Karlsruher Institut für Technologie

Institute for Meteorology and Climate Research
Department of Tropospheric Research

KIT-Campus Nord | IMK-TRO | 76344 Eggenstein-Leopoldshafen

Prof. Sagnik Dey
Editor
Earth System Dynamics

Dr. Assaf Hochman
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen
Telefon: 0721-608-2-4942
Fax: 0721-608-2-4742
E-Mail: assaf.hochman@kit.edu
Web: http://www.imk-tro.kit.edu/
Datum: 1.12.2020

Subject: Submission of our revised manuscript on the predictability of heat waves over the Eastern Mediterranean

Dear Prof. Dey,

Enclosed please find the revised manuscript entitled: "A New View of Heat Wave Dynamics and Predictability over the Eastern Mediterranean", by Assaf Hochman, Sebastian Scher, Julian Quinting, Joaquim G. Pinto and Gabriele Messori.

We have addressed all of the concerns of Reviewer 1 and 2 according to the point by point response.

We would like to thank the editor and anonymous Reviewers for their valuable time and useful contributions, which definitely helped to improve our manuscript. We look forward to receiving feedback on the updated manuscript and would be glad to respond to any further questions and comments.

Sincerely yours,

Dr. Assaf Hochman

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 1:

Reviewer 1: In this paper, the authors employ an approach from dynamical systems theory to quantify the (intrinsic) predictability of atmospheric states based on reanalysis data during cold and hot extremes over the Eastern Mediterranean. This is complemented with GEFS reforecasts, which are used to infer forecast uncertainty, or practical predictability. While the distinction and investigation of practical and intrinsic predictability is not new (e.g. Melhauser & Zhang, 2012), I am not aware of any comparable publications in the context of heatwaves. In addition, a simple Lagrangian model is used to reveal the origin of near-surface air masses during hot and cold extreme events. The resulting paper is nicely structured, not too lengthy and certainly an interesting read. I only have two minor comments and a few additional comments, questions and suggestions, as the manuscript is well written and understandable.

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

Reviewer 1: In Melhauser & Zhang (2012), the classic Lorenz (1969) paper is cited multiple times; whereas intrinsic predictability is first defined as “the extent to which prediction is possible if an optimum procedure is used”, it is then also related to knowledge of the ‘atmospheric state’.

Now, here, the authors do not cite any study when they claim that “As opposed to the practical predictability, the intrinsic predictability only depends on the characteristics of the atmosphere itself.” Even though I understand that authors might want to stick to historical definitions, to me, it makes only little sense to limit the forecasting system to the atmosphere. An increasing amount of evidence shows that the Earth’s surface does not only supply the atmosphere with heat and moisture, but, to some extent, also exerts control over it (e.g. Koster *et al.*, 2010; Dirmeyer *et al.*, 2018). Knowledge of the land (and ocean) surface state thus implies improved predictability up to sub-seasonal timescales (thanks to, e.g., soil memory).

To be clear, I believe the focus on the atmosphere in this study makes sense, but I would still like to question why this ‘intrinsic’ predictability should be purely atmospheric by any means. To me, it seems like Lorenz (1969) emphasized the knowledge of all governing equations as well as observing the initial state, and I do not see why this would not include other components of the Earth System.

Response: We agree with the Reviewer's comment. Indeed, the atmospheric state depends not only on the atmosphere, but is influenced by interactions with land and ocean. However, it is important to note that in many – albeit certainly not all – cases these interactions influence the atmosphere at time scales longer than those we consider in our analysis (e.g., Entin *et al*., 2000), and act as a seasonal-scale preconditioning to extremely high summer temperatures (e.g. the mechanism discussed in the Zampieri *et al*. (2009) study the Reviewer cites in his second comment). We have rephrased the sentences related to this notion. We have specifically clarified that, while we take a predominantly atmospheric perspective, centered on synoptic timescales, there is ample evidence of the importance of surface interactions with other "spheres" of the climate system, not least for controlling atmospheric predictability (such as in the experiments described in the Koster *et al*. (2010) study the Reviewer pointed us to). In this respect, we read and referenced the relevant literature suggested by the Reviewer. The added text can be found in lines: 15-17, 84, 90-92, 419-422 of the revised version of the manuscript.

Reference:

Entin JK, Robock A, Vinnikov KY, Hollinger SE, Liu S, Namkhai A. 2000. Temporal and spatial scales of observed soil moisture variations in the extra tropics. *Journal of Geophysical Research* **105(D9)**: 11865– 11877. https://doi.org/10.1029/2000JD900051

Reviewer 1: Concerning the 2010 heatwave analysis: while it is interesting to show that the air parcels tracked back in time were warmer than on average even 10 days in the past in this specific case, Bieli *et al.* (2015), e.g., already found that high temperatures in the Balkans area tend to be primarily the result of high starting temperatures combined with extensive descent, enabling strong adiabatic heating. The authors also attempt to explain why particularly the metrics calculated on SLP differ so strongly from those calculated on Z500. I am using this opportunity to refer to my previous comment here – perhaps the fact that the evolution of the atmospheric state closer to the surface tends to follow less of a clear pattern, or "the larger spread in dynamical systems properties across the different heat waves for SLP than for Z500", is partially caused by interactions with the land surface. Naturally, these interactions predominantly affect the lowermost parts of the troposphere, and to provide an example, it can actually be seen (if the different units are accounted for) in Fig. 12 of Zampieri *et al.* (2009) that unusually dry soils affect SLP more than Z500 in a modelling experiment. I thus think that the 2010 heatwave part needs a bit more attention, as currently, the main message is that the single case is similar to the climatology with respect to Z500 evolution during heatwaves, but highly different in terms of SLP, and in my opinion, this is not explained sufficiently (see also comments above and concerning L. 296 below).

Response: Thank you for this comment and for pointing us to some useful references we had overlooked. We agree that the discussion of the 2010 heat wave could be extended with respect to the differences between the dynamical systems analysis close to the surface and at 500 hPa. We now discuss the possible influence air-sea interactions may have on the dynamical systems metrics at surface level in the revised manuscript. Soil moisture, on the other hand, may not be an important factor for controlling heat waves over the South-Eastern part of the Mediterranean, but rather more important in the North-Eastern areas (Zittis *et al*., 2014). It may however be possible that low soil moisture in the regions where the air parcels originate influence the intensity of Eastern Mediterranean heat waves. We have added a discussion of this point in the revised manuscript. We have further read the references suggested by the Reviewer and we referred to them in the revised version of the manuscript were applicable. The revised text can be found in lines: 361-364, 386, 417-422 of the revised version of the manuscript.

Reference:

Zittis G, Hadjinicolaou P, Lelieveld J. 2014. Role of soil moisture in the amplification of climate warming in the eastern Mediterranean and the Middle East. *Climate Research* 59: 27-37. https://doi.org/10.3354/cr01205

Reviewer 1: L. 36: Feeling the heat**,** 2018 (commamissing?)

Response: A comma has been added. This can be found in line 37 of the revised version of the manuscript.

Reviewer 1: L. 49: Saaroni and Ziv,2000 (spacemissing?)

Response: A space has been added. This can be found in line 55 of the revised version of the manuscript.

Reviewer 1: L. 126: If the CSI comprises the boundary layer height, then why is it absent from the equation below (L. 130)? Also, in the cited Saaroni *et al.* (2017), the atmospheric boundary layer height itself is barely mentioned, but rather the (height of the) persistent marine inversion. I thus recommend slightly editing (or shortening) this part to further enhance the consistency and clarity of the text.

Response: We agree with the Reviewer. We have reformulated this part of the manuscript according to Saaroni et al. (2017). The revised text can be found in lines: 59 and 135 of the revised version of the manuscript.

Reviewer 1: L. 212: Is there any reason for this choice, i.e. initializing trajectories between the surface and 90 hPa above, other than simplicity? To me, it seems more intuitive to always track the air masses within the atmospheric boundary layer back in time, whose height may vary from day to day, and tends to be (positively) anomalous particularly during heatwaves (this might not be the case in the study area of interest, but was certainly true for the 'epicenter' of the 2010 Russian heatwave; see, e.g. Miralles *et al.*, 2014). However, considering that Dayan *et al.* (2002) demonstrated onlylittle synoptic-scale influence on summer ABL heights compared to the distance to the coast, the usage of a constant layer to be tracked backward might be entirely justified, but perhaps the authors can still elaborate on their choice.

Response: Thank you for this comment. According to recent studies, the planetary boundary layer height in Israel during summer is ~600 – 900m above the surface (Uzan *et al.*, 2016; 2020). Assuming that the pressure decreases by 1 hPa every 8m height difference, 90 hPa corresponds to about 720m, thus 90hPa can be considered a reasonable choice.

We have further evaluated the trajectories that start from 90hPa above the surface and higher (Figure R1). We find that at these levels the picture is somewhat different. Air parcels associated with heat waves are located over the Mediterranean Sea and east of Israel prior to the heat wave (Figure R1a). Air parcels associated with cool days originate mostly over the Atlantic and Europe and are transported downstream (Figure R1b). This further indicates that our trajectory analysis successfully identifies air parcels in the boundary layer, which have a different path than those in the free troposphere. Moreover, the thermodynamic properties of the air parcels do not show any noticeable differences between heat waves and cool days (Figure R1c-f). We have elaborated on the choices we made with regards to the back trajectories' analysis as described above in the revised version of the manuscript. The revised text can be found in lines: 272-275 of the revised version of the manuscript.

Reference

Uzan L, Egert S, Alpert P. 2016. Ceilometer evaluation of the eastern Mediterranean summer boundary layer height – first study of two Israeli sites. *Atmospheric Measurement*

Techniques **9**: 4387–4398. https://doi.org/10.5194/amt-9-4387-2016

Uzan L, Egert S, Khain P, Levi Y, Vadislavsky E, Alpert P. 2020. Ceilometers as planetary boundary layer height detectors and a corrective tool for COSMO and IFS models. *Atmospheric Chemistry and Physics* 20: 12177-12192. https://doi.org/10.5194/acp-20-12177-2020

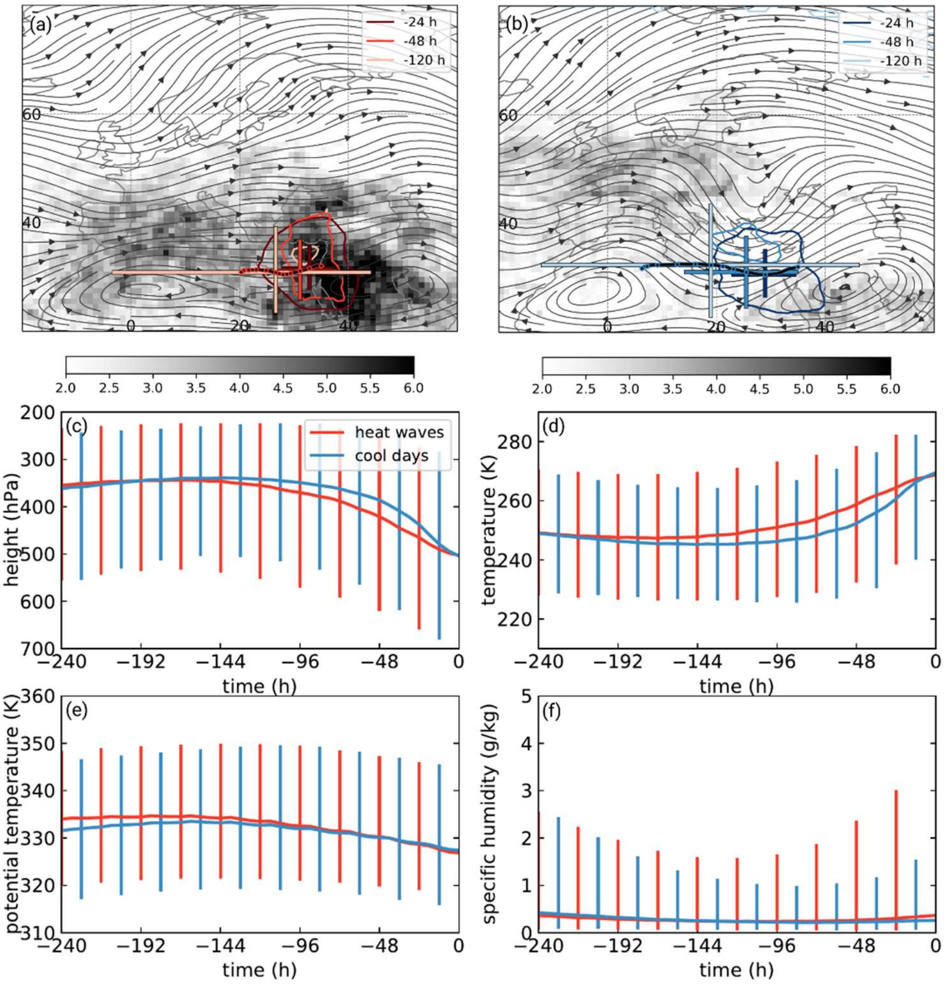
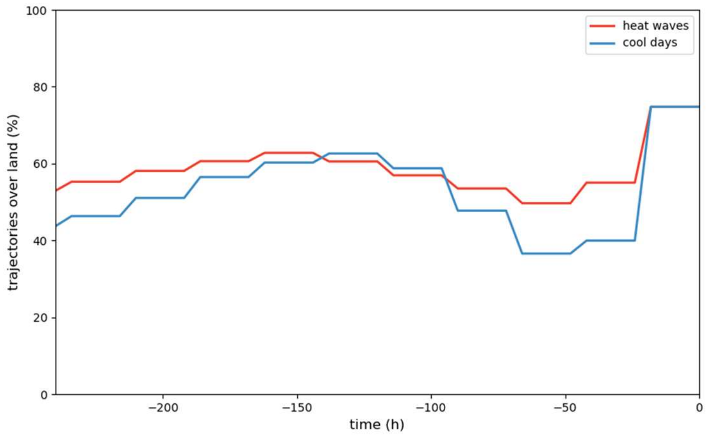


Figure R1 Same as Figure 2, but for trajectories initialized from 90hPa above the surface and higher with the uncertainty around the median trajectories of 25 – 75% of the trajectories.

