Received: 29 June 2023 | Revised: 3 April 2024 | Accepted: 22 April 2024

DOI: 10.1002/qj.4755

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Improving forecasts of precipitation extremes over northern and central Italy using machine learning

Federico Grazzini[1,2] | **Joshua Dorrington**[3] | **Christian M. Grams**[3] | **George C. Craig**[1] | **Linus Magnusson**[4] | **Frederic Vitart**[4]

[1]Ludwig-Maximilians-Universität, Meteorologisches Institut, Munich, Germany

[2]Arpae-SIMC, Regione Emilia-Romagna, Bologna, Italy

[3]Department Troposphere Research, Institute of Meteorology and Climate Research (IMKTRO), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[4]ECMWF, Reading, UK

Correspondence
Federico Grazzini, Meteorologisches Institut, Ludwig-Maximilians-Universität, München, 80333, Germany.
Email: federico.grazzini@lmu.de

Funding information
Deutsche Forschungsgemeinschaft; Helmholtz Young Investigator Group 'Sub-Seasonal Predictability: Understanding the Role of Diabatic Outflow (SPREADOUT)', Grant/Award Number: VH-NG-1243

Abstract

The accurate prediction of intense precipitation events is one of the main objectives of operational weather services. This task is even more relevant nowadays, with the rapid progression of global warming which intensifies these events. Numerical weather prediction models have improved continuously over time, providing uncertainty estimation with dynamical ensembles. However, direct precipitation forecasting is still challenging. Greater availability of machine-learning tools paves the way to a hybrid forecasting approach, with the optimal combination of physical models, event statistics, and user-oriented post-processing. Here we describe a specific chain, based on a random-forest (RF) pipeline, specialised in recognising favourable synoptic conditions leading to precipitation extremes and subsequently classifying extremes into predefined types. The application focuses on northern and central Italy, taken as a testbed region, but is seamlessly extensible to other regions and time-scales. The system is called MaLCoX (Machine Learning model predicting Conditions for eXtreme precipitation) and is running daily at the Italian regional weather service of ARPAE Emilia-Romagna. MaLCoX has been trained with the ARCIS gridded high-resolution precipitation dataset as the target truth, using the last 20 years of the European Centre for Medium-Range Weather Forecasts (ECMWF) reforecast dataset as input predictors. We show that, with a long enough training period, the optimal blend of larger-scale information with direct model output improves the probabilistic forecast accuracy of extremes in the medium range. In addition, with specific methods, we provide a useful diagnostic to convey to forecasters the underlying physical storyline which makes a meteorological event extreme.

KEYWORDS

extreme precipitation, hybrid forecast model, large-scale precursors, machine learning, northern Italy, predictability, warning chain

1 | INTRODUCTION

Italy is among the European nations most exposed to torrential rainfall and flash flooding, due to its geographical conformation and climatological characteristics (Grazzini, 2021). Every year these weather hazards produce huge costs and deadly consequences. The correct prediction of intense meteorological phenomena is one of the main objectives of operational weather services, and this task is even more relevant today with the rapid progression of global warming which amplifies extremes (Seneviratne *et al.*, 2021; Tramblay & Somot, 2018). Numerical weather prediction (NWP) models have improved continuously over time; however, providing accurate, quantitative precipitation forecasts remains challenging. Precipitation is an intermittent and complex phenomenon, often characterised by high spatial variability. Direct predictability of this field is limited compared to the predictability of large-scale circulation patterns in which precipitation events are developing. The broader usage of machine learning in many sectors characterised by high-dimensional data raises the question of whether machine learning can effectively be used for advanced postprocessing of meteorological fields, thereby helping to recover the loss of predictability inherent to surface fields characterised by high variability. Research in this direction is rapidly developing with studies showing how machine learning can be applied to advanced postprocessing of precipitation direct model output (de Sousa Araújo *et al.*, 2022; Espeholt *et al.*, 2022; Frnda *et al.*, 2022; Whan & Schmeits, 2018), or testing the inclusion of large-scale components such as atmospheric rivers (Chapman *et al.*, 2022). The link between extreme precipitation events (EPEs) over the Mediterranean area and large-scale atmospheric flow patterns has long been studied and is consolidated in the literature (Grazzini, 2007; Grazzini & Vitart, 2015; Martius *et al.*, 2008; Rudari *et al.*, 2005). Recently Mastrantonas *et al.* (2022) has shown that inferring EPE probability from predefined specific weather patterns outperforms precipitation output in the medium range, extending the forecasting horizon of the model up to three days in many Mediterranean locations. Starting from this perspective, we consider whether machine-learning methods can be used operationally to improve extreme precipitation predictions in Italy, combining large-scale dynamical predictors with precipitation direct model output. In this work, we describe a random-forest (RF)-based postprocessing chain, called MaLCoX (Machine-Learning model predicting Conditions for eXtreme precipitation), specialized in recognizing favourable synoptic and large-scale conditions leading to precipitation extremes and subsequently classifying the categories proposed by Grazzini *et al.* (2020a). The focus is on northern and central Italy, taken as a preliminary testbed region. The choice of predictors is based largely on previous work by the authors (Grazzini *et al.*, 2020a, 2021). In addition, we include non-local predictors: spatial composites of Euro-Atlantic anomaly patterns in the days preceding Italian EPEs, as described in a recent companion paper (Dorrington *et al.*, 2023). MaLCoX – composed of two modules that detect and classify precipitation extremes respectively – has been implemented semi-operationally at ARPAE-SIMC using as predictors the available fields the institute is receiving in real time from the European Centre for Medium-Range Weather Forecasts (ECMWF) dissemination stream. The goal is to set up an innovative 'warning-bell chain' complementing existing forecasting products, such as direct probabilities and the extreme forecast index (Tsonevsky, 2015). This type of hybrid modelling is relatively novel, with promising application in the field of extreme-event early warnings, potentially anticipating major events at a time-scale of several days, compared to the current standard of 48 hours. To our knowledge, this is the first documented machine-learning application for the prediction of extreme precipitation targetted at the medium range. The paper is organised as follows: in Section 2 we describe the datasets, algorithms, architecture and predictors. In Section 3 we show the results of the comparison against direct model output, while in Section 4 we illustrate the new MaLCoX forecasting tools applied to a recent case study. Finally, in Section 5, we draw our conclusions.

2 | DATA AND METHODS

To introduce the geographic area and its relation with synoptic predictors used in the model, in Figure 1 we show the synoptic situation associated with storm Alex, chosen as an exemplary case. We are not going to discuss the forecast of this event but use it to illustrate the concept of non-local predictors referred to later. A trough is deepening over western Europe, at the leading edge of an incoming Rossby wave packet. At the same time, strong integrated water vapour transport (IVT) is connecting the upstream trough over the US east coast and the deepening trough over Western Europe, forming an atmospheric river on the northern side of the Atlantic ridge. This is a typical and recurrent synoptic situation associated with the strongest EPEs, as discussed in the work of Sioni *et al.* (2023) where the detailed evolution of the two most severe EPEs ever recorded over northern Italy is discussed. On 2 and 3 October 2020 a large EPE was recorded in the area of interest, inside the green box in Figure 1, with

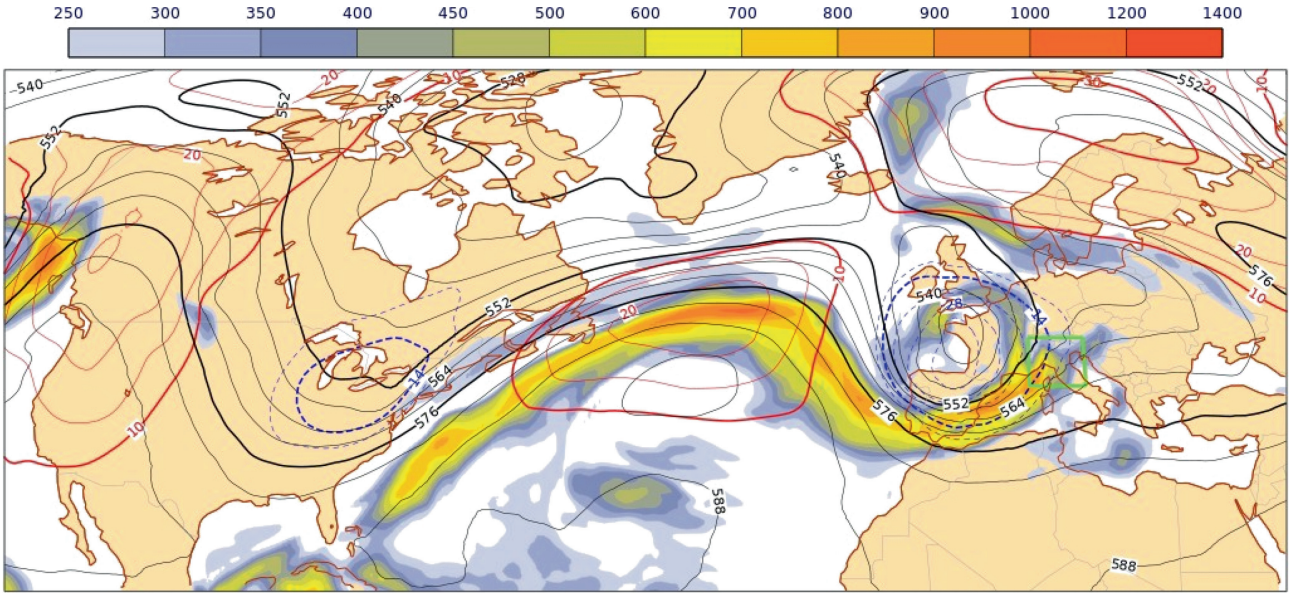

FIGURE 1 Synoptic situation on 02-10-2020 1200 UTC associated with storm Alex. An Extreme Precipitation Event (EPE) occurred inside the test region depicted with the green box between 2 and 3 October 2020. The map displays the geopotential height at 500 hPa which describes the middle-level flow (solid lines) and the associated magnitude of the instantaneous vertically integrated water vapour flow shaded according to the scale above in $kg \cdot s^{-1} \cdot m^{-1}$. Blue dashed isolines and red solid lines show the corresponding 500-hPa anomaly for the seasonal climatological average which is then projected onto the non-dimensional Z500 non-local index which in this flow configuration is 2.4. [Colour figure can be viewed at wileyonlinelibrary.com]

significant damages and flooding in the western Alps, Italy and France, as described in Magnusson *et al.* (2021).

