RESEARCH ARTICLE

Domino: A new framework for the automated identification of weather event precursors, demonstrated for European extreme rainfall

Joshua Dorrington[1] | Christian Grams[1] | Federico Grazzini[2,3] | Linus Magnusson[4] | Frederic Vitart[4,5]

[1]Institute of Meteorology and Climate Research (IMK–TRO), Department Troposphere Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]Ludwig-Maximilians-Universität, Meteorologisches Institut, München, Germany

[3]ARPAE–SIMC, Regione Emilia–Romagna, Bologna, Italy

[4]Forecast Department, Evaluation Section, ECMWF, Reading, United Kingdom

[5]Department, Earth System Predictability Section, ECMWF, Reading, United Kingdom

Correspondence
Joshua Dorrington, Institute of Meteorology and Climate Research (IMK–TRO), Department Troposphere Research, Karlsruhe Institute of Technology (KIT), Karlsruhe 76133, Germany.
Email: joshua.dorrington@kit.edu

Abstract

A number of studies have investigated the large-scale drivers and upstream precursors of extreme weather events, making it clear that the earliest warning signs of extreme events can be remote from the impacted region in both time and space. Integrating and leveraging our understanding of dynamical precursors provides a new perspective on ensemble forecasting for extreme events, focused on building storylines of possible event evolution. This then acts as a tool for raising awareness of the conditions conducive to high-impact weather and providing early warning of their possible development. However, operational applications of this developing knowledge base are limited, perhaps for want of a clear framework for doing so. Here, we present such a framework, supported by open software tools, designed for identifying large-scale precursors of categorical weather events in an automated fashion and reducing them to scalar indices suitable for statistical prediction, forecast interpretation, and model validation. We demonstrate this framework by systematically analysing the precursor circulations of daily rainfall extremes across 18 regional- to national-scale European domains. We discuss the precursor rainfall dynamics for three disparate regions, and show our findings are consistent with, and extend, previous work. We provide an estimate of the predictive utility of these precursors across Europe based on logistic regression, and show that large-scale precursors can usefully predict heavy rainfall between two and six days ahead, depending on region and season. We further show how, for more continental-scale applications, the regionally specific precursors can be synthesised into a minimal set of indices that drive heavy precipitation. We then provide comments and guidance for generalisation and application of our demonstrated approach to new variables, timescales, and regions.

KEYWORDS

extreme precipitation, large-scale precursors, predictability

1 | INTRODUCTION

Understanding the conditioning of extreme weather on the large-scale flow, both locally and upstream, promises a path to improved guidance for extreme event forecasting and identifying windows of forecast opportunity. As examples, high rainfall over North Italy is driven by Rossby-wave trains with North American origin (Grazzini & Vitart, 2015), while Atlantic wave breaking favours Alpine heavy precipitation (Barton et al., 2016; Martius et al., 2008), and atmospheric regime variability modulates both alpine serial cyclone clustering (Barton et al., 2022; Tuel et al., 2022) and European atmospheric river occurrence (Pasquier et al., 2019). Localised temperature and wind extremes are also often associated with hemispheric variability, as revealed by recent studies on compound heatwaves (White et al., 2022), links between cold spells and storm damage (Messori et al., 2016), and the well-documented link between sudden stratospheric warmings and surface cold extremes, modulated by the tropospheric flow (Beerli & Grams, 2019). Even in state-of-the-art forecast systems, skill in predicting extreme precipitation, for example, rarely extends even a week ahead (Gascón et al., 2023; Leon, 2023), limiting the mitigation strategies available to policy makers and industry stakeholders. However, as forecast skill takes longer to decay at larger spatial scales (Lorenz, 1969), large-scale "event-prone" precursors—which may take the form of particular tropospheric circulation patterns or anomalies in slow-varying earth system modes—can be expected to be more predictable than the extreme events themselves. As small-scale processes, long-range teleconnections, and orography are imperfectly represented in models, there is no guarantee that a model will always convert the large-scale conditioning successfully into a surface extreme. However, if the conditioning of an extreme event on a large-scale precursor is sufficiently large—that is, $P(\text{event}|\text{precursor})/P(\text{event}) \gg 1$—then precursors can be used to offer quantitative improvements in forecast skill by using forecasts of precursors to infer the surface extreme statistically.

Qualitatively, a dynamical precursor perspective raises awareness of the necessary conditions for an extreme event, rather than focusing on raw model output. This is useful for building forecast storylines, highlighting important dynamical steps during the unfolding—or not—of an extreme event, a perspective that is already employed for climate projection (Shepherd et al., 2018). As these precursors often occur well before the extreme itself and may be tied to predictability barriers (González-Alemán et al., 2022; Oertel et al., 2023), such an approach can help explain "jumpiness" in situations of high forecast uncertainty, increasing end-user trust (Richardson et al., 2020).

Despite these advantages and the demonstrated value of local flow features in models to predict mesoscale events (Cafaro et al., 2019; Ferrett et al., 2023), there has been relatively little operational use of large-scale flow precursors thus far to improve operational forecasts. One notable exception is the use of regime approaches (Ferranti et al., 2015), which have been shown to boost precipitation forecast skill (Mastrantonas et al., 2021). Here, the approach is typically top-down; regimes are identified in large-scale fields such as geopotential height and the impacts on surface weather in each regime are diagnosed (Grams et al., 2017; van der Wiel et al., 2019). Other large-scale modes such as the Madden–Julian Oscillation (MJO), El Niño–Southern Oscillation (ENSO), or stratospheric polar vortex are also integrated into statistical modelling, especially at seasonal lead times, but are less commonly employed in the extended range. Bottom-up regime approaches exist, as in Bloomfield et al. (2021), which look for the dominant flow regimes that explain the most variability in variables of interest (in Bloomfield et al., energy generation). Of course, there is no guarantee that any pre-established dominant variability modes project cleanly upon a generic extreme event. In fact, if such an event is comparatively rare, as an extreme is by definition, then any predictability coming from dominant variability modes must necessarily be highly diffuse and associated with only minor modifications in occurrence probability. AghaKouchak et al. (2022), in the context of drought prediction, advocated for the bottom-up, event-centric approach of identifying the precursors of specific event types, rather than "thinking platonically", that is, thinking in terms of idealised modes of variability.

In this article, we lay the groundwork needed to make bottom-up eventcentric precursor analysis a routine technique in the toolbox of the operational meteorologist, the applied scientist, or the forecast end user. In Section 2, we introduce our analysis framework for identifying useful precursors, "Domino",[1] supported by a corresponding Python implementation, a fully documented application programming interface (API), and clear worked examples of different use cases.

In the rest of the article, we apply this new framework to analyse the precursors of daily rainfall extremes across Europe. Extreme weather events in Europe are responsible for billions of euros in economic damage and the loss of hundreds of lives every year (EEA, 2022; Hunt & Watkiss, 2011), with 40% of these deaths between 1950 and 2006 being a result of flash floods. Such short-timescale and spatially localised rainfall extremes are particularly challenging to forecast. Our analysis serves as a demonstration of the precursor approach, but also, by providing the first systematic analysis of national-scale precipitation drivers, as an end in its own right.

In Section 3, we discuss the data and event definitions used, the 18 spatial regions we choose, and our selection of Domino parameters. In Section 4, we will discuss the precursor dynamics of three archetypal regions: North Italy, West Iberia, and North Europe, which cover Mediterranean, Atlantic, and continental climates respectively, and demonstrate consistency with prior knowledge as well as noting a number of more novel observations. In Section 5, we will analyse the large-scale precursor patterns in wind, geopotential height, and vapour transport derived for all 18 regions, and discuss commonalities and variations. We will estimate the predictive skill of these precursors using logistic regression and investigate regional and seasonal differences. In Section 6, we show how the regionally specific precursor indices we have computed can be synthesised for larger scale considerations into a minimal set of European rainfall modes using partial least-squares regression and use these to characterise regional variations in rainfall dynamics. Finally, in Section 7, we discuss our conclusions and provide suggestions for the generalisation and extension of this work and

how the framework can best be used for directly applied work.

2 | DOMINO

The aim of the Domino framework is to identify automatically the predictive precursors of categorical, meteorological events. Comprehensive documentation and discussion of the full range of workflow parameters and customisation options can be found in the Domino github repository.[2]

In Section 3, we define concretely the parameters and analysis used in this article for our proof-of-concept application to European extreme rainfall, while in this section we summarise the general purpose and motivation behind each analysis step.

The framework, sketched schematically in Figure 1, is based on two types of input: "events", defined as a categorical time series indicating the class of event at each time step, and "variables": time series of candidate
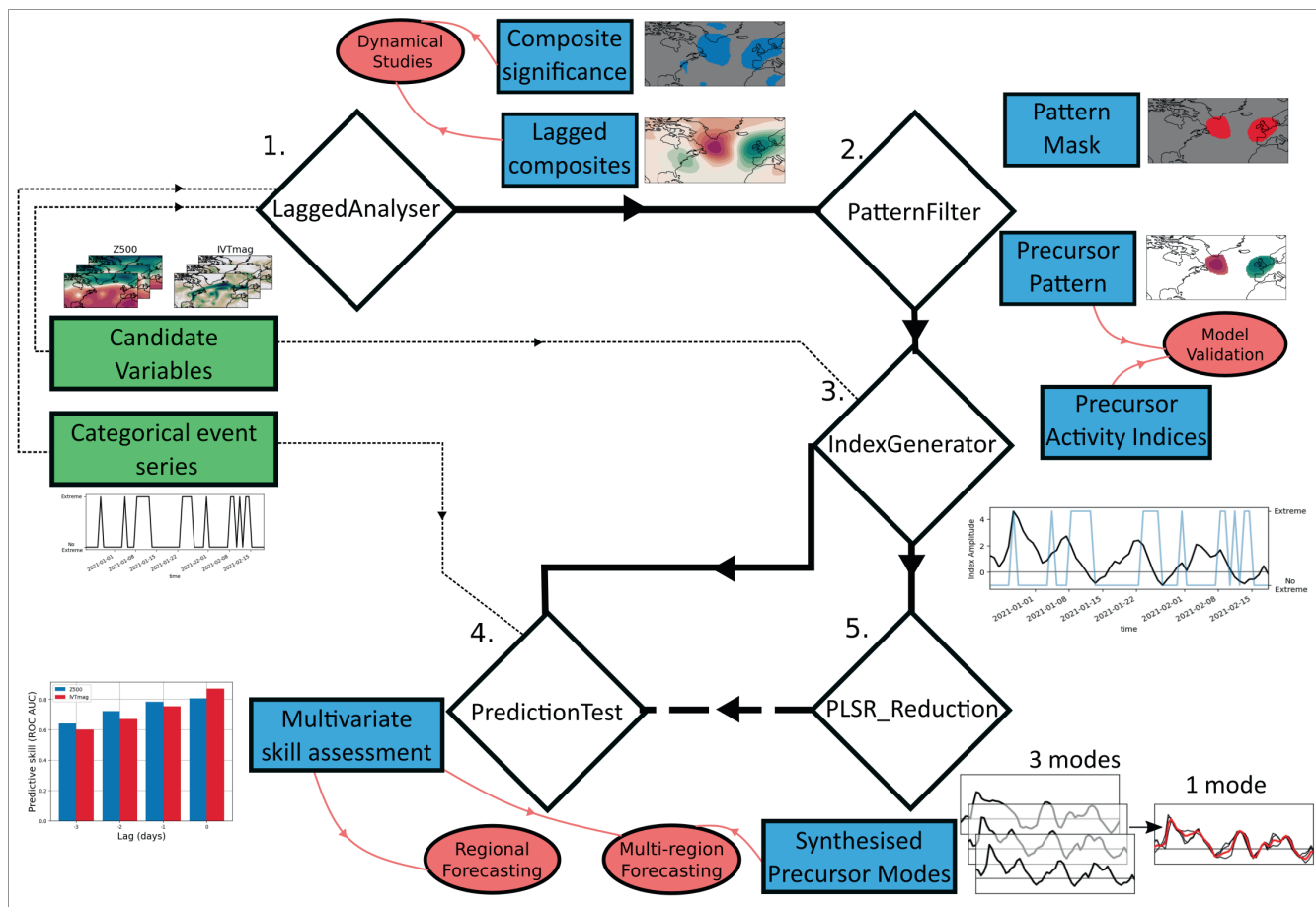
FIGURE 1 A schematic diagram of the Domino workflow. Inputs are shown in green boxes, key analysis steps in white diamonds, output datasets in blue boxes, and different applications that could make use of output data are indicated by red ovals. See accompanying main text for full description. Graphics show visual examples of each data type.

precursors that may play a role in influencing the event and may be time series of scalars, vectors, fields, or higher-dimensional variables. This has similarities to the "large scale meteorological pattern" approach of Grotjahn et al. (2016), which uses nonlocal composites to improve climate projections of North American heat extremes. Our overarching goal is to provide a simple workflow that can replicate, semi-automatically, previous analyses of extreme event precursors, and convert the results into a form that can be used for event prediction. In this way we aim to operationalise current academic insights and streamline the process of obtaining new insights.

