and parametric approaches are

200   here jointly used to discriminate between different TD behaviors.

201

202

203   **Table 1.** Main copulas analyzed in this study and their upper and lower tail-dependence
204   coefficients (λ_L and λ_U, respectively).

Copula	λ_L	λ_U
Gaussian	0	0

Student-t	$2t_{v+1}\left(-\sqrt{v+1}\sqrt{\dfrac{1-\rho}{1+\rho}}\right)$	$2t_{v+1}\left(-\sqrt{v+1}\sqrt{\dfrac{1-\rho}{1+\rho}}\right)$
Gumbel	0	$2-2^{\frac{1}{\theta}}$
Clayton	$2^{\frac{-1}{\theta}}$	0
Joe	0	$2-2^{\frac{1}{\delta}}$
BB7	$2^{\frac{-1}{\theta}}$	$2-2^{\frac{1}{\delta}}$

Even if a copula is selected as the optimal based on the AIC, this does not necessarily exclude that other copulas may perform similarly. For this reason, we introduced a further test based on the relative likelihood criterion (Burnham and Anderson, 2002), $\exp\left(\dfrac{AIC_{min}-AIC_i}{2}\right)$, to establish the likelihood that an AIC value of a given copula (AIC_i) is significantly different than the minimum value (AIC_{min}) obtained for the optimal solution.

2.3 Random forest classification of selected copulas

The interpretation of the selected copula functions may help highlighting the transferability of the observed results over different contexts. For this reason, the observed spatial distribution of the selected copulas is analyzed through a random forest classifier (Breiman, 2001), in order to find evidence of reproducible patterns beyond simple chance.

As input features we consider a set of commonly available variables, such as: ground elevation, annual average temperature, annual total precipitation, precipitation seasonality (ratio between total precipitation in warm and cold months), annual average Normalized Difference Vegetation Index (NDVI), annual average soil moisture, and soil type. As hyperparameters for the random forest, we tuned the number of trees (ntree) and the number of features randomly sampled at each split (mtry) using the "randomForest" R package (Breiman, 2001).

3. Results

A preliminary analysis of the degree of correlation between the monthly standardized 3-month precipitation and soil moisture (analogous to non-parametric SPI-3 and SMA) is tested on the full

225   timeseries of each grid cell using the Kendall's τ, as depicted in Fig. 1 for the entire European
226   domain.

227

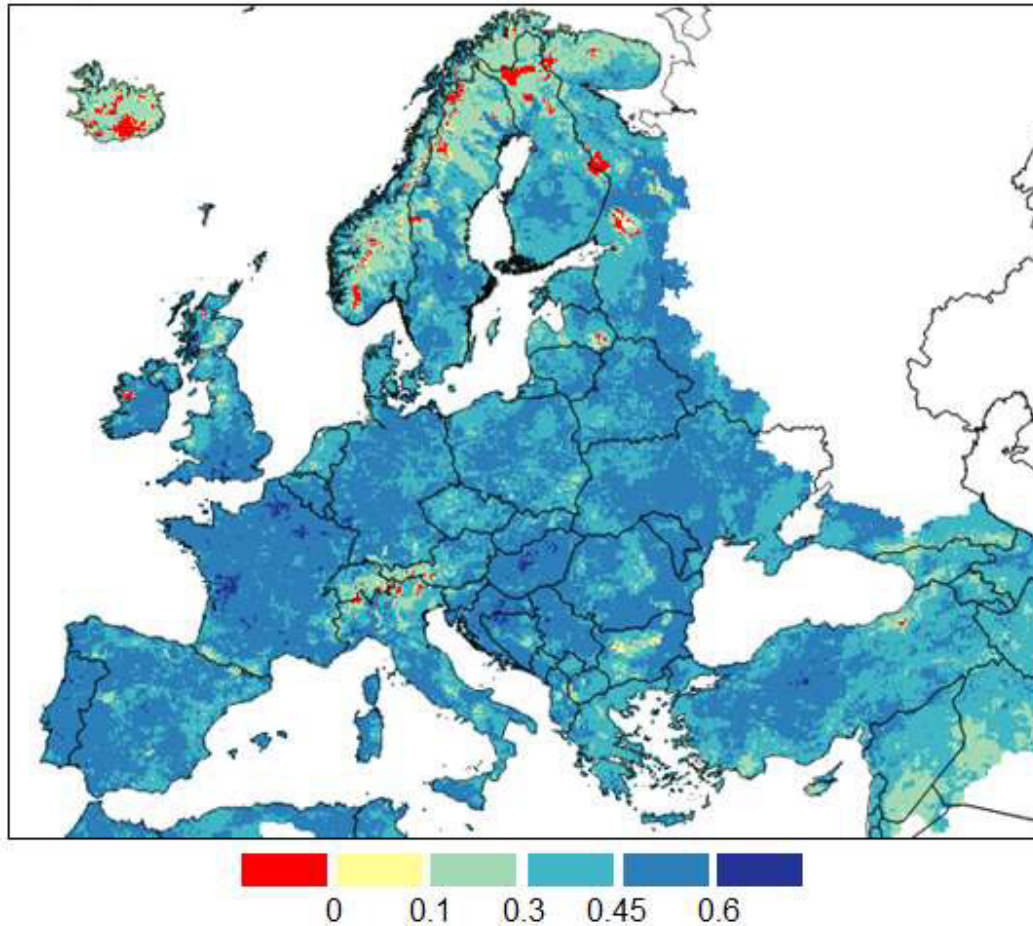



Fig. 1. Spatial distribution of the Kendall's τ between monthly standardized 3-month precipitation
and soil moisture. Roughly, values lower than 0.1 are not statistically significant at $p = 0.05$ (two-tails).

        The results reported in Fig. 1 confirms the expected direct relation between the two
variables, with a relatively homogeneous distribution of medium/high (between 0.3 and 0.5) τ
values (τ = 0.42±0.1). Limited regions with low (and sometimes even slightly negative) τ values
are sporadically observed, mostly over the Alps, Iceland and the coldest regions of the Scandinavia
peninsula. Low correlations over these regions are likely related to the presence of snow coverage
during extended periods of the year. Overall, the observed τ values cannot be considered
statistically significant (at $p = 0.05$) only for less than 2% of the domain.

239   The analysis of the non-parametric tail-dependence values is summarized in the plot

240   depicted in Fig. 2, where the cumulative frequency of the difference between the empirical λ_L and

241   λ_U values is reported. The range of TD values in Fig. 2 for which it is possible to exclude significant

242   asymmetry in the tail dependence coefficients is identified by setting a maximum value for $|\lambda_L -$

243    $\lambda_U|$. To define this threshold, the non-parametric TD coefficients were re-computed on shuffled

244   time series (to artificially reconstruct conditions of null dependence), and the $|\lambda_L - \lambda_U|$ value

245   corresponding to a cumulative frequency of 90% of the grid cells after the shuffling was detected

246   as threshold, corresponding to a value of 0.1. This value can be seen as a lower limit to identify

247   symmetric dependence.

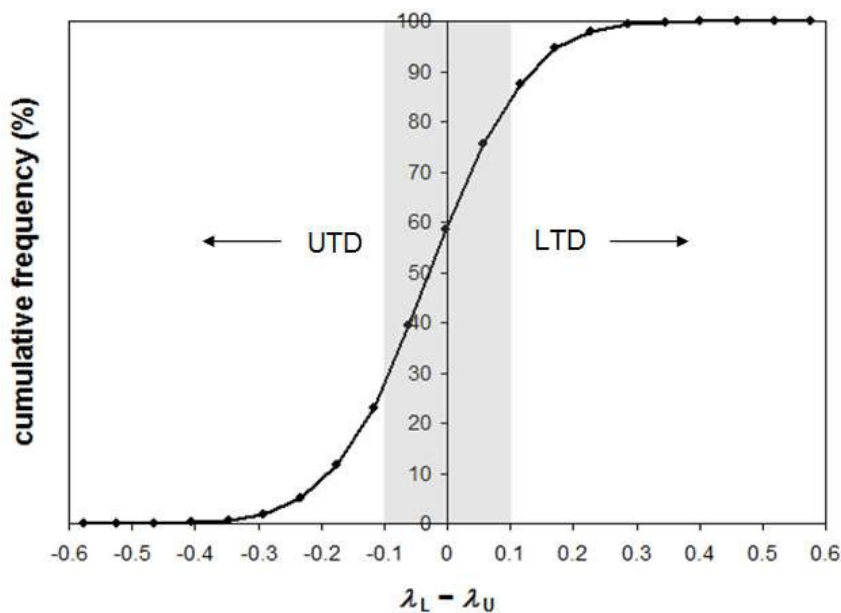
248   The plot in Fig. 2 highlights how the majority (about 50%) of the grid cells can be

249   considered characterized by a symmetric behavior in the tail-dependence coefficients according to

250   the above mentioned criterion ($|\lambda_L - \lambda_U| < 0.1$), whereas the rest of the grid cells are almost equally

251   split between a predominance of the Upper Tail-Dependence (UTD, corresponding to negative

252   differences) or a predominance of Lower Tail-Dependence (LTD, positive differences).