2.1 | Event definition and target variable

Our definition of EPE consists of an extreme precipitation event occurring within the area indicated by the green rectangle in Figure 1. The condition of extreme precipitation yes/no, is obtained through thresholding based on the observational ARCIS dataset. ARCIS is a high-resolution ($5\,km \times 5\,km$) gridded precipitation dataset obtained by the spatialisation of a high-density surface observation network of 11 Italian regions plus several stations of adjacent Alpine regions, described in Pavan *et al.* (2019). Gridded data cover northern and central Italy at daily resolution from 1 January 1961 up to real time. Each day since then has been labelled with EPE yes if the aggregated daily precipitation of one or more of the 94 warning areas into which northern and central Italy are subdivided [see Grazzini *et al.* (2020b) for details] exceeds the 99th percentile of its wet-days climatology of the recent climate period (1991–2020) and the sum of the area above the 99th percentile is greater than $1000\,km^2$. We focus on medium- to large-area extremes, filtering out very localised downpours. The obtained tabular time series of dates with EPE yes/no is our ground truth or target variable. In the recent period (1991–2022), we have observed 782 EPE days, 24.4 ± 6.9 EPEs each year, with a seasonal distribution ranging from about 5% in winter (DJF), spring (MAM) and summer (JJA) and ca 12% in autumn (SON). An equivalent time series of precipitation forecast output, used as a benchmark, is obtained by applying the same thresholding rules to the direct model output daily precipitation, using the 24-hour short-term precipitation forecast from ERA5 reanalyses as base climatology for computing the 99th percentile thresholds of each warning area. This results in lower thresholds, in absolute terms, for the direct model outputs to raise an EPE yes.

2.2 | Machine-learning algorithm description

MaLCox consists of a pipeline of Python modules containing machine-learning models and estimators from the scikit-learn library (Pedregosa *et al.*, 2011). At the core of MaLCoX's algorithms lies the RF method (Breiman, 2001). Random forest is an estimator that fits many decision tree classifiers on various randomly drawn subsamples of the dataset instances and predictors. It uses averaging on the trees to improve the predictive accuracy and control overfitting. Prediction, which can be categorical, probabilistic or continuous regression, is made by aggregating the individual predictions of the ensemble of decision trees. Random forests are already proven to be useful in the context of

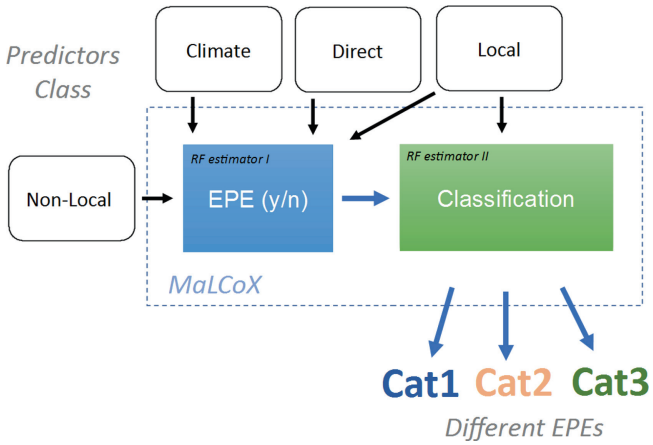

FIGURE 2 Schematic of MaLCoX architecture. EPE, Extreme Precipitation Event. [Colour figure can be viewed at wileyonlinelibrary.com]

severe weather detection (Hill *et al.*, 2020) and are straightforward to implement and optimise. The architecture of MaLCoX can be divided into two major blocks, both using RF estimators, schematised in Figure 2. The first module (EPE module) is designed to predict the probability of an EPE and eventually the total volume of rain (Vol) and the EPE area extension (EPEarea). A second block (Classification module) classifies EPEs into categories defined in Grazzini *et al.* (2020a). Besides different predictors, the two modules differentiate themselves by their different levels of tuning. While the task of EPE classification is relatively simple (we use the default hyperparameters) since it works on a homogeneous set of dates containing only EPE days, the task of the first module (EPE yes/no) is more challenging and requires some specific hyperparameter customisation accounting for EPE rarity. RF produces individual trees from a bootstrap sample of the training data. In learning extremely imbalanced data, as in our dataset, there is a significant probability that a bootstrap sample contains few or even none of the minority class, resulting in an overall poor performance in predicting the positive event (EPE yes).

To alleviate this problem, in the EPE module, we set the option of having balanced subsamples which automatically adjust weights inversely proportional to class frequencies in the bootstrap sample for every tree grown. The other problem to face was avoiding overfitting to improve generalisation. Instead of changing specific hyperparameters, we used cost complexity pruning to control the size of RF trees and prevent overfitting. This pruning technique is parameterised by the cost complexity parameter (ccp_alpha) with values greater than zero progressively increasing the number of nodes pruned. We find the optimal value by testing different values recursively (via GridSearch) using cross-validation on the training dataset. For our application, the choice of ccp_alpha = 0.001 is the best compromise which maximises the validation scores. While the classification module is independent of lead time, the EPE yes/no module has a different RF model for each forecast step, with a different training dataset and the same hyperparameters. We tested also having the predictors and hyperparameters change with lead time, trying a recursive feature elimination with cross-validation. Still, the results were comparable using the same predictors and letting the model decide how to use them. We opted then for this latter choice since having the same number of predictors at all lead times is preferable in terms of the interpretability of the results. The EPE (yes/no) module also includes two additional RF regressor models used to compute the expected rain volume and EPE area, fitted with the same predictors of the classification models. Finally, we complemented the software modules, with the Shapley additive explanations library (Lundberg *et al.*, 2017) specifically designed to help explain the outputs of machine-learning models. It allows us to effectively visualise the contribution of each feature for a particular prediction. The prototype of MaLCoX has been put in a preoperational and testing phase, running daily on ARPAE servers since September 2022.

2.3 | Training and test dataset

On average there are only 24 EPE days for each year, so we need a sufficiently long dataset for the training period to build meaningful statistics. Another requirement, since we are dealing with forecasts of events occurring in different years, is that only small changes in model performance occur over time. These requirements restricted the choice to the ECMWF reforecast, which has a coarse horizontal resolution (18 km), for the investigated period, compared with high-resolution 9 km (HRES) operational runs, but is more stable in terms of model changes. This system is composed of an 11-member ensemble running biweekly for 46 days, with the latest Integrated Forecasting System of ECMWF (IFS) cycle, on the same initial date (day and month) for the previous 20 years (Vitart *et al.*, 2019). We take the 20-year reforecast sets produced on every available date from 2019 to 2021, accumulating statistics over different model cycles. In case of duplicate dates, we keep only the most recent forecast. Another restriction is represented by the availability of the IVT fields. This fundamental predictor, which only became available in June 2018, comes from model cycle 45r1 of the ECMWF IFS system output. To reduce the already large data transfer, we train our model only on the control (CTRL) member of the ensemble reforecast. We also tried training based on the reanalysis but as we want to account for the mean forecast

TABLE 1 Table showing the list of predictors.