1. *Compute lagged composites.* Implemented via the LaggedAnalyser class, time-lagged composites of the input variables are calculated with respect to each event category. Lag-dependent background states and/or seasonal cycles can be computed and subtracted. Synthetic event time series can be provided or generated automatically to serve as the basis for bootstrap significance testing. As such compositing approaches are a valuable analytical tool used in many studies of extreme event precursors, long-range teleconnections, and atmospheric dynamics, we anticipate that the clean, efficient implementation provided here will be of interest to many geoscientists.

2. *Refine precursor patterns.* Implemented via the PatternFilter class, a number of filters can be applied to precursor composites of multidimensional variables, iteratively building a Boolean mask which can be used to focus on features of interest. The motivation is to replicate in an automated way the subjective assessment of interesting features a scientist might make when looking at a composite field. Functions provided include filters for large-amplitude, spatially extended, and statistically significant patterns.

3. *Generate precursor indices.* Implemented via the IndexGenerator class, each precursor pattern can be reduced to scalar time-series of "precursor activity", by projecting anomaly fields onto the masked pattern. Such indices are the automatically generated, eventcentric equivalents of, for example, the well-known North Atlantic Oscillation (NAO), Quasi-Biennial Oscillation (QBO), or nino3.4 indices, defined as spatial averages or differences of spatial averages over particular regions, which provide a valuable scalar summary of important Earth-system variability. As well as computing precursor activity indices from scratch, updating an index using new variable time series and precomputed composites is supported, making it easy to calculate online updates of index time series or project indices into a different dataset, such as in model data.

Although not explored in this work, there is potential for using these indices to validate the eventcentric dynamics of models in much the same vein as model assessments of ENSO (Fredriksen et al., 2020) and the NAO (Simpson et al., 2020).

4. *Evaluate index predictivity.* Implemented via the PredictionTest class, this step provides an estimate of the predictive power of scalar indices for a categorical event. Uni- or multivariate combinations of predictor indices can be specified, as can the statistical model to use, the cross-validation strategy, and the skill score of interest. This can be used for feature selection, allowing for a subset of most predictive indices to be taken forward for additional analysis, or, as in this article, to compare the predictability of precursors in different regions and seasons.

5. *Reduce index dimensionality.* Implemented via the PLSR_Regression class, a set of scalar indices can be reduced to a lower-dimensional set via partial least-squares regression (PLSR), a statistical model that maximises the covariance between the reduced modes and a set of target variables. In this way, we can reduce the dimensionality of our scalar indices in an event-focused way which is well suited to prediction problems. This auxiliary functionality is useful in specific cases where precursors are desired that balance predictive power over different event types or across regions, as discussed in Section 6. The output PLSR modes can then also be passed to PredictionTest, just as can any other scalar index.

The software implementation is built on top of the popular Python package xarray (Hoyer & Hamman, 2017), which provides support for metadata rich, multidimensional array manipulation. Domino makes heavy use of its DataArray and Dataset classes to store and process data, which supports conversion both to and from Pandas Dataframes and iris Cubes, providing cross-compatibility with a range of data pipelines. Implementations of logistic regression and PLSR are taken from scikit-learn (Pedregosa et al., 2011). Via xarray, much of the workflow supports parallel and lazy computation via Dask (Rocklin, 2015), allowing large datasets to be processed quickly.

In this article we focus on the case in which the events have two classes—for example, the occurrence or not of extreme rainfall—but all the algorithmic steps of Domino apply equally to multiclass applications; terciles, deciles, events of multiple type, and so forth. Here we only apply Domino to precursors that are 2D continuous latitude–longitude (lat–lon) fields, but arbitrary dimensioned continuous precursors can be handled, as can categorical scalar precursors. In the latter case, occurrence

FIGURE 2   18 regions used in this article to define spatially aggregated precipitation extremes. Regions are based on catchment data taken from Lehner and Grill (2013), with modifications to account for orographic impacts on the prevailing circulations.
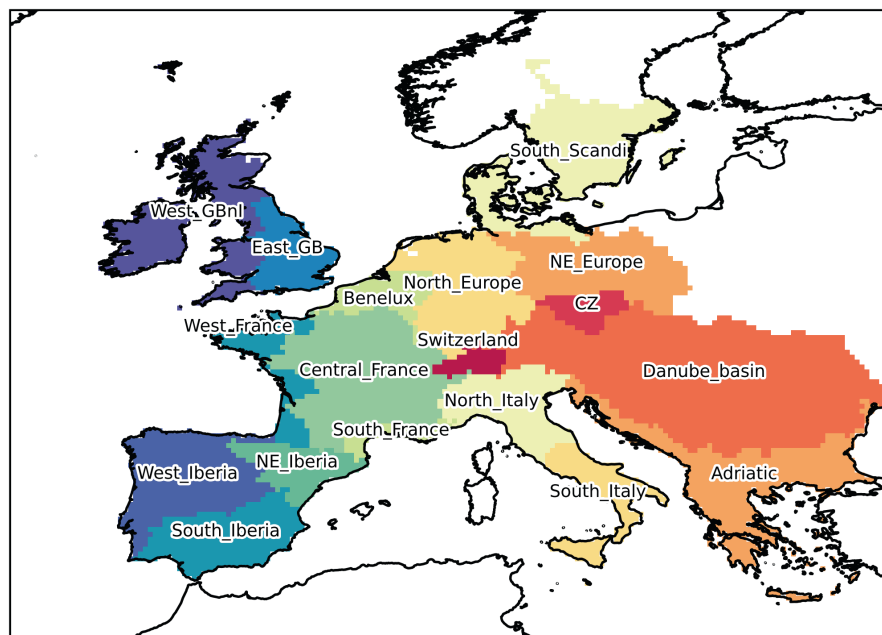
frequency for each category is computed automatically instead of a mean composite.

3 | DATA AND METHODS

3.1 | Data

Both events and precursors are identified in ERA5 reanalysis data (Hersbach et al., 2020) over the 43-year period 1979–2021. Precipitation data were downloaded as three-hourly accumulations at 0.25° resolution, and then accumulated to daily data. Four candidate precursor fields were chosen: 500-hPa geopotential height ($Z500$), 850-hPa zonal wind ($U850$), 300-hPa meridional wind ($V300$), and the magnitude of column-integrated vapour transport (IVTmag). These example fields were chosen to capture the large-scale circulation at a range of tropospheric heights. Each was downloaded as 1200 instantaneous data at 1° resolution.

3.2 | Region definition

Spatially averaged daily rainfall was defined as cosine-latitude weighted averages of precipitation over each of the 18 European regions shown in Figure 2. The boundaries for these regions were based on the HydroSHEDS drainage-basin product (Lehner & Grill, 2013), which avoids the mixing of hydrologically distinct areas that would result from a simple box-average approach. Basins were in some cases merged or split in a subjective fashion based on meteorological considerations

such as the position of mountain ranges and prevailing winds, or to merge very small regions. The areas, chosen to demonstrate the Domino workflow, are not of equal area or population and so are therefore not optimised for direct application. Rather, they serve to cover the majority of Europe collectively in a way that respects the underlying hydrometeorology, and the analysis of their precursors can then inform further studies of particular refined regions.

For each season and region, a categorical heavy rainfall event time series was computed. Events were defined as daily accumulated rainfall exceeding the region's seasonal 90th percentile. This relative definition ensures an equal number of samples in all samples and regions, with the drawback that in some cases such events will not always be high-impact. Supplementary Figure S1 shows the exact threshold for each case and region.

3.3 | Precursor computation

The specifics of how precursor patterns and activity indices were computed are detailed here, and are accompanied by a schematic (Figure 3). Each of the precursor variables ($Z500$, IVTmag, $U850$, and $V300$) was restricted to the extended north Atlantic region [20N–80N,140W–100E] and deseasonalised at a gridpoint level, by calendar date: the mean value of each variable on the same day of the year over the period 1979–2021 was smoothed with a 30-day rolling mean and subtracted from the daily fields.

For each season and region, lagged mean composites of the precursor variables were computed at daily lags from 7 to 0 days before the corresponding heavy rainfall event series. As heavy rainfall events were defined as
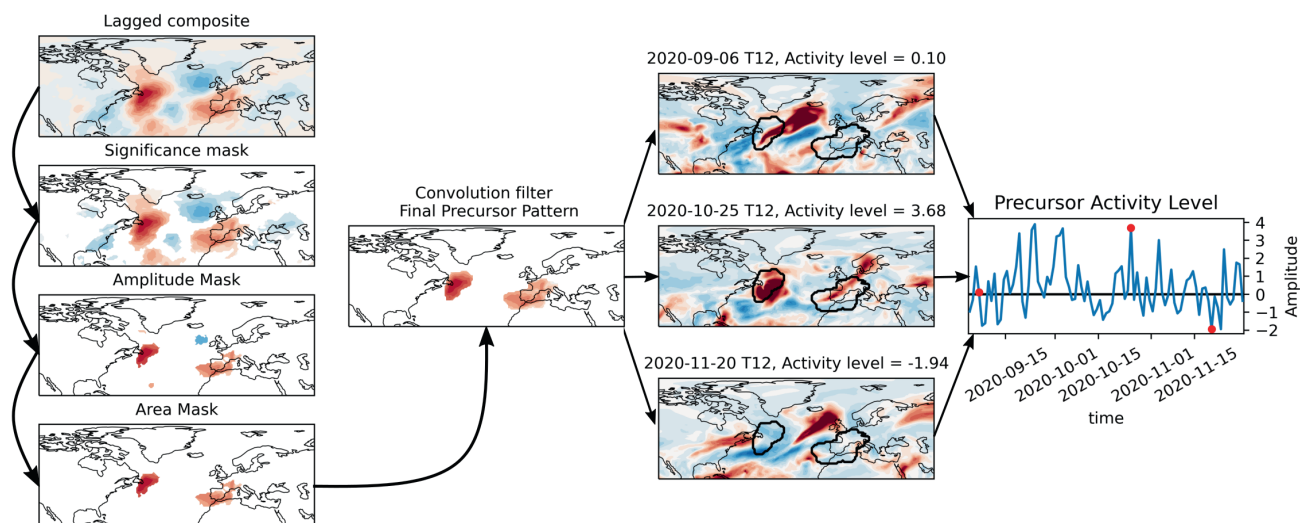
FIGURE 3 A schematic indicating how lagged composites are transformed into precursor patterns and then into precursor indices, demonstrated for SON IVTmag precursors two days prior to North Italian extreme rainfall events. The initial lagged composite is masked based on statistical significance, anomaly amplitude, and spatial extent of unmasked regions, then padded with a five-point 2D convolution. A dot product between this pattern and the daily IVTmag anomaly field discriminates between days that closely resemble the precursor flow and those that do not. When applied to all days in the IVTmag dataset, this dot product defines a daily precursor activity index, which is standardised. The three daily fields used as examples are marked with red dots in the resulting precursor activity time series.

exceedance of the 90th percentile threshold in each case, this represents composites of 385–395 individual days.