253

254

255 **Fig. 2.** Analysis of the frequency of the empirical tail-dependence coefficients. The plot shows the
256 cumulative frequency distribution of the differences between the empirical λ_L and λ_U values
257 computed according to Schmidt and Stadtmueller (2006). The domain with a roughly symmetric
258 behavior ($|\lambda_L - \lambda_U| < 0.1$) is highlighted by the grey box area.

259

The results reported in Fig. 2 were used to divide the entire domain in three categories (symmetric, LTD and UTD) as depicted in Fig. 3. This map shows evidence of some coherent spatial patterns, such as the predominance of LTD in southern France, southern Italy, northern Germany and Denmark, and western Ukraine (among others), and a clustering of UTD in Poland, Czechia, southern Scandinavia, and Greece. The symmetric condition seems overall more spread across the entire domain, also thanks to the higher frequency, with a slightly predominance over northern Europe (i.e., northern Scandinavian peninsula and Iceland).

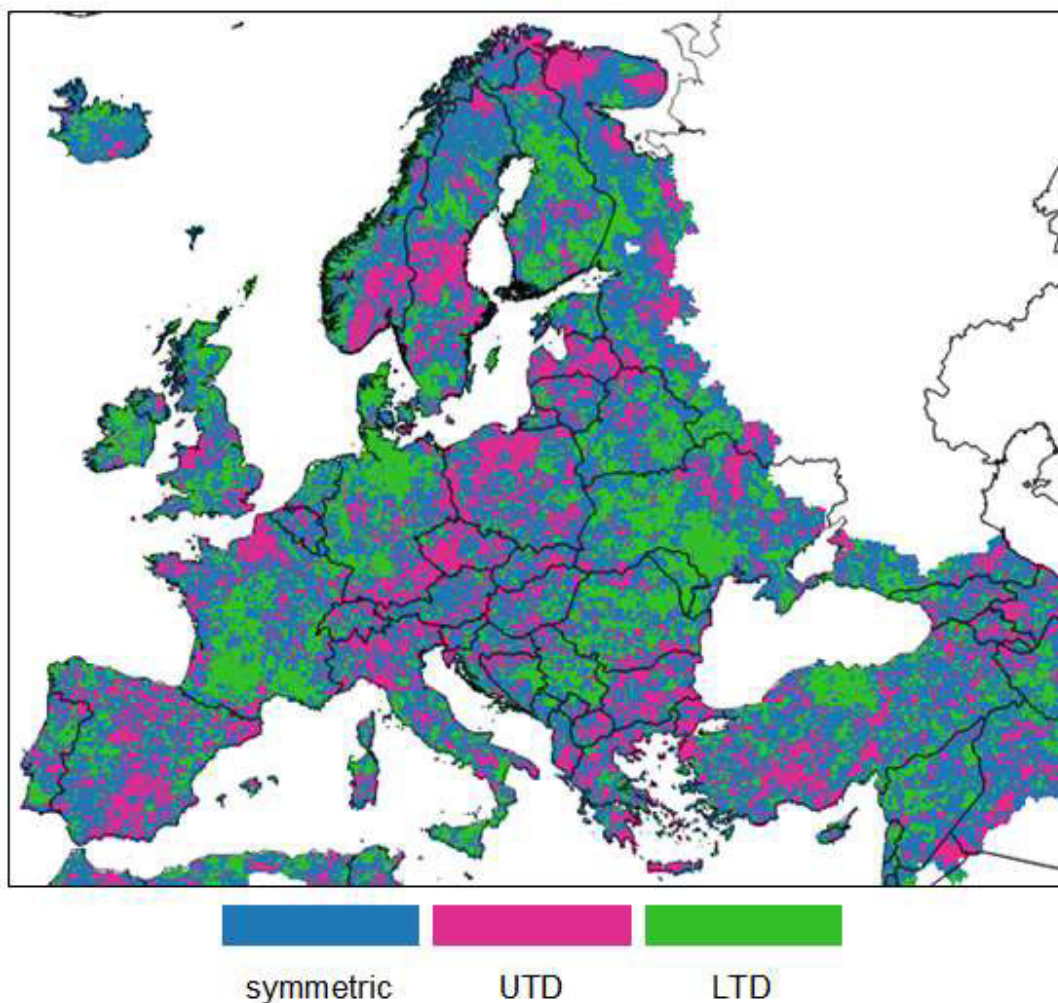



Fig. 3. Spatial distribution of the three categories derived from the differences in the empirical tail-dependence coefficients.

Given the results of the tail-dependence assessment, it is useful to focus the copula parametric analysis on the capability to reproduce such patterns instead of finding the single copula

274 that can perform reasonably well over the entire domain. Indeed, the search for the optimal copula

275 based on the minimum AIC returns the BB7 as the optimal one in about 80% of the domain (not

276 shown). This result is a consequence of the BB7 flexibility (being derived from a combination of

277 two purely asymmetric functions), which allows reproducing both symmetric and asymmetric tail-

278 dependence coefficients according to the values assumed by the two parameters. However, the fact

279 that a single flexible copula works well over a large range of conditions may hide the key spatial

280 patterns observed in the TD analysis. These patterns may be better reproduced by adopting a

281 limited number of more specialized copulas.

282 By limiting the search to a subset of copula functions, comprising only purely symmetric

283 or purely asymmetric tail behaviors, more interesting results are obtained, as summarized by the

284 frequency plot in Fig. 4. The grid cells where symmetric tail behavior copulas are selected as

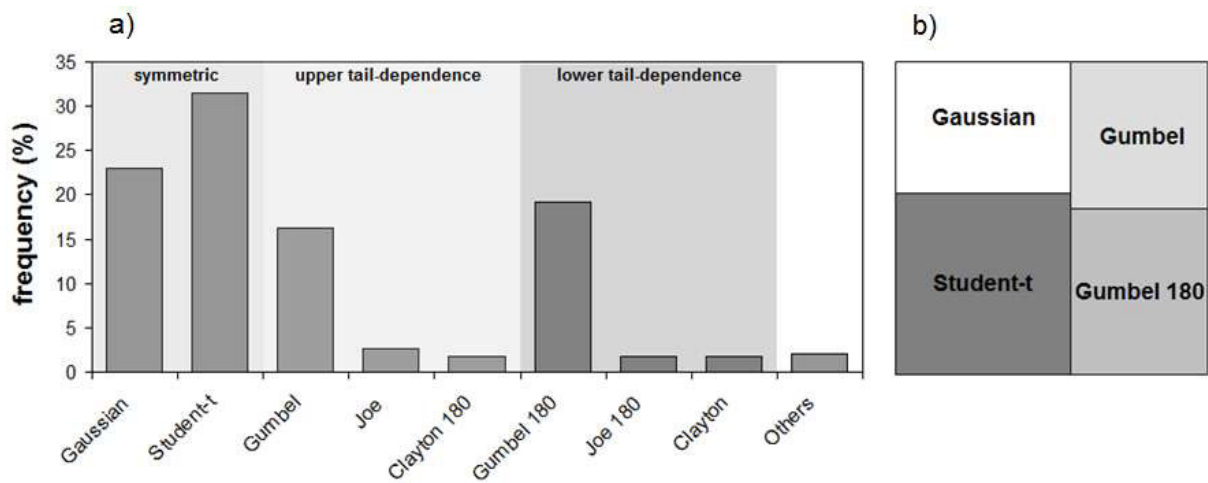
285 optimal are about 55% of the domain (see Fig. 4b), with a predominance of Student-t copula but

286 also with a non-negligible fraction of cells (23%) where the Gaussian (symmetric and without tail-

287 dependence) is chosen (see Fig. 4a). The remaining grid cells are almost equally split between

288 upper and lower tail-dependence, with Gumbel (and its rotated counterpart, Gumbel 180) as the

289 most selected among the asymmetric options.