Variable	Description	Units	RF model	Class
fcst_IVTmag	IVT-normalised anomaly index lead times from 0 to −2 days		EPE (y/n)	Non-Local
fcst_Z500	Geopot-normalised anomaly index at 500 hPa lead times from 0 to −2 days		EPE (y/n)	Non-Local
fcst_V850	850 hPa meridional wind normalised anomaly index at 850 hPa lead times from 0 to −2 days		EPE (y/n)	Non-Local
IVTe	Daily mean of zonal component of integrated water vapour transport	$kg{\cdot}s^{-1}{\cdot}m^{-1}$	EPE (y/n)	Local
IVTn	Daily mean of meridional component of integrated water vapour transport	$kg{\cdot}s^{-1}{\cdot}m^{-1}$	EPE (y/n)	Local
TCWV	Daily mean of total column water vapour	$kg{\cdot}m^{-2}$	EPE (y/n)	Local
MSLP	Daily mean of mean sea level pressure	hPa	EPE(y/n)	Local
V_{olf}	Daily volume of rain over the target domain	m^3	EPE (y/n)	Direct
Juld	Day of the year (Julian day)		EPE (y/n)	Climate
Θ_{e850}	Daily mean of equivalent potential temperature at 850 hPa	K	Classification	Local
$\Delta\Theta_{e500-850}$	Daily minimum of delta Θ_e (500–850) hPa	K	Classification	Local
Θ_{pv2}	Daily mean of Θ on dynamical tropopause (PV = 2)	K	Classification	Local
i_{audmax}	Daily maximum of convective adjustment time-scale	hr	Classification	Local
$CAPE_{dmax}$	Daily maximum of CAPE	$J{\cdot}kg^{-1}$	Classification	Local

Note: Predictors are subdivided according to their type and usage in the two-step blocks of the MaLCoX model.

error of the predictors, which is lead-time-dependent, we opted for training on the reforecast. After removing duplicate reforecast dates, the training dataset consists of ca 5282 days with the number of EPEs slightly changing with forecast steps (357 ± 12). Forecast steps are spanning from +24 hours (D1) to +240 hours (D10). The test dataset is assembled on the same set of predictors extracted from the ECMWF HRES operational daily runs from July 2018 (the first available cycle with IVT as output) until 2022. All the verification dates already included in the training dataset, about 10% of the original available HRES dates, are being excluded from the test dataset to avoid any overlap. At the end, the number of days present in the test dataset is 898, of which 58 ± 4 EPEs, depending on the forecast lead time.

2.4 | Non-local, local, direct and climate predictors

We define four sets of predictors: non-local, local, direct and climate predictors. Table 1 shows the full list of predictors subdivided according to their class and usage in the two modules of the MaLCoX model. The majority of predictors is used in the first step, in the EPE module, used to detect EPE yes/no. The remaining five predictors are used to classify the type of event in three main EPE categories: frontal or orographic uplift of moist statically stable flow (Cat1), stronger frontal uplift of a neutrally moister/warmer stable flow with embedded convection (Cat2), and thermally forced deep convective ascent (Cat3), as proposed in Grazzini *et al.* (2020a). For the training, all the predictors are obtained from the control member of the reforecast, at six-hour intervals up to 15 days and aggregated at daily resolution (last five days not used at the moment). In the test period and in real-time application the predictors are taken instead from the HRES ECMWF forecast every six hours up to day 10 and aggregated at daily resolution, benefitting from higher resolution. This difference between training (low resolution) and testing (high resolution), which makes the comparison unfavourable for MaLCoX, will disappear in the current (after June 2023) IFS model cycles, when the ensemble prediction forecast (ENS) and HRES will have the same resolution and reforecast resolution will also be upgraded accordingly. Let us introduce the first class of predictors: the non-local predictors. This class of predictors is based on the prior systematic identification of precursor patterns, as lagged-anomaly composites of large-scale variables prior EPEs, defined in ERA5 data, following the approach of Dorrington *et al.* (2023) and using the associated *Domino* Python package. Patterns are masked based on the statistical significance, amplitude and spatial extent of anomalies, and then standardised. Finally, time-evolving indices are computed from the

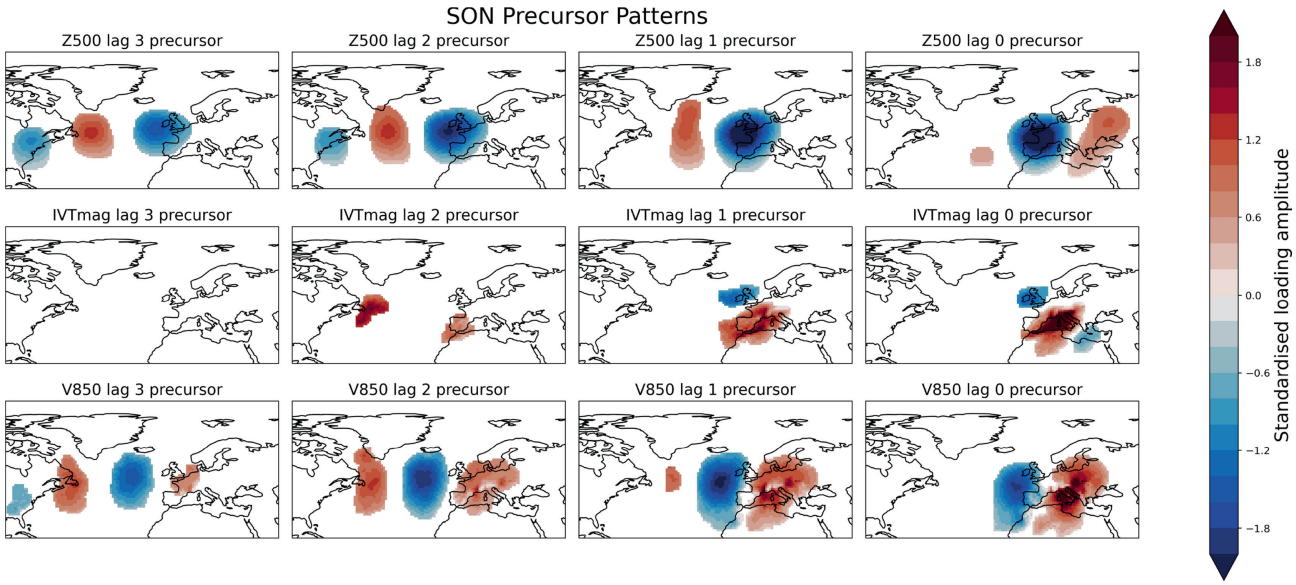

FIGURE 3 Standardised composites of non-local precursors (top row: Z500; middle row: integrated water vapour transport (IVT) magnitude; bottom row: V850) from three days before (lag 3) up to the day (lag 0) of Extreme Precipitation Events (EPEs) occurring in the September–October–November (SON) season. [Colour figure can be viewed at wileyonlinelibrary.com]

spatial patterns for selected variables (Z500, V850 and IVT magnitude) as the projection (scalar product) of the daily deseasonalised field anomaly onto the precursors patterns. In practical terms, the non-local indices summarise the spatial similarity of current synoptic conditions to the typical precursor patterns associated with EPEs n days before (with n from 0 up to 5 days before the event) over the Euro-Atlantic area. Precursor patterns are computed on a seasonal basis accounting for the change of large-scale dynamics during the year. In Figure 3 we show an example of precursors patterns for autumn (SON) EPEs.

The wave pattern associated with EPEs is very evident in the three respective atmospheric variables at day 0 (the same day as the EPE), and it remains coherent while shifting west in the days before the event, up to day 3 in Z500 and V850 and up to day 2 in IVT (Figure 3). Coherency and significance depend on the season, increasing in SON and DJF and decreasing in MAM and JJA (not shown). In the current version of the system, we use non-local precursors up to day 2, to have significant amplitudes for all variables. This means that for each valid date, we have up to three non-local indices for each variable, leading to nine independent non-local indices. For example, the Z500 indices are computed projecting the instantaneous forecast anomaly validating on the composite two days before (lead2), one day before (lead1) and on the day of the event (lead0). So the non-local predictors of each day are not only a function of fields valid at D0 but also reflect the forecast for the days before; they are non-local both in space and time. Qualitatively, it is possible to see in Figure 1 how the anomalies of the wave associated with storm Alex, including high IVT values, match the composites of Figure 3. The resulting non-local indices at lag0 are 2.4 for Z500, 3 for V850 and 3.9 for IVTmag. These values are very high and correspond respectively to 2.4, 3 and 3.9 standard deviations from the mean values observed for autumn EPEs.

Local predictors are a set of thermodynamic and dynamic variables, describing the circulation at the local scale, averaged temporally at the daily resolution, and spatially over the green box of Figure 1. The choice of variables has been made through a combination of established variables described in previous work of the authors (Grazzini *et al.*, 2020a) plus the addition of Θ on the potential vorticity (PV) surface, at a level of the dynamical tropopause (PV = 2) as a tracer of upper-level wave activity. The choice of this variable compared with PV on Θ surfaces instead, as proposed in Grazzini *et al.* (2021), is mostly motivated for practical reasons due to the availability of this field in the operational dissemination already in place at ARPAE. We introduced also a direct model predictor, the daily volume of rain forecast (Volf) over the entire target area, as a feature to convey the explicit prediction of the model. An attempt without direct model output showed worse performance, especially in the short-term forecast. Finally, we introduced the climate predictors class, which in the current configuration is composed only of the day of the year. This variable provides direct information on observed EPE frequency which shows a marked seasonal cycle (see fig. 2 of Grazzini *et al.* (2020a)).