The statistical significance of precursor anomalies was assessed via a bootstrap approach. To account for autocorrelation on heavy rainfall days, 400 synthetic event indices were generated by fitting a Markov chain to the observed event series and resampling from it. Composites of each variable were computed with respect to each synthetic event series, and significance was defined on a gridpoint basis as a value below/above the 10th lowest/highest bootstrap composite (two-sided $p = 0.05$). A Holm–Bonferroni correction was then applied to each variable and lag separately (Holm, 1979), to reduce the false-positive rate. Composites at all time lags were compared with a single set of bootstrap composites computed at lag 0 for reasons of computational efficiency. This amounts to assuming an approximately stationary climate over a 10-day period, which is reasonable.

Lagged composites are transformed into "precursor patterns" by applying a number of filters to the Boolean significance field and then using this to mask the composite. This procedure results in statistically robust spatial anomaly patterns with large amplitudes, smooth edges, and without isolated unmasked points. Concretely, in addition to masking gridpoints with statistically insignificant anomalies, gridpoints with anomalies < 0.25 times the gridpoint climatological standard deviation were masked. Groups of unmasked gridpoints that comprise a connected area of less than 250,000 km^2 (i.e., equivalent to 20 $1 \times 1°$ equatorial gridpoints) were also masked.

Finally, all unmasked gridpoints were convolved with a five-point lat–lon square: all gridpoints within 2 degrees of an unmasked gridpoint were unmasked.

These precursor patterns are then used to compute precursor activity indices, defined as a dot product computed between each deseasonalised variable time series and each corresponding precursor pattern, and then spatially summed using cosine-latitude weighting. This produces a scalar precursor activity index, which is always standardised to 0 mean and standard deviation 1.

3.4 | Skill assessment

To characterise the potential predictability of heavy rainfall precursor activity indices, we use a logistic regression approach. Stratifying by season, region, and precursor variable, in each case our target variable is the corresponding ERA5 Boolean heavy rainfall event index, where we aim to predict the probability of an event occurring based on one or more continuous predictors. The schematic Figure 4 illustrates how this is done. For each time lag, season, region, and large-scale variable we have one precursor pattern (Figure 4a). Each index is offset in time according to its lag, that is, a four-day precursor index calculated from data for January 1, 2000 will be used to predict for January 5, 2000. We train a model for each lag l between 7 and 0 by using all precursor indices that have a lead time $\leq l$: thus the seven-day model is a univariate model using only the lag 7 precursor index, while a one-day
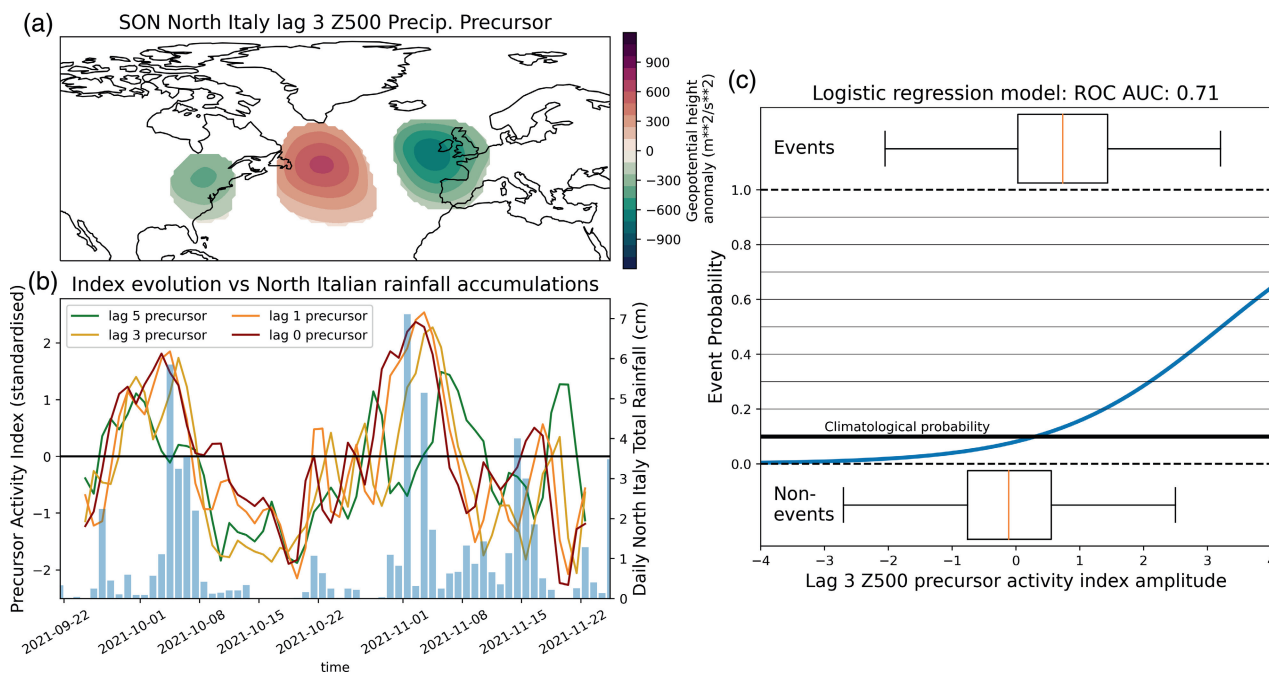
FIGURE 4 A schematic illustrating how the predictive power of precursor activity indices is estimated in a reanalysis framework. (a) For each time lag, region, season, and variable a precursor pattern is defined. Each pattern corresponds to a time-varying activity index, which captures the degree to which the current flow resembles the pattern, as shown in the right panel of Figure 3. (b) These indices co-vary with rainfall to a greater or lesser extent. (c) We fit a logistic regression model with precursor indices for all lags $\geq l$ for a given season, region, and variable, in order to predict categorical heavy rainfall occurrence, and quantify skill using the ROC AUC score (box plots show the distribution of index values during heavy rainfall events or otherwise).

forecast uses all precursors except for the lag 0 precursor. To cross-validate, each year of data was excluded sequentially, and a logistic regression model was fitted to the remaining 42 years to determine $P(\text{extreme}|\text{precursor})$. This was then used to predict the occurrence of daily heavy rainfall events in the excluded year. The probabilistic (receiver-operating-characteristic area-under-curve) ROC AUC skill score (Hanley & McNeil, 1982) was computed for each cross-validation, and the mean value is reported (the interquartile range is shown in Supplementary Figures S3 and S4). This provides a much more realistic estimate of the potential predictability achievable by a perfect forecast than simply fitting a model to the entire event time series. We do not account for autocorrelation in our categorical rainfall events, by, for example, keeping only the first day of multiday rainfall events, as this would reduce sample size by 26% on average across regions and seasons, and we do not wish to penalise the correct prediction of such multiday extremes.

3.5 | Partial least-squares regression

Dimensionality reduction is a common task in atmospheric data analysis, with principal component analysis (PCA) being the most popularly used approach. PCA aims

to find the N-dimensional linear subspace of an initially M-dimensional dataset, which captures the greatest degree of variance. PLSR is similar in that it aims to reduce dimensionality while maximising explained variance, but now by finding the N-dimensional linear subspace of an M-dimensional dataset X, which maximises the cross-covariance with a second L-dimensional dataset Y. That is, it estimates the low-dimensional projection of X that is best at predicting the various target variables of Y. For the mathematical formalism see Abdi (2010) and for algorithmic implementation see scikit-learn (2023). It is closely related to canonical correlation analysis, which has been previously applied to meteorology by, for example, Maldonado et al. (2013). For our example, we demonstrate how PLSR can be applied to SON $Z500$ precursor patterns for different regions and time lags, to obtain a minimal set of modes that balances week-ahead heavy rainfall predictability across Europe. This is described in more detail in Section 6.

4 | DYNAMICS OF RAINFALL PRECURSORS

Having discussed our approach and methodology in general terms, we now apply precursor analysis to
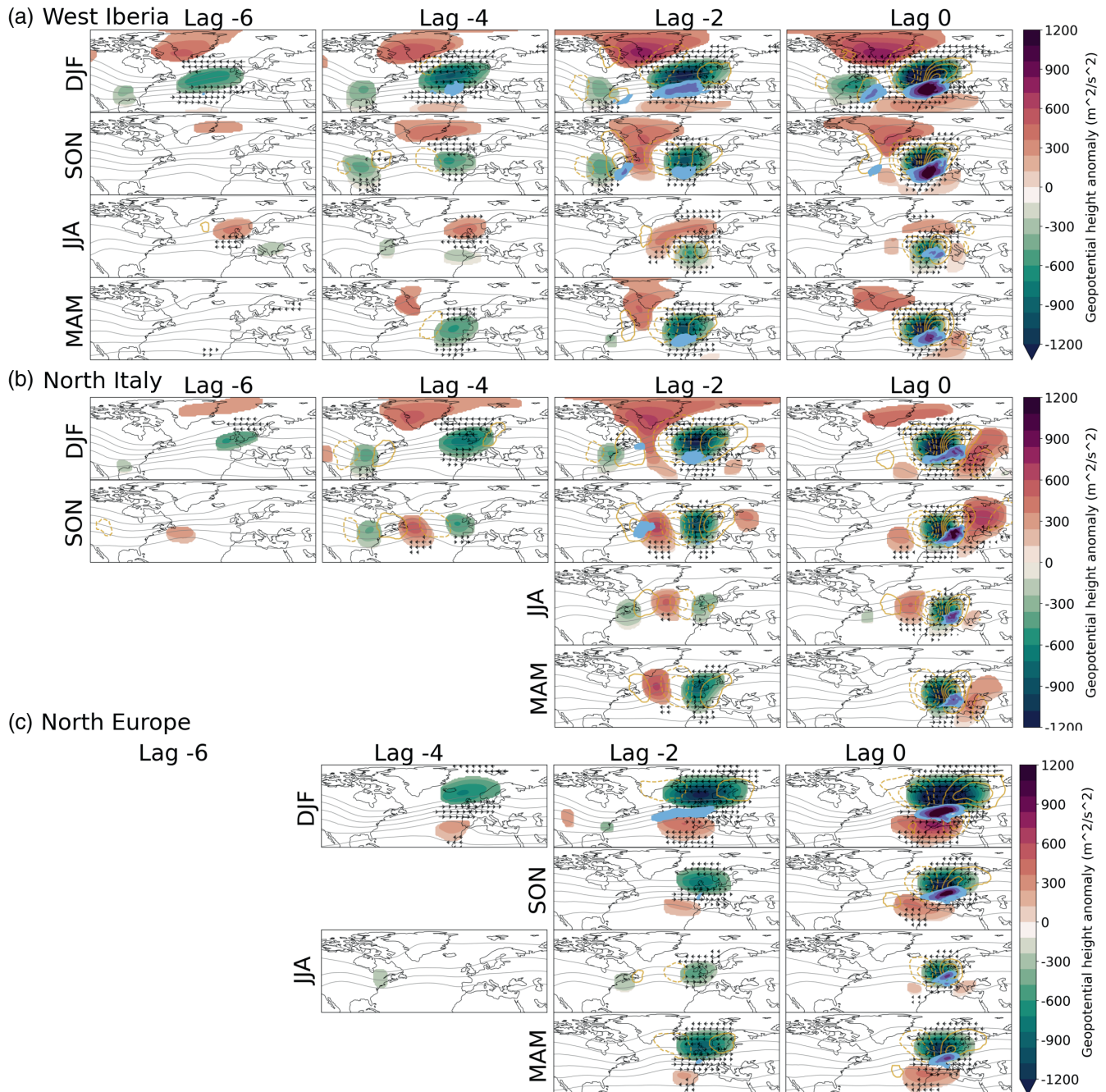
FIGURE 5 Precursor patterns 6, 4, 2, and 0 days before heavy rainfall events in (a) West Iberia, (b) North Italy, and (c) North Europe, as a function of season and computed from ERA5 reanalysis data. Red-to-green shading indicates $Z500$ anomalies as shown by the colour bar. Grey contours show full field $Z500$ for levels with 1e4 spacing between $5.2e5$ and $5.7e5$ m$^2 \cdot$ s^{-2}. Gold contours show $V300$ anomalies at 4 m $\cdot$ s^{-1} intervals ranging from -16 to 16 m $\cdot$ s^{-1}, and excluding the 0 contour. Blue–purple shading shows positive IVTmag anomalies at the 50, 80, 110, and 140 kg m $\cdot$ s^{-1} levels. Arrows indicate anomalies of the $U850$ wind, and should be interpreted qualitatively.

