

RESEARCH ARTICLE

The effects of non-stationarity on SPI for operational drought monitoring in Europe

Carmelo Cammalleri  | Jonathan Spinoni  | Paulo Barbosa  |
Andrea Toreti  | Jürgen V. Vogt 

European Commission, Joint Research
Centre (JRC), Ispra, Italy

Correspondence

Carmelo Cammalleri, European
Commission, Joint Research Centre
(JRC), Via E. Fermi 2749, 21027 Ispra
(VA), Italy.
Email: carmelo.cammalleri@ec.europa.eu

Abstract

It is a good practice to follow common guidelines in the computation of Standardized Precipitation Index (SPI) data sets as part of operational drought monitoring systems. In the European Drought Observatory (<https://edo.jrc.ec.europa.eu/>), reference statistics are computed following the World Meteorological Organization Guidelines on the Calculation of Climate Normals, where a definition of the reference period for monitoring applications is introduced as the most recent 30-year period finishing in a year ending with 0. In this study, the temporal consistency of the SPI time series computed using the fifth-generation European Centre for Medium-range Weather Forecast reanalyses model precipitation data set is tested over Europe to quantify the effect of the transition from one baseline period (1981–2010) to another (1991–2020) and to evaluate the capability of these static baselines to reproduce the behaviour of non-stationary SPIs (nSPIs). The results of the comparison suggest that the threshold commonly used to identify droughts ($SPI = -1$) is only marginally affected by the change in reference period (mean absolute deviation, $MAD = 0.15 \pm 0.1$) for short-term SPI, whereas larger differences (MAD up to 0.6) can be observed over certain areas (i.e., Southern Italy and Eastern Europe) for longer accumulation periods (i.e., SPI-9 or SPI-12). Examples show that changes in drought classification from extremely dry ($SPI < -2$) to moderately dry ($SPI < -1$) are not uncommon, which may lead to misinterpretation by users. Finally, analyses against nSPI highlight an overall good correspondence between stationary and nSPIs, even if both static baselines displayed difficulties in reliably capturing the magnitude of nSPI for the entire 10-year period for which they should be used. In this regard, it has been demonstrated that more spatially uniform results can be achieved with 5-year updates, with a good matching ($MAD < 0.25$) for SPI-1 and an acceptable matching ($MAD < 0.50$) for SPI-12 over more than 80% of Europe.

KEYWORDS

EDO, ERA5, meteorological drought, Standardized Precipitation Index

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *International Journal of Climatology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

The Standardized Precipitation Index (SPI; McKee *et al.*, 1993) is among the most recognized drought indicators. SPI uses precipitation records accumulated over different time scales, usually from 1 to 48 months, to develop a probabilistic assessment of extreme conditions relative to climatology. Thanks to its flexibility, it has been applied in a large variety of conditions, from agriculture to water resources management and other sectors (WMO and GWP, 2016).

Following the outcomes of the Inter-Regional Workshop on Indices and Early Warning Systems for Drought, held in 2009 at the University of Nebraska–Lincoln, the SPI has been endorsed as a standard meteorological drought indicator to be included as one of the tools in all national meteorological and hydrological services (Hayes *et al.*, 2011). To facilitate a consistent implementation of SPI as part of operational systems, The World Meteorological Organization (WMO) has issued guidelines for SPI computation (WMO, 2012), including the preferred probability distribution and minimum length of the time series.

As for all standardized indicators, the length, consistency, and homogeneity of the data set used as reference (i.e., baseline) are of utmost importance (Wu *et al.*, 2005; Mahmoudi *et al.*, 2019). From a purely statistical standpoint, Guttman (1994) suggests that a sample size of at least 20–30 monthly precipitation values is required, with an optimal range being closer to 50–60 values (Guttman, 1994). The preference towards longer reference time series may be in contrast with the hypothesis that samples are independent and identically distributed, a hypothesis that often does not hold true in the case of long time series under non-stationary conditions. Following these considerations, 30 years of continuous monthly precipitation data are suggested in the WMO SPI guidelines (WMO, 2012), and this value seems to be the norm for operational drought monitoring systems, as in the case of the European Drought Observatory (EDO, <https://edo.jrc.ec.europa.eu/>).

In the framework of a non-stationary climate, a mutual inconsistency between the two main purposes of a reference period arises, namely, (1) to provide a stable benchmark for comparison against recent observations and (2) to represent a reliable approximation of the most likely conditions expected in the present or near future. This dichotomy is well-expressed in the WMO Guidelines on the Calculation of Climate Normals (WMO, 2017), where it is stated how these two functions can be simultaneously achieved with a single common reference period only in a stable climate. For variables with consistent trends, the predictive skill of climate normals is greatest if they are updated as frequently as possible, even if some benefits come with a fixed reference benchmark (i.e., temporal consistency and ease of communication).

Accounting for the non-stationarity in the precipitation time series can be seen as the obvious solution to this set of issues, but even if non-stationary analyses of SPI have been the focus of several research studies (e.g., Russo *et al.*, 2013; Li *et al.*, 2015; Wang *et al.*, 2015; Rashid and Beecham, 2019), such methodological approaches are often too complex to be suggested as standard common practice for operational weather and hydrological systems.

As a compromise operational solution, WMO introduced the definition of a climatological standard normal that refers to the most recent 30-year period finishing in a year ending with 0 (1981–2010 at the time of publication of the guidelines, now 1991–2020), which is the recommended reference for monitoring products.

While this solution provides a universal easy-to-implement framework for all weather and hydrological services, ensuring inter-comparability among different data sets, some questions need to be asked on both the reliability and the temporal consistency of the SPI estimates provided under this framework by monitoring systems operating for long time spans. As an example, depending on the magnitude of the trends observed in precipitation, a diminishing adherence to the present condition can be observed for a reference period that is too far in the past, causing biased estimates of SPI and abrupt discontinuities at the time of the change of the standard reference period (i.e., from 1981–2010 in 2020 to 1991–2020 in 2021).

The answers to those questions are specific to the area under investigation and to the precipitation data set considered to be operationally implemented as part of a near-real-time drought monitoring system. The European Centre for Medium-range Weather Forecast (ECMWF) has recently introduced a near-real-time product for its newest reanalysis model, ERA5 (successor of ERA-Interim), available on a daily basis with a 5-day delay (namely ERA5T; Hersbach *et al.*, 2020).

This new data set, even if preliminary (implying the possibility of a future revision in case of later discovered issues), represents a unique source of near-real-time moderate resolution precipitation data, which opens up new possibilities for the operational monitoring of meteorological drought conditions over many regions, including Europe.

The goal of this study is twofold: First, we want to evaluate how smooth the transition is between the two most recent reference periods according to the WMO guidelines (i.e., 1981–2010 and 1991–2020). The evaluation is based on SPI estimates derived over Europe using the ERA5 precipitation data set, as a newly available near-real-time product at 0.25° resolution. Second, the capability of static baselines to reproduce the behaviour of a non-stationary SPI (nSPI) data set is tested, in order to assess the reliability of the assumption that a 10-year update is a good compromise for operational drought

monitoring systems. The results of this study will steer the implementation of an ERA5-based SPI set of products within the EDO system.

2 | MATERIALS AND METHODS

2.1 | ERA5 precipitation data set

ERA5 is the fifth and most recent generation of ECMWF global atmospheric reanalyses (<https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5>), providing a numerical description of the climate by combining a vast number of historical observations (both ground and satellite, around 24 million per day in 2018; Hersbach *et al.*, 2019) and numeric modelling via advanced data assimilation systems (Hersbach *et al.*, 2020).

