This short paper investigates the predictability of the Marine Heatwave that occurred in the Mediterranean Sea in summer 2022. Using short-term (10 days) forecasts from Copernicus Marine Services, it examines if this system is able to successfully forecast the onset, evolution, peaks of this extreme event, first at basin scale, and then at two specific locations of economic importance. The main message is that short-term forecasts exhibit some skill in predicting details of the MHW evolution a few days in advance, which is of importance for guiding management.

This paper addresses an important challenge. Marine Heatwaves have recently received much attention because of their damaging impacts on marine ecosystems; several events have struck the Mediterranean Sea in recent years. It is of great importance to investigate their potential predictability at different lead times. While several papers focused on seasonal forecasts, this study focus on short-term (10 days), which has received little attention so far.

The paper is well written; the approach (first at basin scale, then at local scale) seems relevant to the objectives, and the results are worth being published.

We thank the reviewer for their careful reading and instructive comments which have helped improve the quality of the manuscript.

Yet, I have two major concerns:

- the analyses presented are mostly qualitative, and would deserve more quantification. Sentences like " the area of MHW conditions was also well predicted" (line 116) or "MHW occurrence [...] correctly forecast" (line 137) or "upon visual inspection the forecast temperature was very similar to the observed" are subjective. What is a "good" forecast? How can it be defined? What are the acceptable errors? Spatial correlations, quantification of the differences, or other metrics to better quantify these are needed. Figure 4 is not fully discussed, yet the results do not seem very encouraging for 7-days lead time. Are the forecasts "good" for this lead-time?

  These sentences have been backed up with quantification or reworded. For example, Lines 126-132 now refers to the timing of MHWs being correctly forecast. "Accurate" is reserved for statements backed up by quantification. We have added Table 1 to show

Root-Mean-Square-Differences (RMSD) of the forecasts shown in Figure 4, with values normalised by the standard deviation to show where errors exceed the natural variability. Figure 4 now includes forecasts of MHW area, which allows us to state whether overestimates arise from errors in area or intensity.

Given that these indicators (activity and area) have not been studied in short-term forecasts before, it is not possible to compare. The measures of normalised RMSD show that these indicators are predicted as accurately as atmospheric variables (values <1).

Table 1: Root-Mean-Square Differences of forecasts of summer 2022 MHW activity and atmospheric conditions (Fig, 4). Values in parenthesis are RMSD values normalised by standard deviation over the summer. Differences in MHW activity and area are relative to reprocessed satellite observations, while differences jn T2M anomaly and windspeed are relative to ECMWF analysis. Each column corresponds to a different lead time.

RMSD (Normalised)	Lead: 1 day	Lead: 4 days	Lead: 7 days
MHW Activity	0.16 (0.48)	0.20 (0.59)	0.28 (0.82)
MHW Area	8.88 (0.33)	11.65 (0.43)	16.50 (0.61)
T2M Anomaly	0.18 (0.21)	0.31 (0.38)	0.52 (0.62)
Wind Speed	0.22 (0.18)	0.52 (0.42)	0.94 (0.76)