Reviewer 1: L. 240: It is quite interesting that the specific humidity does not increase nearly as much in the last 48 hours prior to arrival during heatwaves as for cold extremes, but is this only a consequence of different 'inflow', i.e. more trajectories over the Mediterranean Sea? Of course, this cannot be gauged solely by a visual comparison of Figs. 2a & 2b (in which, to me, the trajectory densities shortly before arrival do not seem to differ much), but I would also suspect that additional factors are at play – such as, e.g., enhanced convective activity (and hence moistening of the troposphere).

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 cool days (new Figure S2). It is only at about 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 (new Figure S2). 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 (Fig. 2f). The increase of moisture is most likely related to the passing of the air masses over the Mediterranean Sea. As a caveat, this analysis does not account for local processes such as the convective activity mentioned by the Reviewer. We have revised the text accordingly and added Figure S2 as a Supplementary Figure. The revised text can be found in lines: 305-308 of the revised version of the manuscript.



New Figure S2 The portion of trajectories over land, as defined by the ERA-Interim land-sea mask, for heat waves (red line) and cool days (blue line). The x-axis shows time lags (in h) relative to the first day of CSI ≥ 90% or CSI ≤ 10% and at 12UTC.

Reviewer 1: L. 264: "The build-up towards this type of event is characterized by an increase in θ (decrease in persistence) and a decrease in d (Fig. 4d)." This comment also concerns the Methods section; I think the authors provide a good overview of the two dynamical system metrics, but perhaps it would be helpful to explicitly state that, as explained by Moloney *et al*. (2019), more (expanding) dimensions around the state of interest (or degrees of freedom, I suppose) imply less predictability. Or, in other words, lower d suggests higher predictability – this is actually stated as such on L. 335 in the Summary, but as far as am I concerned, not before. Perhaps it would seem a bit confusing to edit the sentence I am quoting above (L. 264), as in this example (Fig. 4d), persistence decreases, yet predictability as gauged by the local dimension d increases. Still, I believe I am not the only reader who would appreciate a bit more guidance throughout the manuscript.

Response: We have clarified these points and provided a more intuitive qualitative interpretation of the metrics earlier in the manuscript, focusing on their relation with the intrinsic predictability. To that 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 to ensure the reproducibility of our results. The revised text can be found in lines: 94-95, 135-234 and 329-330 of the revised version of the manuscript.

Reviewer 1: L. 280: "The pattern somewhat resembles the temporal evolution of d computed on SLP (cf. Fig. 5e and Fig. 4c), but stands in stark contrast to the pattern computed on Z500". While I agree that there is a stark contrast to the pattern shown in Fig. 4a (Z500), I find this resemblance a bit difficult to spot, as the peaks in d and msl spread appear to be shifted by about one day. Is there any obvious reason for this? Also, I imagine this would look different for shorter or longer lead times, so come to think of it, why not 24 hours less or more? This is not a request to repeat the entire analysis for different lead times (probably out of scope anyways), but I am just curious if the authors looked into this and if so, how much this choice even matters in the first place.

Response: Thank you for this comment. Since the spread and error are computed every 24 hours and the dynamical systems metrics are instantaneous in time (local in phase-space) and computed from the 6-hourly data, we believe that a shift of up to 24-hours may be reasonable. We have further tested a 24-hour shorter lead time and found that the peak in the spread is at about 0 h, resembling the peak in *d* computed on SLP (Figure R3, left panel). We have discussed this point briefly in the revised text, as we have indeed not described in detail the sensitivity of our results to the choice of lead-time. The revised text can be found in lines: 344-348 of the revised version of the manuscript.

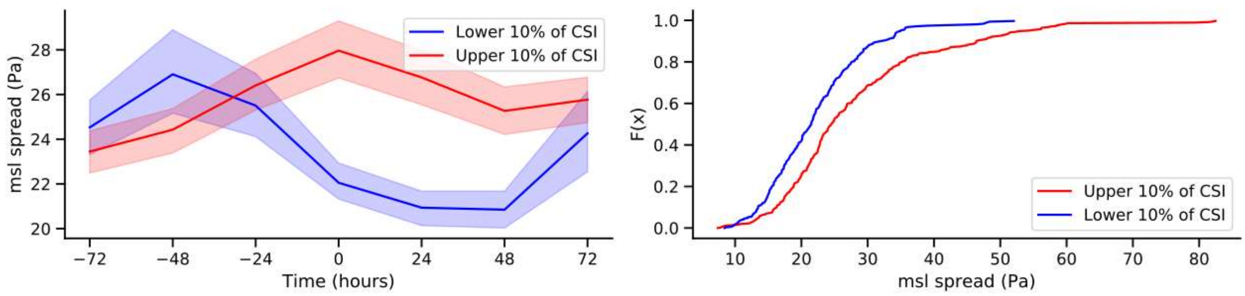


Figure R3 Same as Figure 5e, f but for a 24-hour shorter lead time.

Reviewer 1: L. 296: "We further hypothesize that differences between the single case and the climatology may be related to the relatively small day to day variations during summer over the Eastern Mediterranean, which make it challenging to depict the exact onset of a heat wave". Could you please elaborate how the challenges related to defining the onset of a heatwave could contribute to the extreme differences between the 2010 case and the climatology, but only for SLP and not Z500 – this is not obvious to me.

Response: When comparing the climatology of the temporal evolution of d and θ for Z500 (Fig. 4a) with the single case (Fig. 6b) it is relatively easy to see that in both there is an increase in d and θ as the heat wave develops. On the other hand, when comparing the temporal evolution of d and θ for SLP (Fig. 4c with Fig. 6c), one can see that depicting the exact time the heat wave starts is very important for comparison. Still, in both Figures d increases and θ decreases at some point in the chosen time window, but the timing of these trends is shifted between the climatology (Fig. 4c) and the single case (Fig. 6c). We have clarified this in the revised version of the manuscript. The revised text can be found in lines: 366-371 of the revised version of the manuscript.

Reviewer 1: L. 303: Concerning the anticyclonic wave breaking, if my understanding is correct, then this can be seen in Fig. 8, as the trough east of the ridge centered over European Russia, clearly visible from Fig. 8b onward, is tilted (southwest-northeast; Davini *et al.*, 2012) and advected westward (consistent with the definition given in Quandt *et al.*, 2019). I suggest adding a brief description along these lines for readers unfamiliar with the terminology, this would also prevent readers from overlooking Fig. 8 (which, currently, is only mentioned but not discussed in the main text).

Response: Thank you for this suggestion. The Reviewer is correct with regard to the anticyclonic Rossby wave breaking shown in Fig. 8. We have added some more information on this for the reader and discussed Fig. 8 more in detail in the revised version of the manuscript. The revised text can be found in lines: 376-381 of the revised version of the manuscript.

Reviewer 1: L. 305: (Quandt *et al.*, 2019)**,** played (comma missing?)

Response: A comma has been added. This can be found in line: 378 of the revised version of the manuscript.

Reviewer 1: L. 322 (+ 365, 511): K**ue**ne ⇒ Keune

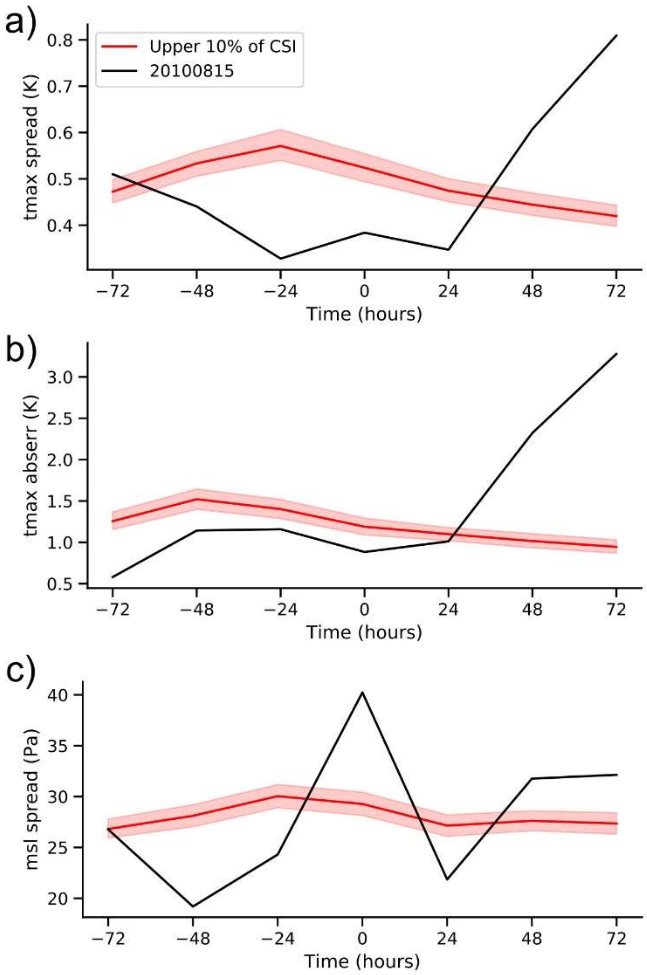
Response: The typo has been corrected. This can be found in lines: 456 and 618 of the revised version of the manuscript.

Reviewer 1: Fig. 5: While a few sentences in the Methods explain what is really shown in Fig. 5, I believe the caption might benefit from a small addition, hinting at the fact that results are plotted for their corresponding initialization times.

Response: We have expanded the caption of Fig. 5 to explain that the results are plotted for initialization time. The added text can be found in lines: 773-774 of the revised version of the manuscript.

Reviewer 1: Fig. 9: I suggest using red colors for the upper 10% of CSI, as in previous figures, and plotting the 2010 heatwave results in black (or any other color than blue) instead.

Response: We have changed the Figure according to the Reviewer's suggestion (see below new Figure 9).



New Figure 9 Forecast spread/skill for the mid-August 2010 heat wave, centered (0 h) on 15.8.2010 at 12UTC (black line). The mean temporal evolution of the ensemble model spread for Tmax (a), SLP (c) and absolute error for Tmax (b) of forecasts with lead-time 69h, initialized at different time lags with respect to the event, computed every 24 hours. The heat waves (upper 10% of CSI - red lines) are displayed for reference. A 95% bootstrap confidence interval for all heatwaves is displayed in shading.

Reviewer 1: References

Bieli, M., Pfahl, S. & Wernli, H. A lagrangian investigation of hot and cold temperature extremes in Europe. Q. J. R. Meteorol. Soc. 141, 98–108 (2015).

Davini, P., Cagnazzo, C., Gualdi, S. & Navarra, A. Bidimensional diagnostics, variability, and trends of northern hemisphere blocking. J. Clim. 25, 6496–6509 (2012).

Dayan, U., Lifshitz-Goldreich, B. & Pick, K. Spatial and structural variation of the atmospheric boundary layer during summer in Israel-profiler and rawinsonde measurements. J. Appl. Meteorol. 41, 447–457 (2002).

Dirmeyer, P. A., Halder, S. & Bombardi, R. On the Harvest of Predictability From Land States in a Global Forecast Model. J. Geophys. Res. Atmos. 123, 13,111-13,127 (2018).

Koster, R. D. *et al.* Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. Geophys. Res. Lett. 37, 1–6 (2010).

Levi, Y., Shilo, E. & Setter, I. Climatology of a summer coastal boundary layer with 1290-MHz wind profiler radar and a WRF simulation. J. Appl. Meteorol. Climatol. 50, 1815–1826 (2011).

Lorenz, E. N. Atmospheric Predictability as Revealed by Naturally Occurring Analogues. J. Atmos. Sci. 26, 636–646 (1969).

Miralles, D. G., Teuling, A. J., Van Heerwaarden, C. C. & De Arellano, J. V. G. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat. Geosci. 7, 345–349 (2014).

Melhauser, C. & Zhang, F. Practical and intrinsic predictability of severe and convective weather at the mesoscales. J. Atmos. Sci. 69, 3350–3371 (2012).

Moloney, N. R., Faranda, D. & Sato, Y. An overview of the extremal index. Chaos 29, (2019).

Quandt, L. A., Keller, J. H., Martius, O., Pinto, J. G. & Jones, S. C. Ensemble sensitivity analysis of the blocking system over Russia in summer 2010. Mon. Weather Rev. 147, 657–675 (2019).

Saaroni, H., Savir, A. & Ziv, B. Synoptic classification of the summer season for the Levant using an ‘environment to climate’ approach. Int. J. Climatol. 37, 4684–4699 (2017).