ERA5 provides hourly estimates of a large number of atmospheric, land, and oceanic climate variables, covering the earth on a 30-km grid and resolves the atmosphere using 137 levels from the surface up to 0.01-hPa (around 80-km). This reanalysis data set covers a period from 1979 onwards (a preliminary version of the data set going back to 1950 is also available), with quality-assured monthly updates published within 3 months of real-time and preliminary daily updates available within 5 days of delay. ERA5 data are distributed through the European Union-funded Copernicus Climate Change Service (<https://climate.copernicus.eu/>).

In this study, hourly total precipitation data at surface level were obtained for the period 1979–2020 over Europe, including only consolidated estimates (i.e., no ERA5T data were used). In order to align this data set with the others available in EDO, precipitation data were cumulated at dekad time scale (roughly 10 days, 3 dekads per month) by simply summing up all the hourly values in a dekad (no missing data were detected in the analysed period), and the dekad precipitation maps were resampled to a regular lat/lon grid of 0.25°.

2.2 | Stationary SPI and nSPI

The SPI is estimated by fitting a set of multi-year precipitation data with the same accumulation period to a probabilistic distribution. The fitted probabilities are then used to transform the precipitation into standard normal values. Commonly used accumulation periods range from monthly to multi-year time scales (e.g., 48 months). Different time scales are used to deal with different drought types, with short-term SPI (i.e., 1–3 months) aiming at modelling meteorological and agricultural droughts, whereas long-term SPI (i.e., 12–48) are mostly used to detect hydrological and groundwater droughts (WMO, 2012).

Guttman (1999) suggests that time scales between 1 and 24 months are recommended when around 50–60 years of data are available since the sample size will be too small and the statistical confidence of the probability estimates on the extremes will be too weak beyond 24 months. Following these considerations, by using a reference period of 30 years, only SPI computed on accumulation periods up to 12 months is analysed here (i.e., SPI-1, 3, 6, 9, 12).

The two-parameter gamma distribution is a commonly chosen parametric family, as it has demonstrated to perform well under a range of conditions, including many accumulation periods and regions across Europe (Stagge *et al.*, 2015). By parameterizing the gamma distribution as a function of a shape, s , and a scale, a , parameter, the distribution function of cumulated precipitation (P) can be expressed as

$$P \sim \text{gamma}(s, a), \quad (1)$$

where the gamma distribution parameters are fitted using the data collected over the reference period using the generalized additive model in location, scale and shape modelling framework (Stasinopoulos and Rigby, 2007). This approach is here used to estimate stationary SPI for two different 30-year baseline periods, 1981–2010 and 1991–2020, defined according to the WMO guidelines (WMO, 2017).

Analogously, the nSPI is obtained by fitting a non-stationary gamma distribution:

$$P \sim \text{gamma}(s_t, a), \quad (2)$$

where the subscript t highlights the dependence of the scale parameter on the year. In this study, a simple linear dependence has been adopted ($s_t = b_1 + b_2 t$). The parameters of the non-stationary gamma distribution are fitted on the entire ERA5 consolidated 40-year time series (1981–2020).

2.3 | Inter-comparison strategy

Even if not strictly necessary for the non-stationary fitting, a preliminary linear trend analysis is performed for each monthly P data set of 40 years (1981–2020), as well as for year total P , with the Theil–Sen estimator (Sen, 1968). This nonparametric approach identifies as robust slope the median value of the slopes between all lines through pairs of points in the sample (780 pairs in this case). The statistical significance of the trend is evaluated through the Mann–Kendall test (two-tailed $p < .05$).

The two main goals of the study are comparing the SPI estimates based on the two statistic baseline periods

and then evaluating these two data sets against the outcomes of the non-stationary analysis. Since in drought applications SPI values are usually classified following the approach introduced by McKee *et al.* (1995), which defines three dry classes (namely moderately, severely, and extremely dry) delimited by SPI values equal to -1 , -1.5 , and -2 , the main analyses performed in this study are focused on these three threshold values for all five SPI accumulation periods (SPI-1, 3, 6, 9, and 12) and 12 monthly fittings.

The effect of changing from one stationary baseline to another (i.e., from 1981–2010 to 1991–2020) is quantified in two ways. First, precipitation corresponding to a threshold value (e.g., SPI-1 = -1) is computed for the first baseline (1981–2010) by inverting the gamma distribution, and then, the SPI value corresponding to this precipitation is computed according to the second baseline (1991–2020). The difference between this second SPI value and the starting threshold (-1 in the example) quantifies the impact of the change in the baseline at that dryness level, which is assessed by two statistical metrics: the mean absolute deviation (MAD) and the mean bias deviation (MBD) derived from the 12 monthly values for each cell separately. Second, the number of changes in drought classes observed between the two SPI data sets in the most recent 5-year period (2016–2020) is computed. This second analysis details the operational impacts of the change in a classified time series of SPI maps, with a specific focus on the years preceding the change of baseline (i.e., the years where most of the differences should be observed). The goal here is to account for the fact that some changes in SPI may not be reflected in a change in drought class (i.e., an SPI-1 of -1.2 with the baseline 1981–2010 and a value of -1.4 in the baseline 1991–2020 will be both classified as moderately dry).

The nSPI fittings are then used to assess the divergence of the two static baselines from the ideal non-stationary behaviour in the 10 years preceding the introduction of a new baseline (i.e., in 2011–2020 for the baseline 1981–2010 and 2021–2030 for the baseline 1991–2020). In this case, the evaluation procedure is based on the temporal evolution of the statistical metrics MAD and MBD, in order to take into account the dependency of the non-stationary gamma parameterization on time.

3 | RESULTS AND DISCUSSION

The trends observed in the ERA5 P data set is at the base of any discrepancy observed among SPI maps derived using different reference periods. As a preliminary analysis, the linear trend for the entire 1981–2020 period has been analysed for each month (i.e., 12 independent analyses

using 40 monthly total values) as well as for the year total P . The maps in Figure 1 report the slope of the linear trend for the year total, according to the Theil–Sen approach (Figure 1a) as well as the percentage of months with statistically significant ($p < 0.05$) increasing (Figure 1b) or decreasing (Figure 1c) trends.

Overall, some clear spatial features can be observed in Figure 1a, including a statistically significant increasing trend in most of the Mediterranean area (i.e., Italy, Greece, and former Yugoslavia) and decreasing trends in Central and Eastern Europe. Similar patterns are present in the two sub-panels, where it is possible to observe how even the areas with the strongest signal only have a relatively small fraction of months (up to 25%) with statistically significant trends ($p < .05$). This latest result should come as no surprise, given the limited length of the time series under consideration.

Trends in reanalysis data sets may be ascribed to a multitude of sources, not only to climatic changes (Bengtsson *et al.*, 2004). As an example, even if the ERA5 data assimilation system is consistent throughout the full time series, changes in both accuracy and availability of assimilated data are present (the number of assimilated observations has increased from an average of 0.75 million per day in 1979 to around 24 million per day in 2018; Hersbach *et al.*, 2019). Independently from the origin of such trends, it is nevertheless relevant to evaluate how they impact SPI estimates based on different baseline assumptions.