European daily heavy rainfall extremes. We begin, in this section, by analysing the multivariate precursor patterns for North Italy, West Iberia, and North Europe. Figure 5 shows precursor patterns for the three regions, 6, 4, 2, and 0 days before heavy rainfall events and across seasons. For each large-scale variable we have a unique precursor pattern for each combination of time lag, season, and region.

$Z500$ precursors indicate the anomalous trough and ridge structure associated with events, $V300$ the upper-level wave dynamics, $U850$ the eddy-driven jet, and IVTmag the role of anomalous moisture transport. Collectively they therefore capture most of the important midlatitude driving dynamics and allow a holistic understanding of each event type.

4.1 | West Iberia

The "West Iberia" region consists of Portugal and coastal Northern Spain, as well as the inland basins of the Douro and Tagus rivers. Precursors extend further back in time in this region than any other considered, as we discuss in Section 5. In this mountainous domain, the majority of rainfall occurs in autumn and winter (Mora & Vieira, 2020) and, while summer rainfall is less frequent and generally dominated by mesoscale processes, the majority of June–July–August (JJA) coastal extreme rainfall events are driven by atmospheric rivers (Ramos et al., 2018). In March–April–May (MAM), snowmelt is an important flood driver, which we neglect here. In general, heavy daily precipitation in this region is known to be dominated by the latitudinal shifting of the jet stream which, when in its southerly state, can guide cyclones and atmospheric rivers to this area (Ramos et al., 2015) and is associated with the negative phase of the NAO (Eiras-Barca et al., 2016; Salgueiro et al., 2013) and cut-off lows to the west (Eiras-Barca et al., 2018).

This pre-existing knowledge is reaffirmed by the large-scale West Iberian precursor patterns shown in Figure 5a. Robust precursors extend back six days in all seasons, demonstrating a substantial conditioning of heavy rain events here on the upstream Atlantic state. In December–January–February (DJF), precursors in the $Z500$ and $U850$ fields indicate southerly jet deflection, accompanied by a negative-NAO-like geopotential height anomaly six days before a rainfall event. Both wind and geopotential anomalies strengthen through to lag 0. This extended persistence of the precursor flow is consistent with the persistent, trimodal regime dynamics of the wintertime Atlantic jet (Parker et al., 2019; Woollings et al., 2010). From lag 4 onwards, a clear zonally oriented IVT precursor emerges to the west of Iberia, indicating increased atmospheric river frequency and the need for not only a displaced jet but also anomalous moisture transport to induce an extreme. While the $U850$ anomaly is initially confined to the east Atlantic, a secondary trough forms over the southeast United States from lag 6, inducing its own local $U850$ anomaly and a diagonally oriented IVT precursor, which suggests a secondary moisture source flowing from the west to the east Atlantic. While the early precursors are predominately zonal, large $V300$ anomalies emerging from lag 4 onward indicate the formation of an upper-level trough over the domain, possibly aided by wave-breaking, as discussed in (Santos et al., 2018). In September–October–November (SON), dynamics are similar, but less zonally symmetric, with clearer upper-level wave features and a more confined trough and associated deep ridge over the mid-Atlantic. Here the role of the secondary cyclone is comparatively more pronounced, perhaps a requirement given the overall weaker moisture transport associated with a weaker low-level jet. In MAM and JJA there are less pronounced large-scale precursors, but a substantial amount of nonlocal conditioning can still be seen, including low pressure over the United States, high pressure over Greenland, and East Atlantic jet anomalies, as well as the high IVT anomalies indicative of increased atmospheric river frequency. The lag 0 $Z500$ precursor indicates that cut-off lows are dominant in summer, while troughs are more common in autumn, in agreement with Pino et al. (2016), who analysed the synoptics of the 24 most severe Iberian floods in the past 150 years. The upstream trough over the Panhandle, visible in $U850$ and $Z500$ precursors for SON and DJF 2–4 days prior to heavy precipitation, is also associated with a recently uncovered pathway: Leeding et al. (2023) have shown that DJF cold air outbreaks over this same US region are associated with increased Iberian precipitation 0–5 days later, and that this trough pattern is associated with further southward jet deflection. Our results suggest that there may also be a role for this teleconnection in SON, which has not previously been explored.

4.2 | North Italy

The North Italian domain suffers from extreme flash flooding events, which are heavily dependent on the large-scale flow and Atlantic Rossby-wave activity (Grazzini et al., 2020a; Grazzini & Vitart, 2015). Highly mountainous and shielded from direct westerly flows by the French Alps, southerly winds are needed to bring large moisture transport into the region, necessitating some level of upstream wave activity. The precursor patterns (Figure 5b) reveal a sharp distinction between the nonlocal conditioning of SON and DJF events, which stretch back to lag 6, and MAM and JJA events, which only emerge at lag 2. Despite this, precursor patterns for the nonwinter seasons are qualitatively rather similar, all showing zonally oriented wave trains with large-scale moisture transport over Southern Europe. For SON, a pre-existing weak ridge over the west Atlantic interferes constructively with the upstream Rossby wave, which becomes clear by lag 4 and is centred at ∼45N. From here the wave amplifies and shows slow eastward movement. High IVT at lag 2 in the west Atlantic may indicate the increased occurrence of extratropical cyclones recurving northeast, which can amplify wave growth and induce Mediterranean cyclone development (Grams et al., 2011; Raveh-Rubin & Flaounas, 2017). The trough to the west of Italy is fully developed by lag 2, with an associated southern European IVT anomaly. Even in JJA, when convective rainfall would be expected to dominate, high IVT over North Italy

is seen, likely as a result of unstable moist air from the Mediterranean advecting and rising over steep orography, as described by (Grazzini et al., 2020b). In DJF, the picture is quite different, with precursors showing a Greenland blocking/NAO− precursor pattern emerging, interacting with an upstream wave train, and leading to a more tripolar high-latitude circulation anomaly up until lag 2. High geopotential heights around the European trough indicate a cut-off structure, and there is a sharp decline in the amplitude of the Greenland blocking anomaly on the day of the North Italian extreme rainfall. This breakdown of the blocking and associated regime transition implies elevated levels of wave breaking over western Europe (Masato et al., 2012), which may aid in deepening and localising the trough and hence amplifying southerly moisture transport. For the North Italian case, these $Z500$, $V300$, and IVTmag precursor indices have been shown in Grazzini et al. (2023) to increase the predictive power of a hybrid dynamical–statistical prediction model, indicating a capability not only to reproduce the known dynamics, but to transform them into a usable source of skill.

4.3 | North Europe

The "North Europe" domain consists of the catchments of the Ems and Weser rivers, as well as the northern portion of the Rhine and Meuse catchments, including large parts of Germany, eastern Belgium, and the northern Netherlands. In contrast to the other two domains considered in this section, this domain, which excludes the German Alps and the Ore mountains to the south and east respectively, features relatively flat terrain and exhibits a more continental climate. Although the heavy rainfall threshold is comparable in all seasons (6 mm·day^{-1}), there are notable seasonal variations in the precursor patterns (Figure 5c). Without strong orographic constraints or direct coastal exposure to Atlantic variability, significant precursors only begin to emerge 4 days before rainfall events during DJF, and only 2 days before in JJA and MAM. In DJF, an initial subtropical ridge off the coast of Morocco develops into a meridional dipole, with low geopotential height over the North Sea and a intensified central low-level jet 4 days prior to heavy rainfall. By lag 2 this anomaly strengthens, with high IVT anomalies coincident with the jet, indicating an influx of Atlantic moisture into the region. At the same time, upper-level wave activity becomes prominent, with some possible indications of upstream forcing from a weak wave-like geopotential anomaly over the United States. By the day of the event itself, the $V300$ field shows a quadrapole structure supporting strong moisture convergence into the region. The low geopotential heights in the precursor pattern are more diffuse than for the other

regions considered. As the domain is more isotropic, the exact placement of weather features over the domain is less strongly constrained, resulting in a greater "blurring" of the precursor pattern. Indeed, cyclones impacting this region are known to evolve along many disparate paths (Hofstätter et al., 2018). A qualitatively similar, but less robust, series of precursors is visible for MAM and SON, with less large-scale conditioning.

Surprisingly, JJA rainfall events have quite extended and remote precursor patterns, with a similar North American low appearing 4 days prior, as was seen for SON West Iberian rainfall events. This is followed by the formation of a localised low over the UK, displaced south compared with other seasons with an accompanying upper-level wave visible at 2 days prior, possibly as a result of downstream development. There is no accompanying southerly high geopotential height as in other seasons, indicating a less zonal flow and a more localised cut-off. The high-pressure precursor over Italy at 1 day prior suggests that a Mediterranean source of moisture is relevant, capturing the footprint of the rare but impactful "Vb" cyclones (Hofstätter et al., 2018).

We have shown for three disparate European regions that our automated precursor detection approach is able to reproduce key insights from previous analyses of regional extreme precursors, and via its systematic application we have been able to discuss interseasonal variability fairly comprehensively, in a way that enables intercomparison between regions. The purported value of this approach is that by conversion of precursor patterns into precursor activity indices we will likewise be able to convert these dynamical insights into improved event predictability.

In the next section, we will show that by converting these precursor patterns into precursor activity indices we can improve event predictability, and zoom out from detailed consideration of a few case study regions to characterise the predictive power of precursor indices for all 18 European regions. Equivalent plots to Figure 5 for all regions can be found in the supplementary material.

5 | EUROPEAN HEAVY RAINFALL PRECURSORS AND PREDICTION

In this section we turn our attention to the precursor activity indices that correspond to each precursor pattern, and show that they have the potential to improve the predictability of heavy rainfall events. Without analysing forecast data, which is the subject of future work, we estimate a lower bound on predictive power by assessing the skill of logistic regression models based on reanalysis precursor indices alone for the prediction of heavy rainfall probability in each region, as discussed in Section 3.1. Heuristically,

we consider a ROC AUC score > 0.6 to indicate some useful amount of predictive power in a precursor.

By using a multivariate or nonlinear model, it is likely that considerably higher skill could be reached, while in a hybrid dynamical–statistical prediction framework the predictive value will be closely tied to the forecast model's ability to identify precursors at extended lead times, as will be discussed in further detail in Section 7. In fact, for the specific case of North Italy, a hybrid dynamical/machine-learning model leveraging these non-local precursors has already been found to boost skill out to day 10 (Grazzini et al., 2023).

For now, we proceed by analysing the univariate logistic-regression-based ROC AUC of precursor indices in the European regions shown in Figure 2. Figure 6a shows the precursor skill averaged over seasons (for all seasons for which a precursor pattern is defined) and showing each region separately. As previous studies on extreme event precursors have focused on targeted regions and are often directed by specialist intuition that precursors should exist, it is by no means obvious that extremes over a generic region will or should have predictive long-range precursors. Through our systematic multiregion analysis, we are now able to comment on this point. There is substantial inter-regional spread in precursor skill; approximately double that of the seasonal spread. Differences are largest at short lead times, with precursor skill between regions converging at lags 3–5 and generally falling below 0.6 during this period.

It is interesting to consider the worst-case scenarios, that is the regions with the lowest skill for each lead time and variable. The regions with the least skilful precursors nevertheless show ROC AUC above 0.7 for day 0 precursors, and maintain non-negligible skill for $Z500$, $U850$, and $V300$ out to lag 2. These least predictable regions are in all cases one of CZ and NE_Europe (see Supplementary Figure S2 for a more detailed presentation), both of which are isolated from the upstream Atlantic flow. Many regions show much greater precursor skill and in the best cases there are indications that even lag 5 precursors have useful levels of skill.