Zampieri, M. *et al.* Hot European Summers and the Role of Soil Moisture in the Propagation of Mediterranean Drought. J. Clim. 22, 4747–4758 (2009)

Response: Thank you for providing these references. We have cited them where applicable and referred to them in the revised version of the manuscript.

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 have addressed 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.

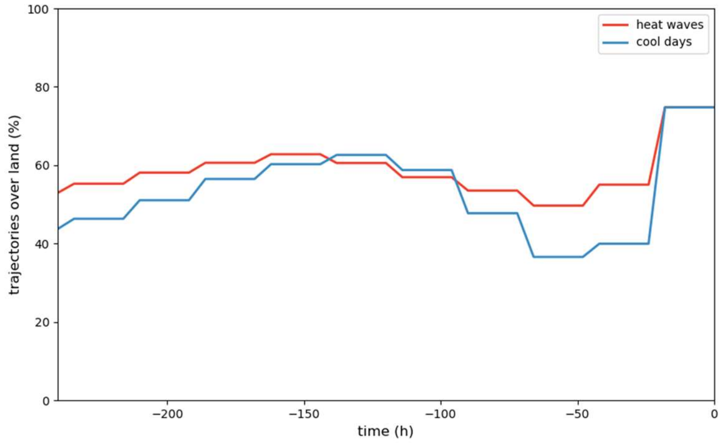
Response: Thank you for this comment. We understand the need for both a clearer intuitive explanation, to aid the understanding of our study, and a more complete analytical derivation of the metrics, to ensure that our study is self-contained and allows reproducibility of our results. To this end, we 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. The revised text can be found in lines: 156-234 of the revised version of the manuscript.

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 have extended 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 the 2003 heatwave that displays the undesirable spread-error characteristics described above. The added text can be found in lines: 442-451 of the revised version of the manuscript.

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 (new Figure S2). 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 (new Figure S2). 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 have revised the text accordingly and have added new Figure S2 as a Supplementary Figure. The revised text can be found in lines: 303-308 of the revised version of the manuscript.



New Figure S2 The portion of trajectories over land, as defined by the ERA-Interim land-sea mask, for heat waves (red line) and cool days (blue line). The x-axis shows time lags (in h) relative to the first day of CSI ≥ 90% or CSI ≤ 10% and at 12UTC.

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

Response: The reference has been cited in the revised version of the manuscript. This can be found in line: 46 of 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 have revised the paragraph according to the Reviewer's suggestion. The revised text can be found in lines: 45-55 of the revised version of the manuscript.

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

Response: In eastern Europe western Russia (Barriopedro et al., 2011; Katsafados et al., 2014). We have updated the text accordingly. The revised text can be found in line: 63 of the revised version of the manuscript.

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

Response: We have addressed this in the text. Indeed, record-breaking here is in terms of temperatures (https://ims.gov.il/sites/default/files/aug10.pdf). The revised text can be found in line: 66 of the revised version of the manuscript.

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

Response: The general framework of the Lagrangian back trajectories. We have clarified this in the revised version of the manuscript. This can be found in lines: 73-74 of the revised version of the manuscript.

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 have introduced it in the paragraph describing the summer climatology of the eastern Mediterranean according to the Reviewer's suggestion. The added text can be found in lines: 48-50 of the revised version of the manuscript.

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 have elaborated 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). The added text can be found in lines: 136-138 of the revised version of the manuscript.

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 have clarified this point in the revised version of the manuscript. The added text can be found in lines: 228-233 of 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 have clarified that the interpolation was done on the horizontal axis in the revised version of the manuscript. This can be found in line: 249 of 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 have rephrased 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). The revised text can be found in lines: 251–255 of the revised version of the manuscript.

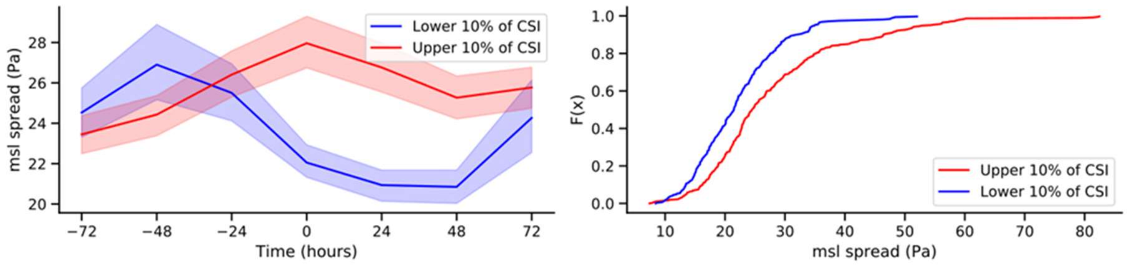


Figure R1 Same as Figure 5e, f but for a 24-hour shorter lead time.

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 now 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. The revised text can be found in lines: 224-227 and 260-261 of the revised version of the manuscript.

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 have specified this in the revised version of the manuscript. The revised text can be found in line: 269-271 of the revised version of the manuscript.

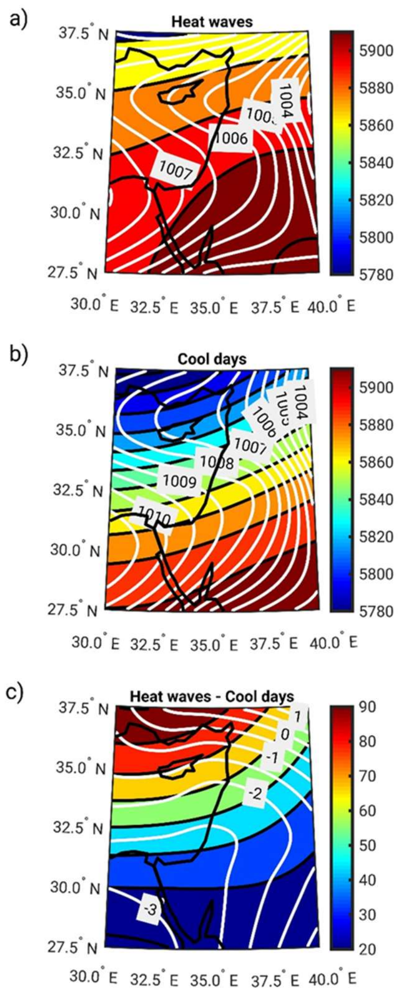
Reviewer 2: L224: "upper level ridge" vs. trough: please write more precisely, that the center of the high is to the southeast of the study area, as it is hard to tell from this map weather the Eastern

Mediterranean is affected by the ridge or the trough to the northwest (it actually seems in between them).

Response: We have clarified this point in the revised version of the manuscript. The revised text can be found in lines: 284-285 of the revised version of the manuscript.

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).

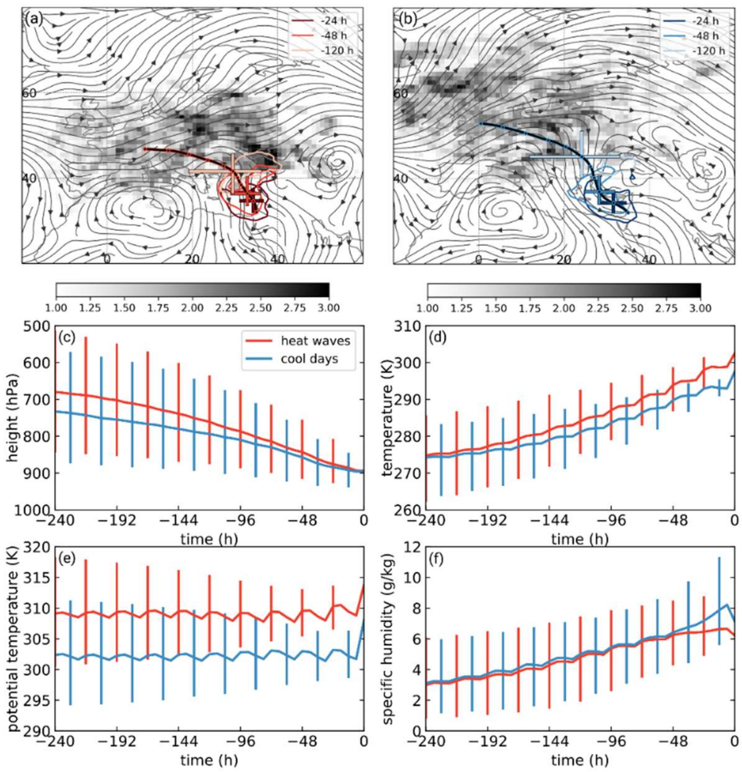
Response: Thank you for this suggestion. We have changed the Figure (see below new Figure 1) so that it will contain the same intervals in Fig. 1a and Fig. 1b according to the Reviewer's suggestion. The revised text can be found in lines: 287-288 of the revised version of the manuscript.



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 have revised 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. The revised text can be found in lines: 750-756 of the revised version of the manuscript.



New Figure 2 Same as the original 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 have added the word x-axis in the revised version of the manuscript. This can be found in lines: 317-319 of 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 have summarized the above points in the revised version of the manuscript. The revised text can be found in lines: 317-319 and 329-330 of 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 have elaborated on the use of both scores to estimate the practical predictability in Sect. 2.4 of the revised version of the manuscript. 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). The added text can be found in lines: 244-246 and 441-446 of the revised version of the manuscript.

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 have extended the description of Figure 8, which describes the large-scale situation during the 2010 heat wave. We especially focused on the effect the Rossby wave breaking may have on the generation of this extreme heat wave. the added text can be found in lines: 376-381 of the revised version of the manuscript.

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 as soon as possible (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.

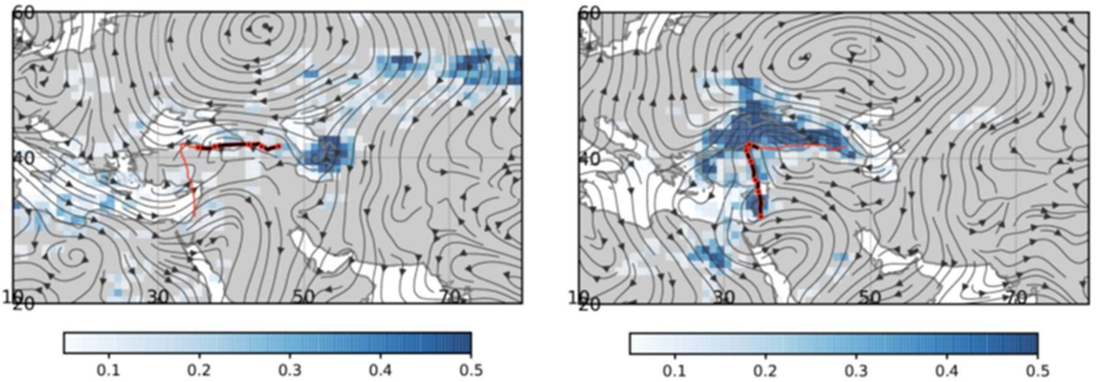
Response: We have revised the Figure according to the Reviewer's suggestion (see new Figure 1 above).

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: The labels were not swapped. 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 have added this in the caption of Figure 3 of the revised version of the manuscript. The added text can be found in line: 758 of the revised version of the manuscript.

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.

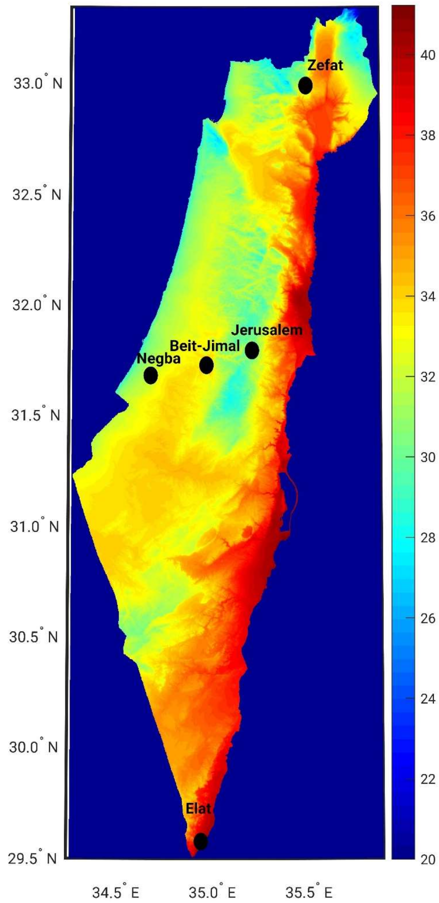
Response: We have made the colors transparent so the Figure is more readable. Please see the new version of Figure 7a, b below.



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 have added 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 have been added according to the Reviewer’s suggestion. This can be found in line: 51 of the revised version of the manuscript.

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

Response: A space has been added. This can be found in line: 55 of the revised version of the manuscript.

Reviewer 2: L89: Please consider deleting the word “thus”.

Response: The word ‘thus’ has been 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 have been added in the revised version of the manuscript, according to the Reviewer’s suggestion. This can be found in line: 152 of the revised version of the manuscript.

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 has been cited in the revised version of the manuscript, according to the Reviewer's suggestion. This can be found in line: 46 of the revised version of the manuscript.

