Response to reviewers
Manuscript to NHSS

Manuscript number: 2024-538
Title: Ready, set, go! An anticipatory action system against droughts
Authors: Gabriela Guimarães Nobre, Jamie Towner, Bernardino Nhantumbo, Célio João da Conceição Marcos Matuele, Isaias Raiva, Massimiliano Pasqui, Sara Quaresima, and Rogério Bonifácio.

General response

We thank the two reviewers and the editor for the effort and time taken to review and process our manuscript. The reviewers provided substantial suggestions to improve the current version of our manuscript. In response, we have revised sections of the manuscript, as outlined in this document. Based on the reviewer's comments and the subsequent revisions, we believe that the manuscript has greatly improved, and trust that we have addressed the concerns of the reviewers. In the following sections, we respond to each of the reviewers' remarks or questions. Our responses are colored in blue.

Reviewer #1:

Reviewer #1: Forecasts for anticipatory actions play a pivotal role in mitigating the impact of natural hazards, particularly in regions vulnerable to frequent disasters like Mozambique. These forecasts not only offer invaluable insights for disaster preparedness but also facilitate timely interventions, ultimately safeguarding lives and livelihoods. I agree with the authors of this manuscript about the academic and social significance of AA-tailored forecasts, particularly in a context as critical as Mozambique. While acknowledging the importance of this study, my primary concern is its reproducibility (the method is described in a long and complex way) and ensuring the correct application of methodology, as clarity remains paramount for effective scientific communication and practical implementation.

We appreciate the reviewer's insightful comments and are pleased that our manuscript is considered relevant to the journal. Below, we offer a detailed summary of the revisions made to enhance the reproducibility of our methods, as well as to improve scientific communication and implementation.

Reviewer #1 comments

The abstract provides significant contextual information, yet falls short in delivering a clear explanation of the "ready set go" process. Consider expanding upon this process, especially in the testing triggers section, perhaps through the inclusion of a flow chart to elucidate the operational workflow. Additionally, the practical implications of the results remain ambiguous, particularly regarding the targeting of the first part of the rainy season. Thirdly, there is a confusing focus on two specific provinces in some manuscript sections while in other sections, results over all of Mozambique are shown. Clarifying these points would enhance the abstract's effectiveness.

R1. We sincerely thank the reviewer for the valuable comments. To enhance the clarity of the manuscript in response to the reviewer's suggestions, we have made the following revisions:

1. Added an explanation of the Ready, Set & Go! system in the abstract for greater clarity.
2. Substantially improved the Methods section, including a detailed description of the methods used for each component of the flowchart in Figure 1.
3. Included two new paragraphs to provide further explanation on the testing of triggers in the section "Testing Several Triggers for the Ready, Set & Go! System."
4. Expanded the explanation of the importance of targeting two windows for anticipatory action in the section "Defining Triggers for Anticipatory Action."
5. Extensively edited the Results section to shift the focus from the province scale to the national scale for scaling up AA activities: (i) Figure 1 is still summarized at the zone level for simplicity, but with an improved description of how the results were obtained; (ii) Figure 2 was revised with an enhanced description of the findings; (iii) Figures 3–9 remain similar, but their descriptions were refined for better clarity.

We hope these extensive revisions have successfully improved the clarity and readability of the manuscript.

The introduction effectively outlines the various natural hazards affecting Mozambique but could benefit from clearer terminology regarding the description of multi hazards (meaning consecutive/compounding ones?). Additionally, it is suggested to specify that "impacts" such as flooding and cyclones are caused by these hazards rather than merely being impacts themselves. Moreover, while the introduction mentions the benefits of early action, it is imperative to address in the discussion how anticipatory actions can mitigate impacts cost-effectively, especially considering the limitations of ‘preparedness possibilities’ with respect to water management.