290

291

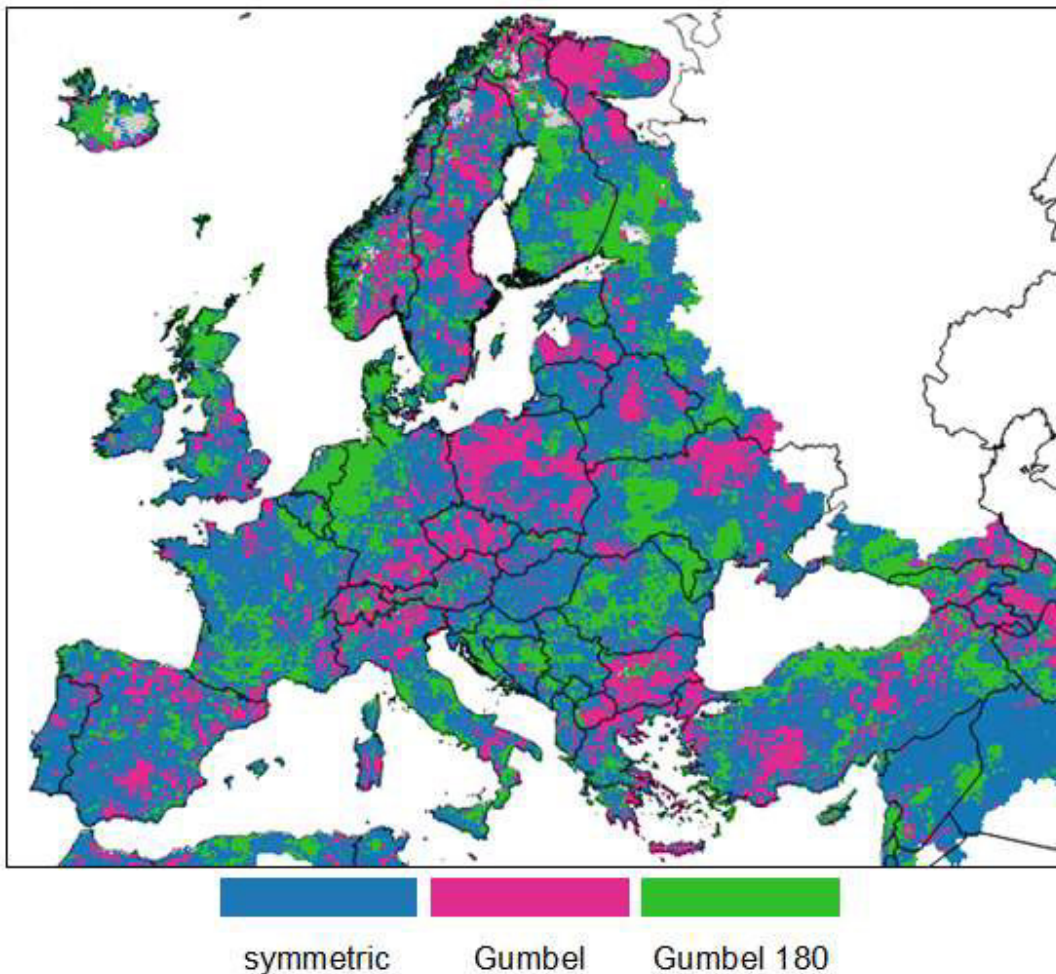
292 **Fig. 4.** Frequency of the optimal copulas based on the minimum AIC. The barplot in panel a)
293 shows the frequency of each copula, while the box in panel b) reports a compact description of the
294 subdivision of the entire domain among the 4 most frequent copulas.

295

296 The spatial distribution of these optimal copulas (Fig. 5) mostly agree with the patterns

297 observed in Fig. 3, supporting the findings on the spatial distribution of TD coefficients. In

298 addition, this result further confirms that a rather limited range of simple copula functions is able
299 to capture the overall dynamics of dependence between precipitation and soil moisture over the
300 entire European domain. Despite the observed spatial clusters in the obtained optimal copulas, the
301 overall patterns observed in Fig. 5 are still rather noisy and may be difficult to interpret. This erratic
302 behavior can be partially explained by the fact that different copulas may perform quite similarly
303 over some grid cells, hence the AIC of the optimal copula (AIC_{min}) may not differ significantly
304 from the AIC of other functions.

symmetric   Gumbel   Gumbel 180

305
306 **Fig. 5.** Spatial distribution of the optimal copulas obtained by minimizing the AIC. The symmetric
307 tail behavior class includes both Gaussian and Student-t copulas.
308

309   To further investigate this hypothesis, we evaluated the possibility to replace the optimal
310 copulas with either a Student-t or a Gumbel (direct and rotated) over the entire domain. The
311 Gaussian copula was excluded from this analysis under the assumption that the no tail-dependence

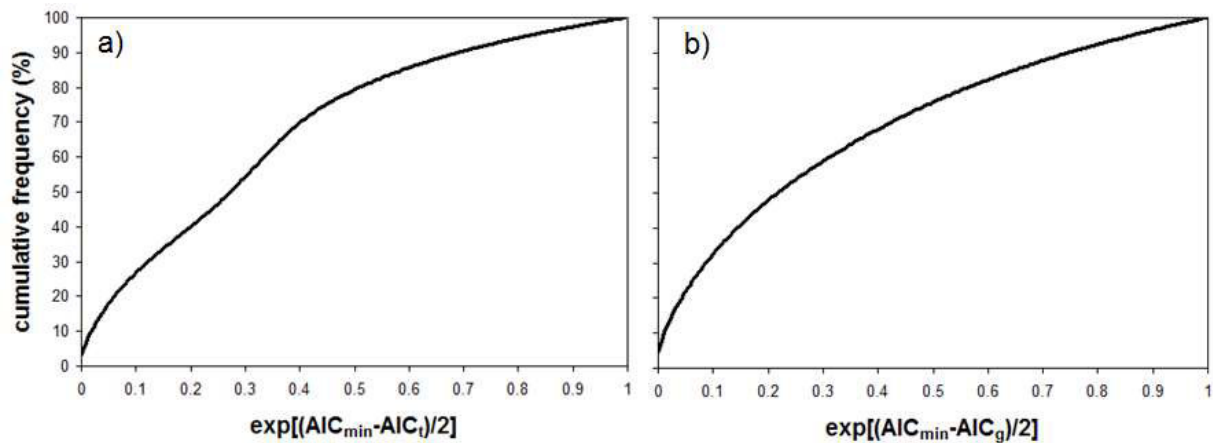
312 of the Gaussian can be adequately reproduced by the Student-t with a small enough tail-

313 dependence. The plots in Fig. 6 reports the relative likelihood for the Student-t (panel a) and

314 Gumbel families (panel b) compared to the locally selected optimal copulas. Low values of this

315 metric correspond to conditions where the optimal copula cannot be replaced by the alternative

316 function (being either the Student-t or the Gumbel).

317

318

319 **Fig. 6.** Frequency analysis of the relative likelihood computed between the optimal AIC (AIC_{min})
320 and: a) Student-t (AIC_t), or b) Gumbel (AIC_g) families. The grid cells where either the Student-t
321 or the Gumbel was already the optimal solution were excluded from the respective frequency
322 analysis.
323

324 The results in Fig. 6 show that, if we assume a relative likelihood of 0.1 as a threshold to

325 detect a statistically significant difference, the Student-t cannot reasonably replace the local

326 optimal copula in about 18% of the entire domain (Fig. 6a), whereas this fraction is about 17% for

327 the Gumbel family (Fig. 6b). It emerges that the Gumbel family is the optimal one in almost the

328 totality (about 99%) of the grid cells where the Student-t is not a suitable replacement of the local

329 optimal, whereas almost only symmetric copulas (63% Student-t and 34% Gaussian) are the

330 optimal functions where the Gumbel family is not a suitable replacement. Overall, these results

331 suggest that the selection of the optimal copula is "univocal" (i.e., cannot be reasonably replaced

332 by another function) in about 35% (18+17) of the domain, whereas either the Student-t or the

333 Gumbel families can be adopted in the remain fraction of the domain with similar performances

334 in terms of AIC (and no clear TD behavior). This analysis also confirms the assumption that all

335 the areas where the Gaussian was chosen as optimal copula can be satisfactory modeled by using

336 the Student-t (i.e. without a statistically significant increase in AIC).