In addition to this simple information, this variable allows MaLCoX to modulate the importance of other

predictors according to the time of the year. We tested the introduction of other slow-varying predictors in the climatological class, accounting for example for the observed warming trend like sea surface temperature (and its deseasonalised anomaly) averaged on the scale of the Mediterranean basin. We found that the Mediterranean averaged sea surface temperature (SST) did not add predictive power to the model and therefore we discarded it. A possible explanation is that SST variability occurs at longer time-scales compared to synoptic disturbances. According to our predictor's correlation matrix (not shown here), there is no significant correlation ($r = 0.03$) between EPE area or rain volume (Vol) and SST (full value and anomaly) averaged over the Mediterranean Sea. This does not mean that SSTs are not influential in EPE genesis, they are indeed important for heat and momentum fluxes. Simply there is no covariance because even if SSTs are high, or anomalous, they will stay high also when the phenomenon is not occurring (especially in summer). Probably latent heat flux (not tried yet) would be a better indicator, with variability on the same temporal scales of EPEs. In addition, water vapour sources from remote areas are sometimes more important for EPE development, as confirmed by other recent studies (Duffourg & Ducrocq, 2011; Khodayar *et al.*, 2021; Khodayar *et al.*, 2022), so that the direct Mediterranean Sea contribution for some EPEs is less important, like in winter.

3 | MODEL EVALUATION

In this section, we discuss the model performance over a large sample of cases. As a main score to evaluate the performance of MaLCoX, against the direct model output of the ECMWF HRES precipitation taken as a benchmark, we used the average precision score (AP) available from the scikit-learn library. Receiver Operator Characteristic (ROC) curves are commonly used to present results for binary classification in machine learning. However, when dealing with highly skewed and imbalanced datasets, like in our application, the AP, which is the area under the precision–recall curve, gives a more informative picture, as discussed in Davis and Goadrich (2006). Precision (P) is a metric that quantifies the number of true positive predictions of minority class (EPE yes) divided by all positive (true plus false), while recall (R), also known as sensitivity, is the fraction of true positive divided by true positives plus false negatives.

AP is calculated as follows:

$$\mathrm{AP} = \sum_n (R_n - R_{n-1}) P_n \tag{1}$$

where P_n and R_n are the precision and recall at the nth threshold.

In addition to AP, we use the Brier score to assess the improvement of the inclusion of the non-local predictors. Brier score or Brier score loss is a negatively oriented score, the smaller the better, which measures the mean squared difference between the predicted probability and the actual outcome.

3.1 | Feature importance

Before focusing on the scores, it is useful to discuss the contribution of each feature class to the model. One of the main reasons for conceiving MaLCoX as a hybrid model is that we can study the relative importance of direct vs large-scale predictors for the correct prediction of EPEs at different forecast ranges. This can be done using the feature importance of the RF estimator. As might be expected, in the first three days of the forecast the direct model output of precipitation (Volf) is the dominant predictor. At short ranges, direct rainfall prediction from state-of-the-art physical models is very well correlated with observed precipitation, especially if temporally and spatially aggregated. As we enter the medium range, precipitation errors grow rapidly and other synoptic variables become more relevant in discriminating days with EPE yes/no. Namely, IVTn (Table 1) is the second most important variable after Volf until D+3, while from D+4 onward the IVTmag (lead0) surpasses IVTn in importance (not shown). To synthesise the relative contribution of the different features, we aggregated the predictors according to their type, indicated in the last column of Table 1, and show how relative importance changes with forecast lead time. This is shown in Figure 4, where we observe a crossing point between the feature importance of the aggregated predictors by type around D+4. From this point onward, the overall effect of non-local predictors (representing the larger-scales) becomes predominant over the other types. Direct model prediction rapidly decreases in importance, reducing to almost climatological value by D+9, while local predictors' importance remains almost constant, mostly supported by the IVTn contribution. The constant increase of climate predictors from D+5 onward is also notable.

3.2 | Performance in the train and test dataset

Once we defined the metrics, we tested two MalCoX versions using different predictors. RF ALL contains all predictors, while RF LOC excludes the non-local predictors.

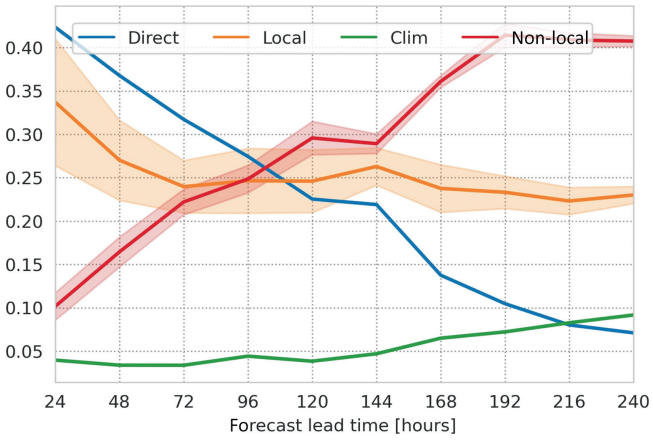

FIGURE 4 Random-forest feature importance evolution with forecast lead time [hours], aggregated by class of predictors according to Table 1. [Colour figure can be viewed at wileyonlinelibrary.com]

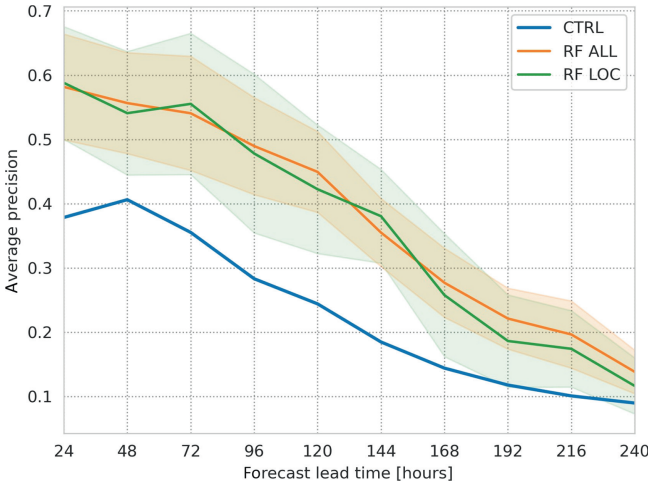

FIGURE 5 Average precision score computed on the training dataset (5282 days, 357 with extreme precipitation events). Scores averaged over five cross-validation folds. [Colour figure can be viewed at wileyonlinelibrary.com]

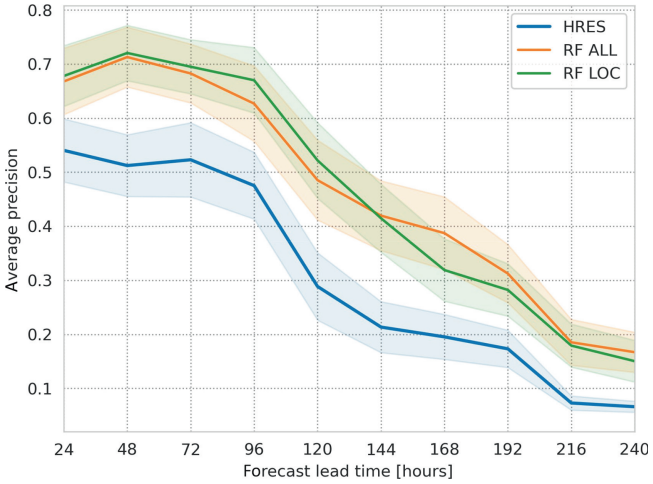

FIGURE 6 Average precision score computed on the test dataset (on average 898 days, 58 with extreme precipitation events). [Colour figure can be viewed at wileyonlinelibrary.com]

The difference between the models is informative as to the importance of non-local predictors. The performance is first assessed on the training dataset with a cross-validation procedure (CV), in which the training set is split into five smaller folds used to compute the validation metrics after being trained on the remaining folds. The result of the probabilistic prediction of the two MaLCoX versions for the train and test are also compared respectively against the categorical prediction obtained from the control forecast (CTRL), shown in Figure 5, and later the HRES forecast in Figure 6. As shown in Figure 5, MaLCoX performance evaluated in the training dataset is always much better than CTRL with the largest gain in the medium range between +96 and +144 hours lead time. Both versions, ALL and LOC, are showing a similar level of skill up to seven days (forecast lead time +168 hours). After, RF ALL tends to perform marginally better, hinting at a positive role of the non-local features for the prediction of EPEs at longer ranges, although there is still a large overlap of confidence intervals which did not allow us to draw firm conclusions.