The patterns observed in Figure 1 are in line with the ones reported by ECMWF for ERA-Interim for the differences between the last and the first decades in the data set (2007–2016 vs. 1979–1988) (<https://climate.copernicus.eu/monthly-summaries-precipitation-relative-humidity-and-soil-moisture>). Some similarities are also observed with the results of other studies on recent climatic trends in SPI over Europe (i.e., Spinoni *et al.*, 2017; Stagge *et al.*, 2017), even if the length of the time series used here is not suited for the analysis of climatic changes in drought occurrence. An additional inter-comparison of the ERA5 outcomes against the E-OBS data set (<https://www.ecad.eu/download/ensembles/download.php>) is reported in Appendix A.

3.1 | Comparison between 1981–2010 and 1991–2020 stationary baselines

Focusing on the modelling of drought conditions, the impact of fitting the gamma distribution over two different baseline periods is tested in terms of differences in SPI dry extremes. The maps in Figure 2 report the MBD (a, c) and MAD (b, d) statistics for a threshold value of -1 in SPI-1 and SPI-12.

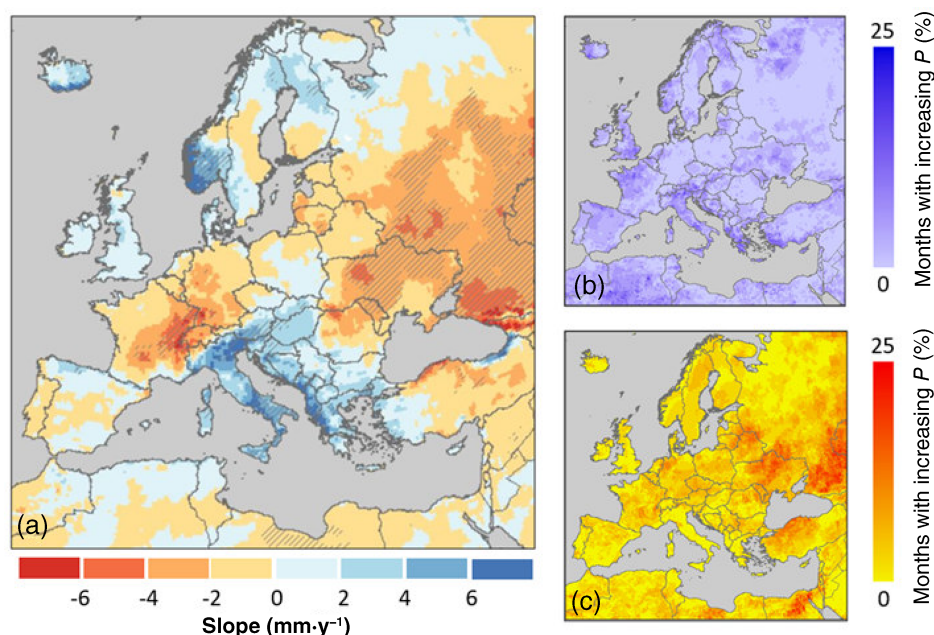


FIGURE 1 Trend analysis over Europe for the period 1981–2020. The main panel (a) shows the linear trend slope (Theil–Sen approach) in year total precipitation, with hatching over regions with a statistically significant trend (two-tailed Mann–Kendall test, $p < .05$); the two sub-panels on the right report the percentage of months with statistically significant increasing (b) or decreasing (c) trends ($p < .05$) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7424)]

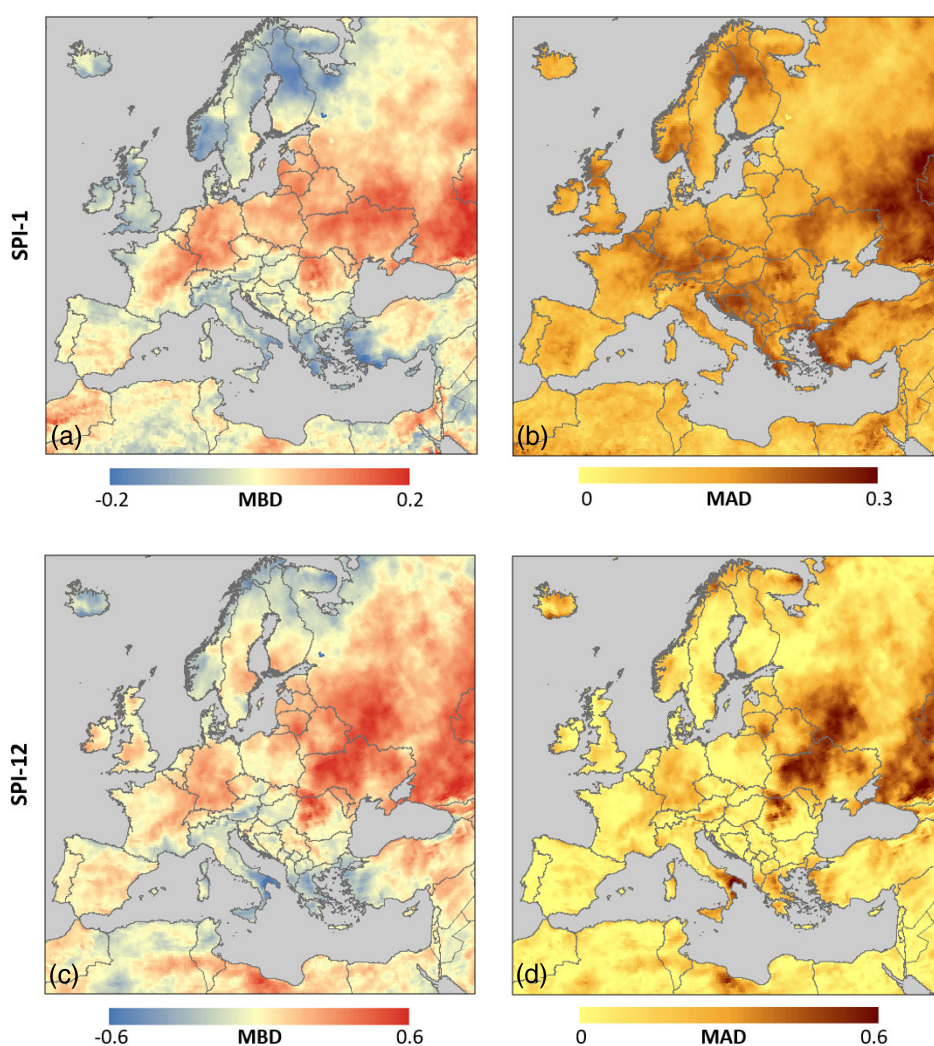


FIGURE 2 Spatial distribution of MBD (a, c) and MAD (b, d) computed on the differences between the baselines 1981–2010 and 1991–2020 for SPI-1 = -1 (upper panels) and SPI-12 = -1 (lower panels) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7424)]

The data of MBD show an overall good resemblance to the slope values in Figure 1a, with positive biases (i.e., less dry with the most recent baseline) in Central and Eastern Europe and negative values (i.e., drier with the most recent baseline) in Northern Europe and the regions surrounding the Adriatic Sea. The MAD maps for SPI-1 (Figure 2b) and SPI-12 (Figure 2d) display some similarities, even if more polarized values can be observed for SPI-12, with several areas with very low differences (e.g., Spain) and others with very large values (Ukraine and Southern Italy). In general, the bias values seem to be larger (in absolute value) for SPI-12 than SPI-1, whereas the MAD has a larger range of variability in SPI-12 compared to SPI-1. This result is better exemplified by the frequency plots in Figure 3 (which report also the other intermediate accumulation periods).

