Title: Can Weather Patterns Contribute to the Prediction

of Winter Floods using Machine Learning?

<u>Response to Reviewer 1</u>

We sincerely thank both reviewers for their thorough and constructive feedback. Their detailed comments have significantly helped us to plan amendments to the manuscript to improve the clarity, framing, and methodological rigor. We appreciate the recognition of the relevance of the topics addressed, as well as the thoughtful critiques regarding the use of weather patterns (WPs), the structure of our modelling framework, and the interpretation of results. In response, we are making substantial revisions to the manuscript. These include a reframing of the paper to better reflect its original contributions, particularly regarding the limited marginal value of WPs in flood magnitude prediction. We will improve the explanation of the methodological design and evaluation strategy and provide a more transparent and fair comparison between regional and national models. We have also clarified definitions, ensured consistent use of terminology, and addressed technical concerns raised in both the major and minor comments. Below, we provide responses to each comment. Throughout this document, our responses will be in blue text, and the reviewer comments in plain text.

R1C1: In the manuscript, if weather patterns contribute to predicting winter flood magnitudes was discussed using machine learning. To my knowledge, flood is mainly caused by intensive rainfall and antecedent soil conditions. In this study, they also considered these two factors and add some other variables. Some main questions are below.

Response: We thank the reviewer for highlighting this fundamental point regarding the drivers of flooding. We fully agree that flood generation is primarily governed by rainfall intensity and antecedent catchment conditions, such as soil moisture. These variables are central components of our study and are included in the feature set used for model training. Our specific focus on weather patterns (WPs) stems from the operational and research relevance of the 30-pattern classification developed by the UK Met Office (Neal et al., 2016). These WPs have been used in prior studies examining their links to precipitation extremes and droughts (e.g., Richardson et al., 2018; 2019; 2020) and are integrated into tools like the Met Office's Fluvial Decider for flood risk assessment. What sets our study apart is that we conduct the first large-scale, data-driven evaluation of the direct relationship between these WPs and observed fluvial flood magnitudes, using NRFA streamflow records. Our goal is to empirically assess whether large-scale atmospheric circulation patterns, often more predictable at longer lead time than local-scale variables, contribute meaningfully to flood prediction frameworks when more immediate forcings are also included. Ultimately, our results show that while WPs capture synoptic-scale atmospheric conditions, their marginal value in flood magnitude prediction is limited. This finding itself is important, as it quantifies the redundancy of large-scale indicators within data-driven hydrological models and helps refine which predictors offer true added value. We focus on weather patterns as the UK Met Office produced a set of 30 weather patterns, see Neal et al (2016). These weather patterns have been used in research associating them with precipitation and drought, and they are employed in a Met Office tool called Decider. We therefore conduct the first study relating these weather patterns directly to fluvial floods using the NRFA streamflow time-series data, to understand the relationship between them and flood magnitudes.

R1C2: Table 2, I cannot understand the relationship between total event count, number of catchments and catchment average event count.

Response: Thank you for this comment, we will improve the table to make this clearer to the reader. The total event count is the total number of fluvial flood magnitude events across all catchments. The number of catchments is the number of unique catchment IDs with events. The catchment average is the total events divided by the number of catchments.

R1C3: Line 154, 'pre-filtered to contain only extreme flood magnitude days', this will not ensure the flood event from beginning to the end.

Response: We appreciate the reviewer's comment and agree that filtering based solely on peak days does not capture the full duration or hydrograph shape of each flood event. However, our study is specifically designed to focus on flood magnitude. By this, we mean the maximum daily flow or peak, associated with each extreme event, rather than the full flood hydrograph or event duration.

R1C4: Line 163, the categorize small, medium and large is not appropriate. Because in hydrology, there is a standard for definition of small, medium and large catchments.

Response: We thank the reviewer for this observation. While we acknowledge that hydrologists often use informal thresholds to classify catchment sizes, to the best of our knowledge there is no single, universally accepted standard for defining small, medium, and large catchments, specifically within the context of UK hydrology and the UKBN2 benchmark dataset. In this study, the terms "small," "medium," and "large" were used descriptively to indicate relative size classes, based on quantiles of the catchment area distribution within the UKBN2 dataset. This was done purely for interpretative convenience, rather than to imply strict hydrological classifications. We will clarify this point in the revised text.

R1C5: Line 278, the WP associated with the most extreme precipitation, does not necessarily translate to the WP associated with extreme flood magnitude days across UK regions.' I cannot understand the intrinsic relations between WP, extreme precipitation and extreme flood magnitude days.

Response: We thank the reviewer for raising this important point. The relationship between WPs, extreme precipitation, and extreme flood magnitudes is not direct or linear, and this is a central theme of our study. While previous research (e.g., Richardson et al., 2018) has shown that certain WPs are associated with heavy precipitation in the UK, extreme precipitation alone does not always result in extreme flooding. Flood magnitudes are influenced by several additional factors, including:

- Antecedent catchment conditions
- Catchment memory (e.g., the timing and accumulation of prior rainfall)
- Catchment-specific properties (e.g., topography, land cover, drainage capacity)
- Temporal structure of rainfall (e.g., short, intense bursts vs. prolonged moderate rainfall)

Even when a WP is associated with high rainfall in aggregate (through conditional probability), regional variability in hydrological response can lead to different WPs dominating the flood magnitude signal in different parts of the country. For example, a WP that causes widespread rainfall may trigger flooding in steep, saturated western upland catchments, but not in drier eastern regions with higher infiltration capacity. Topography plays a key role as well mountainous areas can respond more rapidly and intensely than flatter regions to the same synoptic forcing. Therefore, finding that a WP most associated with extreme precipitation is not necessarily the same WP most associated with peak flood magnitudes highlights the critical modulating role of other hydrological processes. This supports our conclusion that WPs, while useful for understanding large-scale atmospheric conditions, offer limited explanatory power for flood magnitudes when catchment-specific variables are not considered. We will revise the text around Line 278 to clarify this distinction more explicitly in the manuscript

R1C6: Line 363, CEE had the lowest baseline R2 (0.28), and only the final R2 of 0.37 in Generation 6 was statistically significant. Why the precision is so low? Are there any previous hydrological simulation in this region? Please compare this result with previous studies.