R2. We sincerely thank the reviewer for the valuable suggestion. The term "multi-hazards" in this article is aligned with the UNDRR definition: "Multi-hazard means (1) the selection of multiple major hazards that the country faces, and (2) the specific contexts where hazardous events may occur simultaneously, cascadingly, or cumulatively over time, taking into account the potential interrelated effects." While in some instances the term "multi-hazard" simply refers to the multiple major hazards Mozambique faces, it is important to emphasize that these hazards can occur as single or consecutive events (including compounding effects, as per the definition in Ruiter et al., 2020), which may consequently exacerbate losses and damages.

In the revised manuscript, we incorporated key terms from the above definition to clarify the meaning of "multi-hazards." Additionally, despite the still limited evidence on the cost-effectiveness of anticipatory action due to the novel nature of the pilots, we have added a paragraph to the Discussion section that explores how anticipatory actions can mitigate impacts in a cost-effective manner based on available evidence from WFP programmes.

The methodology contains a robust framework, but improvements in clarity, adding relevant citations, and referencing to the framework are needed. Numbering the structure to correspond directly with figures and ensuring consistency between bold text and descriptions in flow charts would enhance readability. Furthermore, some terms require clarification or references, such as "blended precipitation records" and "bilinear interpolations". The core of the analysis revolves around assessing the precision in forecasting precipitation levels one standard deviation below the norm. However, amidst extensive textual explanations, this crucial aspect becomes obscured, overshadowed by peripheral details. Streamlining the method section to prioritize essential components could enhance clarity and comprehension. Also, the nature of the forecasted variables remains ambiguous. Is the SPI3 predicted several months in advance? Including a supplementary list enumerating predictors and predictands would provide invaluable clarification, ensuring transparency and facilitating a deeper understanding of the forecasting methodology.

R3. Thank you for your recommendation. We agree that clarification and restructuring of the Methods section were necessary. In the revised manuscript, we created a subsection for each item highlighted in the flowchart, extensively rewrote sentences, and added details where clarity was lacking. This includes additional explanations, references, and a new figure (Figure 3) that links the lead time of forecasts with the extracted SPIs.

Additionally, it is recommended to utilize the Stagge et al. (2015) approach for SPI calculation to ensure accuracy, particularly in arid regions. Most importantly, the application of a severe drought threshold for extracting drought probabilities based on SPI lacks clarity. Merely selecting a SPI value inherently provides the probability, unless there are issues with distribution fitting, which would signify a methodological problem rather than a data issue. The rationale behind defining and applying danger thresholds for drought events remains ambiguous; why not simply select values like -1 or -2? Without clear explanation, it raises concerns about methodological integrity. Additionally, the utilization of return period as a quality criterion requires further elucidation, as its significance is not apparent (Table 1).

R4. We thank the reviewer for the detailed comment and for referencing the methods proposed by Stagge et al. (2015). While we intend to fully integrate their methodology into WFP products, including testing various probability distributions and addressing zero precipitation issues, I would like to clarify the suitability of the gamma distribution for extracting SPI in Mozambique. In the initial stages of our project, we evaluated the gamma distribution's fit by conducting the Kolmogorov-Smirnov (KS) test (see map example below). This test assessed whether the empirical data matched the gamma distribution. Figure 1 shows the p-values from the KS test, with red areas indicating regions where the null hypothesis was rejected, suggesting a poor fit. However, these red values are infrequent, which strengthens our confidence in the gamma distribution's suitability for this project. Stagge et al. (2015) also recommends the gamma distribution for calculating SPI across various accumulation periods and regions in Europe. Additionally, zero precipitation is not a significant issue in our approach, as the occurrence of 2- and 3-month accumulated rainfall during the rainy season is very rare. Although our system currently uses a simple method for handling zero precipitation, we acknowledge that a more robust approach could be developed for future updates. This recommendation has been added to the Discussion section.