We apply the same metrics to the test dataset, not seen during the training. The test dataset contains a smaller sample size but is still relevant for the significance of the statistics with 898 days on average and 57 EPEs. To estimate the mean and standard deviation of the scores we applied a 100-time resampling procedure with replacement. From Figure 6 we see that the HRES precision score is higher than the CTRL score due to higher resolution and accuracy in predicting EPEs. Better skill of HRES precipitation has a positive influence also on MaLCoX skill which in general is higher compared with results obtained in the training dataset. RF models, even if trained on the CTRL, continue to show higher precision throughout the forecast period compared to the HRES categorical precipitation forecast. This advantage is greatest in the medium range, where the skill gain of using MaLCoX instead of HRES precipitation is equivalent to about three days; the skill of MaLCoX (ALL) at D7 forecast is almost the same as that of HRES direct model output at D+4. Secondly, we see a similar behaviour of the two model configurations observed in the training dataset. RF ALL shows an even larger advantage, although not significant, compared to RF LOC at forecast lead time +144 hours. It should be noted, however, that despite AP being an appropriate score there is a disparity in comparing a probabilistic forecast with a categorical benchmark. A more fair comparison should be done against the probabilistic prediction

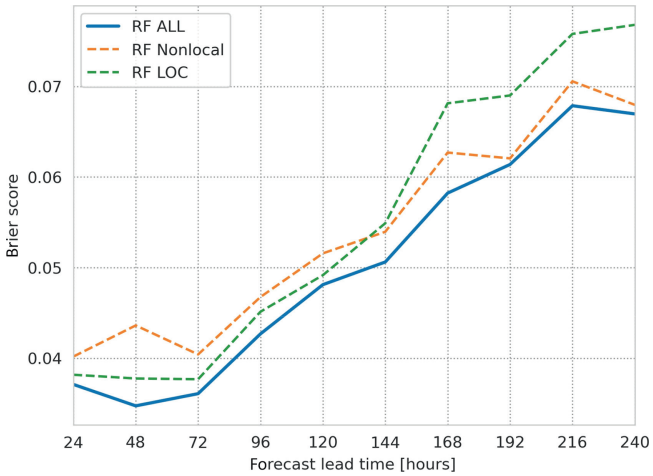

FIGURE 7 Brier score of different Extreme Precipitation Event (EPE) models containing different sets of predictors: ALL, only Local (green curve), only Non-local (orange curve). [Colour figure can be viewed at wileyonlinelibrary.com]

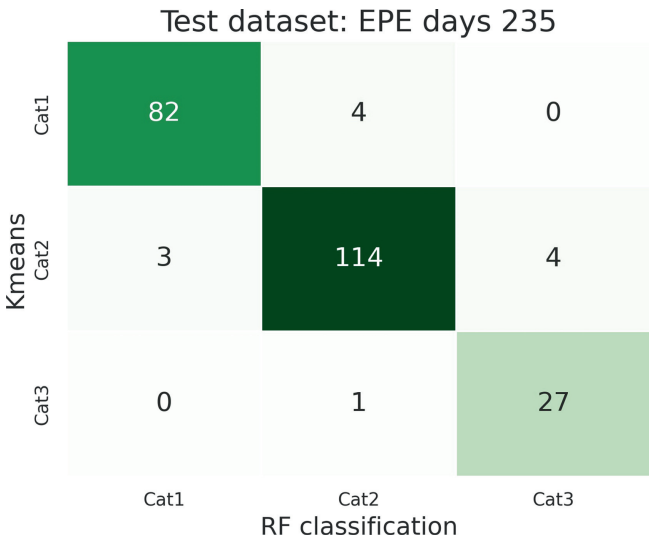

FIGURE 8 Confusion matrix showing the accuracy of prediction of the Extreme Precipitation Events (EPEs) classification module over the test dataset containing 235 EPEs. [Colour figure can be viewed at wileyonlinelibrary.com]

obtained from the full ECMWF ensemble over a long time window, planned for a more comprehensive verification in a future study.

To investigate further the relative importance of the non-local vs local predictors we consider also different skill metrics in the test dataset, among those the Brier score. This skill metric proved particularly effective in highlighting the relative role of different types of predictors that contribute to keeping the performance of MaLCoX consistently higher than the precipitation model output. In Figure 7 we compare the Brier score, which is negatively oriented, of RF ALL predictors against RF LOC and a further version, denominated RF Non-local, which shows only the contributions on non-local predictors. It is interesting to see that the Brier score of RF ALL, up to lead time +144 hours, is very similar to the Brier score of RF LOC (green curve), while after, in the late medium–late range, the level of skill is the same of RF non-local (orange curve). This behaviour seems to reinforce the hypothesis, coming from the feature importance analysis, that non-local predictors contribute more than local predictors from D6 onward.

Finally, we show a measure of the accuracy of the EPE classification model in assigning EPEs (in the forecast) at one of the defined three categories Cat1, Cat2 and Cat3. Note that, differently from the RF models used in the EPE yes/no block, the RF EPE classifier is trained on the analysis fields (1991–2022) using only days where EPEs were observed and a perfect prog approach. Assuming perfect predictors, the skill of the RF category classifier is very good in reproducing the categories obtained by Kmeans clustering, our ground truth, based on the same predictors [see Grazzini *et al.* (2020a) for a description of the clustering]. The confusion matrix in Figure 8 shows that about 87% of the predictions in Cat3 are correct, 95% in Cat2 and 98% in Cat1. The test dataset is obtained from a random sampling of 30% of days in 1991–2022, excluded by the training period.

4 | CASE STUDY: 15 DECEMBER 2022

Besides the verification statistic presented above, in this section we practically demonstrate the usage of MaLCoX, discussing one EPE event, which occurred after MaLCoX's implementation at ARPAE in September 2022. The EPE occurred in December 2022 with three warning areas exceeding their respective 99th percentile of daily precipitation (Figure 9). Localised floods were observed in northern Tuscany due to continuous heavy stratiform rain, enhanced by orographic uplift, consistent with Cat1 classification. The city of Pistoia was particularly hit with 110 mm of rain falling in 18 hours. Among the ranking of past EPEs, this can be considered moderate intensity, and small area (see Figure 12 for comparison with other events). The ground effects were limited due to prevailing antecedent drought conditions, which resulted in good drainage in the mountain basins. Aside from the severity of the ground effects, this is an interesting exemplary case to illustrate how the forecasting system works even with relatively small-amplitude events and to introduce some new forecast tools created to display the information content available with MaLCoX.

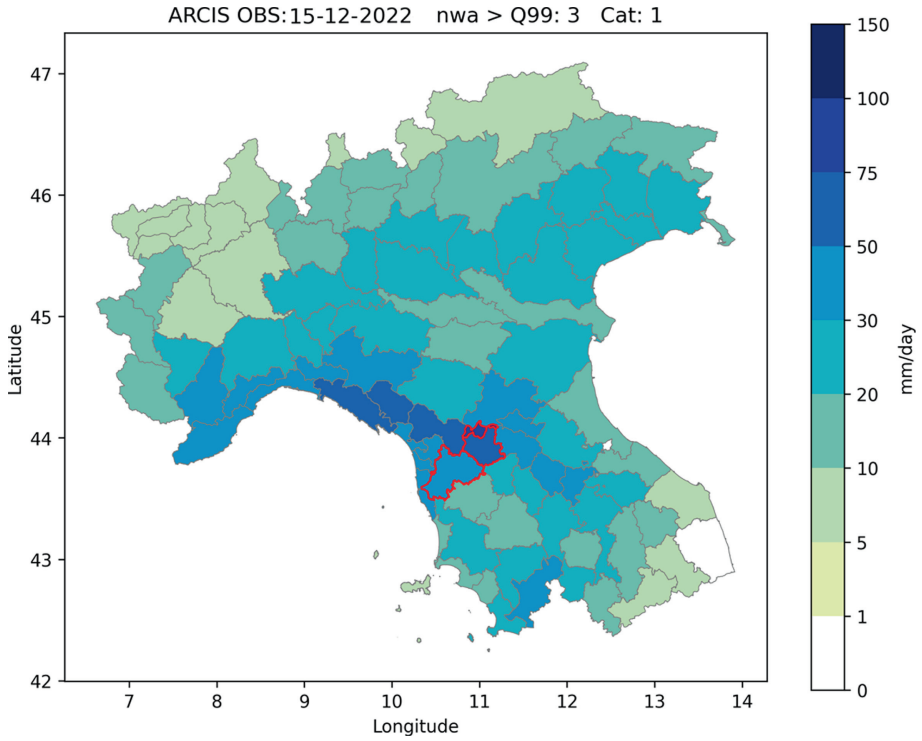

FIGURE 9 Twenty-four-hour observed precipitation during 15-12-2022 from the ARCIS dataset. Color shading shows the precipitation accumulated between 00 and 24 hr (UTC) spatially averaged over warning areas. In the three warning areas with red outlines, the observed daily rainfall exceeded the 99th percentile of the daily accumulated precipitation. [Colour figure can be viewed at wileyonlinelibrary.com]