A New View of Heat Wave Dynamics and Predictability over the Eastern Mediterranean

Assaf Hochman[1], Sebastian Scher[2], Julian Quinting[1], Joaquim G. Pinto[1], Gabriele Messori[2, 3]
[1] Department of Tropospheric Research, Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany.
[2] Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden.
[3] Department of Earth Sciences and Centre of Natural Hazards and Disaster Science (CNDS), Uppsala University, Uppsala, Sweden.

Correspondence to: Assaf Hochman (assaf.hochman@kit.edu)

Abstract. Skillful forecasts of extreme weather events have a major socio-economic relevance. Here, we compare two complementary approaches to diagnose the predictability of extreme weather: recent developments in dynamical systems theory and numerical ensemble weather forecasts. The former allows us to define atmospheric configurations in terms of their persistence and local dimension, which inform on how the atmosphere evolves to and from a given state of interest. These metrics may be used as proxies for the intrinsic predictability of the atmosphere, which only depends on the atmosphere's properties. Ensemble weather forecasts inform on the practical predictability of the atmosphere, which partly depends on the performance of the numerical model used. We focus on heat waves affecting the Eastern Mediterranean. These are identified using the Climatic Stress Index (CSI), which was explicitly developed for the summer weather conditions in this region and differentiates between heat waves (upper decile) and cool days (lower decile). Significant differences are found between the two groups from both the dynamical systems and the numerical weather prediction perspectives. Specifically, heat waves show relatively stable flow characteristics (high intrinsic predictability), but comparatively low practical predictability (large model spread/error). For 500 hPa geopotential height fields, the *intrinsic* predictability of heat waves is lowest at the event's onset and decay. We relate these results to the physical processes governing Eastern Mediterranean summer heat waves: adiabatic descent of the air parcels over the region and the geographical origin of the air parcels over land prior to the onset of a heat wave. A detailed analysis of the mid-August 2010 record-breaking heat wave provides further insights into the range of different regional atmospheric configurations conducive to heat waves. We conclude that the dynamical systems approach can be a useful complement to conventional numerical forecasts for understanding the dynamics and predictability of Eastern Mediterranean heat waves.

1. Introduction

Heat waves are recognized as a major natural hazard (e.g., Easterling *et al*., 2000), causing detrimental socio-economic impacts (e.g., Feeling the heat, 2018) including excess mortality (e.g., Batisti and Naylor, 2009; Benett *et al*., 2014; Peterson *et al*., 2013; Ballester *et al*., 2019), agricultural loss (e.g., Deryng *et al*., 2014) and ecosystem impairment (e.g., Williams, 2014; Caldeira *et al*., 2015). Moreover, heat waves are projected to increase in frequency, intensity and persistence under global warming (e.g., Meehl and Tebaldi, 2004; Stott *et al.,* 2004; Fischer and Schär, 2010; Seneviratne *et al*., 2012; Russo *et al.,* 2014). The Eastern Mediterranean has experienced several extreme heat waves in recent decades (e.g., Kuglitsch *et al*., 2010) and their frequency and intensity are expected to increase in the coming decades (e.g., Giorgi 2006; Seneviratne *et al*., 2012; Lelieveld *et al*., 2016; Hochman *et al*., 2018a) upon a background of regional warming and drying (e.g., Barchikovska *et al*., 2020).

The Eastern Mediterranean climate is characterized by wet conditions and mild air temperatures during the winter season and dry and hot weather conditions during summer (e.g., Goldreich *et al*., 2003; Kushnir *et al*., 2017). The summer season is characterized by very small inter-daily variability, which is attributable to the dominant and persistent influence of the Persian Trough and sub-tropical high-pressure systems. The interaction between these systems leads to persistent north – westerly winds of continental origin blowing across the Aegean Sea. These winds are known since ancient times as 'Etesian winds' (Tyrlis and Lelieveld, 2013). Together with the Mediterranean Sea breeze, moist air can be transported inland (Alpert *et al*., 1990; Bitan and Saaroni, 1992) as far as the Dead Sea (Kunin *et al*., 2019). On the upper levels of the Troposphere, large-scale subsidence is dominant, thus further hampering the development of clouds and precipitation (Rodwell and Hoskins, 1996; Ziv *et al*., 2004). In spite of this generally low variability, heat waves are not infrequent during the summer (Harpaz *et al*., 2014). Still, episodes when the temperature drops to below-normal values do occur, some of which are accompanied by summer rains (Saaroni and Ziv, 2000).

Saaroni *et al*. (2017) have detected weaknesses in the ability of earlier synoptic classifications (Alpert *et al*., 2004a; Dayan *et al*., 2012) to describe local weather conditions during the Eastern Mediterranean summer season. The authors proposed a 'Climatic Stress Index' (CSI), which is a combination of the national heat stress index, used operationally by the Israeli Meteorological Service, and the height of the marine inversion base height (see Sect. 2.2). The authors argued that this novel index improves the classification of heat wave days relative to earlier classifications and additionally links directly to the potential impacts.

A notable heat wave in recent years was the 2010 so-called "Russian heat wave", which caused ~55,000 excess deaths in Eastern Europe and Western Russia (e.g., Barriopedro *et al*., 2011; Katsafados *et al*., 2014). The 2010 Northern Hemisphere summer saw a strong and persistent blocking ridge at 500 hPa over the Middle East and Eastern Europe (e.g., Grumm 2011; Schneidereit *et al*., 2012; Quandt *et al*., 2017), leading to unprecedented temperatures at numerous locations (Barriopedro *et al*., 2011). The Eastern Mediterranean and Israel experienced a record-breaking (in temperature) heat wave during mid-August of that year (https://ims.gov.il/sites/default/files/aug10.pdf), which interestingly coincided with what is considered the decay phase of the Russian heat wave (Quandt *et al*., 2019). In fact, the Zefat Har-Knaan station (Tab. S1; Fig. S1) recorded a temperature of 40.6°C; the highest temperature since 1939, while the Jerusalem station (Tab. S1; Fig. S1) logged a remarkable 41°C, the absolute record for this station since 1942. The ability to predict and issue appropriate warnings for these types of events, and more generally weather events lying in the tails of the respective distributions, is of crucial importance for mitigation of impacts on human life, agriculture and ecosystems (IPCC 2012; Siebert and Evert, 2014; Williams 2014).

A general framework that allows a quantitative understanding of processes leading to extreme temperatures during heat waves is that based on Lagrangian backward trajectories. In this framework, the temperature of an air parcel increases by: (i) adiabatic warming related to descent and (ii) diabatic heating including latent and sensible heat fluxes, short-wave, and long-wave radiation (Holton 2004). Recent studies revealed that extreme temperatures during heat waves are most often a combination of adiabatic warming related to descent and diabatic heating near the surface (e.g., Black *et al*., 2004; Bieli *et al*., 2015; Santos *et al*., 2015; Quinting and Reeder, 2017; Zschenderlein *et al*., 2019). The adiabatic warming is typically associated with upper-level ridges, which promote subsidence. The strongest diabatically-driven heating does not necessarily occur at the location of the heat wave itself, but rather in geographically remote regions (e.g., Quinting and Reeder, 2017; Quinting *et al*., 2018; Zschenderlein *et al*., 2019).

Focusing more directly on the prediction of the evolution of specific atmospheric configurations, which may lead to heat waves, one may consider a partly model-dependent perspective (*practical* predictability) or a model-independent perspective (*intrinsic* predictability; Melhauser and Zhang, 2012). The practical predictability is heavily reliant on the availability of initialization data (Lorenz 1963) and on the correct representation of relevant physical processes in the numerical model being used. However, it also reflects some characteristics of the atmospheric dynamics (e.g., Ferranti *et al*., 2015; Matsueda and Palmer, 2018). An often-used method for quantifying the practical predictability is the spread or skill of ensemble forecasts (e.g., Loken *et al*., 2019).

As opposed to the practical predictability, the intrinsic predictability only depends on the characteristics of the atmosphere itself. However, it is important to note that the atmosphere is influenced and sometimes even controlled by interactions with the land and oceans, albeit mostly at longer time scales than those considered in this study (Entin *et al*., 2000; Koster *et al*., 2010; Dirmeyer *et al*., 2018). Recent developments in dynamical systems theory allow us to quantify the intrinsic predictability of instantaneous atmospheric states using two metrics: persistence (θ^{-1}) and local dimension (d). These reflect how the atmosphere evolves in the neighborhood of a state of interest (Faranda *et al*., 2017a). High (low) θ^{-1} (d) imply high intrinsic predictability, whereas low (high) θ^{-1} (d) suggest low intrinsic predictability. The two forms of atmospheric predictability depend on different factors, and therefore offer different information. While there is some relation between the two (e.g., Scher and Messori, 2018), one should not expect them to always match for individual cases (Hochman *et al*., 2020a).

In the present study, we perform a systematic dynamical systems investigation of the temporal evolution of Eastern Mediterranean summer heat waves, and evaluate whether this may provide insights complementary to a more conventional analysis of the numerical weather forecasts of such events. Specifically, we hypothesize that the dynamical systems analysis captures relevant features of these extremes, such as their persistence, which are not always reflected in the numerical weather forecast. The dynamical systems framework has recently been leveraged for the study of cold spell dynamics (Hochman *et al.*, 2020a).

The paper is organized as follows: Sect. 2 provides a brief description of the methodology, including the used datasets, the CSI index, the dynamical systems and forecast skill metrics and the method for backtracking air parcels. Sect. 3 describes the dynamics of heat waves from both the dynamical system and the numerical weather prediction perspectives and further provides a detailed analysis of the mid-August 2010 heat wave over the Eastern Mediterranean as a case study. Finally, Sect. 4 provides the main conclusions and discusses ideas for future research.

2. Data and methods

2.1 Data

The bulk of our analysis is based on the National Centers for Environmental Prediction/National Center for Atmospheric Research Reanalysis Project (NCEP/NCAR) daily and 6-hourly reanalysis data for 1979 – 2015 (satellite era), on a 2.5° × 2.5° horizontal grid (Kalnay *et al.*, 1996). Faranda *et al.* (2017a) have shown that the conclusions one may infer from the dynamical systems analysis are generally insensitive to the dataset's horizontal spatial resolution, as long as the major structures characterizing the atmospheric field of interest are resolved. On the contrary, the air parcel tracking (Sect. 2.5) requires data on a relatively high horizontal and vertical grid-spacing. Air parcel trajectories are thus computed from 6-hourly ERA-Interim data for 1979 – 2015, on a 1° × 1° horizontal grid and 60 vertical levels (Dee *et al.*, 2011).

The numerical forecasts are acquired from the Global Ensemble Forecast System (GEFS) reforecast v.2 dataset produced by NCEP/NCAR (Hamill *et al.*, 2013). Operational Numerical Weather Prediction (NWP) models are frequently updated. Therefore, archives of operational NWP models are usually inhomogeneous, and thus are not appropriate for studying predictability over long time periods. This problem can be mitigated by using so-called reforecasts. For reforecasts, one fixed version of an NWP model is used in order to create a standardized set of past forecasts. The GEFS reforecast dataset provides a set of daily reforecasts from December 1984 to present. Each reforecast consists of a control forecast and a ten-member ensemble on a 0.5° × 0.5° grid spacing.

Finally, we make use of a homogenized station dataset over Israel to assess the forecasts. Instrumental meteorological records may be influenced by non-meteorological events, such as station relocation, defects in the instrumentation, environmental changes near the station etc. The detrimental effects these may have on the quality of the recorded data can be reduced by homogeneity procedures (Aguilar *et al*., 2003). Our dataset includes five representative, homogenized stations in Israel with an uninterrupted record of maximum temperatures over 1979 – 2015 (Tab. S1, Fig. S1; Yosef *et al*., 2018).

2.2 Heat wave definition according to the Climatic Stress Index (CSI)

Saaroni *et al*. (2017) have proposed a new index for classifying the summer days over the Eastern Mediterranean based on the 'environment to climate' approach (Yarnal 1993; Yarnal *et al*., 2001). The CSI is comprised of the national heat stress index, used operationally by the Israel Meteorological Service, and the marine inversion base height, which is a major factor influencing the summer weather conditions over the Eastern Mediterranean (Ziv *et al*., 2004). The index suits the identification of heat waves as it is does not merely consider the daily temperature, but rather additional variables, e.g., humidity and circulation, which directly relate to the impacts of a heat wave on for example, human physiology (Epstein and Moran, 2006). Saaroni *et al*. (2017) have rigorously evaluated the CSI index with respect to observations and tested a variety of different combinations of predictors, which ultimately resulted in a simple multiple regression equation:

$$CSI = 92.78 + 0.638T_{1000-850} - 0.178\Delta p - 1.08p_{Iraq}$$

Here, $T_{1000-85}$ is the average regional lower-level temperature over [31°N-34°N; 33°E-37°E]. Δp is the average sea level pressure over [36°N-44°N; 42°E-54°E] subtracted from the average sea level pressure over [24°N-29°N; 33°E-37°E], which is an estimate for the intensity of the Etesian winds (see Sect. 1). p_{Iraq} represents the average sea level pressure over northern Iraq [35°N-44°N; 46°E-54°E], which is a proxy for the depth of the Persian Trough.