Summarising another way, Figure 6b shows the ROC AUC for precursors based on each large-scale predictor variable, conditional on season and averaged across the 18 regions. For all variables, precursor skill is highest in DJF, generally followed by SON, with MAM and JJA showing similar, lower skill, with a total interseasonal spread of approximately 1 day of predictability. This corresponds to the known increased importance of large-scale forcing in DJF and the increased role of convective processes in JJA. This skill difference may then derive in part from choice of precursor variables, which emphasises the large-scale flow. While early experiments including convective available potential energy (CAPE) and sea-surface temperature (SST) fields did not reveal skilful precursor patterns, it is possible that a more detailed consideration of such thermodynamic variables might reveal better summertime precursors. The spread in skill between different
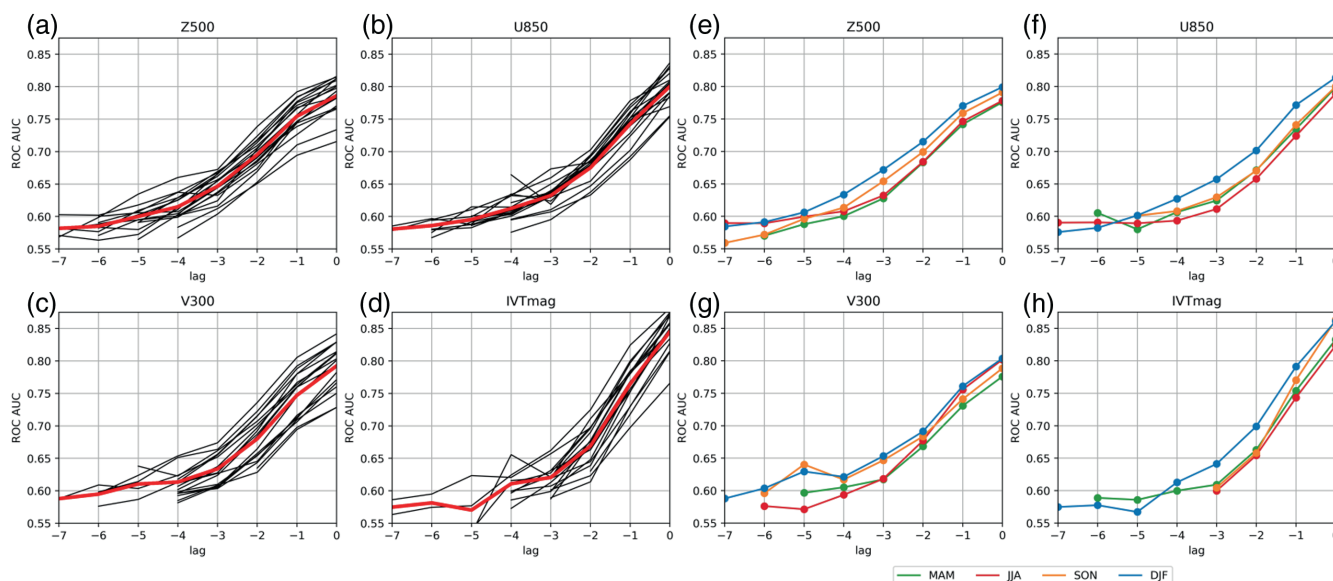
FIGURE 6 (a–d) Predictive skill of precursor indices, as measured by cross-validated ROC AUC score averaged across seasons for each of (a) 500-hPa geopotential height ($Z500$), (b) 850-hPa zonal wind ($U850$), (c) 300-hPa meridional wind ($V300$), and (d) the magnitude of column-integrated vapour transport (IVTmag), with the skill for each region shown by a black line and the mean across both seasons and regions shown in red. (e–h) ROC AUC now averaged across regions and conditioned on season.

cross-validations is shown in supplementary Figures S4 and S5.

IVTmag precursors show the highest skill at short timescales but decay rapidly, with ROC AUC dropping below 0.6 at lags 3 and 4 for nonwinter and wintertime precursors, respectively. Short-lag $Z500$ precursors are comparatively less skilful but maintain predictive power out to lag 5 on average, which is consistent with the general large-scale wave-like precursors often seen for $Z500$, rather than the more isolated streaks of moisture flux seen in IVTmag. $V300$ and $U850$ precursors fall in between these two extremes, although $U850$ is notably more skilful in DJF, in keeping with the dominant role of DJF Atlantic trimodal jet variability (Woollings et al., 2010).

For simplicity, we are working within a categorical event context here, both to identify our precursor indices and to test their predictive skill. However, it is interesting to consider whether the precursor indices are able to assess the amplitude of an extreme event. This is primarily left to future work, but in supplementary Figure S5 we show the value of precursor indices for the most extreme daily rainfall total in each region and season, according to ERA5. There are several notable cases in which these extremes are associated with index values, usually for IVTmag, > 6

standard deviations above the norm, suggesting that this is indeed the case.

The role of spatial variability in shaping precursor skill can be seen more clearly in Figure 7 for $Z500$ precursors. Equivalent skill maps for the other large-scale variables can be found in the supplementary data. In general, the $Z500$ precursor patterns considered represent either upstream wavetrains or blocking events at long lead times and localised troughs at shorter lead times. We see that coastal regions show the longest range precursors, at lags of 6, both on the Atlantic coast, as for Iberian and UK regions, and also along the Eastern Mediterranean, where Italy and the Adriatic coast show early precursors, especially in DJF. Most regions start to show precursors, often skilful ones, from day 4 onwards, often associated with the first indications of local trough formation. More mountainous domains tend to show higher precursor skill, a consequence of the greater constraints placed on the large-scale flow to allow substantial vapour transport into the domain. There is certainly no simple overall relation between location and precursor skill, however: even neighbouring regions can have very different precursor skill and different seasonal variability, even when they have similar precursor patterns, partly as a result
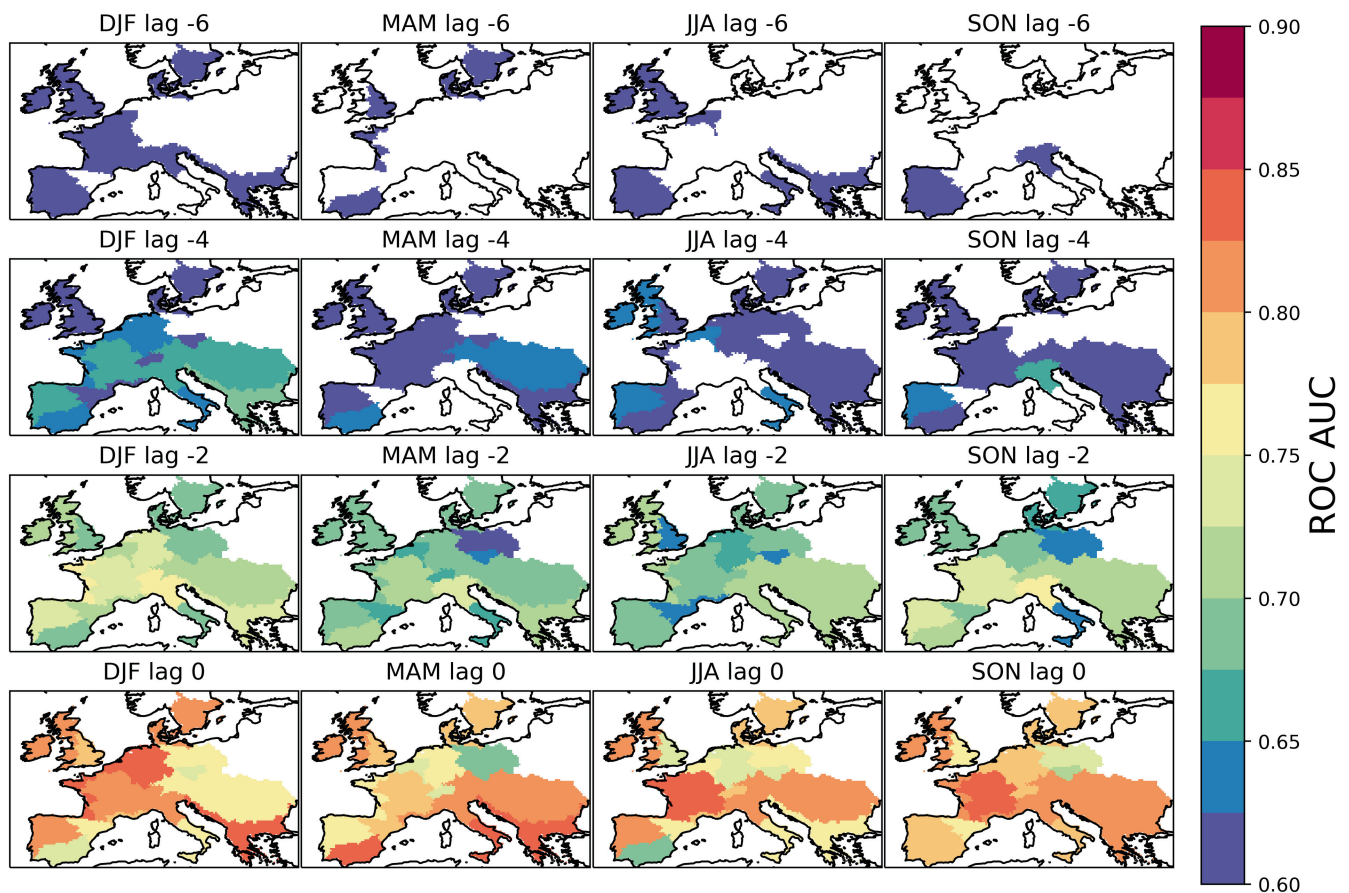

**FIGURE 7**   Z500 precursor ROC AUC for each spatial region, as a function of lag and season.

of differing orography and domain size. This motivates exactly the sort of spatially differentiated approach we have taken here.

6 | SYNTHESISING PRECURSORS ACROSS REGIONS

In the previous sections we have demonstrated the ability of automated precursor detection to reveal the dynamical drivers, and aid the prediction, of extreme events across a large number of European domains. For many applications, such a regional perspective is ideal, but there are also applications for which a broader perspective is useful. For example, operational centres responsible for an entire continent (such as the European Centre for Medium-Range Weather Forecasts (ECMWF) or National Weather Service (NWS)) might wish to monitor the risk of extreme events in geographically and administratively distinct regions of their domain, but without tracking individual precursors for all regions. Entirely equivalently, a national forecaster whose domain features great climactic diversity may wish to apply the precursor approach outlined above to subregions or individual stations, but then also to synthesise extreme event risk holistically. To address this use case,

we now demonstrate how precursors for different regions can be synthesised down to a minimal set of indices that balance predictability in different regions, using PLSR regression.