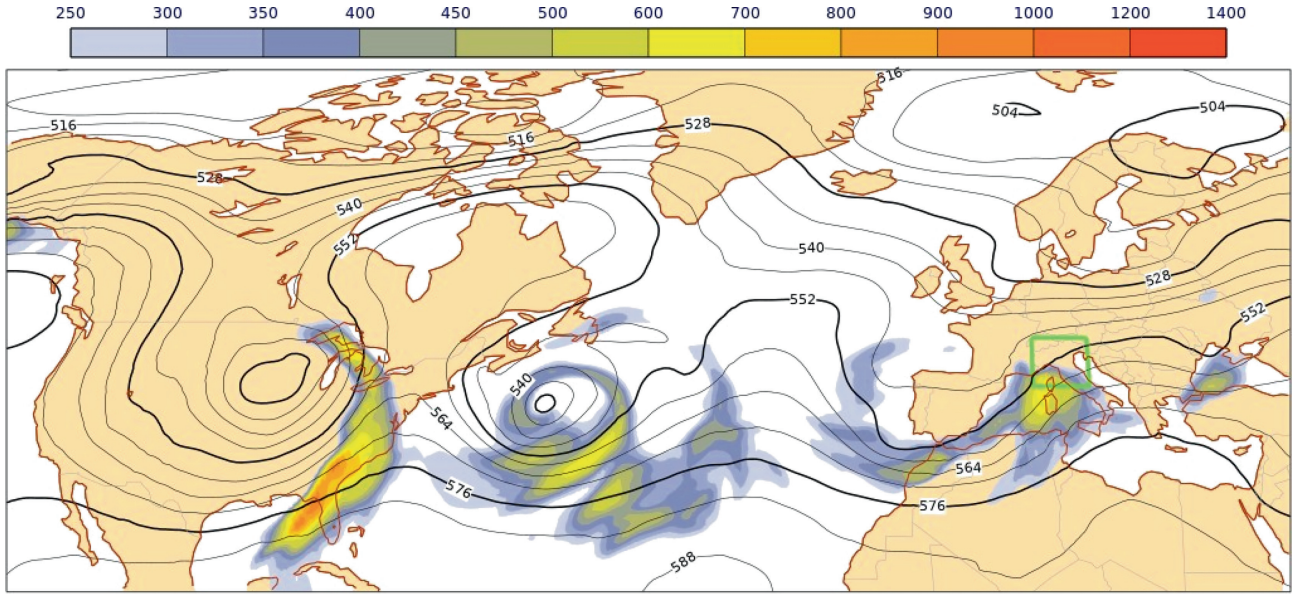

FIGURE 10 Synoptic situation on 15-12-2022 1200 UTC. An Extreme Precipitation Event (EPE) occurred inside the test region depicted with the green box with floods in northern Tuscany. The map displays the geopotential height at 500 hPa (solid lines) and the associated magnitude of the instantaneous vertically integrated water vapour flow shaded according to the scale above in kg·s $^{-1}$·m $^{-1}$. [Colour figure can be viewed at wileyonlinelibrary.com]

The synoptic situation associated with this case study is shown in Figure 10. The EPE was associated with an amplifying Rossby wave centred over the Iberian peninsula, moving east. Further upstream there are regions of enhanced water vapour downstream of two other troughs, flowing in a low-latitude zonally elongated region of high geopotential gradient. This configuration favours continuous replenishment of water vapour from the Atlantic to the Mediterranean basin.

In Figure 11 we show the EPE monitor display of two subsequent MaLCoX forecasts. This is our first alarm bell, showing the probabilistic output for an EPE in a 10-day forecast horizon. MaLCoX probability is displayed and compared against the corresponding probability obtained

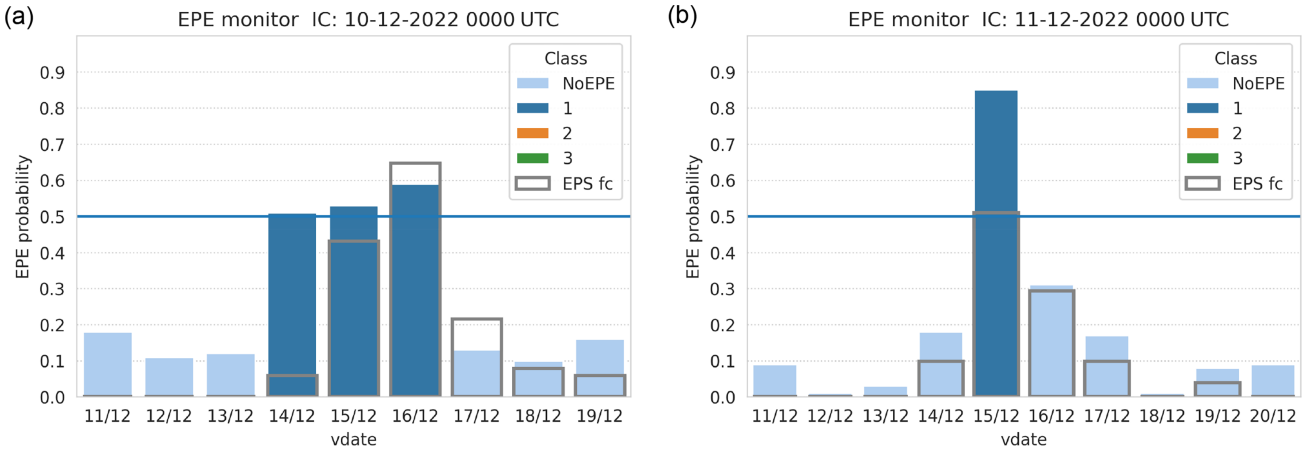

FIGURE 11 Extreme Precipitation Event (EPE) monitor tool used to display MaLCoX predictions daily. The X-axes shows the forecast valid dates (day/month). The bars are showing the EPE probability computed by MaLCoX, coloured according to the predicted category (cyan, no EPE; blue, Cat1; orange, Cat2; green, Cat3). The probability obtained by processing the direct model output of the ensemble prediction system forecast members (empty grey bars) is also shown for comparison. (a) 10 day forecast issued 10 December 2022. (b) 10 day forecast issued 11 December 2022. [Colour figure can be viewed at wileyonlinelibrary.com]

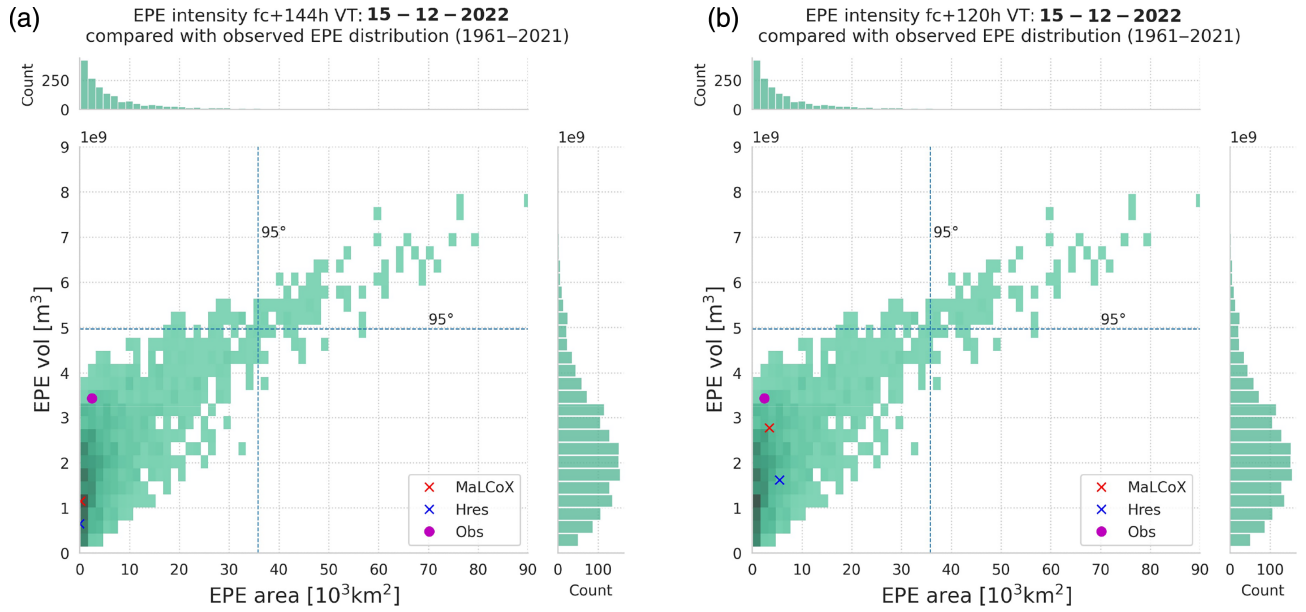

FIGURE 12 Total volume of rain and area extension of the extreme precipitation events (EPEs) predicted by MaLCoX and IFS HRES. The shading of the cells represents the frequency distribution of rain volume and area of previous EPEs observed between 1961 and 2022. Dashed lines divide the plane in areas above or below the 95th percentile of the respective EPE distribution. For comparison, the violet dot shows the observed value. (a) EPE rain volume prediction issued 2022-12-10 and valid for 15 December 2022. (b) EPE rain volume prediction issued 11 December 2022 and valid for 15 December 2022. [Colour figure can be viewed at wileyonlinelibrary.com]

by ENS precipitation direct model output; the colour of the bars corresponds to the EPE category. In Figure 12 we show a display referring to the expected intensity of the event in comparison with all EPEs which occurred before in terms of area above the 99th percentile and total volume of rain. This type of volume–area plot is automatically generated only if MaLCoX or ENS predicts an EPE. As shown by Figures 11 and 12, MaLCoX predicted the likelihood of an EPE Cat1 between 14 and 16 December many days in advance, but only from the forecast starting on 11 December, it started to point to 15 December as the day with the highest rainfall. The MaLCoX forecast issued on 11 December showed a much higher and closer rain volume estimation than the one predicted by ECMWF raw output. Besides these quantitative measures, forecast consistency is another very desirable property of