The analysis described in the next sections is specifically implemented for extremes of the CSI index, i.e., days during which the CSI exceeds the 90th percentile of the July and August climatological distribution (hereafter: 'upper 10% of CSI' or heat waves) versus days when the CSI is below the 10th percentile of the July and August distribution (hereafter 'lower 10% of CSI' or cool days). The onset of a heat wave (cool days) is taken to be the first day in which the CSI exceeds (subceeds) the 90th (10th) percentile threshold at 12UTC (0 h time in the Figures), which ought to roughly match the time of maximum daily temperature. Alpert *et al.* (2004b) have argued that July and August represent the mid-summer months, in which the Persian Trough occurs on more than nine out of eleven days on average. For additional details on the computation of the CSI index and its evaluation, the reader is referred to Saaroni *et al*. (2017).

2.3 Dynamical systems metrics

We view the atmosphere as a chaotic dynamical system, and leverage a recently-developed method combining extreme value theory with Poincaré recurrences (Lucarini *et al*., 2016; Faranda *et al*., 2017a) to estimate the dynamical properties of atmospheric states. The temporal evolution of the atmosphere can be represented as a long trajectory in a suitably defined phase space. When we use temporally discretised data, such as reanalysis data, we are effectively sampling this trajectory for a given time period, for example 6 or 24 hours. An example would be analysing daily latitude-longitude maps of Sea-Level Pressure (SLP) over the Eastern Mediterranean (technically a special Poincaré section of the full dynamics): each 2-D map corresponds to a single point along the aforementioned trajectory, for which we seek to compute instantaneous (in time) and local (in phase space) properties. We specifically consider two metrics, which describe instantaneous atmospheric states: the local dimension d and the persistence θ^{-1}. In order to compute the local dimension and persistence for a given state of interest ξ, which in our example would correspond to a specific SLP map in our dataset, we first define logarithmic returns as:

$$g(x(t),\xi) = -log\big(dist(x(t),\xi)\big)$$

where *dist* is the Euclidean distance between two vectors. Thus, we define g such that it is large when the system is in states close to ξ, and small when the system is in states far from ξ. In other words, g is large whenever the SLP map on a given day resembles the SLP map of the day corresponding to the state of interest ξ.

We next consider all cases in which g exceeds a high threshold $s\ (q,\ \xi)$, where q is a high quantile of the series g itself. Here we select q to be the 98th percentile of g. For these cases, which correspond to days whose SLP map is very similar to that of ξ, we can then define exceedances as:

$$u(t,\xi) = g(x(t),\xi) - s(q,\xi)$$

Given g above the threshold, we compute by how much it is so. The cumulative probability distribution $F\ (u,\ \xi)$ of the exceedances thus defined, converges to the exponential member of the Generalised Pareto Distribution (GPD; Freitas *et al.*, 2010; Lucarini *et al.*, 2012). In other words, given a sufficiently long series of SLP maps, we know that the exceedances u computed from these maps obey the following:

$$F(u,\xi) \simeq e^{\left(-\vartheta(\xi)\frac{u(\xi)}{\sigma(\xi)}\right)}$$

Here ϑ is the extremal index (Moloney *et al.*, 2019), which we calculate using the Süveges Maximum Likelihood estimator (Süveges, 2007), and u and σ are parameters of the distribution, which depend on the chosen ξ. The local dimension is then obtained as:

$$d(\xi) = \frac{1}{\sigma(\xi)}$$

While the persistence is given by:

$$\theta^{-1}(\xi) = \frac{\Delta t}{\vartheta(\xi)}$$

Where Δt is the time interval between successive time steps in our dataset. In practice, we choose each SLP map in the dataset in turn as state of interest ξ, which enables us to obtain a value of d and θ^{-1} for each timestep in our dataset.

In practical terms, d reflects the geometry of the trajectories in a small region (neighbourhood) of the system's phase space around the state of interest. It is therefore related to the number of active degrees of freedom that the system can explore about the state. In other words, it informs on the way the system evolves around the state of interest, and a higher d will correspond to a less predictable evolution of the system. The persistence θ^{-1} quantifies for how long the system resides in the neighbourhood of the state of interest. An infinite persistence implies a fixed point of the system, such that all successive timesteps bring no change to the state of the system. At the opposite end of the spectrum, $\theta^{-1} = 1$ corresponds to a non-persistent state of the dynamics. The dynamical systems persistence tends to be very sensitive to small changes in the state of the system. In atmospheric sciences, persistence is often computed as the residence time of the atmosphere within a given cluster of states. For example, when computing the persistence of weather regimes, one usually counts for how long the atmosphere remains within one given weather regime cluster. However, there are typically a small number of clusters, such that each one contains a relatively large fraction of the total number of timesteps within the dataset. In our case, we define recurrences based on a high threshold such that only a small fraction of timesteps within our dataset qualify as recurrences. By design, θ^{-1} is thus more sensitive to small changes in the atmosphere than the conventional definition of persistence of weather regimes. The two are nonetheless related, and relative differences in θ^{-1} often reflect relative differences in more conventional atmospheric persistence metrics (Hochman *et al.*, 2019).

The above derivations hold under a specific set of conditions, which are seldom satisfied by climate, or indeed any real-world datasets – such as having infinitely long timeseries. However, there are both formal (Caby *et al.*, 2020) and empirical (Messori *et al.*, 2017; Buschow and Friedrichs, 2018) results, which support the application of this framework to natural data. In particular, Buschow and Friederichs (2018) have shown that d successfully reflects the dynamical characteristics of the atmosphere even for datasets where the universal convergence to the exponential member of the GPD is not achieved. Messori *et al*. (2017) have further shown that the persistence estimates for atmospheric data issuing from the Süveges estimator are very stable under bootstrap resampling of the intra and inter-cluster times (*i.e.* the residence times of the trajectory within and without the neighbourhood of the state of interest). For more details on the estimation of the dynamical systems metrics, the reader is referred to Lucarini *et al*. (2016) and Faranda *et al*. (2017a; 2019a).

The dynamical systems perspective has been fruitfully applied to a range of geophysical and other datasets (e.g. Faranda *et al.*, 2019b, c; Brunetti *et al.*, 2019; Faranda *et al.*, 2020; Hochman *et al.*, 2020b; De Luca *et al.*, 2020a; Pons *et al.*, 2020). It has also been explicitly shown that d and θ^{-1} can offer an objective characterization of synoptic systems over different geographical regions, including the Mediterranean (Hochman *et al.*, 2019; De Luca *et al.*, 2020b), the North Atlantic (Faranda *et al.*, 2017a; Messori *et al.*, 2017; Rodrigues *et al.*, 2018) and the full Northern Hemisphere (Faranda et al., 2017b). In this study, we compute d and θ^{-1} for daily and 6-hourly 500 hPa geopotential height (Z500) and SLP fields from the NCEP/NCAR reanalysis, over the Eastern Mediterranean, placing Israel in the middle of the domain (27.5°N-37.5°N; 30°E-40°E; Fig. 1). To understand the differences between heat waves and cool days, we analyse both the CDFs (Cumulative Distribution Functions) and the mean temporal evolution of the two groups of days in terms of d and θ^{-1}. The Wilcoxon Rank-Sum (comparing the medians) and Kolmogorov-Smirnov (comparing the CDFs) tests are used for estimating the differences between the upper and lower 10% of CSI days at the 5% significance level. A bootstrap sampling test with 10^4 samples is used to evaluate the 95% confidence intervals of the mean temporal evolutions.

Previous studies have shown that the dynamical systems metrics d and θ^{-1}, have a strong seasonal cycle (Faranda *et al.*, 2017a, b; Rodrigues *et al.*, 2018; Hochman *et al.*, 2020a). Thus, we remove the seasonal cycle before comparing the various events. 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. In other words, we test whether heat waves are synoptically and dynamically unusual with respect to the cool days in the same part of the season. The seasonal cycle is estimated by averaging the metrics for a given time step (e.g., 15 August at 12UTC) over all years, repeating this for all time steps within the year and ultimately smoothing the series with a 30-day moving average.

2.4 Forecast spread/skill

To obtain an ensemble forecast, a set of numerical forecasts are performed with either different initial conditions, and/or perturbation of physical parametrizations. Ensemble forecasts offer an efficient way of estimating uncertainty by computing the ensemble spread. This is quantified by estimating the standard deviation between ensemble members. The spread can be taken as an indicator of practical predictability: in a perfect ensemble, a small spread would generally indicate we can determine with a good degree of confidence the future weather, while a large spread would point towards a larger uncertainty (e.g., Buizza 1997). This type of approach is commonly used when investigating atmospheric predictability (e.g., Hohenegger *et al.*, 2006; Ferranti *et al.*, 2015), although it does have limitations (e.g., Whitaker and Lough, 1998; Hopson 2014).

An additional frequently used forecast diagnostic is the absolute error, which provides a measure of forecast skill. The correlation between the ensemble spread and skill of the NWP model indicates the extent to which the ensemble can be used to provide an *a priori* estimate of the practical predictability of the atmospheric configuration we are considering. Here, we use the homogeneous station archive mentioned in Sect. 2.1 above as ground truth to estimate the forecasts' absolute error. In order to remove biases due to topographic differences between the model and the stations, the GEFS reforecast gridded data is bilinearly horizontally interpolated to the location of the stations. The bias computed over the whole period is then removed for each station.

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. Since the dynamical systems metrics offer information on the temporal evolution of the atmosphere in the neighbourhood of a given reference state, we argue that using the time of forecast initialization as temporal coordinate when plotting spread and error is most indicative for comparing the dynamical systems and numerical forecasts. In the supplementary material, we also plot the spread/skill for the forecasts initialised 69 h before the marked time. Thus, the plots in the main text show forecast initialisation times, while those in the supplementary material show the forecast valid times. Statistical inference is accomplished by the same tests mentioned in Sect. 2.3.

2.5 Air parcel tracking

In order to identify typical pathways of air masses leading to situations with high and low CSI values, ten-day backward trajectories are computed using the Lagrangian Analysis Tool (LAGRANTO; Wernli and Davies, 1997; Sprenger and Wernli, 2015). The reader is referred to Fig. 2 in Sprenger and Wernli (2015) for a schematic overview of the typical steps taken to compute trajectories. The tracking of temperature and potential temperature along the trajectory further allows to quantify the contribution of adiabatic and diabatic processes to the anomalous temperatures. The vertical and horizontal wind components required for the trajectory computations are acquired from the ERA-Interim reanalysis (Dee *et al*. 2011, Sect. 2.1). The trajectories are initialized at 12UTC from fixed points in the whole study region on the first day of a heat wave or cool days (Fig. 1). In order to analyze the near-surface air masses, i.e. those related to the hot and cool conditions, we consider trajectories that are initialized between the surface and 90 hPa above the surface. According to recent studies, the planetary boundary layer height in Israel during summer is ~600 – 900 m above the surface (Uzan *et al*., 2016; 2020). Assuming hydrostatic balance and thus a pressure decrease of approximately 1 hPa every 8m height difference, 90 hPa corresponds to about 720m. Therefore, this can be considered a reasonable choice.

The trajectories are calculated from 6-hourly ERA-Interim data and remapped to a 1° regular latitude-longitude grid. Thus, the analyzed wind field does not resolve sub grid-scale processes, such as Lagrangian transports due to small convective cells. Also, vertical motion associated with short-lived convection between two-time steps is not accounted for. Still, for a climatological investigation that is the focus of this study, the trajectory calculation is a suitable diagnostic.

3. Results

3.1 Dynamics of heat waves over the Eastern Mediterranean

We first analyze the differences between heat waves (upper 10% of CSI values) and cool days (lower 10% of CSI values). From an atmospheric dynamics' standpoint, the main difference between the two groups is that heat wave days are associated with an upper level ridge, whose center is located in the south – eastern part of the study region (Fig. 1a), while cool days are associated with an upper level trough, whose center is located at the north – western part of the study region (Fig. 1b). The SLP patterns are quite similar in both groups, but the heat waves show lower SLP in the south-west and a higher SLP in the north-east compared to the cool days sample (Fig. 1c). This implies stronger pressure gradients in the cool days' subgroup, leading to enhanced cool air advection from the Mediterranean Sea inland, in comparison to the heat wave days. Furthermore, the above reveals that the large-scale configuration is an important factor in the generation of a heat wave over the Eastern Mediterranean. The backward trajectory air parcel analysis illustrates that the flow preceding an extreme heat wave has a roughly meridional orientation when traveling over the Eastern Mediterranean and originates over the European continent (Fig. 2a). On the other hand, the air parcels for cool days often originate over the Atlantic, and take a more zonal pathway across the Eastern Mediterranean (Fig. 2b). The initial potential temperature of the heat wave air masses is about 7 K higher than that for the cool days (Fig. 2e). The differences in potential temperature between the two groups can mainly be attributed to the more continental origin of the air parcels for the heat waves, thus potentially transporting warmer air masses that descend on their path to the target region. Their descent, which is stronger than for cool days (Fig. 2c), is accompanied by a temperature increase of more than 25 K during the ten-day period (Fig. 2d). The potential temperature remains nearly constant until the final stages of the descent except for the diurnal cycle (Fig. 2e). Thus, we conclude that the extreme heat is related to an adiabatic descent of the air parcels over the Eastern Mediterranean rather than to diabatic heating. In other words, the warm air parcels are transported towards the Eastern Mediterranean with the governing westerlies rather than heated up locally over several days. This supports the findings of Harpaz *et al*. (2014), who argued that extreme summer heat waves over the Eastern Mediterranean are mostly regulated by mid-latitude disturbances rather than by the Asian Monsoon, as previously proposed by Ziv *et al*. (2004). An additional important difference between the two sets of CSI events is that, unlike for the heat waves (Fig. 2a, f), the specific humidity of the cool days increases by 2 g kg^{-1} around t = -48 hours, due to the longer stretch the latter air parcels follow over the Mediterranean Sea (Fig. 2b, f) and perhaps some enhanced convection, which our analysis does not account for. Indeed, comparing the portion of terrestrial back-trajectories for heat waves and cool days (Fig. S2) suggests that for most of the time they are quite similar. It is only 72-24 hours prior to the events that the portion of terrestrial back-trajectories for cool days reaches a minimum and is much lower than for heat waves (Fig. S2).

