esd-2020-37: “A new view of heat wave dynamics and predictability over the Eastern Mediterranean” by Assaf Hochman, Sebastian Scher, Julian Quinting, Joaquim Pinto and Gabriele Messori.

Point by point response to Reviewer 2:

Reviewer 2: The study titled “A New View of Heat Wave Dynamics and Predictability over the Eastern Mediterranean” by Hochman et al., presents a fresh viewpoint on the dynamics and predictability of heat waves in the Eastern Mediterranean by using both dynamical system theory and an ensemble of NWP models. This study continues the line of previous papers by the authors, dealing with the dynamics of cold spells, and weather regimes in the Eastern Mediterranean in general, which is an interesting viewpoint on a well-studied subject. This study stands out, by showing comparisons to ensemble NWP results, which seems really promising and puts this viewpoint in perspective. The study presents the climatology of the heat waves from this viewpoint, along with a complementary analysis of back-trajectories, showing the origins of air parcels – which is interesting on its own. Furthermore, a specific heat-wave case-study is presented and analyzed, showing the evolution of this unique heat wave. To my view, the study addresses relevant and current problems, and shows a novel concept to answering such problems. I think this paper is worth publishing in ESD, after addressing some points, as follows:

Response: Thank you for the positive feedback. We plan on addressing all of the Reviewer’s comments in the revised version of the manuscript as described below.

Reviewer 2: Sect. 2.3, and in general: I feel that there is not enough “intuition” in the description of the dynamical system metrics. For example, it is hard for me to understand what exactly is the meaning of θ^{-1} (“quantifies the persistence of the system in the neighborhood of the state of interest, and tends to be very sensitive to small changes in the state of the system”). I do realize this notion was already mentioned in quite a few papers in recent years, however, I feel that me (and the other readers) of this paper can benefit from a more intuitive explanation of the metrics. Moreover, the possibility of other readers to repeat such an analysis is limited by the fact you refer to other studies on how to estimate the parameters. I believe the parameter estimation should be further detailed, at least in the supplementary material, including the report on the errors in estimating the parameters.

Response: Thank you for this comment. We understand the need for both a clearer intuitive explanation, to aid the understanding of our study, and a more complete analytical derivation of the metrics, to ensure that our study is self-contained and allows reproducibility of our results. To this end, we have completely re-structured Sect. 2.3 to provide both a clearer intuitive explanation of the metrics and a more detailed description of the mathematical background. We also reference earlier uncertainty estimates for the calculation of the parameters.

Reviewer 2: It could be worthy to expand much the discussion about the differences between the intrinsic and the practical predictability. Such a discussion could be exampled and explained using the presented case study, by showing how exactly can you elevate the dynamical system theory in predicting this heatwave better than using only the ensemble of NWP. Could this be performed and displayed in the paper?

Response: Thank you for this important comment. The practical predictability relies on the performance of a numerical forecast model. As such, it blends model and data assimilation biases with the intrinsic characteristics of the atmospheric flow. Moreover, when restricting the analysis to a single realization of an ensemble forecast, such as in our case study, even a perfect ensemble may not provide a good skill-spread relationship. That is, even a perfect ensemble may have a spread that does not reflect the actual forecast error. In the specific case we analyze here, Tmax spread and error were well correlated, and matched an increase in local dimension. In other cases, the ensemble forecast can have, *posteriori*, a bad spread-error relationship. In these cases, local dimension and/or persistence trends that seem to contradict a low ensemble spread may serve as warning of a potentially poor spread-error relationship. We will extend the discussion on how exactly can the dynamical systems metrics be used to better predict heat waves, and support this by providing an example of a heatwave that displays the undesirable spread-error characteristics described above.

Reviewer 2: Another point which I think you should address, is the portion of the back-trajectories which is terrestrial vs. the marine portion. It seems to me there could be much of a difference between the heat waves and the cold days. Am I correct?

Response: Thank you very much for this suggestion. Our additional analysis reveals that for most of the time, the portion of terrestrial back-trajectories is similar for heat waves and cold days (Figure R1). It is only 72 to 24 hours prior to the events that the portion of terrestrial (marine) back-trajectories is lower (higher) for cool days than for heat waves (Figure R1). This is in line with the evolution in specific humidity along the trajectories, which increases more strongly for the cool days than for the heat waves (Figure 2f). The increase of moisture is most likely related to the passing of the air masses over the Mediterranean Sea. We will revise the text accordingly and consider adding Figure R1 as a Supplementary Figure.