Response: We thank the reviewer for this important observation. The relatively low R^2 values for the Central and Eastern England (CEE) region reflect genuine modelling challenges in this area, which are likely driven by a combination of hydrological, data, and methodological factors. First, the CEE region is characterized by predominantly low-relief terrain, permeable soils, and mixed land use, leading to slower and more diffuse runoff responses to precipitation. These characteristics make peak flood magnitudes less directly tied to single rainfall events, and thus more difficult to capture using event-scale predictors alone. We are not aware of any published regional flood magnitude prediction models specifically for CEE. Higher uncertainty in this region could be explained by the baseflow-dominated hydrology and the difficulty in defining extreme event thresholds. Considering this, our R^2 value of 0.37, though modest, is consistent with regional hydrological characteristics and represents a statistically significant improvement over the baseline model. We will add further discussion in the manuscript to contextualize the CEE performance within existing literature and acknowledge the challenges of modelling in low-gradient, infiltration-dominated catchments.

R1C7: Line 370, 'The SE region's relatively lower sensitivity to antecedent precipitation and hydrometeorological inputs suggests that other factors, such as urbanization and engineered drainage systems, may dominate flood generation.' However, when you select the watersheds, they are not influenced by human activities.

Response: We appreciate this clarification and will revise the text to remove any implication that urbanization is a known factor in these specific sites.

R1C8: When using SHAP, you need to explain the definition of aridity, runoff ratio….

Response: We thank the reviewer for pointing this out. We agree that clear definitions of all predictor variables used in the SHAP analysis are essential for interpretability, especially for features like aridity index and runoff ratio that may vary slightly in definition depending on the context. In response, we will add a table to the manuscript (in the Methods or Supplementary Material) listing all variables included in the SHAP analysis, along with:

- Full variable names
- Units
- Definitions
- Data sources

For clarity, we define these variables as follows:

- Aridity Index: The ratio of potential evapotranspiration (PET) to precipitation (P) over the climatological period. A higher value indicates drier climatic conditions relative to available rainfall.
- Runoff Ratio: The ratio of mean annual streamflow (Q) to mean annual precipitation (P) at the catchment scale. This reflects how efficiently rainfall is converted to runoff, influenced by soil type, land cover, and infiltration capacity.

We will ensure these, and all other input variables are clearly described in the manuscript to improve reproducibility and interpretation.

R1C9: Line 490, 'The SHAP summary plot further supports the limited contribution of the WPs'. In traditional flood analysis, rainfall and soil moisture are the main contributors. We never consider WPs. In this study, WPs are focused, but still limited contribution. What is the innovation of this study?

Response: We thank the reviewer for this important question. Yes, traditional flood analysis rightly emphasizes precipitation and antecedent soil moisture as the dominant drivers of flood events. Our study does not challenge this understanding, instead, it aims to quantify whether atmospheric circulation patterns (in our case, synoptic scale WPs) provide any additional explanatory or predictive value when these direct drivers are already included in the model. The innovation of our study lies in the following key contributions:

1. First empirical test of WPs for flood magnitudes at national scale: While WPs have been linked to rainfall and drought in past studies (e.g., Neal et al., 2016; Richardson et al., 2018), this is the first study to evaluate their direct relationship with flood magnitudes across a large sample of UK catchments using machine learning and observational streamflow data.
2. Quantifying redundancy of large-scale indicators: By showing that WPs have limited marginal importance once local meteorological and catchment-scale features are included, we provide empirical evidence that these large-scale predictors are largely redundant in data-driven flood models. This is a useful insight for future feature selection and model design.
3. Challenging assumptions in operational tools: WPs are already being used in tools such as the Met Office's Decider. Our findings critically examine these assumptions and help define the limits of what WP-based reasoning can offer, especially for flood magnitudes rather than occurrence or likelihood.
4. Contribution to the growing interface between climate diagnostics and hydrology: The study advances understanding of how far ahead predictive signals can be leveraged, which is useful both for operational forecasting and long-term planning under climate variability.

In summary, while WPs were shown to contribute little additional predictive power in this specific context, our clear empirical evaluation of their role, and demonstration of their limited value, constitutes a meaningful and novel contribution to the field. We will revise the manuscript discussion to make this contribution more explicit.

R1C10: Line 501, ‘Interestingly, precipitation on the day of the event consistently ranks higher than antecedent precipitation’, actually, this is a common sense.

Response: We thank the reviewer for this observation. We agree that the dominant role of event-day precipitation in flood generation is well known in hydrology and should not be described as “interesting” in a way that implies novelty. We will revise the manuscript text at Line 501 to remove the word “interestingly” and instead frame the observation as a validation of model realism, consistent with standard hydrological expectations.

Thank you for your review!