From a dynamical systems point of view, the upper and lower 10% of CSI also exhibit substantial differences. Fig. 3 shows a phase-plane diagram for d and θ computed on Z500 and SLP for the heat waves and cool days. θ is significantly lower at both levels for heat waves with respect to cool days, i.e., the former are generally more persistent systems. Statistically significant differences in the median local dimensions (d) of the two groups are found only for the Z500 variable, for which the heat waves typically display a lower local dimension (d) than the cool days (Fig. 3a). The clear separation between the two groups, especially at upper level (cf. Fig. 3a and Fig. 3b) correlates well with the atmospheric dynamics' viewpoint, which also shows more pronounced differences at Z500 (Fig. 1). This points to the importance of using different variables at different pressure levels to obtain a comprehensive picture of the dynamics of heat waves.

Fig. 4 displays the average temporal evolution of d and θ during the selected events, again computed for Z500 and SLP. Zero on the x-axis denotes the first day of the event at 12UTC, whereas, a value of zero on the y-axis implies that the events are not different from the climatology of the days they occurred in. Substantial differences are found between the time evolutions of the upper and lower 10% of the CSI events. For Z500, the temporal evolution of d and θ for heat waves are in phase with each other, and show a minimum with below climatology values in the 24 h preceding the event onset (Fig. 4a). While there is still a considerable spread around the mean, even the upper bounds of our confidence intervals are well below zero in the build-up to the events. Instead, cool days display weak positive anomalies of d and θ, but these are almost never significantly different from 0 (Fig. 4b). The dynamical systems metrics computed on SLP provide a completely different picture: heat waves typically display a weak above-climatology d, which increases towards the event onset and then decreases (Fig. 4c). θ displays a slightly below-climatology persistence (i.e. positive anomalies) and decreases towards the event onset (Fig. 4c). However, the very large spread in the composite evolution, and in particular in d, suggests some caution in over-interpreting the details of these evolutions. Cool days are characterized by higher positive anomalies of d and θ in the days preceding the event. The build-up towards this type of event is characterized by an increase in θ (decrease in persistence) and a decrease in d (decrease in active degrees of freedom; Fig. 4d). The cool days also appear to have a more coherent evolution (lower spread around the mean) than the heat waves for SLP.

The differentiation between the two samples is thus more pronounced when computing the metrics on Z500 than on SLP (Fig. 4), as also shown in the daily distributions (Fig. 3). Moreover, the variability in the temporal evolution of the dynamical systems metrics is smaller in Z500 than in SLP (Fig. 4). This points to: i) coherent, and very different, upper level conditions, which engender the two sets of CSI days; and ii) a comparatively wide range of possible near-surface patterns leading to severe heat waves. The latter may be explained by the fact that, given initially warm upper-level air parcels and upper-level subsidence leading to rapid adiabatic warming, the occurrence of a heat wave is then relatively insensitive to the details of the surface conditions (e.g., Baldi *et al*., 2006; Harpaz *et al*., 2014). Our general understanding of the synoptic conditions at surface levels further suggests that the delicate interplay between the Persian Trough and Subtropical High systems (Alpert *et al*., 1990) may contribute to the large spread of both heat waves and cool days regarding the dynamical systems metrics computed on SLP.

We analyze next numerical ensemble forecasts from the GEFS reforecast dataset for both sets of events. Substantial differences are again found between the two groups (Fig. 5). Both the ensemble spread and the absolute error are significantly higher for heat waves than for cool days (Fig. 5). The model spread and absolute error increase before the onset of the heat wave, peaking at around 24-48 hours negative lags (Fig. 5). This pattern stands in stark contrast to the temporal evolution of d computed on Z500 (cf. Fig. 5a, c, e and Fig. 4a), but somewhat resembles the evolution of d computed on SLP, albeit with a ~24 hours shift in time (cf. Fig. 5e and Fig. 4c). Such a shift may be explained by the fact that the spread/skill of the ensemble forecasts are computed every 24 hours, while the dynamical systems metrics are instantaneous in time (local in phase-space) and computed from 6-hourly data. The reforecasts computed for the individual stations (not shown) resemble the average forecast spread/skill (Fig. 5). The corresponding plots for forecast valid time (see Sect. 2.4), are provided in Fig. S3.

3.2 Analysis of the mid-August 2010 heat wave over the Eastern Mediterranean

The mid-August 2010 heat wave over the Eastern Mediterranean was a severe heat wave, which lies in the upper 0.3% of the CSI distribution. A detailed analysis of the heat wave highlights both similarities and differences with the climatology of the heat wave days (Sect. 3.1). The Z500 and SLP patterns for 15th August 2010 are comparable with the average configuration of a heat wave, but show a stronger upper level ridge and meridionally-oriented isobars (cf. Fig. 6a and Fig. 1a). From a dynamical systems point of view, the 2010 heat wave was also an uncommon extreme, especially for the metrics computed on Z500. The dynamical systems metrics' anomalies computed on this field range between -0.9 and -1.4 for *d,* and -0.14 and -0.2 for θ (Fig. 6b). This situates the heat wave in the lower 10% of the respective distributions (see also red dots in Fig. 3a). During its evolution, the event displays an increase in both *d* and θ computed on Z500 and a decrease (increase) in θ *(d)* computed on SLP (Fig. 6b, c). While the Z500 *d* and θ evolution is roughly comparable to that identified for heat wave days (cf. Fig. 4a and 6b), the SLP *d* and θ evolutions show profound differences. This may simply reflect the larger spread in dynamical systems properties across the different heat waves for SLP than for Z500, which is likely to be partially modulated by interactions between the surface and the atmosphere. Naturally, these interactions predominantly affect the lowermost parts of the Troposphere. We further hypothesize that differences between the single case and the climatology may be related to the relatively small day to day variations during summer over the Eastern Mediterranean, which make it challenging to depict the exact onset of a heat wave. Indeed, when comparing the climatology of the temporal evolution of *d* and θ for Z500 (Fig. 4a) with the single case (Fig. 6b) it is relatively easy to see that there is an increase in both *d* and θ as the heat wave develops. On the other hand, when comparing the temporal evolution of *d* and θ for SLP (Fig. 4c with Fig. 6c), one can see that depicting the exact time the heat wave starts is very important for comparison. In both Figures, *d* increases and θ decreases at some point in the chosen time window, but the timing of these trends is shifted between the climatology (Fig. 4c) and the single case (Fig. 6c).

The 2010 heat wave was also uncommon in terms of the large-scale flow and Lagrangian trajectories (Fig. 7). Between -10 to -5 days prior to the event, the majority of air parcels were transported in an easterly flow on the southern flank of an anticyclone located over Russia. Thus, air parcels came from the Zagros Plateau of Northern Iran, rather than from central Europe as in the climatology (cf. Fig. 7a, b and Fig. 2a). Indeed, Zaitchik *et al*. (2007) have argued that the Zagros Plateau has a strong influence on extreme summertime heat waves over the Eastern Mediterranean. Here we show that the anti-cyclonic wave breaking of the blocking regime over Russia, which interestingly is related to the decay phase of the Russian 2010 heat wave

(Quandt *et al*., 2019), played an important role in transporting the warm air masses from Northern Iran towards the Eastern Mediterranean and Israel (Fig. 7a, b and Fig. 8). This is realized through the trough east of the blocking ridge centered over European Russia, which is tilted southwest - northeast and advected westward (Fig. 8b - d, Davini *et al*., 2012; Quandt *et al.*, 2019). For the last five days prior to the heatwave (Fig. 7b), the parcel's trajectories resemble more closely the climatology of heat waves (Fig. 2a). Reflecting the different advection pathways, the initial potential temperature and temperature of the air parcels are respectively about 2K and 7K higher than the climatology of heat waves (cf. Fig. 7d, e and Fig. 2d, e). Accordingly, the hot air masses in the mid-August 2010 heat wave are transported to the Eastern Mediterranean and undergo adiabatic heating, rather than being heated up locally. This is in line with the climatology discussed in Sect. 3.1, and heat waves in other parts of the world (e.g., Bieli *et al*., 2015; Quinting and Reeder, 2017; Zschenderlein *et al*., 2019).

Fig. 9 shows the temporal evolution of the forecast spread/skill for the mid-august 2010 heat wave compared to the heat wave climatology. Throughout the lead up and early phases of the event, the forecast displays a lower spread and error than for other heat waves. A large decrease in the practical predictability occurs as the event develops, i.e., an increase in the spread/skill for maximum temperature (Fig. 9a, b). This mirrors the increase in d and θ computed on Z500 and for d computed on SLP (cf. Fig. 9a, b with Fig. 6b, c). Indeed, the decay phase of the Russian heat wave was characterized by low practical predictability (Matsueda 2011), which may have influenced the predictability over the Eastern Mediterranean. However, it should be noted that the spread computed on maximum temperature for the mid-August 2010 heat wave does not correlate well with the spread computed on SLP (cf. Fig. 9a with Fig. 9c). Moreover, some striking differences are displayed between the ensemble forecast of this single event and the climatology of forecasts for heat waves. These discrepancies may be related to the fact that we are analyzing a single event, whose error may not reflect the practical predictability of the atmosphere even for a perfect ensemble (e.g., Whitaker and Lough, 1998; Buizza *et al*., 2005; Hopson 2014). The corresponding plots for forecast valid time (see Sect. 2.4), are provided in Fig. S4.

4. Summary and conclusions

Heat waves are a major weather-related hazard, especially in an era of rapid climate change. We define heat waves over the Eastern Mediterranean according to a state-of-the-art 'Climatic Stress Index' (CSI; Saaroni *et al*., 2017), developed specifically for the region's summer weather conditions. We use a combination of dynamical systems theory, numerical weather forecasts and air parcel back-trajectories to investigate the evolution and predictability characteristics of heat waves (high CSI) and cool days (low CSI) for the region.

The main conclusions are as follows: significant differences are found between heat waves and cool days from both a dynamical systems and numerical weather prediction perspectives. Heat waves show relatively low practical predictability (large model spread/low skill) in the ensemble reforecast dataset used here, in spite of the relatively stable flow characteristics (high intrinsic predictability) compared to the cool days. When considering Z500, the *intrinsic* predictability of heat waves over the Eastern Mediterranean is highest, i.e., low local dimension (d) and high persistence (low θ), in the 24 h preceding the onset of the event, and lowest in the decay phase of the event. Indeed, Lucarini and Gritsun (2020) recently argued that atmospheric blocking over the Atlantic also displays such characteristics. The persistent upper level ridge that characterizes the heat waves may explain the high intrinsic predictability during the onset phase. The dynamical systems metrics computed on SLP show a different temporal evolution to their Z500 counterparts, emphasizing the different characteristics of the atmospheric flow at the different vertical levels. Specifically, there is a very large spread across different heat wave events. We argue that this may be associated with the delicate interplay between the Subtropical High and the Persian Trough at surface levels (Alpert *et al*., 1990), which can lead to a range of different SLP configurations all leading to a heat wave. This may further be reasonably attributed to the interactions between the land/sea surface and the atmosphere, which mainly influence the lower parts of the Troposphere. However, it is important to note that in many – albeit certainly not all – cases these interactions influence the atmosphere at time scales longer than those we consider in our analysis (e.g., Entin *et al*., 2000), and act as a seasonal-scale preconditioning to extremely high summer temperatures (Zampieri *et al*., 2009; Zittis *et al*., 2014).

Based on the Lagrangian air parcel analysis, we conclude that the physical processes governing Eastern Mediterranean summer heat waves relate to adiabatic descent of the air parcels over the region rather than diabatic heating, in agreement with previous findings (e.g., Harpaz *et al*., 2014). In other words, the air parcels are transported horizontally and vertically towards the Eastern Mediterranean with the governing westerlies rather than heated up locally over consecutive days. We further conclude that the origin of the air parcels over land in the days before the onset of a heat wave plays an important part in its generation.

A detailed analysis of the record-breaking mid-August 2010 heat wave provides further insights in this respect, by underscoring how the parcels, which contributed to the heat wave, were warmer than those of the climatology of heat waves as early as 10 days prior to the event. Interestingly, the onset of the heat wave over the Eastern Mediterranean was related to the decay phase of the Russian heat wave (Quandt *et al*., 2019) and we conclude that the anti-cyclonic Rossby wave breaking over Russia contributed to the onset of the Eastern Mediterranean heat wave. The 2010 heat wave showed both differences and similarities to other heat waves, highlighting the range of possible atmospheric and dynamical developments leading to high CSI values. This is compounded by the general difficulty of analyzing the life-cycle of heat waves, since there is little agreement as to what exactly a heat wave is and when it starts and ends (Shaby *et al*., 2016).