The frequency plot for MBD (Figure 3a) shows how the bias values are mostly zero-centred for all SPIs, even if a slight tendency towards positive values can be noticed with increasing accumulation. The spread of the bias data increases with the accumulation period, with MBD up to ± 0.6 for SPI-12. The frequency of MAD values (Figure 3b) shows that for larger accumulation periods, there is a simultaneous shift of the peak value towards zero and a lengthening of the right tail of the distribution.

This opposite behaviour in MAD values can be mostly explained by the combined effects of the variable trends observed in different months on the aggregated precipitation. Similar to the differences observed between trends in monthly, seasonal, and annual precipitation in Caloiero *et al.* (2018), there is a compensation effect that results in smaller differences at large aggregation scale compared to short accumulation periods in the areas with opposite trends in different months (e.g., Western France, where dekads with both increasing and

decreasing P are reported in Figure 1b,c). On the opposite, in the areas with consistent trends (e.g., Southern Italy, where only months with increasing P are reported in Figure 1b), the effects are amplified at larger accumulation periods, which result in the high MAD values observed in the tail of the distribution for SPI-9 and SPI-12.

The results for the other SPI threshold values (-1.5 and -2) are rather similar to the ones discussed for -1 , as summarized by the data in Table 1. This table shows a strong linear relationship between the MBD and MAD values discussed for SPI = -1 and the results obtained for the other two thresholds, as expected due to the relationship between the extreme values modelled according to the same gamma distribution.

The slopes of the linear relationship (b , forced through the origin due to negligible intercepts) are always greater than 1, and they increase with the increase (in absolute value) of the SPI threshold. As a result, the differences shown for SPI = -1 are the smallest that one can observe for cells considered under drought conditions (i.e., SPI < -1). In addition, the high values of the Pearson correlation coefficient ($r > 0.95$ in all cases) confirm that the results discussed for SPI = -1 can be easily extrapolated to the other two thresholds. Following these two considerations, the successive analyses against the nSPI will mostly focus on SPI = -1 only.

While the analysis of MAD and MBD for specific threshold values gives an overall assessment of the similarity between the SPI obtained with the two baselines, the effect of moving from one baseline to another in an operational drought monitoring system can be also quantified by the number of changes in drought classes in the SPI data in the years preceding the change of the baseline (i.e., between 2016 and 2020 in this study). These data are normalized as the average number of changes per year,

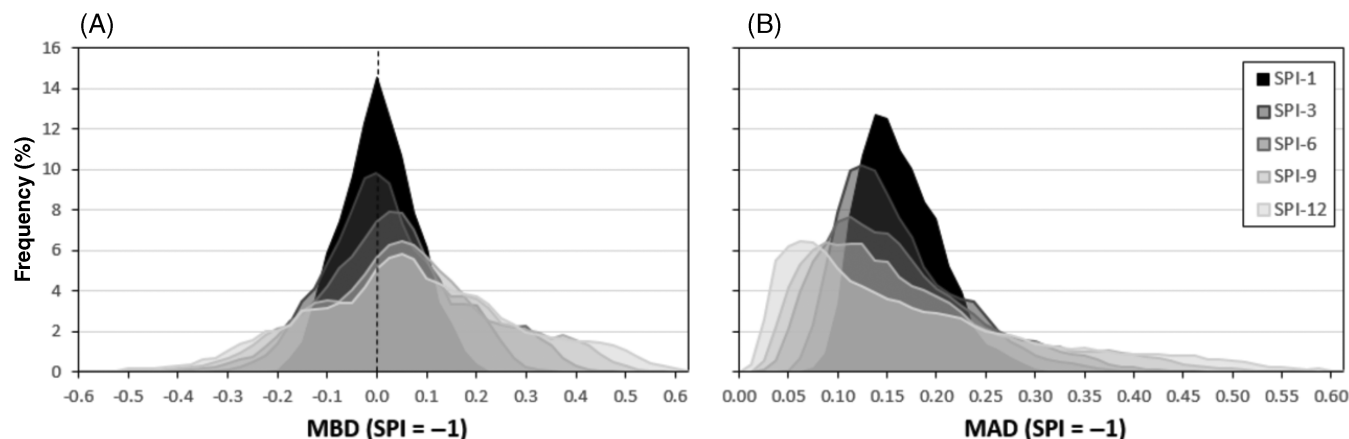


FIGURE 3 Frequency distribution of MBD (a) and MAD (b) computed between the baselines 1981–2010 and 1991–2020 for an SPI value of -1 at different accumulation periods

TABLE 1 Summary of the relationship between MBD and MAD values for SPI = −1 and the ones for SPI = −1.5 and −2

Variable		vs. −1.5		vs. −2	
		<i>b</i>	<i>r</i>	<i>b</i>	<i>r</i>
SPI-1	MBD	1.10	0.97	1.19	0.93
	MAD	1.19	0.99	1.41	0.99
SPI-3	MBD	1.08	0.98	1.17	0.96
	MAD	1.16	0.99	1.35	0.99
SPI-6	MBD	1.08	0.97	1.16	0.95
	MAD	1.14	0.99	1.30	0.99
SPI-9	MBD	1.08	0.97	1.17	0.95
	MAD	1.12	0.99	1.25	0.98
SPI-12	MBD	1.09	0.97	1.18	0.95
	MAD	1.10	0.98	1.22	0.97

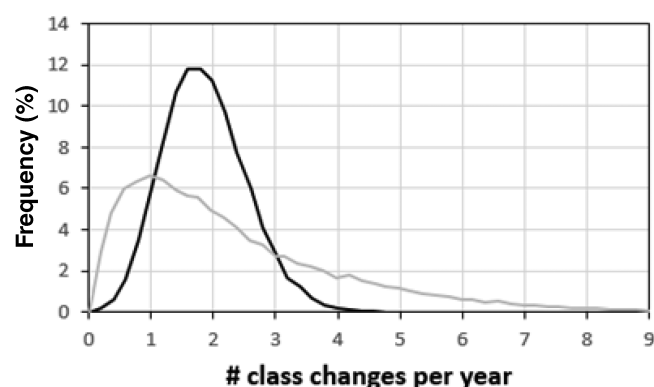


FIGURE 4 Frequency distribution of the average number of changes per year in drought classes in the period 2016–2020 for SPI-1 (black line) and SPI-12 (grey line)

as summarized by the frequency distributions in Figure 4. Note that a certain number of changes (e.g., 3) may occur in a single month (e.g., an extremely dry condition become near-normal) or spread through different months (e.g., extremely dry to severely dry in 3 different months). However, single-class changes constitute the majority of the cases, and changes are never more than 3 in the same month.

The plot in Figure 4 (focused on SPI-1 and SPI-12 only) shows a strong similarity between these data and the frequency distribution of MAD (see Figure 3b), with the frequencies for SPI-1 more bell-shaped compared to the SPI-12. Overall, all the cells in the study areas have 2 ± 2 changes per year in SPI-1, with no strong spatial patterns (not shown). Differently, the results for SPI-12 show a large fraction of the domain with less than 1 change per year, but also a significant fraction with more than 5 changes per year.

The spatial distribution of these changes for SPI-12 is depicted in Figure 5, showing both the overall total changes (a) as well as the results separated between changes to wetter (b) or dryer (c) classes. The main map highlights some clear hotspots in terms of changes related to the shift from one baseline to the other, most notably Northern Scandinavia, Eastern Ukraine, Southern Russia, and the Carpathians, with changes up to 9 per year (i.e., on average, 75% of the monthly maps have one change in drought class).