337

Student-t   Gumbel   Gumbel 180

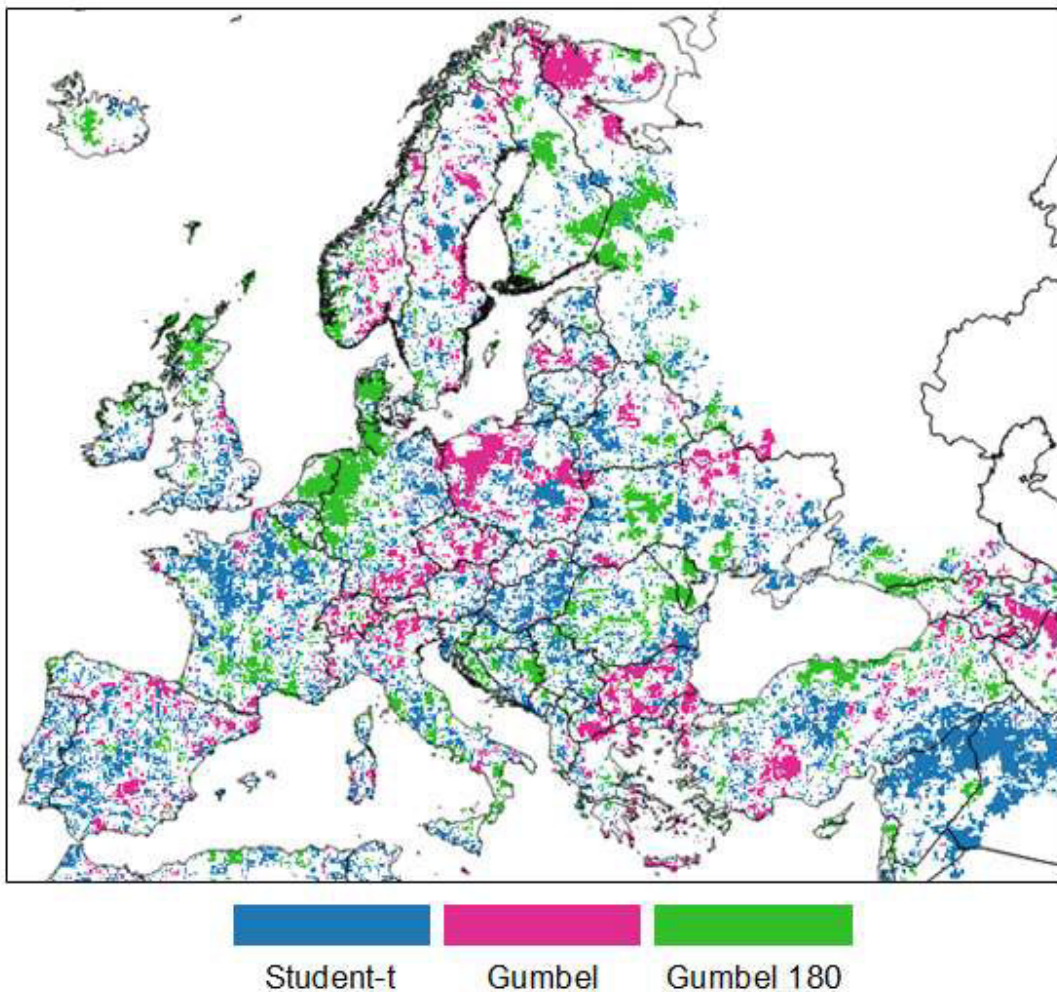



Fig. 7. Spatial distribution of the grid cells where the selection of the optimal copula is "univocal" according to the relative likelihood criterion.

The "univocal" areas derived from the previous analysis are mapped in Fig. 7, highlighting some of the more consistent spatial clusters already observed in both Figs. 3 and 5, as well as a large fraction of cells in northern Europe where a "univocal" optimal copula cannot be selected. These grid cells with "univocal" copula are used as a starting point for the random forest classification, given the robustness in their signal, and the agreement in the outcome of both parametric and non-parametric TD behaviors.

A sample corresponding to 25% of the "univocal" grid cells (about 8% of the entire domain) was used to train the random forest, adopting a number of trees (ntree) of 80 and a single feature randomly sampled at each split (mtry = 1). The training size and the minimum values of

351 hyperparameters were chosen to reduce the problem of overfitting. Among the possible features,

352 three variables were selected by analyzing the variable importance plots, as well as the ease of

353 access: annual average temperature, annual total precipitation, and precipitation seasonality. The

354 trained classifier was then applied to the testing subset (the remaining 75% of the "univocal" grid

355 cells) and the outcomes were analyzed by mean of a confusion matrix, which results are

356 summarized in Table 2. Overall, the obtained classification has a very satisfactory matching with

357 the test subset, with a general high accuracy (ACC = 0.86) and with all the metrics pointing toward

358 a significant improving in the performance compared to the reference No-Information-Rate (NIR)

359 (i.e., small p-values) and a high probability to have correct modeled values compared to simple

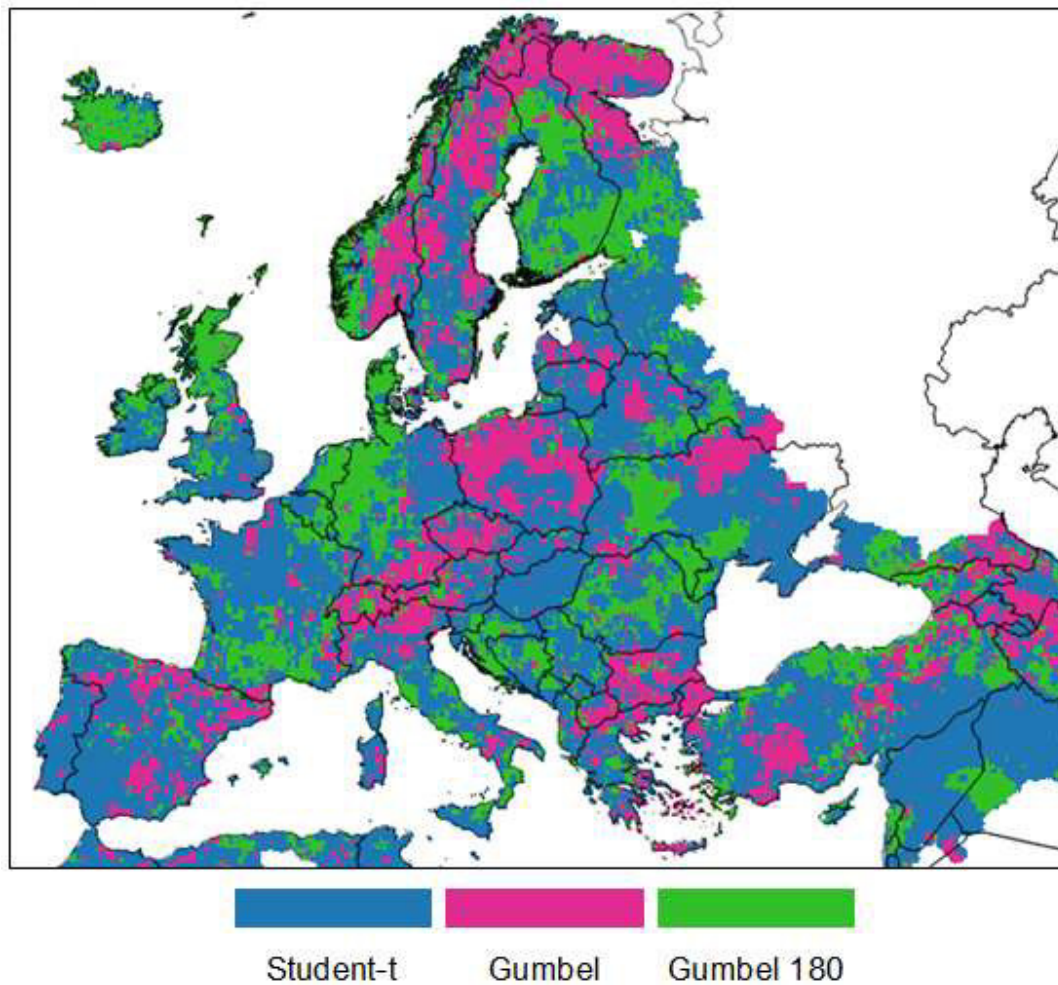
360 chance (i.e., high Cohen's K).

361

362 **Table 2.** Summary of the confusion matrix analysis applied to the trained random forest on the
363 testing subset.

Accuracy (ACC)	0.86
No-Information-Rate (NIR)	0.50
p-value (ACC > NIR)	$< 2.2 \times 10^{-16}$
McNemar's test p-value	3.44×10^{-5}
Cohen's kappa statistic (K)	0.78

364

Student-t   Gumbel   Gumbel 180

365

366 **Fig. 8.** Map of the optimal copula as modeled by the trained random forest classifier.

367

368   Finally, the trained classifier was applied to the entire dataset to obtain a classification of

369 the European domain in term of the expected optimal copula and the corresponding TD behavior.