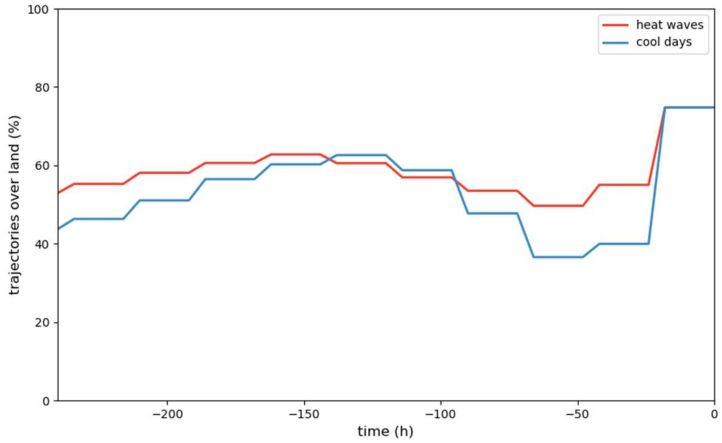


Figure R1 The portion of trajectories over land for heat waves (red line) and cool days (blue line).

Reviewer 2: L44: "(e.g., Goldreich et al., 2003)". Consider citing (Kushnir et al., 2017).

Response: The reference will be cited in the revised version of the manuscript.

Reviewer 2: L48-L52: some of this description seems appropriate to describe the generally mild temperatures and small inter-daily variability which was mentioned at the beginning of the paragraph. Please consider having some of this description moved to the beginning of the paragraph, and leave out only the part which is unique to low-temp days (possibly emphasize the role of the upper-level trough).

Response: We will revise the paragraph according to the Reviewer's suggestion.

Reviewer 2: L58: "~55,000 excess deaths": where?

Response: In eastern Europe western Russia (Barriopedro et al., 2011; Katsafados et al., 2014). We will update the text accordingly.

Reviewer 2: L62: "record-breaking": in terms of duration? Extent? Temperatures? Please address this in the text.

Response: We will address this in the text. Indeed, record-breaking here is both in terms of temperatures and duration (https://ims.gov.il/sites/default/files/aug10.pdf).

Reviewer 2: L69: "A framework": which framework? Is it yours or in general?

Response: The general framework of the Lagrangian back trajectories. We will clarify this in the text.

Reviewer 2: L133: The term "Etesian winds" was not introduced before. To me it seems like a good idea to present it in the paragraph describing the summer climatology of the Eastern Mediterranean.

Response: We appreciate this term may not be known to all readers, and we will introduce it in the paragraph describing the summer climatology of the eastern Mediterranean according to the Reviewer's suggestion.

Reviewer 2: Sect. 2.2: please elaborate on why the CSI is better at describing heat waves than, e.g., the temperature alone. This could be done by using examples or just a further explanation on other effects these heat waves consist of.

Response: We will elaborate on the advantages of the CSI index with respect to temperature alone. Indeed, this index incorporates the inversion base height and the heat stress. It therefore includes also humidity and circulation rather than only temperature (Saaroni *et al*., 2017). This means that the CSI relates more directly to the impacts of a heatwave on, for example, human physiology, than a conventional temperature-based measure (Epstein and Moran, 2006).

Reference:

Epstein Y, Moran DS. Thermal Comfort and the Heat Stress Indices. Vol. 44, Industrial Health. 2006.

Reviewer 2: L174-179: Does this seasonal cycle related to the synoptic-scale circulation? If it is, I am not sure why it is reasonable to subtract it from the data.