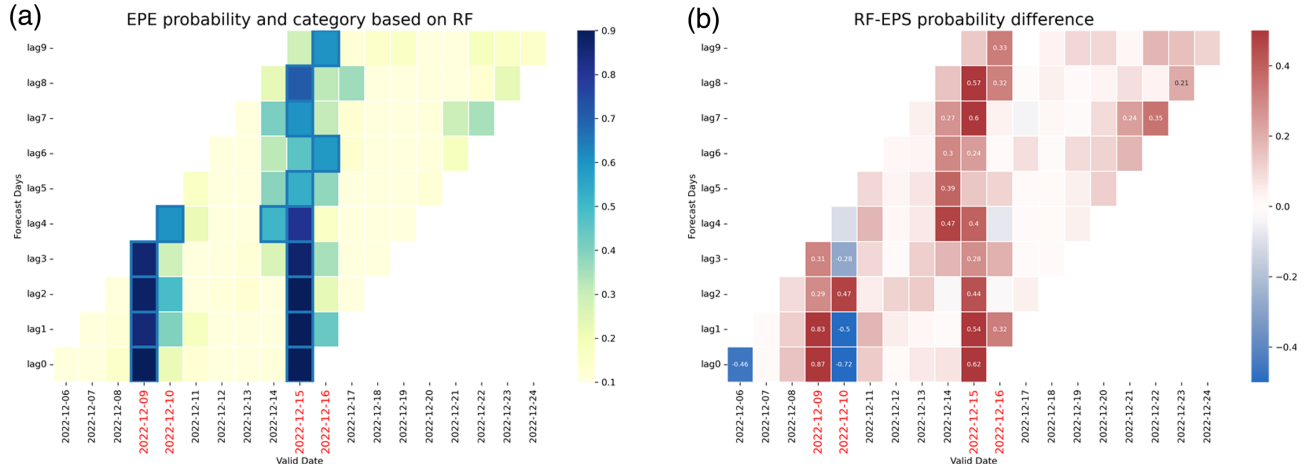

FIGURE 13 Extreme Precipitation Event (EPE) heatmap for 20 days covering the case study. (a) EPE probability predicted by MaLCoX is shaded according to the legend, and the predicted category is indicated by the tick outline (blue, Cat1; orange, Cat2; green, Cat3). Dates marked in red indicate days with observed Extreme Precipitation Events (EPEs). (b) Difference in probability, MaLCoX minus ensemble prediction forecast (ENS), according to legend. In boxes without annotation the difference is smaller in absolute terms than 0.2. RF, stands for random forest module (of MaLCoX) [Colour figure can be viewed at wileyonlinelibrary.com]

a reliable forecast system, particularly to increase trust. A further assessment of consistency can be obtained with a heatmap view of EPE probability. In Figure 13a, we show the predicted probability and the EPE category by MaLCoX for a running window of 20 validating dates (X-axes) and forecast at different lead times (Y-axes). MaLCoX became certain of the 15th day from the D+4 forecast (lag3) while at longer lead times the forecast was showing some inconsistency, alternating between days 14 and 16 as possible EPEs. In Figure 13b we show a systematic difference of EPE probability predicted by MaLCoX minus the probabilities predicted by the ENS. MaLCoX correctly raised the probabilities (red boxes) compared to the ENS on the days with observed EPEs, except for a short-range forecast valid on 10 December. A general increase prevailed on the other days, but was non-significant, with differences smaller than 20% (boxes without annotation).

To explain the behaviour of MaLCoX and to convey the physical process which made the precipitation extreme in the forecast we adopted the Shapely additive explanation package (Lundberg *et al.*, 2017), already used for a similar application related to atmospheric composition forecast (Vega García & Aznarte, 2020). The waterfall representation, displayed in Figure 14, shows the contribution of each feature in the two successive forecasts, both validating on 15 December 2022.

Starting from the background expectation $E[f(X)]$, which is defined by the average frequency of the EPE in the forecast, 6% at this lead time, the waterfall shows the incremental increase or decrease of the model output adding features until it reaches the outcome $[f(X)=1]$ (EPE yes). In both forecasts, MaLCoX predicted an EPE but the decision was led by different predictors. In the forecast initiated on 10 December (D+5) (Figure 14a) the most important predictor is the IVTmag non-local index (lead time 0) followed by the day of the year (Juld) and Z500 non-local index (lead time 0), while in the forecast of next day (11 December) (Figure 14b) the total rain volume predicted by HRES becomes the most important feature, followed by IVTmag non-local index (lead time 0) and V850 non-local index (lead time 0). In Figure 14a it is interesting to note the slightly negative contribution of Volf due to HRES predicting a smaller rain amount compared with a typical EPE. Despite the lack of sufficient explicit rain, MaLCoX was still leading to an EPE due to the contribution of the large-scale component namely the non-local class. From D+4 and closer to the event, the ranking of predictors changes and Volf becomes the driving source of information with non-local predictors progressively going down in the ranking as the lead time reduces. Despite the predictors ranking changing case by case, this example illustrates well the typical complementary value of local and non-local predictors acting when the predictability of direct precipitation output is low. This improves the forecast skill in the medium range and the consistency, preventing sudden jumps in the forecast.

5 | CONCLUSION

In this paper, we present a hybrid dynamical-statistical model (MaLCoX) which uses sequential RF models to combine different classes of forecast atmospheric predictors and improve predictions of EPEs. These types

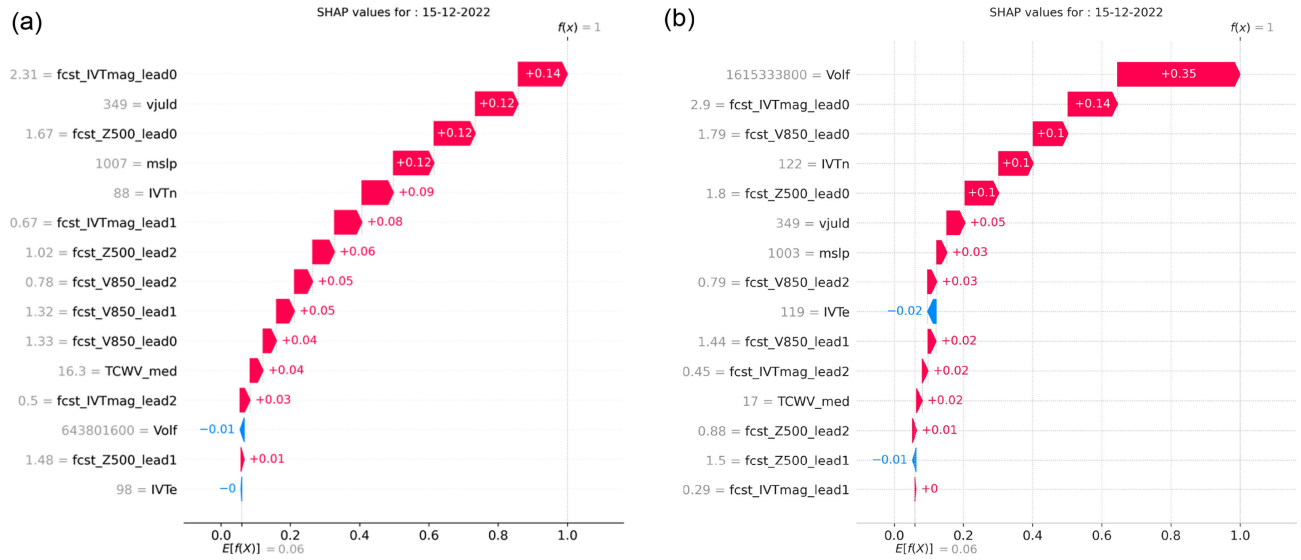

FIGURE 14 SHAP waterfall illustrating the additive contribution of each feature for two consecutive MaLCoX forecasts both validating on 15 December. In both forecasts, MaLCoX predicted an Extreme Precipitation Event (EPE) but the result was driven by different predictors. The bottom of a waterfall plot starts as the climatological expected value of the model output for EPE = yes, and then each row shows how the positive (red) or negative (blue) contribution of each feature moves the value from the background value to the model output for this prediction. On the Y-axes predictors' names are ranked according to their influence and their value is shown with grey numbers. (a) D+5 forecast initiated on 10 December 2022. (b) D+4 forecast initiated on 11 December 2022. [Colour figure can be viewed at wileyonlinelibrary.com]

of hybrid models represent an emerging class in the spectrum of weather forecasting systems. While previous studies have already reported hybrid predictions of short-term precipitation forecast, to our knowledge, no analogous examples of medium-range forecasting of extreme events exist, which is the target of our work. In the medium range, non-local predictors – scalar normalised indices representing large-scale anomalies preceding EPEs in the Euro-Atlantic sector – are an important source of skill and an innovation of our approach. MaLCoX, trained with a low-resolution dataset with a 20-years equivalent period of 10-day forecasts obtained from the ECMWF reforecast dataset, shows a better performance than the high-resolution ECMWF operational forecast in the prediction of EPEs over northern–central Italy, with a gain of about three days of forecast skill in the medium range.