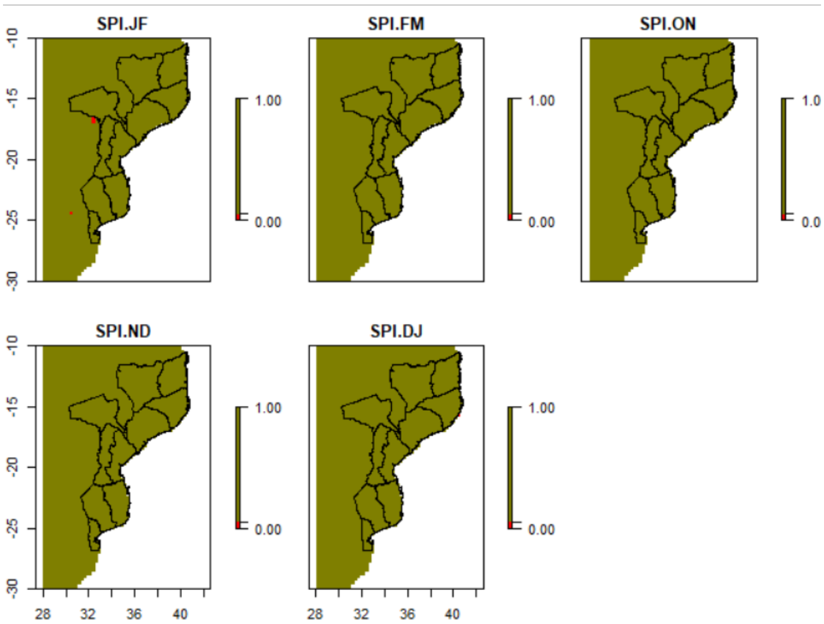

Figure 1 P-value of the kolmogorov-smirnov test for SPI2 indicators.

Furthermore, we would like to emphasize that the most appropriate way to derive SPI thresholds for identifying severe droughts is through a historical analysis that examines the relationship between drought events and their socio-economic impacts. Unfortunately, comprehensive, downscaled impact data is largely unavailable, particularly in African countries, which limits our ability to refine thresholds and indicators and weakens the connection between drought conditions and past impacts. In the absence of such data, we base the choice of threshold levels on frequencies that are suitable for anticipatory action (AA) operations in the region. Typically, AA programs target hazards that occur once every three to six years on average. Periodically implementing AA pilots is crucial for improving program effectiveness. As such, thresholds for AA operations should not be set too low, since severe drought events are rare. Therefore, the choice of a threshold of -1, rather than -2, is primarily driven by the needs of AA operations and is also linked to the predictability of the event. The more severe the event, the less predictable it is, so this threshold reflects a balance between allowing periodic interventions and relying on reasonably accurate forecast information. To clarify this, we have strengthened the Methods section, particularly in the part titled "Define Danger Threshold for Identifying Past Drought Events." Regarding the return period within the quality criteria list, we agree with the reviewer that its relevance needs further clarification. This metric calculates the frequency of theoretical anticipatory action (AA) interventions from 1993 to 2021 based on the selected trigger. It helps determine whether the empirical frequency of AA interventions—resulting from both hits and false alarms—aligns with the frequency of the threshold for severe droughts. To enhance understanding, we have added a sentence to clarify this metric in the section "Sensitivity Analysis Including Four Scenarios," where the statistical metrics are described.

Another significant issue arises from bias correction based on SPI between CHIRPS and forecast reanalysis data. The process of SPI fitting involves converting values to standardized ones, reflecting standard deviations from a normal distribution. However, using two time series of SPI values for bias correction raises questions about information loss and overall methodological coherence. Furthermore, the inclusion of ENSO in bias correction methodology lacks justification; while rainfall patterns may vary under different ENSO states, assumptions about forecast biases in these states are not clearly articulated. The paragraph on bias correction fails to provide convincing rationale, particularly concerning the absence of transfer functions calculated over raw precipitation data. This oversight could potentially undermine the validity of the approach. Clarification on these aspects is imperative for ensuring methodological soundness and reproducibility.