Response: We show that the seasonal cycle of the dynamical systems metrics is related to the synoptic scale circulation. Since we are comparing individual days/events during different parts of the summer season, it is better to de-seasonalize the data in order to study the anomalies. I.e., we test whether heat waves are synoptically and dynamically unusual with respect to the other days in the same part of the season. If we did not do this, the seasonal cycle would be "embedded" into our anomalies, biasing the results depending on which part of the summer season the selected episodes occurred in. As a practical example, imagine a scenario where the seasonal cycle of *d* peaks during month "x" of summer, and is lowest during month "x+2". We next consider a heatwave during month "x" which has a *d* in-line with climatology, and a heatwave in month "x+2" which has a *d* above climatology for that month, but lower than the *d* of the heatwave during month "x". If we did not de-seasonalize *d* we would draw the incorrect conclusion that the first heatwave has an unusually high *d* and that the second heatwave has a low *d*. We will clarify this point in the revised version of the manuscript.

Reviewer 2: L190: Could you write explicitly if this interpolation is done on the horizontal axis only or on the vertical axis as well?

Response: We will clarify that the interpolation was done on the horizontal axis in the revised version of the manuscript.

Reviewer 2: L191: I did not understand what the reason was for choosing 69 hours as the lead time. Please elaborate.

Response: We will rephrase the paragraph in the methods section to clarify the choice of 69 hours. Indeed, The GEFS reforecasts are initialized at 00UTC and are available at three-hour intervals. Since our analysis focuses on heat waves, we estimate the spread/skill for maximum temperature and SLP at a lead-time of 69 hours, while the maximum temperature is defined between 45 h and 69 h. Given the three-hour interval of the forecast data, and bearing in mind that each station's maximum temperature is recorded between 20UTC and 20UTC of the next day, this time-window roughly corresponds to the definition of maximum temperature for the station data. We did however test a 24-hour shorter lead time and found that our main conclusions remain the same (see for example Figure R1).

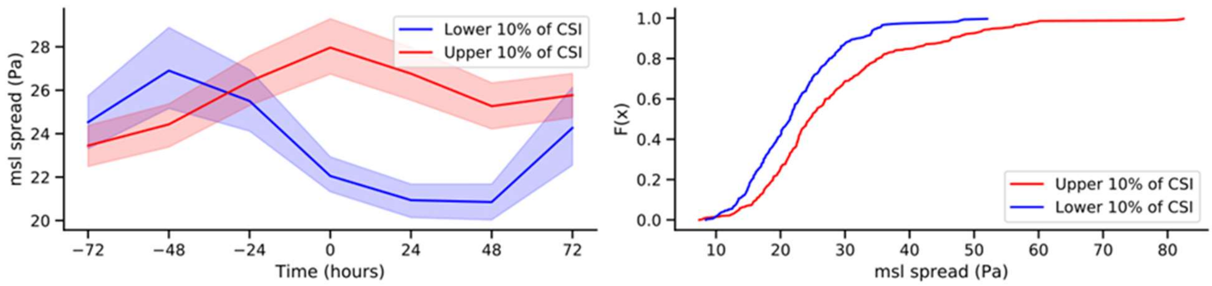


Figure R2 Same as Figure 5e, f but for a 24-hour shorter lead time.

Reviewer 2: L200: the bootstrapping and the statistical tests are already mentioned elsewhere, and to my opinion should not be detailed twice. However, it will be good if you could explain what was the variable on which the bootstrapping was applied on, and how many repetitions were made.

Response: Thank you for this suggestion. We will mention the statistical tests used only in Section 2.3 and elaborate that we bootstrapped the mean of the events and that the number of repetitions in the bootstrapping test is 10^4.

Reviewer 2: L209-210: where do you start the trajectories from? Are they spread all over the domain? Is it only from the 5 stations?

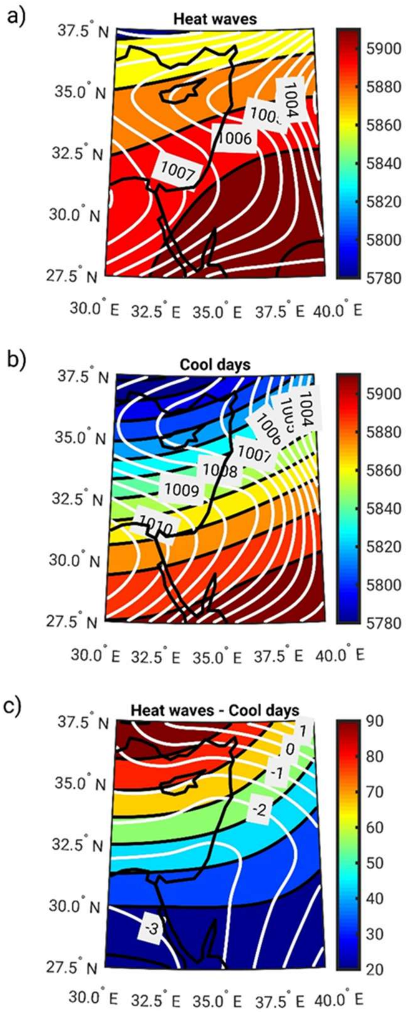
Response: The trajectories are initialized from fixed points in the whole domain. We will specify this in the revised text.