Our task is simplified in many cases by the fact that precursor patterns for extremes in neighbouring regions can often be very similar. This is demonstrated for lag 2 Z500 precursors in Figure 8, which is representative of many precursor patterns, suggesting that it should be possible to reduce the dimensionality of an initially large number of precursor activity indices. By doing so, we aim to obtain a few general modes relevant for many regions. Concretely, in our case we see that seasonal differences for particular regions are often very pronounced, while qualitative similarities between regional precursor patterns for a given season are easy to see in Figure 8: the positive-NAO-like pattern characteristic of DJF precursors in northwest domains and the negative-NAO-like patterns in Mediterranean domains, JJA southeasterly wave visible in a large number of regions, and more zonally propagating wave that dominates SON central and southern Europe precursors with only slight phase differences. Further, as can be seen in Figure 5, the precursor patterns for sequential time lags can also be very similar.
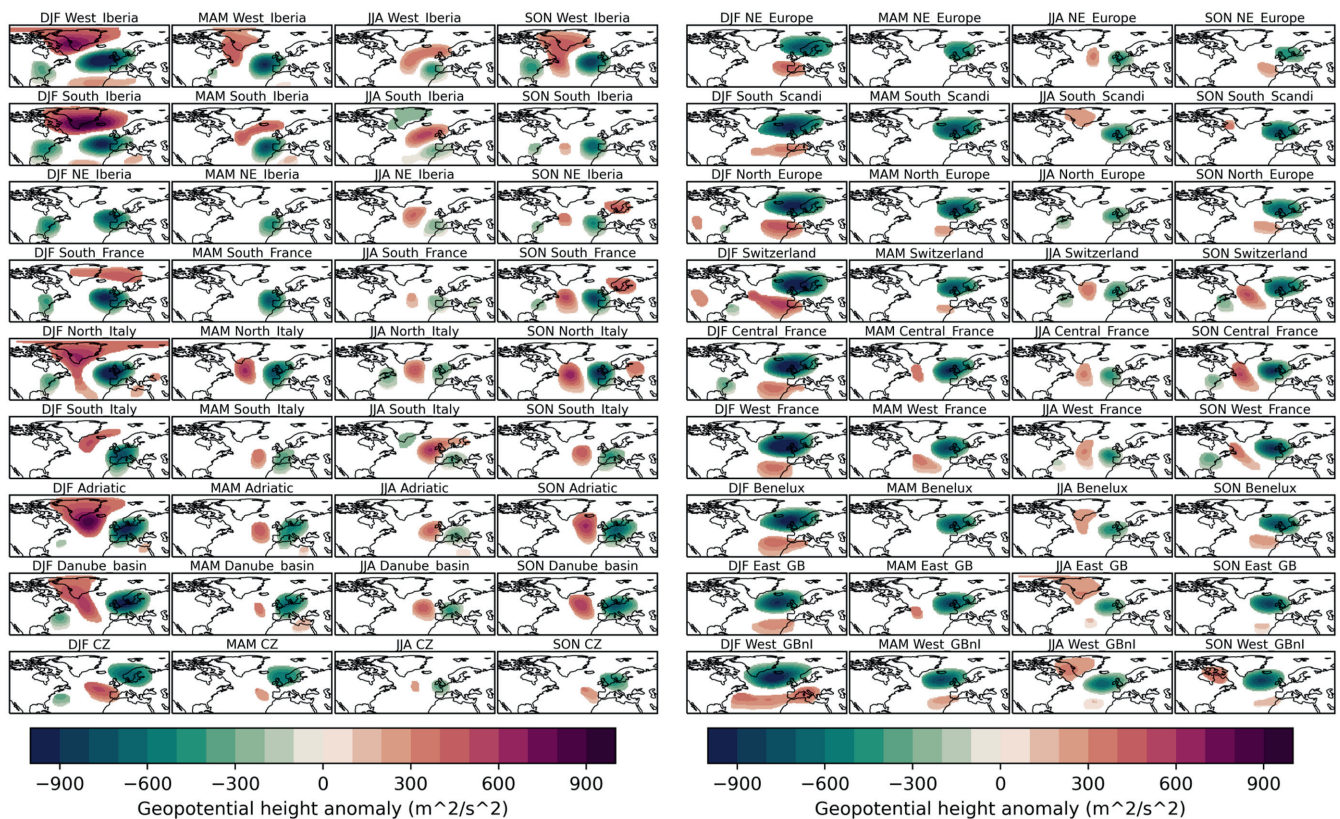
FIGURE 8 Lag 2 Z500 anomaly precursor patterns for each region and season. Clear similarities can be seen between many neighbouring regions.

Without carrying out an exhaustive synthesis for all variables and seasons, we demonstrate how a minimal set of precursor indices can be produced for $Z500$ SON precursors, with a method that can be easily applied to any other season or variable of interest. $Z500$ precursors in SON generally show high predictive power at long lags, while large-scale SON dynamics are not as comprehensively studied as those of DJF, making this a natural combination to trial our approach on.

We apply PLSR as discussed in Section 3, searching for the n orthogonal linear combinations (PLSR modes) of our predictor indices that explain the most variance in our event variables. As our predictors, we take SON $Z500$ precursor indices spanning from lags 5 to 0 for each of the 18 regions (108 variables total, reduced to 88 after discarding blank precursor patterns). For the target variables we use the heavy rainfall event index for each of the 18 regions, but, as we seek modes that capture precursors at different lead times, that is, we want to include lag covariance, we duplicate each event index and offset it in time between 0 and 5 days, producing a corresponding 108-dimensional event index as our target variable.

Figure 9a shows $Z500$ anomaly patterns corresponding to the first eight PLSR modes. As for empirical orthogonal functions (EOFs), the sign of the anomaly patterns is arbitrary, each representing a symmetric mode of variability. Unlike EOFs of $Z500$, however, which tend to be dominated by large-scale dipole patterns, the resulting PLSR modes clearly project onto small-scale flow features that are important for extreme European precipitation events. Larger scale trough/ridge patterns are visible in PLSR modes 1 and 3, zonal wave trains in modes 2, 4, and 5, and cut-off lows/highs in modes 7 and 8.

In exchange for reducing 88 initial precursors to 8 modes, the interpretation of PLSR mode impact on rainfall becomes more complex: positive values of any given mode will increase heavy rainfall occurrence in some regions and at certain lead times over the following five days and decrease it in others. Furthermore, as the PLSR indices are uncorrelated to each other by construction, the covariability of the indices prior to an event carries additional information. This can be seen in Figure 9b, which shows the average value of each PLSR index in the days preceding and following SON heavy rainfall events in each region. For North Italy as an example, pronounced high or low values of almost all modes occur in the lead up to a heavy rainfall event: high modes 1 and 6 combining to approximate the lag 4 wavetrain precursor seen in Figure 5b, low
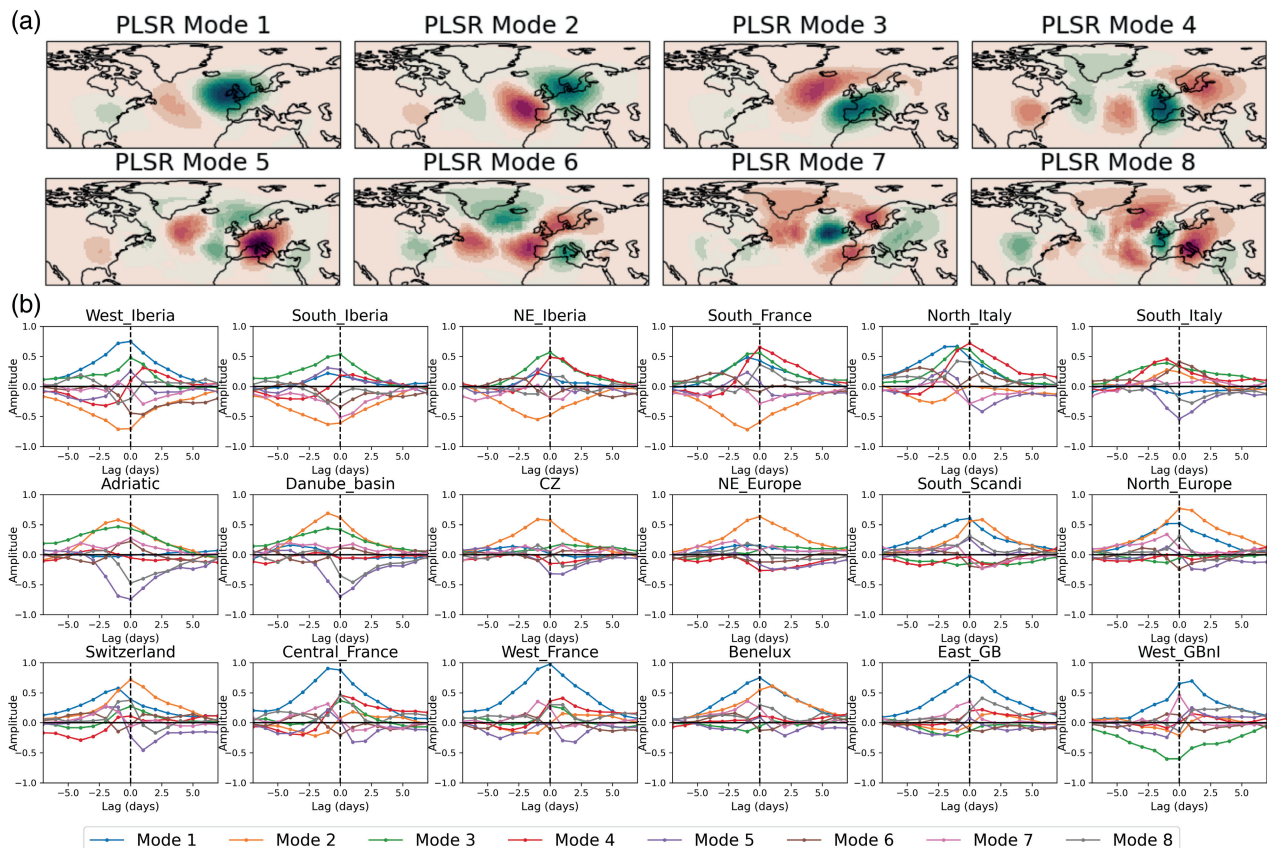
FIGURE 9 (a) $Z500$ anomaly patterns corresponding to each PLSR mode of SON $Z500$ precursors across lags 5 to 0. (b) The mean composite of each PLSR mode relative to occurrence of a SON heavy rainfall event in each region.

values of mode 2 indicating the movement of the wave eastward by lag 2, and strong amplification of mode 4 as the wave reaches its final position just to the west of Italy.

The reduced interpretability of the PLSR modes makes them better suited for use in statistical prediction models, rather than manual operational monitoring. Indeed, we find that the predictive power of the PLSR indices compares favourably with the predictive power of the regionally specific precursor indices considered previously. Figure 10a shows the ROC AUC of statistical predictions of SON regional heavy rainfall using the same cross-validated logistic regression approach as in Section 3.4, but, rather than using a single regionally specific precursor index as our predictor, we use between 1 and 12 PLSR indices as predictors. The difference between the PLSR-based skill and the skill found for SON $Z500$ precursors for each region is shown in Figure 10b. In general we find that the regional indices are more skilful than predictions based on < 5 PLSR indices, with little improvement seen from using more than 8 PLSR indices. This is particularly the case when looking at lag 0 patterns, where the trough driving rainfall in each region is most localised and so least well approximated by general modes. However, overall we see that the degrees of freedom in our precursors can be reduced by a factor of 10 without reducing statistical skill. We therefore suggest that PLSR reduction may be a useful post-processing step for multidomain applications to simplify workflows and for deep learning approaches.

7 | DISCUSSION AND CONCLUSION

The "quiet revolution" of steadily improving weather forecasts has been achieved through the gradual synthesis and operational application of improved meteorological understanding (Bauer et al., 2015). In this article we have presented a novel framework, termed "Domino", for identifying the large-scale precursor patterns of extreme weather events and reducing those to data-efficient, predictive scalar activity indices. In doing so we provide a starting point for the operationalisation of research into the conditioning of extreme events on large-scale dynamics.

7.1 | Summary of results

We have tested the "Domino" precursor identification framework by analysing regional daily heavy precipitation across western and central Europe and across seasons. Estimating predictive skill using only a simple univariate regression framework, we have shown that in all cases at least some degree of usable predictability (ROC AUC > 0.6) is present in large-scale precursors based on geopotential height, wind fields, and vapour transport patterns. This conditioning is in general lowest in JJA, when mesoscale dynamics are more dominant, and in flatter, more continental domains. Even in these cases, predictive precursors are seen three days ahead, while in DJF, where conditioning on large scales is strongest, some domains indicate borderline-useable skill 6 days ahead. We have demonstrated that the precursor patterns identified are consistent with the previously known dynamical precursors for three diverse regions and, by virtue of our systematic approach, are able to characterise them comprehensively. Finally, we have demonstrated that many precursor patterns are similar between different domains; the same large-scale dynamics favour rainfall in multiple regions. We show how this can be used to simplify the analytical framework without giving up regional specificity by synthesising the precursor indices for different regions together into a lower-dimensional set of predictive modes, using partial least-squares regression. For the specific case of SON $Z500$ precursors, we show that a 10 times reduction in dimensionality can be achieved with only a minor decrease in predictive skill and a more complex physical interpretation, a trade-off that may be useful for certain applications. This synthesis is enabled by the use of precursor patterns to generate scalar precursor indices, and indicates the value of this novel approach.

7.2 | Caveats

Naturally, there are limitations of the Domino framework which must be discussed. Most obviously, as a bottom-up data-driven method, the precursor patterns will be sensitive to changes in the event time series, with longer time series providing more robust results. The convolution filter applied during pattern computation counteracts this to an extent, by smoothing the pattern mask. In this article we have used $\mathcal{O}(300)$ events for each region, which was found in early tests to provide qualitatively identical results for different data subsamples.