A third innovative contribution is its interpretability: explaining the reason for a predicted outcome in terms of the relative contribution of different predictors. Understanding why a model makes a certain prediction can be as crucial as the prediction's accuracy, especially when forecasters have to face rare extreme conditions and they need to gain trust in model output. Combining the results obtained by feature importance, scores on training and test datasets and results of feature attribution on single case studies, we show that predictors act in complementary ways throughout the forecasting period. At short forecast lead times, the hybrid model is mostly informed by the explicit precipitation field (total volume of rain) and local synoptic features, like IVTn. In the medium range, and especially after lead time + 144 hours, the cumulative effects of non-local predictors become predominant together with local synoptic features. The four-day forecast horizon is on average the crossing point where the redistribution of weights among the different feature components occurs. This information also has implications on predictability suggesting that, for the medium range and even longer ranges, flow-pattern characteristics remain predictable and the information that can be extracted is useful to infer the likelihood of an EPE. Finally, we want to stress the positive role played by the non-local predictors class, here newly tested, in reducing the loss of predictability at longer time ranges. The results shown here refer to the first operational version that will be updated with further refinements. In particular, in the future upgrade of MaLCoX we plan to increase even further the training dataset using all ensemble members of the ENS reforecast as independent realisation and to expand the prediction to all individual warning areas.

5.1 | Scripts and datasets

The table containing observed aggregated daily precipitation over northern and Central Italy warning areas

(1961–2022) and days with EPE (yes/no) is available on the open data repository of ARPAE Emilia-Romagna: https://dati.arpae.it/dataset/serie-giornaliera-di-eventi-estremi-di-precipitazione-sul-centro-nord-italia.

FUNDING INFORMATION

Deutsche Forschungsgemeinschaft; Transregional Collaborative Research Centre, Project T2, Helmholtz Young Investigator Group 'Sub-Seasonal Predictability: Understanding the Role of Diabatic Outflow' (SPREADOUT; Grant VH-NG-1243).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in [Serie giornaliera di eventi estremi di precipitazione sul centro-nord Italia] at ARPAE Emilia-Romagna repository https://dati.arpae.it/dataset/serie-giornaliera-di-eventi-estremi-di-precipitazione-sul-centro-nord-italia. These data were derived from observational ARCIS dataset available in the public domain: https://www.arcis.it/wp/.

ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

ORCID

Federico Grazzini https://orcid.org/0000-0002-3435-2376
Joshua Dorrington https://orcid.org/0000-0003-3802-457X
Christian M. Grams https://orcid.org/0000-0003-3466-9389
George C. Craig https://orcid.org/0000-0002-7431-8164
Linus Magnusson https://orcid.org/0000-0003-4707-2231
Frederic Vitart https://orcid.org/0000-0001-8485-7981

REFERENCES

Breiman, L. (2001) Random forests. *Machine Learning*, 45, 5–32.

Chapman, W.E., Monache, L.D., Alessandrini, S., Subramanian, A.C., Martin Ralph, F., Xie, S.P. et al. (2022) Probabilistic predictions from deterministic atmospheric river forecasts with deep learning. *Monthly Weather Review*, 150, 215–234 https://journals.ametsoc.org/view/journals/mwre/150/1/MWR-D-21-0106.1.xml

Davis, J. & Goadrich, M. (2006) The relationship between precision-recall and ROC curves. In: Cohen, W., & Moore, A., (Eds.) *Proceedings of the 23rd International Conference on MachineLearning*. Pittsburgh, PA: Omni Press.

de Sousa Araújo, A., Silva, A.R. & Zárate, L.E. (2022) Extreme precipitation prediction based on neural network model – a case study for southeastern Brazil. *Journal of Hydrology*, 606, 127454.

Dorrington, J., Grams, C., Grazzini, F., Magnusson, L. & Vitart, F. (2023) Domino: a new framework for the automated identification of weather event precursors, demonstrated for European extreme rainfall. *Quarterly Journal of the Royal Meteorological Society*. 150(759), c776–795. Available from: https://doi.org/10.1002/qj.4622

Duffourg, F. & Ducrocq, V. (2011) Origin of the moisture feeding the heavy precipitating systems over southeastern France. *Natural Hazards and Earth System Science*, 11, 1163–1178.

Espeholt, L., Agrawal, S., Sønderby, C., Kumar, M., Heek, J., Bromberg, C. et al. (2022) Deep learning for twelve hour precipitation forecasts. *Nature Communications*, 13, 5145. Available from: https://doi.org/10.1038/s41467-022-32483-x

Frnda, J., Durica, M., Rozhon, J., Vojtekova, M., Nedoma, J. & Martinek, R. (2022) ECMWF short-term prediction accuracy improvement by deep learning. *Scientific Reports*, 12, 7898. Available from: https://doi.org/10.1038/s41598-022-11936-9

Grazzini, F. (2007) Predictability of a large-scale flow conducive to extreme precipitation over the western Alps. *Meteorology and Atmospheric Physics*, 95, 123–138. Available from: https://doi.org/10.1007/s00703-006-0205-8

Grazzini, F. (2021) *Extreme precipitation in northern Italy*. Ph.D. thesis. Munich: Ludwig Maximilian University of Munich. http://nbn-resolving.de/urn:nbn:de:bvb:19-282191

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. Available from: https://doi.org/10.1002/qj.3635

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 https://doi.org/10.1007/s42865-020-00018-1

Grazzini, F., Fragkoulidis, G., Teubler, F., Wirth, V. & Craig, G.C. (2021) Extreme precipitation events over northern Italy. Part II: dynamical precursors. *Quarterly Journal of the Royal Meteorological Society*, 147(735), 1237–1257. Available from: wiley.com/doi/10.1002/qj.3969.

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

Hill, A.J., Herman, G.R. & Schumacher, R.S. (2020) Forecasting severe weather with random forests. *Monthly Weather Review*, 148, 2135–2161 https://journals.ametsoc.org/view/journals/mwre/148/5/mwr-d-19-0344.1.xml

Khodayar, S., Davolio, S., Di Girolamo, P., Lebeaupin Brossier, C., Flaounas, E., Fourrie, N. et al. (2021) Overview towards improved understanding of the mechanisms leading to heavy precipitation in the western Mediterranean: lessons learned from HyMeX. *Atmospheric Chemistry and Physics*, 21, 17051–17078.

Khodayar, S., Pastor, F., Valiente, J.A., Benetó, P. & Ehmele, F. (2022) What causes a heavy precipitation period to become extreme? The exceptional October of 2018 in the Western Mediterranean. *Weather and Climate Extremes*, 38, 100493.

Lundberg, S.M., Allen, P.G. & Lee, S.-I. (2017) A unified approach to interpreting model predictions. In: *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, Morehouse Lane, Red Hook, NY: Cambridge University Press. pp. 4768–4777 https://github.com/slundberg/shap

Magnusson, L., Hewson, T. & Lavers, D. (2021) Windstorm Alex affected large parts of Europe. *ECMWF Newsletter N*, 166, 4–5.

Martius, O., Schwierz, C. & Davies, H.C. (2008) Far-upstream precursors of heavy precipitation events on the alpine southside. *Quarterly Journal of the Royal Meteorological Society*, 134, 417–428. Available from: https://doi.org/10.1002/qj.229

Mastrantonas, N., Magnusson, L., Pappenberger, F. & Matschullat, J. (2022) What do large-scale patterns teach us about extreme precipitation over the Mediterranean at medium and extended-range forecasts? *Quarterly Journal of the Royal Meteorological Society*, 148, 875–890.

Pavan, V., Antolini, G., Barbiero, R., Berni, N., Brunier, F., Cacciamani, C. et al. (2019) High resolution climate precipitation analysis for north-central Italy, 1961–2015. *Climate Dynamics*, 52, 3435–3453.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2011) Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830 http://jmlr.org/papers/v12/pedregosa11a.html

Rudari, R., Entekhabi, D. & Roth, G. (2005) Large-scale atmospheric patterns associated with mesoscale features leading to extreme precipitation events in northwestern Italy. *Advances in Water Resources*, 28, 601–614.

Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A. et al. (2021) Weather and climate extreme events in a changing climate. In: *Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge: Cambridge University Press, pp. 1513–1766.

Sioni, F., Davolio, S., Grazzini, F. & Giovannini, L. (2023) Revisiting the atmospheric dynamics of the two century floods over north-eastern Italy. *Atmospheric Research*, 286, 106662 https://linkinghub.elsevier.com/retrieve/pii/S0169809523000595

Tramblay, Y. & Somot, S. (2018) Future evolution of extreme precipitation in the Mediterranean. *Climatic Change*, 151, 289–302.

Tsonevsky, I. (2015) New EFI parameters for forecasting severe convection. *ECMWF Newsletter N.144*, 27–32.

Vega García, M. & Aznarte, J.L. (2020) Shapley additive explanations for NO2 forecasting. *Ecological Informatics*, 56, 101039.

Vitart, F., Balsamo, G., Bidlot, J.-R., Lang, S., Tsonevsky, I., Richardson, D. et al. (2019) Use of ERA5 to initialize ensemble Re-forecasts. Tech. Rep. 841, ECMWF, Reading. https://www.ecmwf.int/node/18872

Whan, K. & Schmeits, M. (2018) Comparing area probability forecasts of (extreme) local precipitation using parametric and machine learning statistical postprocessing methods. *Monthly Weather Review*, 146, 3651–3673 https://journals.ametsoc.org/view/journals/mwre/146/11/mwr-d-17-0290.1.xml

How to cite this article: Grazzini, F., Dorrington, J., Grams, C.M., Craig, G.C., Magnusson, L. & Vitart, F. (2024) Improving forecasts of precipitation extremes over northern and central Italy using machine learning. *Quarterly Journal of the Royal Meteorological Society*, 1–15. Available from: https://doi.org/10.1002/qj.4755