Reviewer 2: L224: "upper level ridge" vs. trough: please write more precisely, that the center of the high is to the southeast of the study area, as it is hard to tell from this map weather the Eastern Mediterranean is affected by the ridge or the trough to the northwest (it actually seems in between them).

Response: We will clarify this point in the text.

Reviewer 2: L224-226: please make sure the SLP intervals are the same in panels a and b, as it is hard to understand which of these situations is associated with a deeper/shallower longer/shorter Persian trough (might be worth mentioning this as well).

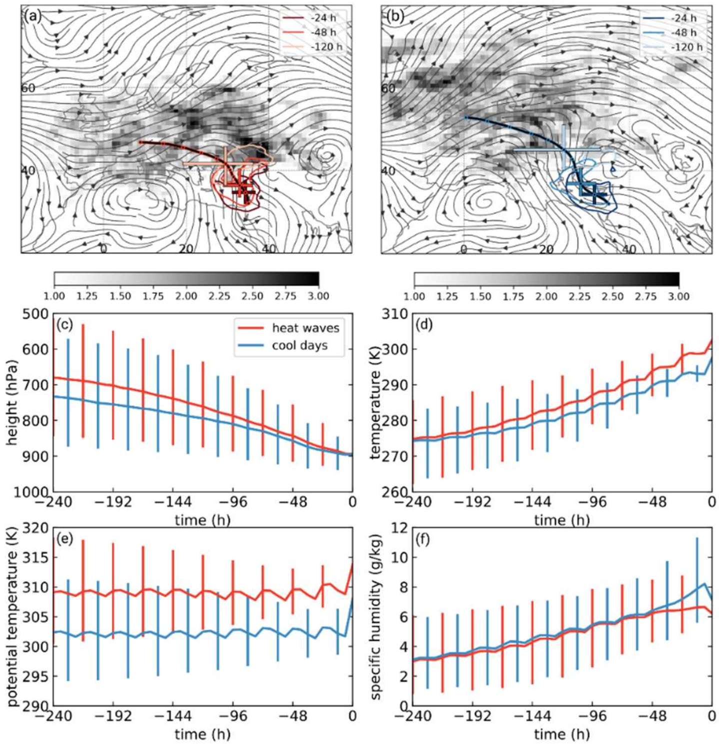
Response: Thank you for this suggestion. We will change the Figure (see below new Figure 1) so that it will contain the same intervals in Fig. 1a and Fig. 1b according to the Reviewer's suggestion.



New Figure 1 Mean sea level pressure (SLP in hPa, white contours) and 500 hPa geopotential height (Z500 in m, shaded in color) for the 10% of days with the highest (heat waves) and lowest (cool days) ‘Climatic Stress Index’ (CSI) values. (a) Heat wave days mean composite (b) Cool days mean composite (c) Heat waves minus cool days.

Reviewer 2: L227-242: I am not convinced that the median back-trajectory is a good representation of the paths of the air parcels. For example, the median track for the cool days is out of the highest density region (if I understand the plot correctly). This means, it could also be some compromise between trajectories passing over the Black Sea, and trajectories passing over the Mediterranean. Could you please explain why the median is a good representation? Could you convince me (and the readers) why should those maps not be read differently? For example, one can argue the main difference between the trajectories is that during heat waves more tracks are arriving after the passage over continental regions (Turkey), while during cold spells, tracks are arriving from the Black and Aegean seas.