370 This map, reported in Fig. 8, shows a strong resemblance to both the empirically-derived map in

371 Fig. 3 and the optimal AIC fitting in Fig. 5. Beside this overall agreement, some notable

372 discrepancies can be observed over northern Scandinavia and Iceland, two regions where low

373 Kendall's τ and a small fraction of "univocal" selected copulas were already identified.

374

375 **4. Discussion**

376 The overarching goal of the study is to investigate the joint probability of two standardized

377 variables aiming at capturing agricultural drought conditions, hence the overall agreement between

378 these two quantities is a fundamental prerequisite. A direct relationship between standardized 3-

month cumulated precipitation and soil moisture is expected, since both SPI-3 and SMA are similarly-used agricultural drought indices, and this can support the identification of the most suitable set of copula families (Salvadori et al., 2007; Genest et al., 2007). This direct relationship is overall confirmed by the positive Kendall's τ values estimated over most of the domain ($\tau = 0.42\pm0.1$). Moderately high correlation values of standardized precipitation and soil moisture were estimated also in other studies. Kwon et al. (2018) reported Pearson's r values between 0.4 and 0.6 for 55 stations in South Korea, albeit with seasonal patterns; Gaona et al. (2022) found similar values over the Ebro basin with both land-surface modeled and satellite soil moisture, and Sepulcre-Cantó et al. (2012) obtained an average value of r of about 0.6 over nine stations across Europe.

Sehler et al. (2019) studied the correlation between remote sensing-based precipitation and soil moisture, finding moderate correlation over southern Europe, and a weak (often not significant) correlation in central Europe. However, central Europe is close to the upper limit of the analyzed remote sensing products, which can explain such low performance. Limited correlation even among different soil moisture products has been observed in northern Europe in other studies (Almenda-Martín et al., 2022), confirming the difficulty to model soil moisture dynamics over this region.

The obtained values for the Kendall's τ fall in a somewhat optimal range for the analysis of the joint probability, since they are statistically significant almost everywhere (i.e., the two indices are to a certain degree consistent) but not too high to make meaningless any joint use of the two datasets (i.e., the two indices are too similar and provide the same information).

The outcome of the tail-dependence analysis is even more interesting, given the role that such metric plays in the detection of extreme events (and in particular the low-tail for droughts). The TD investigation is sometimes overlooked in the development of multivariate drought indices, where previous studies often focused on optimizing the copula to the local data without analyzing the implicit assumption on the TD, the consistency with the non-parametric TD, and the implications of the associated dependence. Previous studies on the joint probability of precipitation and soil moisture are rather scarce, and TD is rarely the focus of such analyses or, at least, limited to specific areas and/or conditions.

As an example, Manning et al. (2018) performed a very detailed analysis over 11 FluxNet sites in Europe on the role of precipitation and evapotranspiration on soil moisture drought, based

on pair copula constructions, but the authors did not provide any indication on which bivariate copula was the optimal one for each site. Kwon et al. (2018) reported that Frank copula was the most frequent optimal choice in their study over South Korea. However, some clear spatial patterns observed in their outcomes were not discussed, with Frank being the selected copula mostly in the central area of the domain, but with Gumbel and Student-t performing the best in the southern and eastern coasts, respectively.

Dash et al. (2019) found Frank (among the Archimedean copulas) working the best for 3-month precipitation and soil moisture over an Indian basin; while Hao and AghaKouchak (2013) highlighted the good performance of Frank and Gumbel in five regions of California, even if neither Gaussian nor Student-t were considered. In all these applications, no specific considerations on the TD behaviors were reported, even if a common trend seems to be the good performance of Frank copula. This is in contrast with our results, where the Frank was very rarely selected as optimal (less than 1% of the domain). A possible explanation of these results may be our focus on empirical marginal frequencies rather than theoretical ones, given the well-documented increasing uncertainty in parametric fitting in the tails (Farahmand and AghaKouchak, 2015; Laimighofer and Laaha, 2022). As a possible confirmation of this hypothesis, a good performance of Gumbel and Gaussian has been observed over Iran by Bateni et al. (2018), similarly to our results, when a non-parametric form for SPI and SSI (Standardized Soil Moisture Index) was used.

The absence of a strict standard procedure to investigate tail-dependence may be another factor affecting the limited focus on the topic in many studies on multivariate drought indices. Non-parametric TD has the clear advantage to avoid any alteration of the data due to the fitting procedure, but the outcomes in this study also show a high degree of spatial noise likely due to the intrinsic nature of non-parametric analyses, the large uncertainty in non-parametric methods (Serinaldi et al., 2015), as well as the effects of the limited sample size (for this last issue see also the illustration 3.18 in Salvadori et al., 2007). The threshold used here to define a symmetric behavior, based on a random shuffling of the data, seems to successfully overcome the difficulty to define a self-consistent maximum difference in TD values, but it cannot be seen as a reliable approach to easily identify TD symmetry without the support of further evidence (e.g., by theoretical analyses).

440   In this regard, the fitting of parametric copula functions returns spatial patterns in TD

441   coefficients similar to the ones obtained with the non-parametric approach. However, the absence

442   of "univocal" fittings can be observed for large areas, as well as some contrasting results compared

443   to the non-parametric TD especially over northern Europe (areas with low correlation). The grid

444   cells where a given copula clearly outperforms the alternative options is limited to roughly 1/3 of

445   the domain, further stressing the evidence that clear-cut outcomes are difficult to infer from a

446   single methodology. Thus, it seems reasonable to state that only a critical concerted analysis of

447   both parametric and non-parametric TDs can return robust practical indications based on a

448   converge of evidence.

449   A clear outcome of our study is the predominance of regions with symmetric tail-

450   dependence coefficients, where the Student-t copula is suitable to reproduce the joint probability

451   of standardized precipitation and soil moisture. An even split of the remaining domain between

452   areas with either lower or upper tail-dependence is also observed, where the Gumbel copula (in

453   either is direct or 180 rotated forms) is proven to be a suitable option. These results are crucial in

454   defining the role of standardized precipitation and soil moisture datasets in detecting drought

455   events, and to which extent they can work in synergy in a drought monitoring system. While the

456   correlation between the two datasets highlights the extent of their overall agreement, which in this

457   study was somewhat uniform across most of the domain (τ ranging between 0.3 and 0.5), very

458   different degrees of tail-consistency can be obtained for similar Kendall's τ if the TDs differ

459   substantially. Regions with higher LTD will have a higher agreement in the detection of drought

460   extremes compared to areas with a UTD predominance, hence a low number of false alarm and a

461   higher signal-to-noise ratio may be expected.

462   To further explore this behavior, the time series of standardized variables were converted

463   in binary vectors based on the commonly used standardized drought threshold of -1 (corresponding

464   to an empirical frequency of 0.16). On these data, the pair-wise binary correlation coefficient, $\rho(-1)$,

465   was computed separately for the grid cells with LTD and UTD. Results are shown in Fig. 9,

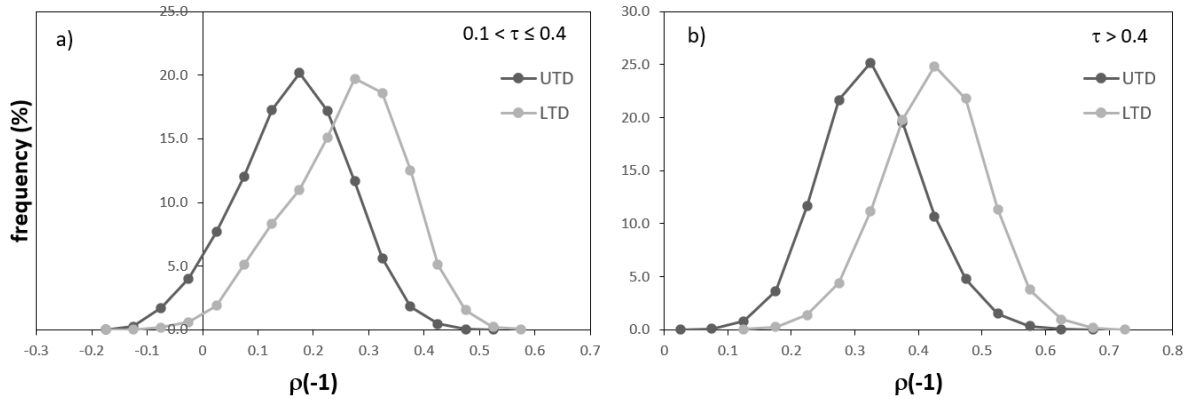
466   for grid-cells with low ($0.1 < \tau \leq 0.4$, panel a) and high ($\tau > 0.4$, panel b) overall correlation,

467   respectively. They show a net increase in the pairwise binary correlation for the grid cells with

