Interactive comment on “Seasonal forecast verification and application in times of change” by Yoav Levi and Itzhak Carmona

Reply to Anonymous Referee #1

We appreciate the time and effort invested by the referee in reviewing this manuscript. We thank the reviewer for the helpful comments which we will be address to improve the manuscript, as highlighted below in blue:

To facilitate the assessment of risks, seasonal forecasts are usually expressed as the deviation relative to a 30-yr average condition for end-users. However, the climate and socio-economic changes during this period may enable the deviation questionable. To eliminate these changes, this paper introduces a method, which uses the previous year’s season condition as a reference for the next season, to evaluate the skill and usefulness of the ECMWF system 4 (Sys4) seasonal June-July-August (JJA) temperature forecasts, wherein the ECMWF ERA-Interim reanalysis is employed as the surrogate observations. The results may be valuable for the policy-makers who are interested in the application of seasonal forecasts. The paper is organized logically and well writing. However, I still have a few concerns that need to be addressed.

We are grateful for these comments which confirms us in our concern about using seasonal forecasts.

Major comments:

1. Policy-makers or end-users may have great interests on the impacts of extreme event to the socio-economy, agriculture, and environment. In this paper, however, the metric of “Fiasco score” was constructed based on the above-normal and below-normal condition which does not simply equal to extreme event. The skill of Sys4 in terms of capture the extremes event should be added or discussed in the context.

One of the limitation of current seasonal forecast is that the forecast is given as an average value for 3 months. Obviously a situation with an extreme above normal episode (precipitation or high temperatures) can be followed by an extreme opposite situation causing the whole period to be "normal".

Although desired by end users currently there is no evidence that long range forecast can predict short term extreme events. In order to evaluate the skill for extreme events the daily or even hourly data are needed, while monthly averages are presented here.

This comment raises an interesting idea to check the Sys4 skill to forecast the most extreme season in the hindcast period (the coldest and warmest season). This can hopefully be performed.

2. As Figure 6 shows, the forecast skill derived by the new method (i.e., using the previous year's season as a reference for the next season forecast) tends to be lower than that derived from the traditional 30 year reference period. As well, the authors concluded that the new method is obviously not suggested to replace the robust traditional 30 year reference period (see Line 241-242). If so, why you attempt to use it?

The traditional 30 year reference period is the WMO guideline for climatologists. As a precaution measure we suggest to use both the 30 years climatology and the previous year as a reference for the next year forecast. However, there are relatively large areas over land that the new method over performs the 30 years reference period as indicated in Figure 6 and more visually in Figure 4. In order to be more decided we will add a case study to demonstrate the problem of using the 30 years climatology.

3. In my view, the precipitation or streamflow forecasts, as compared with temperature, may have more significance on end-users practice. Why you put the primary attention just on temperature, rather than precipitation?

Crop yield is also sensitive to temperatures and usually droughts come together with warm temperatures anomalies. However the main reason we choose the temperature field is the aim to check the conception that the temperature forecast skill is higher than precipitation. As shown in the paper a large source of temperature forecast skill originates from climate change.

Furthermore, the precipitation variance is usually larger than temperature and unlike temperatures which is usually distributed normally, precipitation may be right-skewned as there is a limit at zero and a long tail of high precipitation events. This may cause a situation where the seasonal precipitation value is below average but the tercile category will be above average. Therefore, the method for precipitation should be tested separately with even more caution.

Specific comments:

1. Section 2 – Model data. Please provide the archive website for ECMWF system 4 data and ERA-Interim reanalysis, respectively.

The ECMWF system 4 data is not open freely and only ECMWF members can retrieve the data. The ERA-Interim reanalysis is now open source so a link to the download site will be added.

2. Line 94-95. Please add an example table to explicitly illustrate the contingency table, so that the readers are more easily to understand it.

A new figure with a case study will hopefully illustrates the contingency table and the idea of the "Fiasco score".

3. Line 101-102. Please provide more details (e.g., equations) for the calculation of AUROC.

From the new case study that will be added also the AUROC will be calculated in order that readers can more easily understand the method.