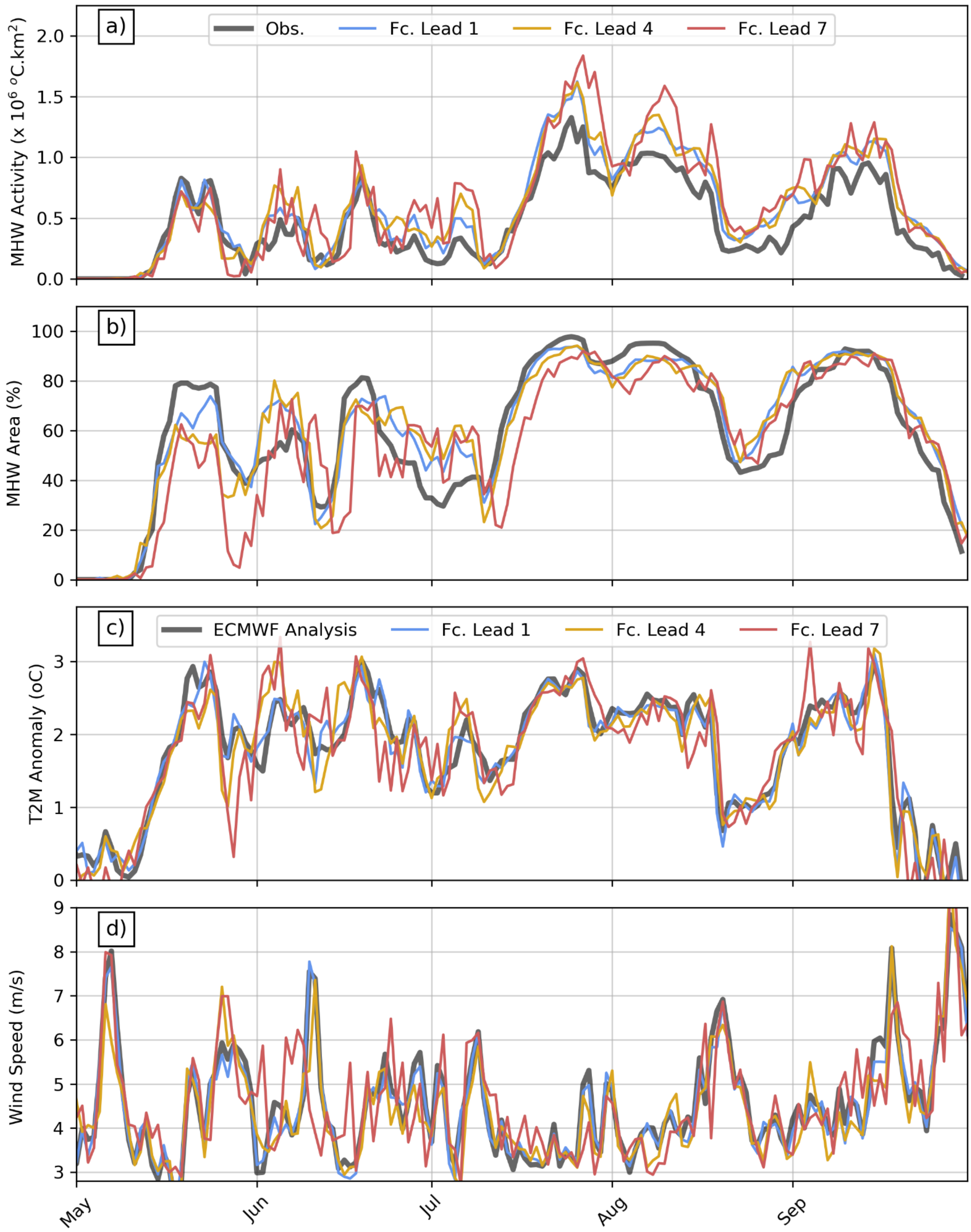

Figure 4: Effect of lead time and atmospheric forcings on forecasts of MHW activity and area. Comparison between reprocessed satellite observations and forecasts of 2022 MHW activity (a) and area (b). Each forecast time series corresponds to a different lead time (I.e. how many days in advance the forecast was made). Forecasts of MHW activity was calculated for forecasts initiated every day; the lead time from each forecast was extracted to construct the time series. Area-averaged 2m temperature anomaly (c) and wind speed (d) from the ECMWF analysis and forecasts used to force the MedFS system. All time series correspond to the Western Mediterranean Sea (Fig. 2).

- The analyses remain descriptive, without analyses on the processes that can lead to improved or degraded forecasting capability. Although I understand that a full characterization of the 2022 MHW drivers with a heat budget may be beyond the scope of this paper, it would be useful to provide more hints on the underlying processes, so that the scope of this study would not be restricted to this particular event but would provide information useful for other upcoming events.

  To Figure 4 we have added forecasts from the ECMWF system, used to forecast the MedFS system, of T2M anomaly and wind speed. We highlight that Figure 4 now corresponds to the Western Mediterranean, not the full basin, to facilitate a study of drivers. We believe that area averages of T2m and wind speed over the whole basin will lead to masking signals; the Western Med is chosen as a compromise between studying drivers and representing the extent of the MHW.

In particular, one interesting result is the contrast between the two periods in early summer and late summer (May-mid-June versus July-August-September) as seen in Figure 1b. The operational system exhibit better forecasting skill during the early period, whereas it fails to predict the sharp increase in activity (intensity) in mid-July, early August, early September. Looking at the three specific periods to examine which processes in the real world led to MHW decrease instead of intensification, would help to understand these false alarms. Line 134, the authors suggest as potential candidates "cloud cover or winds". Would it be possible to investigate this?

As mentioned above, we include an analysis of forcing forecasts of T2m and wind speed. Line 252-262: *"The decreases in skill with lead time can partly be explained by the decrease in skill of the ECMWF atmospheric forecasts used to force the Mediterranean Sea forecasts system. T2M and wind speed correlate strongly and significantly with the MHW activity (correlation values of 0.89 and 0.50 with the ECWMF analysis respectively), evidencing their role in MHW formation. Errors of forecasts of T2M and wind speed grow with time but do not exceed natural variability at lead time 7 (Table 1). In the first half of the summer, forecasts at lead time 7 of both T2M and wind speed are frequently out of phase with the observed changes. In fact, the underestimations of MHW area in this period occur simultaneously with underestimations of T2M. For example, the underestimation of MHW area at the end of May, by an area of roughly 30% of the western Mediterranean, corresponds to* *overestimations of wind speed by up to 1 m/s and temperature anomalies roughly 1°C weaker than observed. However, the overestimation of activity in July and August, found to be linked to overestimations of SST, does not correspond to overestimations of T2M, implying that other phenomena are not well represented. It should be noted that the use of area-averaged atmospheric variables may hide sub-regional scale processes which impact the MHW location and intensity."*

Finally, some details are missing. The area chosen for the definition of the Ligurian Sea and Gulf of Taranto (shown in Figure 3) are not given. In Figure 4, it should be stated that the Figure is the basin average.

We thank the reviewer for spotting these missing details. The areas have been added to Figure 2. Figure 4 now corresponds to MHW activity (integrated over the western Mediterranean Sea) and MHW area. The atmospheric variables in Fig. 4c and 4d are area-averaged.

Minor comments:

-line 24: McAdam et al. 2023 reference is missing

We have now added this reference.

-line 34: I disagree with the statement that 5 days of longer are harmful to marine life. This is completely species-dependent. The Hobday et al. (2016) definition is just a practical one, without reference to ecosystems impacts.

We have updated this sentence to highlight the reviewers point: *"The definition of MHWs assumes persistent conditions are harmful to marine life if the duration is 5 days or longer (Hobday et al., 2016), although this number is quite arbitrary and in principle should be species-dependent. "*

- line 74: no trend is removed. This choice may be usefully justified, given the recent debates from the community about removing or not a trend for MHW detection

We have added to description of how MHWs are calculated: Line 98: *"Although there are benefits of detrending SST prior to detecting MHWs (Amaya et al., 2023), we chose not to detrend in order to present the true values of temperature because they are of more relevance to species impacts (e.g. Galli et al., 2017)."*