468   LTD (of about 0.15 in both cases) compared to the cases with UTD, even if the overall correlation

469   is comparable. This increase in $\rho(-1)$ translates in a stronger agreement in the detection of extremes

470   when a low tail-dependence is observed, resulting in a more robust detection of the drought

471 conditions thanks to the concurrency of extreme conditions in both drought indices (i.e.,
472 convergence of evidence).
473

474
475 **Fig. 9.** Frequency distribution of the pairwise binary correlation between standardized
476 precipitation and soil moisture lower than -1, computed separately for grid cells with UTD (dark
477 grey lines) and LTD (light grey lines). Panel a) reports the results for the grid cells with low overall
478 correlation ($0.1 < \tau \le 0.4$), while panel b) reports the results for the grid cells with high correlation
479 ($\tau > 0.4$).
480

481 Regions such as southern France, northern UK, northern Germany and Denmark (where a
482 strong LTD is observed, see Fig. 8) are appropriate candidates for a robust assessment of
483 agricultural drought conditions based on a joint precipitation-soil moisture index, whereas some
484 regions in central Europe (i.e., Poland, Czechia, Switzerland) may not equally benefit from the use
485 of a joint index due to the lower importance of LTD.
486 Overall, the parametric copula fittings confirm most of the non-parametric TD patterns
487 suggesting that a parametric approach is suitable for an operational implementation of a
488 precipitation-soil moisture joint drought index over most of Europe. This implies that the proposed
489 procedure, based on the combination of parametric and non-parametric analyses, can be considered
490 a reliable tool to provide meaningful insight on the potential application of joint probability as
491 detector of extreme droughts.
492 At first glance, it may seem difficult to assign an explanation for the observed spatial
493 patterns in LTD and UTD. However, the proven possibility to reasonably reconstruct these spatial
494 patterns with a random forest classifier, starting from only a small sample of robust training data
495 (less than 10% of the domain) and with commonly available driving features, suggests that the
496 observed clusters are unlikely to be caused only by chance and that hidden structures may be

497 present and may be further explored. This result is encouraging for an extension of the derived
498 approach to other regions of the world.
499
500 **5. Summary and Conclusions**
501 The use of combined indices based on copula seems a promising development in the field of
502 drought detection and monitoring. In this study, we analyzed the joint probability of two variables
503 commonly used in agricultural drought analyses: the empirical frequencies of 3-month cumulated
504 precipitation and soil moisture. We focus on the probabilistic characteristics being key for
505 agricultural drought studies.
506   The overall agreement in the marginal probability of the two standardized variables
507 suggests that they are indeed valid candidates for the development of a joint drought index over
508 the European domain. However, an in-depth analysis of the tail-dependence, derived with both
509 non-parametric and parametric approaches, shows some clear spatial patterns, which have direct
510 repercussion on the capability of such data to provide robust and coherent estimates of drought
511 extremes. In this regard, regions such as southern France, northern UK, northern Germany, and
512 Denmark may benefit more from the joint use of the two standardized variables thanks to the
513 observed strong low tail-dependence (i.e., increasing agreement on the left tail extremes). The joint
514 dependence of standardized precipitation and soil moisture is well reproduced by using three
515 common copulas (Student-t, Gumbel and 180 rotated Gumbel), with spatial patterns that were
516 successfully reconstructed with a random forest classification, suggesting the presence of a
517 structure in the outcomes not related to chance.
518
519 **Code availability:** The codes used for this analysis can be provided upon request via the
520 corresponding author.
521
522 **Data availability:** All the data used in this study can be accessed and retrieved through the
523 European   Drought   Observatory   (EDO)   web   portal
524 (https://edo.jrc.ec.europa.eu/gdo/php/index.php?id=2112).
525

526 **Author contribution:** CC designed the experiments, with inputs from AT and CDM. CC
527 developed the codes and performed the analyses. CC prepared the manuscript, which was
528 expanded and revised by all co-authors.
529
530 **Competing interests:** At least one of the (co-)authors is a member of the editorial board of
531 Hydrology and Earth System Sciences.

532 **References**

533 Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence,

534   Ins.   Math.   Econ.,   44(2),   182-198,

535   https://doi.org/https://doi.org/10.1016/j.insmatheco.2007.02.001, 2009.

536 Aghakouchak, A., Ciach, G., Habib, E.: Estimation of tail dependence coefficient in rainfall

537   accumulation   fields,   Adv.   Water   Resour.,   33,   1142-1149,

538   https://doi.org/10.1016/j.advwatres.2010.07.003, 2010.

539 Almenda-Martín, L., Martínez-Fernández, J., Piles, M., González-Zamora, A., Benito-Verdugo,

540   P., Gaona, J.: Influence of atmospheric patterns on soil moisture dynamics in Europe, Sci. Tot.

541   Environ., 846, 157537, https://doi.org/https://doi.org/10.1016/j.scitotenv.2022.157537, 2022.

542 Anderson, M. C., Hain, C., Wardlow, B., Pimstein, A., Mecikalski, J. R., Kustas, W. P.: Evaluation

543   of drought indices based on thermal remote sensing of evapotranspiration over the continental

544   United   States,   J.   Climate,   24(8),   2025-2044,

545   https://doi.org/https://doi.org/10.1175/2010JCLI3812.1, 2011.

546 Arnal, L., Asp, S. -S., Baugh, C., de Roo, A., Disperati, J., Dottori, F., Garcia, R., Garcia Padilla,

547   M., Gelati, E., Gomes, G., Kalas, M., Krzeminski, B., Latini, M., Lorini, V., Mazzetti, C.,

548   Mikulickova, M., Muraro, D., Prudhomme, C., Rauthe-Schöch, A., Rehfeldt, K., Salamon, P.,

549   Schweim, C., Skoien, J. O., Smith, P., Sprokkereef, E., Thiemig, V., Wetterhall, F., Ziese,

550   M.:. EFAS upgrade for the extended model domain – technical documentation, JRC Technical

551   Reports, EUR 29323 EN, Publications Office of the European Union, Luxembourg, 58 pp,

552   https://doi.org/https://doi.org/10.2760/806324, 2019.

553 Bachmair, S., Tanguy, M., Hannaford, J., Stahl, K.: How well do meteorological indicators

554   represent agricultural and forest drought across Europe? Environ. Res. Lett., 13, 034042,

555   https://doi.org/10.1088/1748-9326/aaafda, 2018.

556 Bateni M. M., Behmanesh J., De Michele C., Bazrafshan J., Rezaie H.: Composite

557   agrometeorological drought index accounting for seasonality and autocorrelation, J. Hydrol.

558   Eng., 23(6), 04018020, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001654, 2018.

559 Box G. E. P., Jenkins G. M.: Time Series Analysis, Forecasting and Control, Holden-Day, San

560   Francisco, pp. 64-65, 1976.

561 Breiman,   L.:   Random   forests,   Machine   Learn.,   45,   5-32,

562   https://doi.org/10.1023/A:1010933404324, 2001.

563 Brown, J. F., Wardlow, B. D., Tadesse, T., Hayes, M. J., Reed, B. C.: The Vegetation Drought
564   Response Index (VegDRI): A new integrated approach for monitoring drought stress in
565   vegetation, GISci. Remote Sens., 45(1), 16-46, https://doi.org/10.2747/1548-1603.45.1.16,
566   2008.

567 Burnham, K. P., Anderson, D. R.: Model Selection and Multimodel Inference: A practical
568   information-theoretic approach, Springer-Verlag, 488 pp, 2002.

569 Cammalleri, C., Micale, F., Vogt, J.: On the value of combining different modelled soil moisture
570   products for European drought monitoring, J. Hydrol., 525, 547-558,
571   https://doi.org/10.1016/j.jhydrol.2015.04.021, 2015.

572 Cammalleri, C., Micale, F., Vogt, J.: A novel soil moisture-based drought severity index (DSI)
573   combining water deficit magnitude and frequency, Hydrol. Process., 30(2), 289-301,
574   https://doi.org/10.1002/hyp.10578, 2016.

575 Cammalleri, C., Vogt, J.V., Bisselink, B., de Roo, A.: Comparing soil moisture anomalies from
576   multiple independent sources over different regions across the globe, Hydrol. Earth Syst. Sci.,
577   21, 6329-6343, https://doi.org/10.5194/hess-21-6329-2017, 2017.

578 Cammalleri, C., Arias-Muñoz, C., Marinho Ferreira Barbosa, P., De Jager, A., Magni, D., Masante,
579   D., Mazzeschi, M., Mccormick, N., Naumann, G., Spinoni, J. and Vogt, J.: A revision of the
580   Combined Drought Indicator (CDI) as part of the European Drought Observatory (EDO), Nat.
581   Haz. Earth Syst. Sci., 21(2), 481-495, https://doi.org/10.5194/nhess-21-481-2021, 2021a.