Response: Thank you for this comment. We will revise Figure 2 so that it will contain the uncertainty around the median (New Figure 2). Indeed, we show that the uncertainty is relatively small in the -48 hours prior to arrival of air parcels in the Eastern Mediterranean. This exemplifies that the median trajectory is a reasonable choice to make. As a caveat, this does not rule out a bimodal distribution of the trajectories, although the streamlines may be partly used to evaluate whether this is likely in the specific cases we consider here.



New Figure 2 Same as Figure 2, but now also displaying the uncertainty of 25 – 75% around the median trajectories in crosses at different time lags.

Reviewer 2: L251: "Zero. . ." please add one of the following, or a similar description: x-axis / Time / abscissa.

Response: We will add one of the above in the revised version of the manuscript.

Reviewer 2: L267-275: Could you please give an intuition about the numbers shown in Fig. 4? For example, what does zero on the y-axis means? What is the difference between an increase of d and an increase of theta?

Response: The numbers shown in Fig. 4 are deviations from the climatology, and thus should be interpreted in a relative sense. A value of zero on the y-axis implies that the events we chose are not different from the climatology of the days they occurred in.

Concerning the second part of the Reviewer's question, d and θ relate to different aspects of the atmosphere's intrinsic predictability. The local dimension is a proxy for the "complexity" (here we use the term in a very broad and mathematically imprecise sense) of the evolution of the atmosphere about a given reference state. The persistence tells us how rapidly the evolution described by d happens. While the two metrics are overall correlated, there are cases where they diverge. In such cases, there is no exact rule to determine which of the two will "dominate", and an evaluation must be made in relation to our physical understanding of the weather phenomenon being studied. We will summarize the above points in the revised version of the manuscript.

Reviewer 2: L277-283: It is not clear to me what can we learn from the abserr graphs. Is this the error computed relative to the stations?

Response: Indeed, this is the error computed relative to the stations. Both the spread and absolute error relate to the practical predictability as defined in Sect. 1 and in Sect. 2.4. We will elaborate on the use of both scores to estimate the practical predictability in Sect. 2.4 of the revised text. Specifically, the correlation between the ensemble spread and skill of the NWP model indicates how well the ensemble describes a priori the practical predictability of the atmospheric configuration we are considering (Whitaker and Loughe, 1998).

Reference:

Whitaker JS, Loughe AF. 1998. The relationship between ensemble spread and ensemble mean skill. *Monthly Weather Review* **126**: 3292–3302. https://doi.org/10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2

Reviewer 2: L306: What do we learn from Fig. 8? Please enhance its description or cut it out of the main body (it could be transferred to the supplementary).

Response: We plan on extending the description of Figure 8, which describes the large-scale situation during the 2010 heat wave. We will especially focus on the effect the Rossby wave breaking may have on the generation of this extreme heat wave.

Reviewer 2: L390: Could you also provide a table showing the d and Theta for the analyzed times?

Response: We will add the computed d and θ metrics to the KIT open data repository (https://www.bibliothek.kit.edu/english/kitopen.php).

Reviewer 2: Figure 1: The color scale of panels a and b is not the same (panel a uses green colors in the middle of the SLP range, while panel b uses only yellows). Please match the color scales. Furthermore, please either write the interval of the SLP contours or add labels to some of them, so it could be easier to compare between the plots.

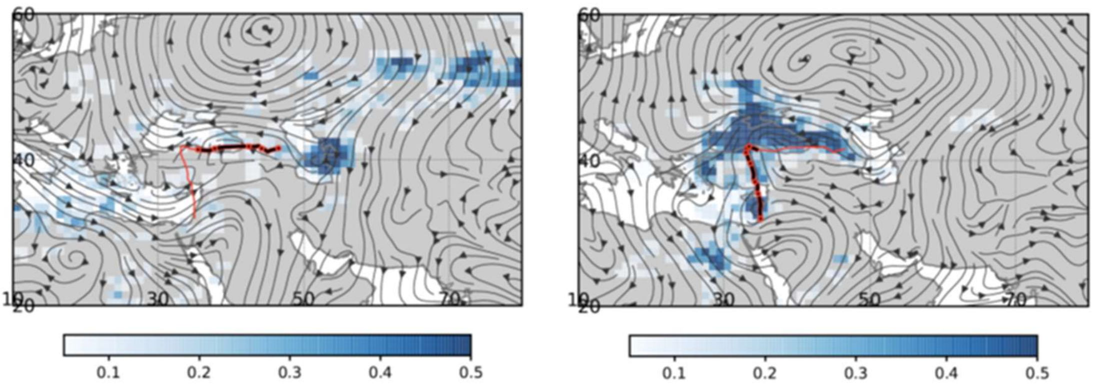
Response: We will revise the Figure according to the Reviewer's suggestion (see new Figure 1 above).