The two maps on the sub-panels (Figure 5b,c), which split the changes between wetter or dryer conditions, well resemble the spatial patterns observed on the statistically significant monthly trends (Figure 1b,c), with changes to wetter classes mostly located in central and Eastern Europe and changes to dryer classes predominant in the Mediterranean and Northern Europe. The maximum number of changes observed in the two sub-plots are roughly the same as the ones observed in the main map for total changes, highlighting how the hotspots in Figure 5a are mostly constituted by homogeneous (either to wetter or to dryer) changes, with no areas with a mixture of a large number of changes in both directions.

Overall, the results reported in Figures 4 and 5 show somewhat limited effects on short-term SPI (i.e., SPI-1) but potentially major discrepancies for large accumulation periods (i.e., SPI-12) over some specific regions. To better illustrate the magnitude of such changes in an operational drought monitoring system, an example for SPI-12 (as the accumulation period mostly affected) is reported in Figure 6 for the month of July 2020 (one of the months with the largest total number of changes). This figure shows the SPI-12 computed using the two baselines 1981–2010 (Figure 6a) and 1991–2020 (Figure 6b) as well as the map reporting the number of class changes between the two SPIs (Figure 6c) and the nSPI for comparison (Figure 6d).

These maps highlight how even if the general patterns are not influenced by the change in the baseline, some differences can be observed over Northern Italy and Finland (less wet according to the 1991–2020 baseline) and Ukraine and Southern Russia (less dry in the map based on 1991–2020 baseline). Changes up to two classes (from extremely dry/wet to moderately dry/wet) can be observed over Northeast Italy and the Polesia region. The magnitude of these differences may be small compared to the uncertainty related to the choice of the input precipitation data set (Kurnik *et al.*, 2011; Golian *et al.*, 2019; Hoffmann *et al.*, 2020), but they still can be easily misinterpreted as variation in drought severity by users who assume homogeneity in the reference period in a continuous monitoring data set. Finally, a visual comparison of

FIGURE 5 Spatial distribution of the total number of changes per year in drought classes for SPI-12 (a) as well as values separated between changes to wetter (b) or dryer (c) classes [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7424)]

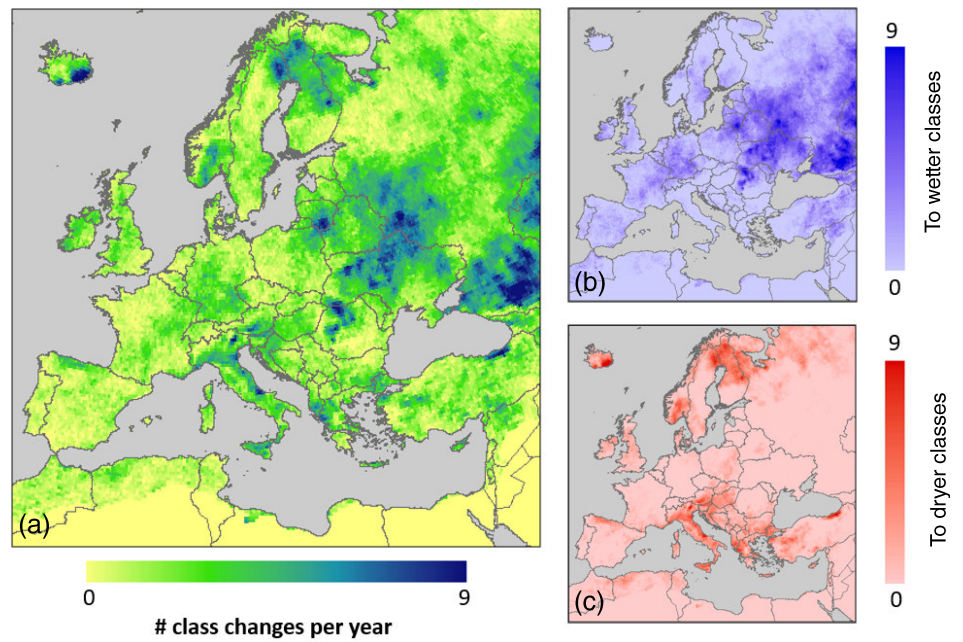
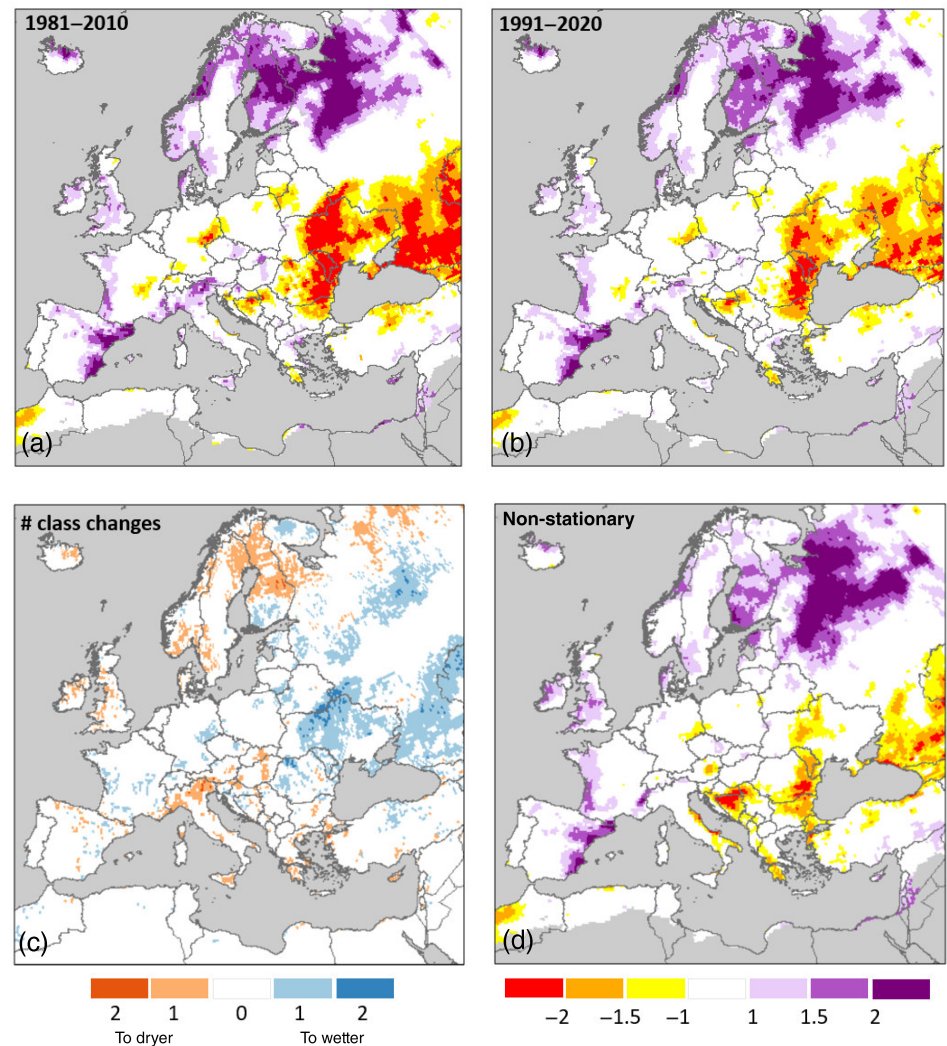


FIGURE 6 Examples of SPI-12 map for July 2020 computed using the two baselines 1981–2010 (a) and 1991–2020 (b) as well as the number of class changes between the two SPIs (c) and the non-stationary SPI (d) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7424)]



the two maps with the one in Figure 6d highlighting how the most recent baseline better resemble the expected behaviour according to the nSPI, as discussed in the next section.