582 Cammalleri, C., Spinoni J., Barbosa, P., Toreti, A., Vogt, J. V.: The effects of non-stationarity on
583   SPI for operational drought monitoring in Europe, Int. J. Climatol., 21, 1-13,
584   https://doi.org/10.1002/joc.7424, 2021b.

585 Carrão, H., Russo, S., Sepulcre-Canto, G., Barbosa, P.: An empirical standardized soil moisture
586   index for agricultural drought assessment from remotely sensed data, Int. J. Appl. Earth Obs.
587   Geoinf., 48, 74-84, https://doi.org/10.1016/j.jag.2015.06.011, 2016.

588 Chen, L., Guo, S.: Copulas and Its Application in Hydrology and Water Resources. Springer
589   Water, 290 pp, 2019.

590 Dash, S. S., Sahoo, B., Raghuwanshi, N. S.: A SWAT-Copula based approach for monitoring and
591   assessment of drought propagation in an irrigation command, Ecol. Eng., 127, 417-430,
592   https://doi.org/ 10.1016/j.ecoleng.2018.11.021, 2019.

593   De Michele, C., Salvadori, G.: A generalized Pareto intensity-duration model of storm rainfall
594   exploiting 2-copulas, J. Geophys. Res. Atmos., 108 (D2), 4067,
595   https://doi.org/10.1029/2002JD002534, 2003.

596   De Roo, A. P. J., Wesseling, C., van Deusen, W.: Physically based river basin modelling within a
597   GIS: The LISFLOOD model, Hydrol. Process., 14, 1981-1992, https://doi.org/10.1002/1099-
598   1085(20000815/30)14:11/12<1981::AID-HYP49>3.0.CO;2-F, 2000.

599   Dißman, J., Brechmann, E. C., Czado, C., Kurowicka, D.: Selecting and estimating regular vine
600   copulae and application to financial returns, Comput. Stat. Data Analysis, 59, 52-69,
601   https://doi.org/10.1016/j.csda.2012.08.010, 2013.

602   Dixit, S., Jayakumar, K. V.: Spatio-temporal analysis of copula-based probabilistic multivariate
603   drought index using CMIP6 model, Int. J. Climatol., 42(8), 4333-4350,
604   https://doi.org/10.1002/joc.7469, 2021.

605   Dutra, E., Viterbo, P., Miranda, P. M. A.: ERA-40 reanalysis hydrological applications in the
606   characterization of regional drought, Geophys. Res. Lett., 35(19), L19402,
607   https://doi.org/10.1029/2008GL035381, 2008.

608   Farahmand, A., AghaKouchak, A.: A generalized framework for deriving nonparametric
609   standardized drought indicators, Adv. Water Resour., 76, 140-145,
610   https://doi.org/10.1016/j.advwatres.2014.11.012, 2015.

611   Frahm, G., Junker, M., Schmidt, R.: Estimating the tail-dependence coefficient: properties and
612   pitfalls, Insur. Math. Econ., 37(1), 80-100, https://doi.org/10.1016/j.insmatheco.2005.05.008,
613   2005.

614   Gaona, J., Quintana-Seguí, P., Escorihuela, M. J., Boone, A., Llasat, M. C.: Interactions between
615   precipitation, evapotranspiration and soil-moisture-based indices to characterize drought with
616   high-resolution remote sensing and land-surface model data, Nat. Hazard Earth Syst. Sci., 22,
617   3461-3485, https://doi.org/10.5194/nhess-22-3461-2022, 2022.

618   Genest, C., Favre, A. C., Béliveau, J., Jacques, C. : Metaelliptical copulas and their use in
619   frequency analysis of multivariate hydrological data, Water Resour. Res., 43(9), 1-12,
620   https://doi.org/10.1029/2006WR005275, 2007.

621   Halwatura, D., McIntyre, N., Lechner, A. M., Arnold, S.: Capability of meteorological drought
622   indices for detecting soil moisture droughts, J. Hydrol. Reg. Studies, 12, 396-412,
623   https://doi.org/10.1016/j.ejrh.2017.06.001, 2017.

Hao, Z., AghaKouchak, A.: Multivariate Standardized Drought Index: A parametric multi-index model, Adv. Water Resour., 57, 12-18, https://doi.org/10.1016/j.advwatres.2013.03.009, 2013.

Hao, Z., Singh, V. P.: Drought characterization from a multivariate perspective: A review, J. Hydrol., 527, 668-678, https://doi.org/10.1016/j.jhydrol.2015.05.031, 2015.

Ji, L., Peters A. J.: Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices, Remote Sens. Environ., 87, 85-98, https://doi.org/10.1016/S0034-4257(03)00174-3, 2003.

Joe, H.: Dependence Modeling with Copulas, CRC Press, Taylor and Francis, 480 pp, 2015.

Kanthavel, P., Saxena, C. K., Singh, R. K.: Integrated drought index based on vine copula modelling, Int. J. Climatol., 42(16), 9510-9529, https://doi.org/10.1002/joc.7840, 2022.

Kao, S. C., Govindaraju, R. S.: A copula-based joint deficit index for droughts, J. Hydrol., 380, 121–134, https://doi.org/10.1016/j.jhydrol.2009.10.029, 2010.

Kwon, M., Kwon, H. -H., Han, D.: Spatio-temporal drought patterns of multiple drought indices based on precipitation and soil moisture: A case study in South Korea, Int. J. Climatol., 39(12), 4669-4687, https://doi.org/ 10.1002/joc.6094, 2019.

Laimighofer, J., Laaha, G.: How standard are standardized drought indices? Uncertainty components for the SPI & SPEI case, J. Hydrol., 613(A), 128385, https://doi.org/10.1016/j.jhydrol.2022.128385, 2022.

Manning, C., Widmann, M., Bevacqua, E., van Loon, A. F., Maraun, D., Vrac, M.: Soil moisture drought in Europe: A compound event of precipitation and potential evapotranspiration on multiple time scales. J. Hydrometeorol., 19(8), 1255-1271, https://doi.org/10.1175/JHM-D-18-0017.1, 2018.

Mishra, A. K., Singh, V. P.: A review of drought concepts, J. Hydrol., 391, 202-216, https://doi.org/10.1016/j.rse.2016.02.064, 2010.

Mo, K. C., Lettenmaier, D. P.: Objective drought classification using multiple land surface models, J. Hydrometeorol., 15, 990-1010, https://doi.org/10.1175/JHM-D-13-071.1, 2013.

Mo, K. C., Lyon, B.: Global meteorological drought prediction using the North American multi-model ensemble, J. Hydrometeorol., 16, 1409–1424, https://doi.org/10.1175/JHM-D-14-0192.1, 2015.

654 Mohammed, S., Alsafadi, K., Enaruvbe, G. O., Bashir, B., Elbeltagi, A., Széles, A., Alsalman, A.,
655   Harsanyi, E.: Assessing the impacts of agricultural drought (SPI/SPEI) on maize and wheat
656   yields across Hungary, Sci. Rep. 12, 8838, https://doi.org/ 10.1038/s41598-022-12799-w,
657   2022.

658 Nelsen, R. G.: An introduction to copulas, Springer Series in Statistics, Springer-Verlag, New
659   York, 272 pp, https://doi.org/10.1007/0-387-28678-0, 2006.

660 Panu, U. S., Sharma, T. C.: Challenges in drought research: Some perspectives and future
661   directions, Hydrol. Sci. J., 47, S19-S30, https://doi.org/10.1080/02626660209493019, 2002.

662 Pieper, P., Düsterhus, A., Baehr, J.: A universal Standardized Precipitation Index candidate
663   distribution function for observations and simulations, Hydrol. Earth Syst. Sci., 24, 4541-
664   4565, https://doi.org/10.5194/hess-24-4541-2020, 2020.

665 Poulin, A., Huard, D., Favre, A. C., Pugin, S.: Importance of tail dependence in bivariate frequency
666   analysis, J. Hydrol. Eng., 12(4), 394-403, https://doi.org/10.1061/(ASCE)1084-
667   0699(2007)12:4(394), 2007.

668 Quiring, S. M., Papakryiakou, T. N.: An evaluation of agricultural drought indices for the Canadian
669   prairies, Agr. Forest Meteorol., 118(1-2), 49-62, https://doi.org/10.1016/S0168-
670   1923(03)00072-8, 2003.

671 Ravelo, A. C., Decker, W. L.: The probability distribution of a soil moisture index, Agr. Meteorol.,
672   20(4), 301-312, https://doi.org/10.1016/0002-1571(79)90004-9, 1979.