We conclude that the instantaneous dynamical systems metrics of local dimension (d) and persistence (θ^{-1}) provide complementary information on extreme summer heat waves compared to the conventional analysis of numerical weather forecasts. The discrepancy between the practical and the intrinsic predictability of the heat waves reflects this complementarity. For example, we interpret a very persistent system as being intrinsically highly predictable, yet the numerical forecasts we analyse display larger spread and error for the more persistent atmospheric configurations. In this respect, having an *a priori* measure of the persistence of an atmospheric configuration from dynamical systems can be a useful complement to the numerical forecast. Specifically, 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, even a perfect ensemble may not provide a good skill-spread relationship (Hopson 2014). That is, even a perfect ensemble may have a spread that does not always reflect the actual forecast error (Whitaker and Lough, 1998). In the specific case of the heat waves we analyze here, the spread and skill were well correlated for maximum temperature, but this is not a universal rule. For example, the mid-august 2003 heat wave had a very low spread (Tmax spread = 0.2 K, cf. Fig. 5b) and an above average error (Tmax absolute error = 1.1 K, cf. Fig. 5d). However, both d and θ computed on Z500 display a strong increase (d = -1.7 to 0.3, θ = -0.2 to 0.1 over the considered time window, not shown; cf., Fig. 4a and Fig. 6b), pointing to a decrease in intrinsic predictability. In such cases, local dimension (d) and/or persistence (θ^{-1}) trends that seem to contradict a low ensemble spread may serve as a warning of a potentially poor spread-skill relationship.

As a caveat, the comparison of the practical and intrinsic predictability still carries some interpretation challenges. Although the differences between the two can be partly ascribed to the different characteristics of the two measures, they may also be subject to the shortcomings of the GEFS ensemble data. In particular, the spread of the GEFS ensemble data, as most NWP ensemble forecasts, does not always reflect the practical predictability of the atmospheric flow (e.g., Whitaker and Lough, 1998; Keune *et al*., 2014). Moreover, our interpretation of the dynamical systems metrics may also be imperfect. Indeed, the metrics provide local information in phase space, while the spread and error of an ensemble forecast presumably reflect the longer-term evolution of the atmospheric flow. Similar interpretation challenges for the practical vs. intrinsic predictability have emerged when studying cold spells over the Eastern Mediterranean (Hochman *et al*., 2020a).

Notwithstanding these ongoing challenges, we believe that the novel view presented here, which leverages a dynamical systems approach for diagnosing extreme weather events, outlines an important avenue of research. We trust that it may be successfully applied to other regions and weather extremes in the future.

Acknowledgements

We would like to thank Mr. Yitzhak Yosef, head of the climate research department at the Israeli Meteorological Service (IMS), for providing homogeneous high-quality continuous station data over Israel. We further thank Dr. Noam Halfon of the IMS for providing averaged summer temperatures computed from the IMS Climatic Atlas. We thank the National Centers for Environmental Prediction/National Center for Atmospheric Research for the Reanalysis data and the Global Ensemble Forecast System reforecasts. We are grateful to Heini Wernli and the Atmospheric Dynamics group at ETH Zurich for sharing the ERA-Interim data. This paper is a contribution to the Hydrological cycle in the Mediterranean Experiment (HyMeX) community.

Funding

AH is funded by the German Helmholtz Association (ATMO Program). The contribution of JQ was funded by the Helmholtz-Association (grant VH-NG-1243). JGP thanks AXA Research Fund for support (https://axa-research.org/en/project/joaquim-pinto). GM was partly supported by the Swedish Research Council Vetenskapsrådet (grant no. 2016-03724).

Author contributions

All authors have contributed to the conceptual development of the study. AH and GM analyzed the data from a dynamical systems perspective. SS analyzed the forecast model data. JQ computed the air parcel backward trajectories. AH drafted the first version of the manuscript. All authors contributed through discussions and revisions.

Data availability

The paper and/or the supplementary materials contain or provide instructions to access all the data needed to evaluate the conclusions drawn in the paper. Additional data is available from the corresponding author upon request.

Competing interests

The authors declare no competing interests.

References

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. Guide-lines on climate metadata and homogenization. WCDMP-No. 53, WMO-TDNo. 1186, World Meteorological Organization, Geneva, Switzerland.

Alpert P, Abramsky R, Neeman BU. 1990. The prevailing summer synoptic system in Israel - Subtropical High, not Persian Trough. *Israel Journal of Earth Sciences* 39: 93–102.

Alpert P, Osetinsky I, Ziv B, Shafir H. 2004a. Semi-objective classification for daily synoptic systems: Application to the Eastern Mediterranean climate change. *International Journal of Climatology* 24*: 1001-1011.* https://doi.org/10.1002/joc.1036

Alpert P, Osetinsky I, Ziv B, Shafir H. 2004b. A new seasons' definition based on the classified daily synoptic systems, an example for the Eastern Mediterranean. *International Journal of Climatology* 24: 1013–1021. https://doi.org/10.1002/joc.1037

Baldi M, Dalu G, Marrachi G, Pasqui M, Cesarone F. 2006. Heat waves in the Mediterranean: A local feature or a larger-scale effect? *International Journal of Climatology* 26: 1477-1488. https://doi.org/10.1002/joc.1389

Ballester J, Achebak H, Herrmann F, Robine J, Rodó X. 2019. Trends in heat-attributable vulnerability and mortality in Europe: role of macroeconomic growth during the Great Recession. *Environmental Epidemiology* 3: 21. https://doi.org/10.1097/01.ee9.0000605844.32533.87

Barcikowska MJ, Kapnick SB, Krishnamurty L, Russo S, Cherchi A, Folland CK. 2020. Changes in the future summer Mediterranean climate: contribution of teleconnections and local factors. *Earth System Dynamics* 11: 161–181. https://doi.org/10.5194/esd-11-161-2020

Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R. 2011. The hot summer of 2010: redrawing the temperature record map of Europe. *Science* 332 (6026): 220-224. https://doi.org/10.1126/science.1201224

Battisti DS, Naylor RL. 2009. Historical warnings of future food insecurity with unprecedented seasonal heat. *Science* 323(5911): 240–244. https://doi.org/10.1126/science.1164363

Bennett JE, Blangiardo M, Fecht D, Elliott P, Ezzati M. 2014. Vulnerability to the mortality effects of warm temperature in the districts of England and wales. *Nature Climate Change* 4 (4): 269–273. https://doi.org/10.1038/nclimate2123

Bieli M, Pfahl S, Wernli H. 2015. A Lagrangian investigation of hot and cold temperature extremes in Europe. *Quarterly Journal of the Royal Meteorological Society* 141: 98–108. https://doi.org/10.1002/qj.2339

Bitan A, Saaroni H. 1992. The horizontal and vertical extension of the Persian Gulf trough. *International Journal of Climatology* 12: 733–747. https://doi.org/10.1002/joc.3370120706

Black E, Blackburn M, Harrison RG, Hoskins BJ, Methven J. 2004. Factors contributing to the summer 2003 European heatwave. Weather, 59, 217–223, doi:10.1256/wea.74.04.

Buizza R. 1997. Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Monthly Weather Review* 125(1): 99-119. https://doi.org/10.1175/1520-0493(1997)125%3C0099:pfsoep%3E2.0.co;2

Brunetti M, Kasparian J, Vérard C. 2019. Co-existing climate attractors in a coupled aqua-planet. *Climate Dynamics* ***53*****(9-10)**: 6293-6308. https://doi.org/10.1007/s00382-019-04926-7

Caby T, Faranda D, Vaienti S, Yiou P. 2020. Extreme value distributions of observation recurrences. *arXiv preprint arXiv:2002.10873*.

Caldeira MC, Lecomte X, David TS, Pinto JG, Bugalho MN, Werner C. 2015. Synergy of extreme drought and plant invasion reduce ecosystem functioning and resilience. *Scientific Reports* 5:15110. https://doi.org/10.1038/srep15110

Davini P, Cagnazzo C, Gualdi S, Navarra A. 2012. Bidimensional diagnostics, variability, and trends of northern hemisphere blocking. *Journal of Climate* **25**: 6496–6509. https://doi.org/10.1175/JCLI-D-12-00032.1

Dayan U, Tubi A, Levy I. 2012. On the importance of synoptic classification methods with respect to environmental phenomena. *International Journal of Climatology* 32*:* 681-694. https://doi.org/10.1002/joc.2297

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, Van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm, EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, De Rosnay P, Tavolato C, Thépaut JN, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137: 553-597. https://doi.org/10.1002/qj.828

De Luca P, Messori G, Pons FME, Faranda D. 2020a. Dynamical systems theory sheds new light on compound climate extremes in Europe and Eastern North America. *Quarterly Journal of the Royal Meteorological Society* https://doi.org/10.31223/osf.io/qdn5w

De Luca P, Messori G, Faranda D, Ward PJ, Coumou D. 2020b. Compound warm–dry and cold–wet events over the Mediterranean. *Earth System Dynamics* **11(3)**: 793-805. https://doi.org/10.5194/esd-11-793-2020

Deryng D, Conway D, Ramankutty N, Price J, Warren R. 2014. Global crop yield response to extreme heat stress under multiple climate change futures. *Environmental Research Letters* 9(3): 034011. https://doi.org/10.1088/1748-9326/9/3/034011

Dirmeyer PA, Halder S, Bombardi R. 2018. On the harvest of predictability from land states in a global forecast model. *Journal of Geophysical Research Atmosphere* **123(13):** 111-127. https://doi.org/10.1029/2018JD029103

Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO. 2000. Climate extremes: Observations, modelling and impacts. *Science* 289(5487): 2068-2074. https://doi.org/10.1126/science.289.5487.2068

Entin JK, Robock A, Vinnikov KY, Hollinger SE, Liu S, Namkhai A. 2000. Temporal and spatial scales of observed soil moisture variations in the extra tropics. *Journal of Geophysical Research* **105(D9)**: 11865– 11877. https://doi.org/10.1029/2000JD900051

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

Faranda D, Messori G, Yiou P. 2017a. Dynamical proxies of North Atlantic predictability and extremes. *Scientific Reports* 7: 412782017b. https://doi.org/10.1038/srep41278

Faranda D, Messori G, Alvarez-Castro MC, Yiou P. 2017b. Dynamical properties and extremes of Northern Hemisphere climate fields over the past 60 years. *Nonlinear Processes in Geophysics* 24(4): 713-725. https://doi.org/10.5194/npg-24-713-2017

Faranda, D, Messori G, Vannistem S. 2019a. Attractor dimension of time-averaged climate observables: insights from a low-order ocean-atmosphere model. *Tellus A* 71(1): 1554413. https://doi.org/10.1080/16000870.2018.1554413

Faranda D, Alvarez-Castro MC, Messori G, Rodrigues D, Yiou P. 2019b. The hammam effect or how a warm ocean enhances large scale atmospheric predictability. *Nature communications* 10(1): 1316. https://doi.org/10.1038/s41467-019-09305-8

Faranda D, Sato Y, Messori G, Moloney N, Yiou P. 2019c. Minimal dynamical systems model of the northern hemisphere jet stream via embedding of climate data. *Earth System Dynamics* https://doi.org/10.5194/esd-2018-80-rc1

Faranda D, Messori G, Yiou P. 2020. Diagnosing concurrent drivers of weather extremes: application to warm and cold days in North America. *Climate Dynamics* https://doi.org/10.1007/s00382-019-05106-3

Feeling the heat. 2018. *Nature Climate Change* 8: 347. https://doi.org/10.1038/s41558-018-0169-y

Ferranti L, Corti S, Janousek M. 2015. Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society* 141: 916-924. https://doi.org/10.1002/qj.2411

Fischer EM, Schär C. 2010. Consistent geographical patterns of changes in high-impact European heatwaves. *Nature Geoscience* 3: 398–403. https://doi.org/10.1038/ngeo866

Freitas ACM, Freitas JM, Todd M. 2010. Hitting time statistics and extreme value theory. *Probability Theory and Related Fields* 147: 675–710. https://doi.org/10.1007/s00440-009-0221-y

Giorgi F. 2006. Climate change hot spots. *Geophysical Research Letters* 33: L08707. https://doi.org/10.1029/2006gl025734

Goldreich Y. 2003. The Climate of Israel: Observation, Research and Application. Springer: Israel, Kluwer Academic/Plenum Publishers: Netherland, 298pp. https://doi.org/10.1007/978-1-4615-0697-3_2

Grumm RH. 2011. The Central European and Russian Heat Event of July–August 2010. *Bulletin of the American Meteorological Society* October 2011: 1285-1296. https://doi.org/10.1175/2011bams3174.1

Hamill TM, Bates GT, Whitaker JS, Murray DR, Fiorino M, Galarneau TJ, Zhu Y, Lapenta W. 2013. NOAA's Second-Generation Global Medium-Range Ensemble Reforecast Dataset. *Bulletin of the American Meteorological Society* 94: 1553–1565. https://doi.org/10.1175/bams-d-12-00014.1

Harpaz T, Ziv B, Saaroni H, Beja E. 2014. Extreme summer temperatures in the East Mediterranean – dynamical analysis. *International Journal of Climatology* 34(3): 849–862. https://doi.org/10.1002/joc.3727

Hochman A, Mercogliano P, Alpert P, Saaroni H, Bucchignani E. 2018a. High-resolution projection of climate change and extremity over Israel using COSMO-CLM. *International Journal of Climatology* 38(14): 5095-5106. https://doi.org/10.1002/joc.5714

Hochman A, Harpaz T, Saaroni H, Alpert P. 2018b. Synoptic classification in 21st century CMIP5 predictions over the Eastern Mediterranean with focus on cyclones. *International Journal of Climatology* 38(3): 1476-1483. https://doi.org/10.1002/joc.5260

Hochman A, Harpaz T, Saaroni H, Alpert P. 2018c. The seasons' length in 21st century CMIP5 projections over the Eastern Mediterranean. *International Journal of Climatology* 38(6): 2627-2637. https://doi.org/10.1002/joc.5448

Hochman A, Alpert P, Harpaz T, Saaroni H, Messori G. 2019. A new dynamical systems perspective on atmospheric predictability: eastern Mediterranean weather regimes as a case study. *Science Advances* 6(5): 1 – 10. https://doi.org/10.1126/sciadv.aau0936

Hochman A, Scher S, Quinting J, Pinto JG, Messori G. 2020a. Dynamics and Predictability of Cold Spells over the Eastern Mediterranean. *Climate Dynamics*. https://doi.org/10.1007/s00382-020-05465-2

Hochman A, Alpert P, Kunin P, Rostkier-Edelstein D, Harpaz T, Saaroni H, Messori G. 2020b. The dynamics of cyclones in the 21st century; the eastern Mediterranean as an example. *Climate Dynamics* 54(1-2): 561-574. https://doi.org/10.1007/s00382-019-05017-3

Hohenegger C, Lüthi D, Schär C. 2006. Predictability mysteries in cloud-resolving Models. *Monthly Weather Review* 134: 2095–2107. https://doi.org/10.1175/mwr3176.1

Holton JR. 2004. An Introduction to Dynamic Meteorology. Elsevier, London, UK.