R5. We thank the reviewer for his/her comment. Below we provide clarification on the bias correction approach:

1. ENSO-Informed Bias Correction: Our method employs an ENSO-informed quantile mapping transfer function to correct SPI forecasts based on SPI reference values conditioned on ENSO states. This approach ensures that bias correction considers variations in SPI quantities according to the climatology of different ENSO phases, effectively integrating relevant global processes. For Mozambique, El Niño typically reduces rainfall in the southern and central regions, while La Niña usually increases rainfall in the northern and central areas. Internal tests revealed that skill in forecasting severe droughts improved only when incorporating ENSO information into the bias correction process. The effectiveness of bias correction is thus validated by a tangible gain in forecasting skill. The figure below shows a comparison of bias correction results for SPI forecasts issued in July, with and without ENSO information.
2.

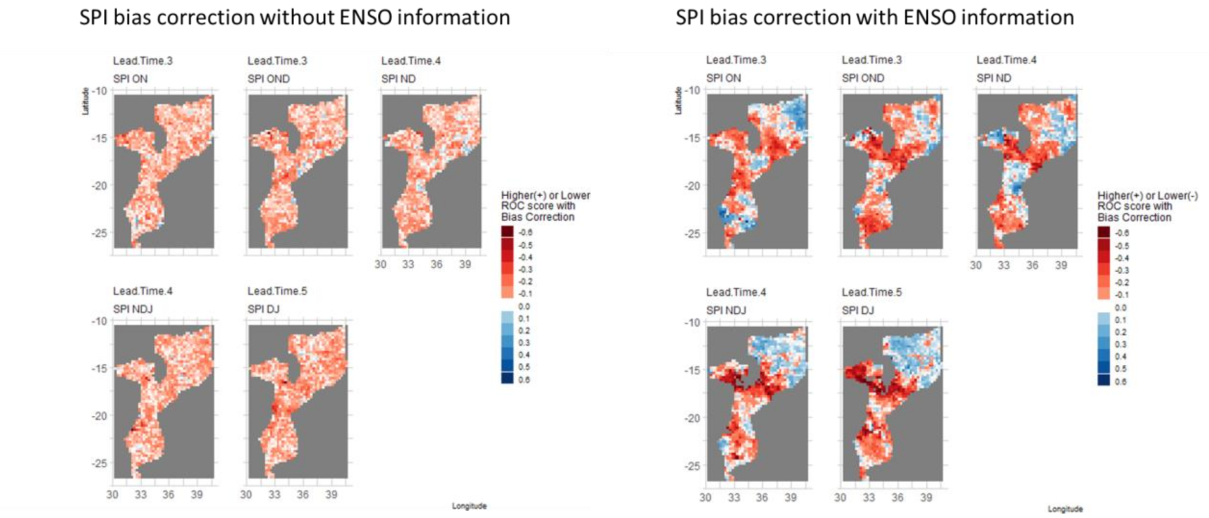

Figure 2 Differences between the AUROC for severe events using bias correction methodology minus the AUROC score from raw forecasts. Regions in blue show the added value of bias correction, whereas in red regions with decreased skill due to BC. The plot shows the skill of the forecast issued in July. Maps on the left were produced without ENSO information whereas the right maps were produced with ENSO information.

2. Transfer Function Based on SPI Rather than Raw Precipitation: We opted to target the SPI indicator directly with the transfer function to enhance drought detection accuracy by aligning SPI forecasts more closely with observed SPI climatology. This approach aims to produce SPI forecasts that better reflect historical patterns and trends. This is crucial for our system, which issues alerts based on negative SPI anomalies rather than raw rainfall

amounts. Our initial internal tests revealed significant improvements in drought predictability when using a transfer function that directly links SPI forecasts to SPI observations, as opposed to the traditional method of bias-correcting monthly raw rainfall forecasts before converting them to SPI values. The figure below illustrates the results of bias correction using a transfer function based on monthly rainfall before SPI conversion. We found that using an SPI-based transfer function offers greater potential for enhancing severe drought forecasts.