3.2 | Comparison against nSPI

In this section, results on the comparison between the two static baselines against the non-stationary fittings are reported, assuming the nSPI as the reference. Following the outcomes of the previous inter-comparison, the differences between the stationary and non-stationary baselines are evaluated at different time horizons after the end of the stationary baseline period (i.e., 2011 is +1 for the baseline 1981–2010). The plots in Figure 7 focus on

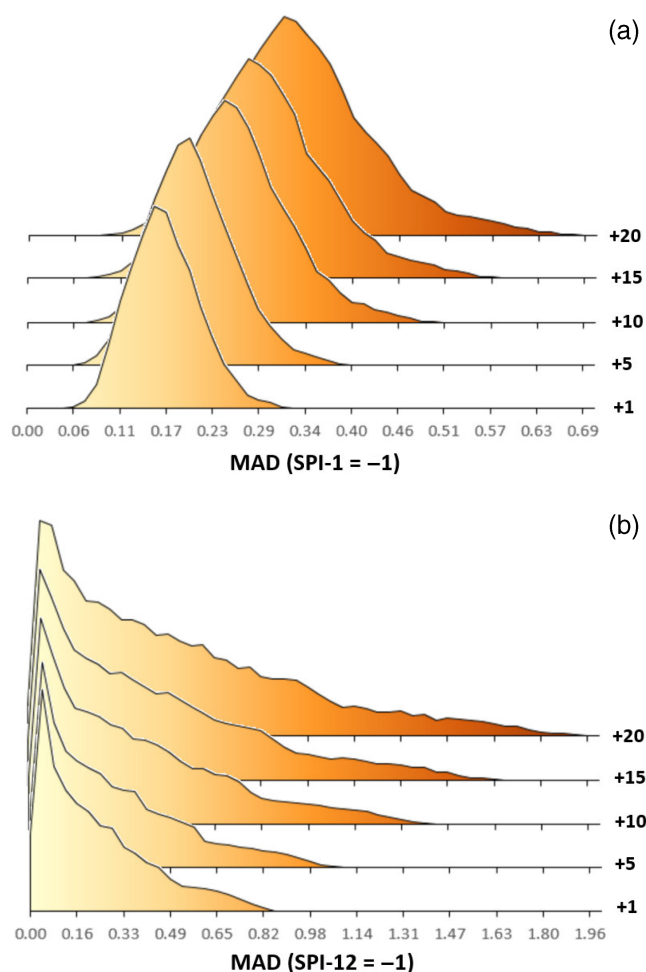


FIGURE 7 Frequency distribution of MAD values computed between the baselines 1981–2010 and the non-stationary fittings for SPI-1 (a) and SPI-12 (b) equal to -1 . The different curves represent different time horizons (i.e., +1 is 2011, 1 year after the end of the reference period for the baseline 1981–2010) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

the MAD values for $\text{SPI} = -1$ obtained by comparing the 1981–2010 baseline with the non-stationary fitting for both SPI-1 (Figure 7a) and SPI-12 (Figure 7b). The results for the baseline 1991–2020 are analogous to the ones reported in Figure 7 (not shown); hence, the results can be discussed without losing generality.

Similar to what was observed in the plots in Figure 3b, the frequency plots assume a roughly bell-shaped distribution for SPI-1 and long right tails for SPI-12. For SPI-1, the MAD values have a general increase the farther we move from the baseline period, with both peak and range increasing with time ($\text{MAD} = 0.15 \pm 0.12$ at +1, up to 0.33 ± 0.30 at +20). In contrast, for SPI-12, the low MAD values (peak) do not show any time dependency, while the tail is much longer for time horizons farther in time (MAD up to 2 for the time horizon +20). This behaviour is explained again by the fact that MAD values are small and time independent where the linear trend is negligible, but higher and increasing with time in the areas with strong trends.

If we compare the data in Figure 7 with the ones in Figure 3b, we can observe how the magnitude of the MAD values is comparable only for the time horizons up to +5. This means that as soon as we move more than 5 years away from the end of the reference period, the differences between the stationary and non-stationary baselines become larger than the differences between two consecutive stationary baselines (10 years apart).

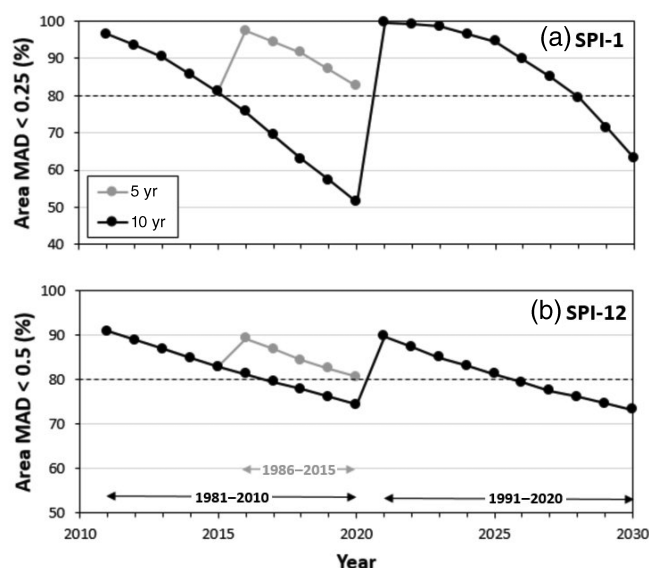


FIGURE 8 Percentage of domain with MAD under 0.25 for SPI-1 (a) and 0.50 for SPI-12 (b) for an SPI value of -1 . The black lines represent the results for the standard 10-year updates, whereas the grey lines represent an additional 5-year update

Following this observation, a final analysis on the capability of the static baselines to reliably reproduce the behaviour of the nSPI for up to 10 years after the end of the reference period is performed. This analysis focuses on both the 1981–2010 and 1991–2020 baselines as well as an additional intermediate update after 5 years (1986–2015). The plots in Figure 8 summarize the temporal dynamic of the observed differences by reporting the fraction of the domain with MAD values lower than the threshold values 0.25 for SPI-1 = −1 (Figure 8a) and 0.5 for SPI-12 = −1 (Figure 8b). Two different values are used, given the large difference in the magnitude of MAD values in SPI-12 and SPI-1 (i.e., all the domains have MAD < 0.5 for SPI-1 and only a small fraction of the domain have MAD < 0.25 for SPI-12).

These results clearly show the effect of updating the baseline, with maximum percentages right after the adoption of a new baseline (values greater than 90% in 2011 and 2021) and decreasing values reaching a minimum right before the new update (minimum in 2020 and 2030 for both SPI-1 and SPI-12). The steeper reduction in SPI-1 compared to SPI-12 is likely due to the fact that very large differences in SPI-12 are limited only to a smaller portion of the domain (i.e., only the tail of the distribution), whereas the difference in SPI-1 increases with time for the entire domain.

For both SPI-1 and SPI-12, the percentage falls under 80% around 5 years after a new baseline is adopted, suggesting that a possible improvement can be achieved by increasing the frequency in which the baseline is updated. The grey line in Figure 8 highlights the benefits of this change, with percentages that stay above 80% during the entire period 2011–2020. It should be noted that such a test cannot be performed in the second decade (2021–2030) due to the obvious unavailability of data in the period 2021–2025. Overall, by increasing the frequency in which the baseline is updated, the magnitude of the abrupt jump is significantly reduced, especially for SPI-1 (from 50% for a 10-year update to 20% for a 5-year update for the change in 2020/2021).