Secondly, as the framework is based on lagged composites of potentially predictive variables over all events, it is not possible to distinguish distinct dynamical pathways leading to the same event; instead the precursor pattern will contain a superposition of both. In the pathological case where two precursor pathways have opposite circulations they would cancel, producing no significant composite. This can be mitigated by subsetting the extreme events if such pathways are suspected a priori (e.g., by splitting by season, as done here), or by choice of variables (e.g., wave envelope amplitude versus $V300$ to prevent phase
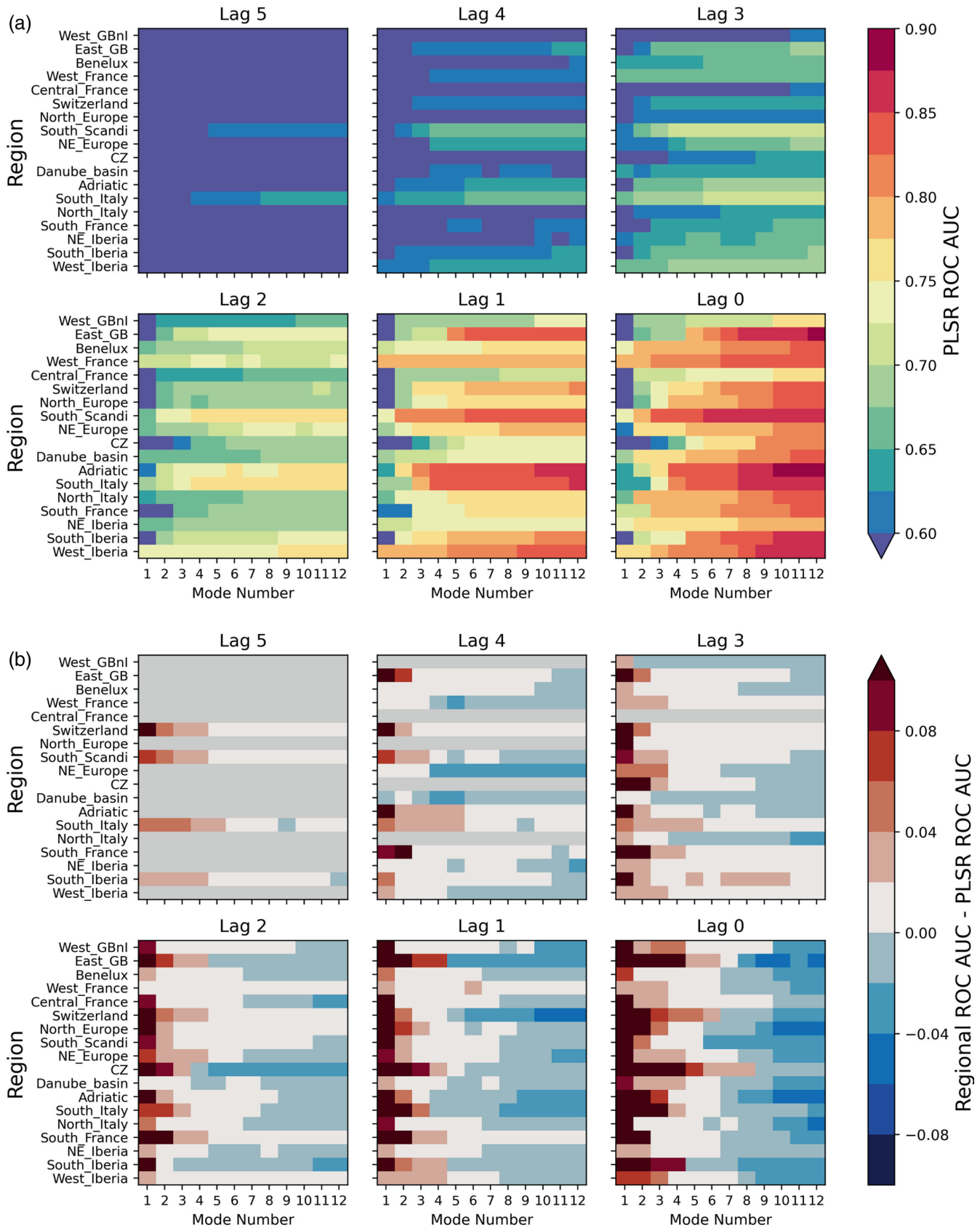
FIGURE 10 (a) ROC AUC skill score for SON *Z*500 PLSR modes used to predict SON heavy rainfall via logistic regression and (b) the difference between PLSR skill and the corresponding SON *Z*500 precursor activity index skill in each region as a function of region and number of modes used. In (b), red values indicate that regional indices are more predictive, while blue values indicate that PLSR modes are more predictive and lag–region combinations with no defined *Z*500 SON precursor pattern are greyed out.

cancellation). Another option is to compare the variance of the composite with the climatological value, an analysis supported by Domino, but which has not been included in this work for simplicity. A related point is the implicit stationarity assumption at play. Nonstationary dynamics can essentially be understood as multiple pathways occurring preferentially during different time periods. Applying Domino to rolling time windows of the event time series could reveal such underlying nonstationarity.

7.3 | Extension

In this section, we present methodological advice for extending the work done in this article to different events, regions and timescales.

In this article we have provided a relatively straightforward proof-of-concept application; there are multiple refinements that could be made for real-world flood-prediction applications. Firstly, events could be defined using rain-gauge data or local high-resolution gridded rainfall datasets in place of the ERA5 precipitation used here. Spatial averaging can be avoided entirely to provide extra information on local extremes: precursor indices could be computed for each rain gauge or grid-point independently, and then synthesised via the PLSR approach described above. For a consideration of true flood risk instead of simply high precipitation risk, hydrological discharge data could be used to define events. In this case, including land-surface precursor variables such as soil saturation and snow accumulation is likely to be of value.

Extending beyond the logistic regression used here, the scalar precursor indices are well suited for input into more advanced statistical and machine-learning models, already providing a low-dimensional summary of the flow. While raw data sometimes prove just as effective as physically refined data for deep-learning applications, at the least, precursor analysis can assist variable selection by highlighting the nonlocal regions and timescales that are relevant for a particular event. In this case, it may be useful to lower the predictive threshold used to define a "useful" precursor, in order to include more marginal or conditional sources of skill.

Relatedly, the predictability of events we report here should not be understood as inherent, but contingent on the precursor variables we have chosen. For summer rainfall, predictive thermodynamic precursors may exist for some regions, based on variables such as CAPE, static stability, and T2m, although the predictability of such processes is inherently much more limited than the large-scale dynamics we focus on here.

Generalisation to different regions is trivial, simply requiring different event data. However, relevant precursor variables will of course be quite different in tropical or subtropical regions, where the thermodynamic variables mentioned above may contribute more significantly, and in strongly continental climates, where land-surface variable are likely more relevant. As a coastal, midlatitude region, the variables considered here may extend most straightforwardly to the west-coast United States Grotjahn et al. (2016), perhaps also with the addition of Pacific SSTs.

It is similarly easy to move beyond rainfall extremes; wind extremes or extreme waves are both impactful on daily timescales and could be treated in much the same way as rainfall. For longer timescale events such as heatwaves, or even relevant nonextremes such as predicting monthly rainfall terciles, the main methodological differences are a result of reduced sample size. When computing the pattern mask, a reduced amplitude mask and increased convolution filter may increase sensitivity to weak sources of seasonal-to-subseasonal predictability while helping to maintain robust patterns. Variables such as sea-ice extent, SSTs, and stratospheric wind are natural candidates for such a study.

Finally, beyond prediction, automated precursor analysis has value for both the fundamental study of dynamics and model validation purposes. Just as circulation regimes are used to explore the dynamics of surface teleconnections (Domeisen et al., 2020), investigate sources of model bias (Wandel, 2023), and judge the performance of climate models (Dorrington et al., 2022), "event-prone regimes" and corresponding indices could be used to summarise the representation of different dynamical processes in models. Conversely, events could be defined in terms of dynamical features—such as sudden stratospheric warmings, regime transitions, or wave-breaking events—and their precursor and "postcursor" indices can be used to help understand the variability between events by, for example, clustering in the space of indices.

7.4 | A path to operationalisation

Our motivating hypothesis is that the predictive precursors we identify have potential operational use. Given the timescale of our precursors (i.e., 2–6 days ahead) and the quality of modern numerical weather prediction systems, a purely statistical prediction based on precursors will not outperform dynamical forecasts directly (although this may be different if applied on S2S or seasonal timescales). Instead quantitative value may come from hybrid forecasting: using the dynamical forecast to predict precursor patterns at extended-range lead times and then inferring rainfall probability even further ahead

from the precursors. A complete roadmap for real-world application is as follows.

- Define an event time series based on high-quality observational data.
- Compute reanalysis precursors for this event class in large-scale variables and fit a statistical relation between precursors and events.
- Project precursors onto ensemble forecast data and use those to produce probabilistic hybrid predictions of event occurrence.
- Compare direct and precursor-based forecast skill to assess whether quantitative skill improvements are seen.

We consider quantitative skill improvements to be highly plausible, due to weak wave propagation and coarse representation of orography and near-surface processes in models, and Grazzini et al. (2023) have recently shown the assumption to hold for the North Italy region. Even qualitatively, precursors hold narrative value for forecasters, helping them to understand the range of possibilities and unfolding storylines in a forecast ensemble and so boosting their awareness of plausible weather beyond the limits of formalised skill. In a sense, analysing the unfurling preconditions to extreme events can serve to precondition the forecaster, allowing them to interpret and act on probable extreme events more quickly as they emerge at shorter lead times. In our own continuing work, we aim to enact this roadmap fully for extreme rainfall, and to develop an early-warning system for extreme events that summarises and digests the behaviour of large forecast ensembles.

AUTHOR CONTRIBUTIONS

Joshua Dorrington: conceptualization; data curation; investigation; methodology; resources; software; validation; visualization; writing – original draft; writing – review and editing. **Christian Grams:** conceptualization; funding acquisition; project administration; supervision; writing – review and editing. **Federico Grazzini:** conceptualization; writing – review and editing. **Linus Magnusson:** conceptualization; writing – review and editing. **Frederic Vitart:** conceptualization; writing – review and editing.

DATA AVAILABILITY STATEMENT

ERA5 data used in this study are freely available from the Copernicus Data Store. Software tools required to duplicate or extend this work are available as part of the Domino package at https://github.com/joshdorrington/domino.

A compressed folder of further supplementary figures is available at https://doi.org/10.5281/zenodo.8087482.

ENDNOTES

[1] For the curious, the name "Domino" was inspired by the imagery of a row of dominoes falling in causal sequence, which serves as a (admittedly optimistically deterministic) metaphor for the sequence of meteorological developments that increase the likelihood of a weather event of interest.
[2] https://github.com/joshdorrington/domino.

ORCID
Joshua Dorrington https://orcid.org/0000-0003-3802-457X
Christian Grams https://orcid.org/0000-0003-3466-9389
Federico Grazzini https://orcid.org/0000-0002-3435-2376
Linus Magnusson https://orcid.org/0000-0003-4707-2231
Frederic Vitart https://orcid.org/0000-0001-8485-7981

REFERENCES

Abdi, H. (2010) Partial least squares regression and projection on latent structure regression (PLS regression). *WIREs Computational Statistics*, 2, 97–106.

AghaKouchak, A., Pan, B., Mazdiyasni, O., Sadegh, M., Jiwa, S., Zhang, W. et al. (2022) *Status and prospects for drought forecasting: opportunities in artificial intelligence and hybrid physical statistical forecasting*. Philosophical Transactions of the Royal Society A, 380(2238), 20210288.

Barton, Y., Giannakaki, P., von Waldow, H., Chevalier, C., Pfahl, S. & Martius, O. (2016) Clustering of regional-scale extreme precipitation events in southern Switzerland. *Monthly Weather Review*, 144, 347–369.

Barton, Y., Rivoire, P., Koh, J., Ali, S., Kopp, J.M. & Martius, O. (2022) On the temporal clustering of European extreme precipitation events and its relationship to persistent and transient large-scale atmospheric drivers. *Weather and Climate Extremes*, 38, 100518.

Bauer, P., Thorpe, A. & Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55.