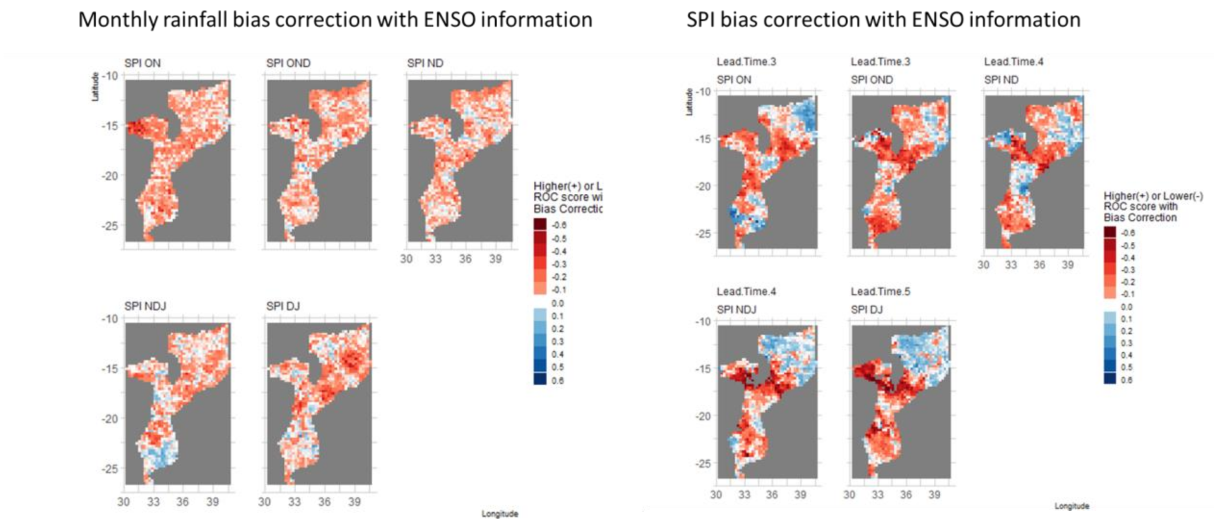

Figure 3 The plot displays differences between the AUROC (Area Under the Receiver Operating Characteristic) scores for severe events using bias correction (BC) methodology versus the AUROC scores from raw forecasts. Regions highlighted in blue indicate an improvement in forecast skill due to bias correction, whereas regions in red show a decrease in skill. The maps reflect the forecast skill for predictions issued in July. The maps on the left were generated using ENSO information based on the monthly rainfall transfer function, while the maps on the right were produced using ENSO information based on the SPI transfer function.

To enhance the clarity of the bias correction methodology, we have revised the flowchart for the ENSO-informed quantile mapping (figure 4) and extensively edited the section titled "Bias Correction of Ensemble Forecasted SPI 2 and SPI 3." Additionally, we have improved the Discussion section to emphasize potential future improvements to the methods.

In the results in line 359, there appears to be an error in the sentence structure. The reference to "frequency" lacks clarity; it's unclear whether it pertains to the occurrences of values below -1. Given that <-1 represents the definition of severe drought, consistency in its frequency is expected across regions. Regarding Figure 4, clarification is needed on whether the counts on the y-axis represent aggregated data across ensemble members. Additionally, the observation that only 24% demonstrate skill improvement with bias correction raises questions about the efficacy of this effort. It may be beneficial to revisit this finding for accuracy. Furthermore, the persistent low hit rate in multiple regions for the "ready set go" approach, as depicted in Figure 7, suggests room for improvement. Consider exploring potential adjustments to enhance the effectiveness of this method.

Please reflect on the identified limitations in the discussion, as I feel the most important bottlenecks, such as on SPI usage, bias correction methodologies, distribution fitting techniques, and threshold selection, are barely discussed. Consider placing greater emphasis on these potential methodological challenges rather than ENSO-related variability.