4 | SUMMARY AND CONCLUSIONS

In operational drought monitoring, the introduction of a standard reference period is important to allow inter-comparability between different data sets, but to keep-up with the non-stationarity of the climate, reference periods are also regularly revisited (i.e., every 10 years according to WMO standards for normals).

While the WMO approach seems a sounding compromise between a more sophisticated non-stationary analysis and a rigid fixed reference period, it may introduce

artefacts in the continuity of SPI time series in case of large trends. The inter-comparison performed in this study between two ERA5-based SPI data sets, one computed on the 1981–2010 reference period and one on the most recent 1991–2020 baseline, suggests that for the 1-month SPI (i.e., SPI-1), the usual threshold adopted to identify a drought event (SPI = −1) is only marginally affected by the change in reference period (MAD = 0.15 ± 0.1). Generally, short-term SPI seems to be less affected by the non-stationarity, mainly due to the limited magnitude of the trends observed in the last 40 years over most regions.

At longer accumulation periods (i.e., SPI-9 or SPI-12), the differences start to become more noticeable, at least over some specific regions of the domain (i.e., Southern Italy and Eastern Europe). Over these areas, the combined effect of several months with coherent (either increasing or decreasing) statistically significant trends has an amplifying effect on the differences observed for long-term SPIs. When this occurs, noticeable discontinuities in the SPI time series can be observed when moving from one reference period to the other, with up to 75% of the monthly SPI values affected by at least 1 change in drought classification. Examples showed that changes in drought classification from extremely dry (SPI < −2) to moderately dry (SPI < −1) are not uncommon and may lead to misinterpretation by users of the monitoring system.

Overall, the magnitude of the differences observed between the two stationary baselines, as well as against the non-stationary reference, can be considered limited in many cases. Indeed, similarities between stationary and nSPI time series have been observed in other studies (i.e., Li *et al.*, 2015), which is an outcome that is generally confirmed by our analysis. However, local differences may be quite large depending on the observed trend. It is interesting to highlight that the differences between non-stationary and static baselines are generally larger than the differences observed between the two stationary references. This result can be partially explained by the 20-year overlap (out of 30 years) between the two samples used in the static fittings, whereas the non-stationary analysis not only uses the full 40-year data set but its parameters are also time dependent.

Overall, the SPI time series computed following the WMO standards hold well over Europe in the majority of the cases, even if the results highlighted some sensible areas where both static baselines have difficulties in reliably capturing the magnitude of nSPI for the entire 10-year period for which they should be used, with part of the domain experiencing large differences right before the switch to a new baseline. In this regard, the performances of both baselines seem better (i.e., smaller differences) in the first half of the period for all the

accumulation periods. The obtained results, while confirming a general good consistency between the nSPI data set and the most recent static baseline, also suggest that an update frequency of 10 years for the baseline may not be enough to keep up the pace of the changes observed over some areas in the study domain.

In order to achieve a more spatially uniform adherence of the static SPI with the non-stationary reference, updates every 5 years may be more suitable over the study domain, with a good matching (MAD <0.25) for SPI-1 and acceptable matching (MAD <0.50) for SPI-12 over more than 80% of Europe during the entire period. This outcome needs to be carefully considered in any future operational implementation of ERA5-based SPI products over Europe.

AUTHOR CONTRIBUTIONS

Carmelo Cammalleri: Conceptualization; Data curation; Formal analysis; Methodology; Software; Visualization; Writing-original draft; Writing-review & editing. **Jonathan Spinoni:** Methodology; visualization; writing – review and editing. **Paulo Barbosa:** Methodology; project administration; supervision; writing – review and editing. **Andrea Toreti:** Methodology; supervision; writing – review and editing. **Juergen Vogt:** Methodology; supervision; writing – review and editing.

ORCID

Carmelo Cammalleri  <https://orcid.org/0000-0003-4834-7508>

Jonathan Spinoni  <https://orcid.org/0000-0002-8903-085X>

Paulo Barbosa  <https://orcid.org/0000-0003-3023-3502>

Andrea Toreti  <https://orcid.org/0000-0002-1983-2523>

Jürgen V. Vogt  <https://orcid.org/0000-0003-2955-9484>

REFERENCES

- Arnone, E., Pumo, D., Viola, F. and Noto, L.V. (2013) Rainfall statistics changes in Sicily. *Hydrology and Earth System Sciences*, 17, 2449–2458. <https://doi.org/10.5194/hess-17-2449-2013>.
- Bengtsson, L., Hagemann, S. and Hodges, K.I. (2004) Can climate trends be calculated from reanalysis data? *Journal of Geophysical Research – Atmospheres*, 109(D11), 1–8. <https://doi.org/10.1029/2004JD004536>.
- Caloiero, T., Caloiero, P. and Frustaci, F. (2018) Long-term precipitation trend analysis in Europe and in the Mediterranean basin. *Water and Environment Journal*, 32(3), 433–445. <https://doi.org/10.1111/wej.12346>.
- Golian, S., Javadian, M. and Behrangi, A. (2019) On the use of satellite, gauge, and reanalysis precipitation products for drought studies. *Environmental Research Letters*, 14, 075005. <https://doi.org/10.1088/1748-9326/ab2203>.
- Guttman, N.B. (1994) On the sensitivity of sample L moments to sample size. *Journal of Climate*, 7(6), 1026–1029.
- Guttman, N.B. (1999) Accepting the Standardized Precipitation Index: a calculation algorithm. *Journal of the American Water Resources Association*, 35(2), 311–322.
- Hayes, M., Svoboda, M., Wall, N. and Widhalm, M. (2011) The Lincoln Declaration on Drought Indices: universal meteorological drought index recommended. *Bulletin of the American Meteorological Society*, 92(4), 485–488.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. <https://doi.org/10.1002/qj.3803>.
- Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Muñoz Sabater, J., Nicolas, J., Radu, R., Schepers, D., Simmons, A., Soci, C. and Dee, D. (2019) Global reanalysis: goodbye ERA-Interim, hello ERA5. *ECMWF Newsletter*, 159, 17–24. <https://www.ecmwf.int/sites/default/files/elibrary/2019/19001-newsletter-no-159-spring-2019.pdf>.
- Hoffmann, D., Gallant, A.J.E. and Arblaster, J.M. (2020) Uncertainties in Drought From Index and Data Selection. *Journal of Geophysical Research – Atmospheres*, 125(18), e2019JD031946. <https://doi.org/10.1029/2019JD031946>.
- Kurnik, B., Barbosa, P. and Vogt, J. (2011) Testing two different precipitation datasets to compute the standardized precipitation index over the Horn of Africa. *International Journal of Remote Sensing*, 32(21), 5947–5964. <https://doi.org/10.1080/01431161.2010.499380>.
- Li, J.Z., Wang, Y.X., Li, S.F. and Hu, R. (2015) A Nonstationary Standardized Precipitation Index incorporating climate indices as covariates. *Journal of Geophysical Research – Atmospheres*, 120, 12082–12095. <https://doi.org/10.1002/2015JD023920>.
- Mahmoudi, P., Rigi, A. and Miri Kamak, M. (2019) Evaluating the sensitivity of precipitation-based drought indices to different lengths of record. *Journal of Hydrology*, 579, 124–181. <https://doi.org/10.1016/j.jhydrol.2019.124181>.
- Mavromatis, T. and Voulanas, D. (2021) Evaluating ERA-Interim, Agri4Cast, and E-OBS gridded products in reproducing spatiotemporal characteristics of precipitation and drought over a data poor region: The Case of Greece. *International Journal of Climatology*, 41(3), 2118–2136. <https://doi.org/10.1002/joc.6950>.
- McKee, T.B., N.J. Doesken and J. Kleist, 1993. The Relationship of Drought Frequency and Duration to Time Scales. Proceedings of the 8th Conference on Applied Climatology, 17–22 January, Anaheim, CA. American Meteorological Society.
- McKee, T. B., N. J. Doesken, and J. Kleist, 1995. Drought monitoring with multiple time scales, paper presented at 9th Conference on Applied Climatology, American Meteorological Society, Dallas, TX.
- Rashid, M.M. and Beecham, S. (2019) Development of a non-stationary Standardized Precipitation Index and its application to a South Australian climate. *Science of the Total Environment*, 657, 882–892. <https://doi.org/10.1016/j.scitotenv.2018.12.052>.