Reviewer 2: Figure 3: either the SLP and Z500 labels were swapped or their mentioning in the figure caption. Please also explain what is represented by each dot. Is it the 12 UTC d and theta from the NCEP for each of the analyzed days? If so, please write something in that spirit.

Response: We do not believe the labels were swapped, but we will recheck this and fix the labels accordingly. Regarding the second question from the Reviewer, the dots represent the daily values of d and θ for days either exceeding the 90% of CSI or below the 10% of CSI. We will add this in the caption of Figure 3.

Reviewer 2: Figure 7a, b: Could you please make the blue colors somewhat transparent? It is harder to read the map in the opaque form of the trajectory's densities.

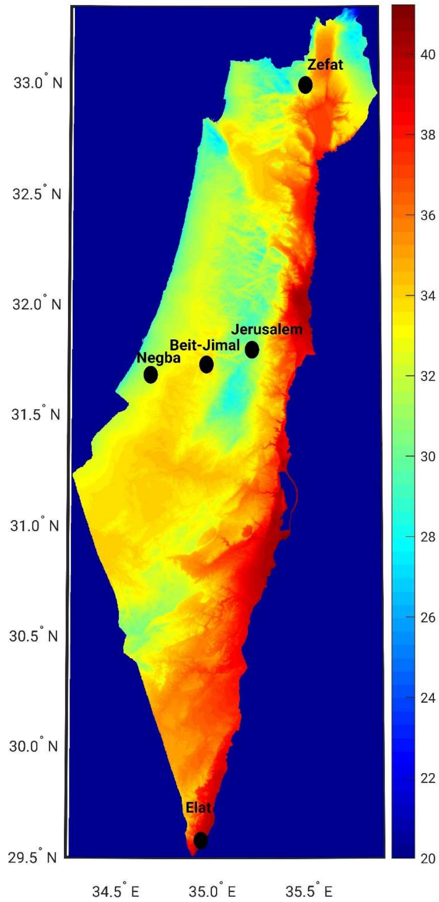
Response: We will make the colors transparent so the Figure is more readable. Please see the new version of Figure 7a, b below.



New Figure 7a, b Same as Figure 7a, b, but with transparent blue colors.

Reviewer 2: Figure S1: Please add either topography or some measure of the summer-climate (average temp. / max temp.) to the map. In this way the readers could assess why are the 5 stations are representative of the climate.

Response: We will add the average summer (July-August) temperature to the Figure, in order to better show the representative nature of the stations. Please see new Figure S1 below.



New Figure S1 The five homogenized stations on top of the average summer (July-August) temperature over Israel for 1995-2009 (https://ims.gov.il/he/climateAtlas; shading in color - ˚C).

Reviewer 2: Technical corrections L46: "On the upper levels": please consider adding the words "of the troposphere".

Response: The words will be added according to the Reviewer's suggestion.

Reviewer 2: L49: "Saaroni and Ziv,2000": please add a space before the "2000".

Response: A space will be added.

Reviewer 2: L89: Please consider deleting the word "thus".

Response: The word 'thus' will be deleted in the revised version of the manuscript, according to the Reviewer's suggestion.

Reviewer 2: L141: ". . . nine out of eleven days": please add "on average".

Response: The words will be added in the revised version of the manuscript, according to the Reviewer's suggestion.

Reviewer 2: References: Kushnir, Y., Dayan, U., Ziv, B., Morin, E. and Enzel, Y.: Climate of the Levant: phenomena and mechanisms, in Quaternary of the Levant: environments, climate change, and humans, edited by Y. Enzel and B.-Y. Ofer, pp. 31–44, Cambridge University Press, Cambridge, UK., 2017.

Response: Thank you for providing this reference. It will be cited in the revised version of the manuscript, according to the Reviewer's suggestion.