Overall, enhancing clarity and addressing the identified concerns would significantly strengthen the manuscript for publication.

R6. We thank the reviewer for the comment. We have revised the text related to the former Figure 4 (now Figure 5 in the revised manuscript) and restructured the discussion to better address the improvements in bias correction and the overall methodology. Additionally, we have added a sentence in the Results section regarding the impact of the bias correction. The manuscript has been extensively revised, and we hope these changes enhance the clarity and reproducibility of the methods.

Reviewer #2:

The Manuscript is excellent, and explores a valuable strategy that is relevant to many Anticipatory Action projects, showing that a "double checking" trigger is valuable at improving triggering skill. I recommend that the authors only perform minor revisions to make the paper more valuable and impactful for those who may cite it.

We thank the reviewer for the positive feedback. We are pleased to hear that the reviewer recognizes the relevance of this manuscript.

ECMWF 5.1--The authors utilize ECMWF 5.0, evaluate its skill, and highlight the importance of validation. If I understand right, ECMWF 5.1 has just been released with no hindcasts between the end of the training period of 2016 and something like 2022. For validation for humanitarian use, it is important for ECMWF to generate the hindcasts to address that gap. It would be valuable for the authors to discuss this issue, as it may provide motivation for ECMWF and other models to generate hindcasts for validation. Of course if they are aware of, or uncover an existing plan for the generation of the data, that would be valuable for the reader to know.

R1. Thank you for your comment. The ECMWF SEAS5.1 model was released in November 2022, including an extended hindcast dataset back to 1981, which can be freely accessed from the Copernicus Climate Data Store. The findings in this manuscript are based on SEAS5, but an update to extend the verification period to 1981 is planned for next year. We have added a note regarding the availability of SEAS5.1 to the Methods section.

2) Rainfall validation and crop yields, humanitarian drought impacts. The authors touch on the difference between equating drought to rainfall and further relating it to forage, cropping losses, or humanitarian impacts. Because there is such a large gap in the relationship of rainfall and drought impacts (perhaps as large as the error between forecasts and rainfall), it would be important for the authors to discuss this in more depth. For example, the drought of 2015 in nearby Malawi (a year mentioned in the introduction of the paper), where rainfall in a water stress model was used for triggering ARC national insurance, there was a globally publicized gap between the rainfall formulas and agricultural losses, leading to articles in venues such as the Economist. There is a large literature on this gap for index insurance, it directly using rainfall in more detailed formulae than forecast anticipatory action projects. It would be valuable for the authors to cite some of the literature on rainfall basis risk/loss mismatches to highlight the importance of investigations that compliment rainfall forecast skill with rainfall relationships will losses.

R2. Thank you for this valuable comment. We agree that rainfall-based thresholds do not always align well with drought-related losses, as the relationship between drought risk and impact is often location-specific, non-linear, and influenced by non-climatic factors such as vulnerability. Ideally, selecting anticipatory action (AA) thresholds and indicators should ideally involve historical analysis of the connection between drought events and socio-economic impacts like crop yields, income losses, health outcomes, and food security. Unfortunately, comprehensive, downscaled impact data is largely unavailable, especially across African countries, which limits our ability to refine thresholds and indicators and to establish strong links between drought conditions and past impacts. We have addressed this issue in the revised manuscript by emphasizing the need for future efforts to refine thresholds and strengthen the relationship between physical drought hazards and expected impacts. Additionally, we have included several studies highlighting the gap between agricultural losses and rainfall to expand on this important issue.

3) AA projects relevance. This paper provides a valuable set of insights for anticipatory action projects, well beyond the case study. The authors should mention that there is a family of project (eg WFP, OCHA) that their findings could be relevant to.

R3. We thank the reviewer for the valuable suggestion. We have added an overview from the Anticipation Hub detailing the number of organizations and countries currently engaged in ongoing anticipatory action pilot programs in the introduction section.