- Russo, S., Dosio, A., Sterl, A., Barbosa, P. and Vogt, J. (2013) Projection of occurrence of extreme dry-wet years and seasons in Europe with stationary and nonstationary Standardized Precipitation Indices. *Journal of Geophysical Research – Atmospheres*, 118, 7628–7639. <https://doi.org/10.1002/jgrd.50571>.
- Sen, P.K. (1968) Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of American Statistical Association*, 63(324), 1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>.
- Spinoni, J., Naumann, G. and Vogt, J.V. (2017) Pan-European seasonal trends and recent changes of drought frequency and severity. *Global and Planetary Change*, 148, 113–130. <https://doi.org/10.1016/j.gloplacha.2016.11.013>.
- Stagge, J.H., Kingston, D.G., Tallaksen, L.M. and Hannan, D.M. (2017) Observed drought indices show increasing divergence across Europe. *Scientific Reports*, 7, 14045. <https://doi.org/10.1038/s41598-017-14283-2>.
- Stagge, J.H., Tallaksen, L.M., Gudmundsson, L., Van Loon, A.F. and Stahl, K. (2015) Candidate distributions for climatological drought indices (SPI and SPEI). *International Journal of Climatology*, 35, 4027–4040. <https://doi.org/10.1002/joc.42.67>.
- Stasinopoulos, D.M. and Rigby, R.A. (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, 23(7), 1–46. <https://doi.org/10.18637/jss.v023.i07>.
- Wang, Y., Li, J., Feng, P. and Hu, R. (2015) A time-dependent drought index for non-stationary precipitation series. *Water Resources Management*, 29(15), 5631–5647. <https://doi.org/10.1007/s11269-015-1138-0>.
- World Meteorological Organization, 2012: Standardized Precipitation Index User Guide (WMO-No. 1090), Geneva.
- World Meteorological Organization, 2017: Guidelines on the Calculation of Climate Normals (WMO-No. 1203), Geneva.
- World Meteorological Organization and Global Water Partnership. (2016) Integrated Drought Management Programme (IDMP), Integrated Drought Management Tools and Guidelines Series 2. In: Svoboda, M. and Fuchs, B.A. (Eds.), *Handbook of Drought Indicators and Indices*. Geneva: World Meteorological Organization and Global Water Partnership.
- Wu, H., Hayes, M.J., Wilhite, D.A. and Svoboda, M.D. (2005) The effect of the length of record on the Standardized Precipitation Index calculation. *International Journal of Climatology*, 25(4), 505–520.

How to cite this article: Cammalleri, C., Spinoni, J., Barbosa, P., Toreti, A., & Vogt, J. V. (2022). The effects of non-stationarity on SPI for operational drought monitoring in Europe. *International Journal of Climatology*, 42(6), 3418–3430. <https://doi.org/10.1002/joc.7424>

APPENDIX

The maps in Figure A1 report the slope of the linear trend analysis for the year total precipitation, as calculated for ERA5 (Figure A1a, same as in Figure 1a) and E-OBS (Figure A1b) data sets. In general, the two maps share several similarities, including the predominance of decreasing trends in Central and Eastern Europe and a statistically significant ($p < 0.05$) increasing trend in the Mediterranean. The outcome for ERA5 appears overall smoother compared to E-OBS, likely due to the effects of the physically consistent numeric modelling in ERA5. In spite of the general similarities, a more in-depth inspection highlights some striking differences over specific regions, as the opposite trends observed over most of Greece, Norway, and Iceland, even if trends are significant for both data sets only over Greece.

Some studies have observed inconsistencies among different gridded precipitation data sets at local and regional scales as well as discordant results compared to local-scale analyses. As an example, a recent study from Mavromatis and Voulanas (2021) highlighted how trends in ERA5 data may be unreliable over some regions of Greece, where the E-OBS data set seems to perform better. They suggest that a possible cause of this behaviour can be related to the assimilation scheme, given that the two products use similar stations. Other studies based on local observations have highlighted mostly decreasing trends in year total P over Southern Italy (Arnone *et al.*, 2013), an opposite behaviour to what is reported in both ERA5 and E-OBS. In this case, an uneven temporal distribution of ground observations may be seen as a potential cause.

In general, while the broad spatial patterns seem consistent between the two data sets, the small trends observed over some areas during this short time span, the local discrepancies between the two outcomes, and the potential inconsistencies against other regional studies hint at the influence of other sources for trend in addition to climate, like temporal changes in data availability for assimilation/interpolation. For the sake of this study, it is also important to highlight how no major differences are observed between the overall magnitude of the observed trends, with similar range of variability in the slope values obtained for both data sets.

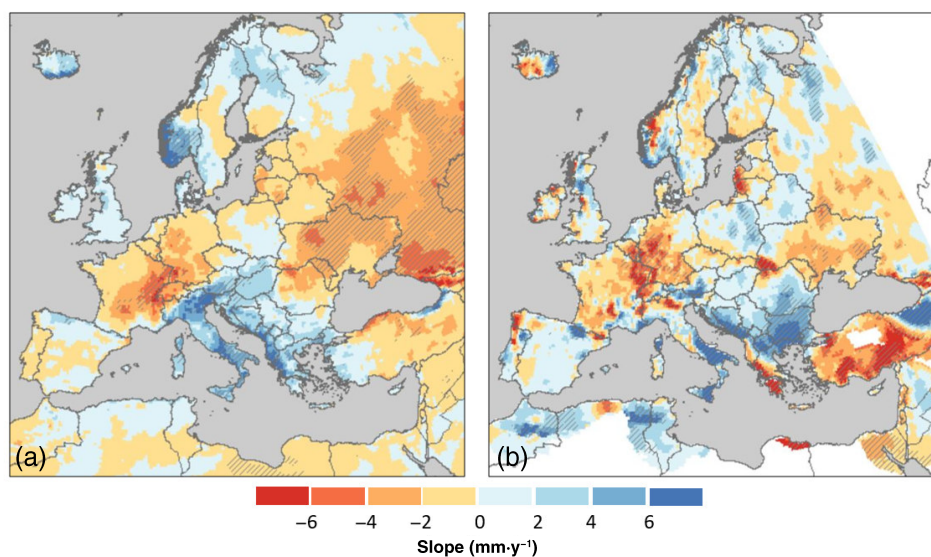


FIGURE A1 Linear trend slope (Theil–Sen approach) in year total precipitation over Europe for the period 1981–2020: (a) results for ERA5 (same as in Figure 1a) and (b) E-OBS. Hatching highlights statistically significant trends (two-tailed Mann–Kendall test, $p < .05$) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7424)]