Beerli, R. & Grams, C.M. (2019) Stratospheric modulation of the large-scale circulation in the Atlantic–European region and its implications for surface weather events. *Quarterly Journal of the Royal Meteorological Society*, 145, 3732–3750.

Bloomfield, H.C., Brayshaw, D.J., Gonzalez, P.L.M. & Charlton-Perez, A. (2021) Pattern-based conditioning enhances sub-seasonal prediction skill of European national energy variables. *Meteorological Applications*, 28, e2018.

Cafaro, C., Frame, T.H.A., Methven, J., Roberts, N. & Bröcker, J. (2019) The added value of convection-permitting ensemble forecasts of sea breeze compared to a Bayesian forecast driven by

the global ensemble. *Quarterly Journal of the Royal Meteorological Society*, 145, 1780–1798. https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3531

Domeisen, D.I.V., Grams, C.M. & Papritz, L. (2020) The role of North Atlantic–European weather regimes in the surface impact of sudden stratospheric warming events. *Weather and Climate Dynamics*, 1, 373–388.

Dorrington, J., Strommen, K. & Fabiano, F. (2022) Quantifying climate model representation of the wintertime euro-Atlantic circulation using geopotential-jet regimes. *Weather and Climate Dynamics*, 3, 505–533.

EEA. (2022) Economic losses and fatalities from weather- and climate-related events in Europe—European Environment Agency. https://www.eea.europa.eu/publications/economic-losses-and-fatalities-from

Eiras-Barca, J., Brands, S. & Miguez-Macho, G. (2016) Seasonal variations in North Atlantic atmospheric river activity and associations with anomalous precipitation over the Iberian Atlantic margin. *Journal of Geophysical Research: Atmospheres*, 121, 931–948.

Eiras-Barca, J., Lorenzo, N., Taboada, J., Robles, A. & Miguez-Macho, G. (2018) On the relationship between atmospheric rivers, weather types and floods in Galicia (NW Spain). *Natural Hazards and Earth System Sciences*, 18, 1633–1645.

Ferranti, L., Corti, S. & Janousek, M. (2015) Flow-dependent verification of the ECMWF ensemble over the euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, 141, 916–924.

Ferrett, S., Methven, J., Woolnough, S.J., Yang, G.-Y., Holloway, C.E. & Wolf, G. (2023) Hybrid dynamical–statistical forecasts of the risk of rainfall in Southeast Asia dependent on equatorial waves. *Monthly Weather Review*, 151, 2139–2152 https://journals.ametsoc.org/view/journals/mwre/151/8/MWR-D-22-0300.1.xml

Fredriksen, H.-B., Berner, J., Subramanian, A.C. & Capotondi, A. (2020) How does El Niño–southern oscillation change under global warming—a first look at CMIP6. *Geophysical Research Letters*, 47, e2020GL090640.

Gascón, E., Montani, A. & Hewson, T.D. (2023) Post-processing output from ensembles with and without parametrised convection, to create accurate, blended, high-fidelity rainfall forecasts.

González-Alemán, J.J., Grams, C.M., Ayarzagüena, B., Zurita-Gotor, P., Domeisen, D.I.V., Gómara, I. et al. (2022) Tropospheric role in the predictability of the surface impact of the 2018 sudden stratospheric warming event. *Geophysical Research Letters*, 49, e2021GL095464.

Grams, C.M., Beerli, R., Pfenninger, S., Staffell, I. & Wernli, H. (2017) Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, 7, 557–562.

Grams, C.M., Wernli, H., Böttcher, M., Čampa, J., Corsmeier, U., Jones, S.C. et al. (2011) The key role of diabatic processes in modifying the upper-tropospheric wave guide: a North Atlantic case-study. *Quarterly Journal of the Royal Meteorological Society*, 137, 2174–2193.

Grazzini, F., Craig, G.C., Keil, C., Antolini, G. & Pavan, V. (2020a) Extreme precipitation events over northern Italy. Part I: a systematic classification with machine-learning techniques. *Quarterly Journal of the Royal Meteorological Society*, 146, 69–85.

Grazzini, F., Dorrington, J., Grams, C.M., Craig, G.C., Magnusson, L. & Vitart, F. (2023) Boosting forecast of precipitation extremes over N-Italy using machine learning. In Preparation.

Grazzini, F., Fragkoulidis, G., Pavan, V. & Antolini, G. (2020b) The 1994 Piedmont flood: an archetype of extreme precipitation events in northern Italy. *Bulletin of Atmospheric Science and Technology*, 1, 283–295.

Grazzini, F. & Vitart, F. (2015) Atmospheric predictability and Rossby wave packets. *Quarterly Journal of the Royal Meteorological Society*, 141, 2793–2802.

Grotjahn, R., Black, R., Leung, R., Wehner, M.F., Barlow, M., Bosilovich, M. et al. (2016) North American extreme temperature events and related large scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, 46, 1151–1184. Available from: https://doi.org/10.1007/s00382-015-2638-6

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29–36.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049.

Hofstätter, M., Lexer, A., Homann, M. & Blöschl, G. (2018) Large-scale heavy precipitation over Central Europe and the role of atmospheric cyclone track types. *International Journal of Climatology*, 38, e497–e517.

Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.

Hoyer, S. & Hamman, J. (2017) Xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, 5, 10.

Hunt, A. & Watkiss, P. (2011) Climate change impacts and adaptation in cities: a review of the literature. *Climatic Change*, 104, 13–49.

Leeding, R., Riboldi, J. & Messori, G. (2023) On Pan-Atlantic cold, wet and windy compound extremes. *Weather and Climate Extremes*, 39, 100524.

Lehner, B. & Grill, G. (2013) Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27, 2171–2186.

Leon, J.A.P. (2023) Scale-dependent verification of precipitation and cloudiness at ECMWF. https://www.ecmwf.int/en/newsletter/174/earth-system-science/scale-dependent-verification-precipitation-and-cloudiness

Lorenz, E.N. (1969) The predictability of a flow which possesses many scales of motion. *Tellus*, 21, 289–307.

Maldonado, T., Alfaro, E., Fallas-López, B. & Alvarado, L. (2013) Seasonal prediction of extreme precipitation events and frequency of rainy days over Costa Rica, Central America, using canonical correlation analysis. *Advances in Geosciences*, 33, 41–52.

Martius, O., Schwierz, C. & Davies, H.C. (2008) Far-upstream precursors of heavy precipitation events on the alpine south-side. *Quarterly Journal of the Royal Meteorological Society*, 134, 417–428.

Masato, G., Hoskins, B.J. & Woollings, T.J. (2012) Wave-breaking characteristics of midlatitude blocking. *Quarterly Journal of the Royal Meteorological Society*, 138, 1285–1296.

Mastrantonas, N., Herrera-Lormendez, P., Magnusson, L., Pappenberger, F. & Matschullat, J. (2021) Extreme precipitation events in the Mediterranean: spatiotemporal characteristics and connection to large-scale atmospheric flow patterns. *International Journal of Climatology*, 41, 2710–2728.

Messori, G., Caballero, R. & Gaetani, M. (2016) On cold spells in North America and storminess in western Europe. *Geophysical Research Letters*, 43, 6620–6628.

Mora, C. & Vieira, G. (2020) The climate of Portugal. In: Vieira, G., Zêzere, J.L. & Mora, C. (Eds.) *Landscapes and landforms of Portugal*. Cham: Springer International Publishing, World Geomorphological Landscapes, pp. 33–46.

Oertel, A., Pickl, M., Quinting, J.F., Hauser, S., Wandel, J., Magnusson, L. et al. (2023) Everything hits at once: how remote rainfall matters for the prediction of the 2021 north American heat wave. *Geophysical Research Letters*, 50, e2022GL100958.

Parker, T., Woollings, T., Weisheimer, A., O'Reilly, C., Baker, L. & Shaffrey, L. (2019) Seasonal predictability of the winter North Atlantic oscillation from a jet stream perspective. *Geophysical Research Letters*, 46, 10159–10167.

Pasquier, J.T., Pfahl, S. & Grams, C.M. (2019) Modulation of atmospheric river occurrence and associated precipitation extremes in the North Atlantic region by European weather regimes. *Geophysical Research Letters*, 46, 1014–1023.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2011) Scikit-learn: machine learning in Python. Machine learning in PYTHON.

Pino, D., Ruiz-Bellet, J.L., Balasch, J.C., Romero-León, L., Tuset, J., Barriendos, M. et al. (2016) Meteorological and hydrological analysis of major floods in NE Iberian Peninsula. *Journal of Hydrology*, 541, 63–89.

Ramos, A.M., Martins, M.J., Tomé, R. & Trigo, R.M. (2018) Extreme precipitation events in summer in the Iberian Peninsula and its relationship with atmospheric rivers. *Frontiers Earth Science*, 6.

Ramos, A.M., Trigo, R.M., Liberato, M.L.R. & Tomé, R. (2015) Daily precipitation extreme events in the Iberian Peninsula and its association with atmospheric rivers. *Journal of Hydrometeorology*, 16, 579–597.

Raveh-Rubin, S. & Flaounas, E. (2017) A dynamical link between deep Atlantic extratropical cyclones and intense Mediterranean cyclones. *Atmospheric Science Letters*, 18, 215–221.

Richardson, D.S., Cloke, H.L. & Pappenberger, F. (2020) Evaluation of the consistency of ECMWF ensemble forecasts. *Geophysical Research Letters*, 47, e2020GL087934.

Rocklin, M. (2015) Dask: parallel computation with blocked algorithms and task scheduling. In Python in science conference, 126–132. Austin, Texas.

Salgueiro, A.R., Machado, M.J., Barriendos, M., Pereira, H.G. & Benito, G. (2013) Flood magnitudes in the Tagus River (Iberian Peninsula) and its stochastic relationship with daily North Atlantic oscillation since mid-19th century. *Journal of Hydrology*, 502, 191–201.

Santos, M., Fragoso, M. & Santos, J.A. (2018) Damaging flood severity assessment in northern Portugal over more than 150 years (1865–2016). *Natural Hazards*, 91, 983–1002.

scikit-learn. (2023) Sklearn.cross_decomposition.PLSRegression. https://scikit-learn/stable/modules/generated/sklearn.cross_decomposition.PLSRegression.html

Shepherd, T.G., Boyd, E., Calel, R.A., Chapman, S.C., Dessai, S., Dima-West, I.M. et al. (2018) Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, 151, 555–571.

Simpson, I.R., Bacmeister, J., Neale, R.B., Hannay, C., Gettelman, A., Garcia, R.R. et al. (2020) An evaluation of the large-scale atmospheric circulation and its variability in CESM2 and other CMIP models. *Journal of Geophysical Research: Atmospheres*, 125, e2020JD032835.

Tuel, A., Schaefli, B., Zscheischler, J. & Martius, O. (2022) On the links between sub-seasonal clustering of extreme precipitation and high discharge in Switzerland and Europe. *Hydrology and Earth System Sciences*, 26, 2649–2669.

van der Wiel, K., Bloomfield, H.C., Lee, R.W., Stoop, L.P., Blackport, R., Screen, J.A. et al. (2019) The influence of weather regimes on European renewable energy production and demand. *Environmental Research Letters*, 14, 094010.

Wandel, J.L. (2023) *Representation of warm conveyor belts in sub-seasonal forecast models and the link to Atlantic-European weather regimes*. Karlsruhe, Germany: KIT Scientific Publishing.

White, R.H., Kornhuber, K., Martius, O. & Wirth, V. (2022) From atmospheric waves to heatwaves: a waveguide perspective for understanding and predicting concurrent, persistent, and extreme extratropical weather. *Bulletin of the American Meteorological Society*, 103, E923–E935.

Woollings, T., Hannachi, A. & Hoskins, B. (2010) Variability of the North Atlantic eddy-driven jet stream. *Quarterly Journal of the Royal Meteorological Society*, 136, 856–868.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Dorrington, J., Grams, C., Grazzini, F., Magnusson, L. & Vitart, F. (2024) Domino: A new framework for the automated identification of weather event precursors, demonstrated for European extreme rainfall. *Quarterly Journal of the Royal Meteorological Society*, 1–20. Available from: https://doi.org/10.1002/qj.4622