673 Rembold, F., Meroni, M., Urbano, F., Csak, G., Kerdiles, H., Perez-Hoyos, A., Lemoine, G., Leo,
674   O., Negre, T.: ASAP: A new global early warning system to detect anomaly hot spots of
675   agricultural production for food security analysis, Agr. Syst. 168, 247–257,
676   https://doi.org/10.1016/j.agsy.2018.07.002, 2019.

677 Ribeiro, A., Pires, C.: Seasonal drought predictability in Portugal using statistical–dynamical
678   techniques, Phys. Chem. Earth, 94, 155–166, https://doi.org/ 10.1016/j.pce.2015.04.003,
679   2016.

680 Sadri, S., Pan, M., Wada, Y., Vergopolan, N., Sheffield, J., Famiglietti, J. S., Kerr, Y., Wood, E.
681   F.: A global near-real-time soil moisture index monitor for food security using integrated
682   SMOS and SMAP, Remote Sens. Environ., 246, 111864,
683   https://doi.org/10.1016/j.rse.2020.111864, 2020.

684 Salvadori G., De Michele C.: Frequency analysis via copulas: Theoretical aspects and applications
685   to hydrological events, Wat. Resour. Res., 40, W12511,
686   https://doi.org/10.1029/2004WR003133, 2004.

687 Salvadori, G., De Michele, C., Kottegoda, N. T., Rosso, R.: Extremes in Nature: An approach
688   using Copulas, Water Science and Technology Library Series, vol. 56, Springer, Dordrecht,
689   292 pp, ISBN: 978-1-4020-4415-1, 2007.

690 Schmidt, R., Stadtmueller, U.: Non-parametric estimation of tail dependence, Scandinav. J. Stat.,
691   33(2), 307-335, https://doi.org/10.1111/j.1467-9469.2005.00483.x, 2006.

692 Sehler, R., Li, J., Reager, J. T., Ye, H.: Investigating relationship between soil moisture and
693   precipitation globally using remote sensing observations, J. Cont. Water Res. Edu., 168(1),
694   106-118, https://doi.org/10.1111/j.1936-704X.2019.03324.x, 2019.

695 Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B.,
696   Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A
697   review, Earth-Sci. Rev., 99, 125-161, https://doi.org/10.1016/j.earscirev.2010.02.004, 2010.

698 Sepulcre-Cantó, G., Horion, S., Singleton, A., Carrão, H., Vogt, J.: Development of a combined
699   drought indicator to detect agricultural drought in Europe, Nat. Hazard Earth Syst. Sci., 12,
700   3519-3531, https://doi.org/10.5194/nhess-12-3519-2012, 2012.

701 Serinaldi, F.: Analysis of inter-gauge dependence by Kendall's τ_K, upper tail dependence
702   coefficient, and 2-copulas with application to rainfall fields, Stoch. Environ. Res. Risk
703   Assess., 22(6), 671–688, https://doi.org/10.1007/s00477-007-0176-4, 2008.

704 Serinaldi, F., Bárdossy, A., Kilsby, C. G.: Upper tail dependence in rainfall extremes: would we
705   know it if we saw it? Stoch. Environ. Res. Risk Assess., 29, 1211-1233, https://doi.org/
706   10.1007/s00477-014-0946-8, 2015.

707 Sheffield, J., Goteti, G., Wen, F., Wood, E. F.: A simulated soil moisture based drought analysis
708   for the United States, J. Geophys. Res., 109, D24108, https://doi.org/10.1029/2004JD005182,
709   2004.

710 Sheffield, J., Wood, E. F.: Characteristics of global and regional drought, 1950–2000: Analysis of
711   soil moisture data from off-line simulation of the terrestrial hydrologic cycle, J. Geophys.
712   Res., 112, D17115, https://doi.org/10.1029/2006JD008288, 2007.

713 Sims, A. P., Niyogi, D. S., Raman, S.: Adopting drought indices for estimating soil moisture: A
714   North Carolina case study, Geophys. Res. Lett., 29(8), 24-1-24-4,
715   https://doi.org/10.1029/2001GL013343, 2002.

716 Sivakumar, M. V. K., Motha, R. P., Wilhite, D. A., Wood, D. A.: Agricultural Drought Indices,
717   Proceedings of the WMO/UNISDR Expert Group Meeting on Agricultural Drought Indices,
718   2-4 June 2010, Murcia, Spain: Geneva, Switzerland: World Meteorological Organization,
719   AGM-11, WMO/TD No. 1572, WAOB-2011, 197 pp, 2011.

720

721 Soľáková, T., De Michele, C., Vezzoli, R.: Comparison between parametric and nonparametric
722   approaches for the calculation of two drought indices: SPI and SSI, J. Hydrol. Eng., 19(9),
723   04014010, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000942, 2014. Stagge, J. H.,
724   Tallaksen, L. M., Gudmundsson, L., van Loon, A. F., Stahl, K.: Candidate distributions for
725   climatological drought indices (SPI and SPEI), Int. J. Climatol., 35, 4027-4040,
726   https://doi.org/10.1002/joc.4267, 2015.

727 Stoica, P., Selen, Y.: Model-order selection: a review of information criterion rules, IEEE Signal
728   Proces. Mag., 21(4), 36-47, https://doi.org/10.1109/MSP.2004.1311138, 2004.

729 Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R.,
730   Palecki, M., Stooksbury, D.: The drought monitor, Bull. Am. Meteorol. Soc., 83, 1181-1190,
731   https://doi.org/10.1175/1520-0477-83.8.1181, 2002.

732 Thielen, J., Bartholmes, J., Ramos, M. -H., De Roo A. P. J.: The European flood alert system –
733   part 1: concept and development, Hydrol. Earth Syst. Sci., 13, 125-140,
734   https://doi.org/10.5194/hess-13-125-2009, 2009.

735 Thieming, V., Gomes, G. N., Skøien, J., Ziese, M., Rauthe-Schöch, A., Rustemeier, E., Rehfeldt,
736   K., Walawender, J. P., Kolbe, C., Pichon, D., Schweim, C., Salamon, P.: EMO-5: a high-
737   resolution multi-variable gridded meteorological dataset for Europe, Earth Syst. Sci. Data, 14,
738   3249-3272, https://doi.org/10.5194/essd-14-3249-2022, 2022.

739 Tian, L., Yuan, S., Quiring, S. M.: Evaluation of six indices for monitoring agricultural drought in
740   the south-central United States, Agr. For. Meteorol., 249, 107-119,
741   https://doi.org/10.1016/j.agrformet.2017.11.024, 2018.

van der Wiel, K., Batelaan, T. J., Wanders, N.: Large increases of multi-year droughts in north-western Europe in a warmer climate, Clim. Dynam., 60, 1781-1800, https://doi.org/10.1007/s00382-022-06373-3, 2022.

Vicente-Serrano S. M., Beguería, S., López-Moreno, J. I.: A Multi-scalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index – SPEI, J. Climate, 23, 1696-1718, https://doi.org/10.1175/2009JCLI2909.1, 2010.

Wang, H., Rogers, J. C., Munroe, D. K.: Commonly used drought indices as indicators of soil moisture in China, Hydrometeorol., 16(3), 1397-1408, https://doi.org/10.1175/JHM-D-14-0076.1, 2015.

Wilhite, D. A., Glantz, M. H.: Understanding the drought phenomenon: The role of definitions, Water International, 10(3), 111–120, 1985.

World Meteorological Organization (WMO): Standardized Precipitation Index User Guide (WMO n. 1090), Geneva, 24 pp, 2012.

World Meteorological Organization (WMO), Global Water Partnership (GWP): Handbook of Drought Indicators and Indices (M. Svoboda and B.A. Fuchs). Integrated Drought Management Programme (IDMP), Integrated Drought Management Tools and Guidelines Series 2, Geneva, 52 pp, 2016.

Xia, Y., Ek, M. B., Peters-Lidard, C. D., Mocko, D., Svoboda, M., Sheffield, J., Wood, E. F.: Application of USDM statistics in NLDAS-2: optimal blended NLDAS drought index over the continental United States, J. Geophys. Res. Atmos., 119(6), 2947-2965, https://doi.org/10.1002/2013JD020994, 2014.

Yuan, X., Wood, E. F.: Multimodel seasonal forecasting of global drought onset, Geophys. Res. Lett., 40, 4900–4905, https://doi.org/10.1002/grl.50949, 2013.

Zargar, A., Sadiq, R., Naser, B., Khan, F. I.: A review of drought indices, Environ. Rev., 19, 333-349, https://doi.org/ 10.1139/A11-013, 2011.