IPCC. 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner GK, Allen SK, Tignor M, Midgley PM (Eds.) Available from Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge CB2 8RU ENGLAND, 582 pp. https://doi.org/10.1017/cbo9781139177245

Hopson TM. 2014. Assessing the ensemble spread–error relationship. *Monthly Weather Review* 142: 1125–1142. https://doi.org/10.1175/MWR-D-12-00111.1

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G,Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-Year reanalysis project. *Bulletin of the American Meteorological Society* 77*:* 437-471.

Katsafados P, Papadopoulos A, Varlas G, Papadopoulou E, Mavromatidis E. 2014. Seasonal predictability of the 2010 Russian heat wave. *Natural Hazards and Earth System Sciences* 14: 1531–1542. https://doi.org/10.5194/nhess-14-1531-2014

Keune J, Ohlwein C, Hense A. 2014. Multivariate Probabilistic Analysis and Predictability of Medium-Range Ensemble Weather Forecasts. *Monthly Weather Review* 142: 4074 – 4090. https://doi.org/10.1175/mwr-d-14-00015.1

Koster RD, *et al.* 2010. Contribution of land surface initialization to sub-seasonal forecast skill: First results from a multi-model experiment. *Geophysical Research Letters* **37**: 1–6. https://doi.org/10.1029/2009GL041677

Kuglitsch F, Toreti A, Xoplaki E, Della-Martta P, Zerefos CS, Türkeş M, Luterbacher J. 2010. Heat Wave Changes in the Eastern Mediterranean since 1960. *Geophysical Research Letters* 37(4). https://doi.org/10.1029/2009gl041841

Kunin P, Alpert P, Rostkier-Edelstein D. 2019. Investigation of sea-breeze/foehn in the Dead-Sea valley employing high-resolution WRF and observations. *Atmospheric Research* 229: 240-254. https://doi.org/10.1016/j.atmosres.2019.06.012

Kushnir Y, Dayan U, Ziv B, Morin E, Enzel Y. 2017. 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.

Lelieveld J, Proestos Y, Hadjinicolaou P, Tanarhte M, Tyrlis E, Zittis G. 2016. Strongly increasing heat extremes in the Middle East and North Africa (MENA) in the 21st century. *Climatic Change* 137(1-2): 245-260. https://doi.org/10.1007/s10584-016-1665-6

Loken ED, Clark JC, Xue M, Kong F. 2019. Spread and skill in mixed and single physics convection allowing ensembles. *Weather and forecasting* 34(2): 305-330. https://doi.org/10.1175/waf-d-18-0078.1

Lorenz EN. 1963. Deterministic non-periodic flow. *Journal of Atmospheric Science* 20*:* 130-141.

Lorenz EN. 1969. Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences* **26**: 636–646. https://doi.org/10.1175/1520-0469(1969)26%3C636:APARBN%3E2.0.CO;2

Lucarini V, Faranda D, Wouters J. 2012. Universal behaviour of extreme value statistics for selected observables of dynamical systems. *Journal of Statistical Physics* 147: 63–73. https://doi.org/10.1007/s10955-012-0468-z

Lucarini V, Faranda D, Freitas ACM, Freitas JM, Holland M, Kuna T, Nicol M, Todd M, Vaienti S. 2016. Extremes and Recurrence in dynamical systems, in Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts, pp. 126–172, Wiley, Hoboken, N. J. https://doi.org/10.1002/9781118632321.ch6

Lucarini V, Grisun A. 2020. A new mathematical framework for atmospheric blocking events. Climate Dynamics 54: 575-598. https://doi.org/10.1007/s00382-019-05018-2

Matsueda M. 2011. Predictability of Euro-Russian blocking in summer of 2010. *Geophysical Research Letters* 38: L06801. https://doi.org/10.1029/2010gl046557

Matsueda M, Palmer TN. 2018. Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts. *Quarterly Journal of the Royal Meteorological Society 144*(713): 1012-1027. https://doi.org/10.1002/qj.3265

Meehl GA, Tebaldi C. 2004. More intense, more frequent, and longer lasting heatwaves in the 21st century. *Science* 305: 994–997. https://doi.org/10.1126/science.1098704

Melhauser C, Zhang F. 2012. Practical and intrinsic predictability of severe and convective weather at the mesoscales. *Journal of the Atmospheric Sciences* **69(11)**: 3350–3371. https://doi.org/10.1175/JAS-D-11-0315.1

Messori G, Caballero R, Faranda D. 2017. A dynamical systems approach to studying midlatitude weather extremes. *Geophysical Research Letters* 44: 3346–3354. https://doi.org/10.1002/2017gl072879

Moloney NR, Faranda D, Sato Y. 2019. An overview of the extremal index. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29(2): 022101. https://doi.org/10.1063/1.5079656

Peterson TC, Heim RR, Hirsch R, et al. 2013. Monitoring and understanding changes in heat waves, cold waves, floods and droughts in the United States: State of knowledge. *Bulletin of the American Meteorological Society* 94: 821–834. https://doi.org/10.1175/bams-d-12-00066.1

Pons FME, Messori G, Alvarez-Castro MC, Faranda D. 2020. Sampling hyperspheres via extreme value theory: implications for measuring attractor dimensions. *Journal of Statistical Physics* **179**: 1698-1717. https://doi.org/10.1007/s10955-020-02573-5

Quandt LA, Keller JH, Martius O, Jones SC. 2017. Forecast variability of the blocking system over Russia in summer 2010 and its impact on surface conditions. *Weather Forecasting* 32: 61–82. https://doi.org/10.1175/WAF-D-16-0065.1

Quandt LA, Keller JH, Martius O, Pinto JG, Jones SC. 2019. Ensemble sensitivity analysis of the blocking system over Russia in summer 2010. *Monthly Weather Review* 147 (2): 657–675. https://doi.org/10.1175/mwr-d-18-0252.1

Quinting JF. Reeder MJ. 2017. Southeastern Australian heat waves from a trajectory viewpoint. *Monthly Weather Review* 145: 4109–4125. https://doi.org/10.1175/MWR-D-17-0165.1

Quinting JF, Parker T, and Reeder MJ. 2018. Two synoptic routes to subtropical heat waves as illustrated in the Brisbane region of Australia. *Geophysical Research Letters* 45: 10,700– 10,708. https://doi.org/10.1029/2018GL079261

Robine,JM, Cheung SLK, Roy SL, Oyen HV, Griffiths C, Michel JP, Herrmann FR. 2008. Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes Rendus Biologies* 331 (2): 171 – 178. https://doi.org/10.1016/j.crvi.2007.12.001

Rodrigues D, Alvarez-Castro MC, Messori G, Yiou P, Robin Y, Faranda D. 2018. Dynamical Properties of the North Atlantic Atmospheric Circulation in the Past 150 Years in CMIP5 Models and the 20CRv2c Reanalysis. *Journal of Climate* 31: 6097–6111. https://doi.org/10.1175/jcli-d-17-0176.1

Rodwell MJ, Hoskins B. 1996. Monsoons and the dynamic of deserts. *Quarterly Journal of the Royal Meteorological Society* 122: 1385–1404. https://doi.org/10.1002/qj.49712253408

Russo S, Dosio A, Graversen RG, Sillmann J, Carrao H, Dunbar MB, Vogt JV. 2014. Magnitude of extreme heatwaves in present climate and their projection in a warming world. *Journal of Geophysical Research-Atmospheres* 119: 1–13. https://doi.org/10.1002/2014jd022098

Saaroni H, Ziv B. 2000. Summer rain episodes in a Mediterranean climate – the case of Israel: climatological-dynamical analysis. *International Journal of Climatology* 20: 191–209. https://doi.org/10.1002/(sici)1097-0088(200002)20:2%3C191::aid-joc464%3E3.0.co;2-e

Saaroni H, Savir A, Ziv B. 2017. Synoptic classification of the summer season for the Levant using an ‘environment to climate’ approach. *International Journal of Climatology* 37(13): 4684-4699. https://doi.org/10.1002/joc.5116

Santos, JA, Pfahl S, Pinto JG, Wernli, H. 2015. Mechanisms underlying temperature extremes in Iberia: a Lagrangian perspective. *Tellus Series A: Dynamic Meteorology and Oceanography* 67: 26032. https://doi.org/10.3402/tellusa.v67.26032

Scher S, Messori G. 2018. Predicting weather forecast uncertainty with machine learning. *Quarterly Journal of the Royal Meteorological Society* 144: 2830–2841. https://doi.org/10.1002/qj.3410

Scher S, Messori G. 2019. How global warming changes the difficulty of synoptic weather forecasting. *Geophysical Research Letters* 46: 2931– 2939. https://doi.org/10.1029/2018gl081856

Schneidereit A, Schubert S, Vargin P, Lunkeit F, Zhu X, Peters DHW, Fraedrich K. 2012. Large-scale flow and the long-lasting blocking high over Russia: Summer 2010. *Monthly Weather Review* 140(9): 2967-2981. https://doi.org/10.1175/mwr-d-11-00249.1

Seneviratne SI, Nicholls N, Easterling D, Goodess CM, Kanae S, Kossin J, Luo Y, Marengo J, McInnes K, Rahimi M, Reichstein M, Sorteberg A, Vera C, Zhang X. 2012. Changes in climate extremes and their impacts on the natural physical environment. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 109-230.

Shaby BA, Reich BJ, Cooley D, Kaufman CG. 2016. A Markov-Switching model for heat waves. *The annals of Applied Statistics* 10(1): 74-93. https://doi.org/10.1214/15-aoas873

Siebert S, Ewert F. 2014. Future crop production threatened by extreme heat. *Environmental Research Letters* 9: 041001. https://doi.org/10.1088/1748-9326/9/4/041001

Slingo Y, Palmer T. 2011. Uncertainty in weather and climate prediction. *Philosophical Transactions of the royal Meteorological Society* 369: 4751-4767.

Sprenger M, Wernli H. 2015. The LAGRANTO Lagrangian analysis tool—version 2.0. *Geoscientific Model Development* 8: 2569–2586. https://doi.org/10.5194/gmd-8-2569-2015

Stott PA, Stone DA, Allen MR. 2004. Human contribution to the European heatwave of 2003. *Nature* 432: 610–614. https://doi.org/10.1038/nature03089

Süveges M. 2007. Likelihood estimation of the extremal index. *Extremes* 10(1–2): 41–55. https://doi.org/10.1007/s10687-007-0034-2

Tyrlis E, Lelieveld J. 2013. Climatology and dynamics of the summer Etesian winds over the Eastern Mediterranean. *Journal of the Atmospheric Sciences* **70**: 3374–3396. https://doi.org/10.1175/JAS-D-13-035.1

Uzan L, Egert S, Alpert P. 2016. Ceilometer evaluation of the eastern Mediterranean summer boundary layer height – first study of two Israeli sites. *Atmospheric Measurement Techniques* **9**: 4387–4398. https://doi.org/10.5194/amt-9-4387-2016

Uzan L, Egert S, Khain P, Levi Y, Vadislavsky E, Alpert P. 2020. Ceilometers as planetary boundary layer height detectors and a corrective tool for COSMO and IFS models. *Atmospheric Chemistry and Physics* 20: 12177-12192. https://doi.org/10.5194/acp-20-12177-2020

Wernli H. Davies HC. 1997. A lagrangian-based analysis of extratropical cyclones. I: The method and some applications. *Quarterly Journal of the Royal Meteorological Society* 123: 467-489. https://doi.org/10.1002/qj.49712353811

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

Williams CA. 2014. Heat and drought extremes likely to stress ecosystem productivity equally or more in a warmer, CO2 rich future. *Environmental Research Letters* 9: 101002. https://doi.org/10.1088/1748-9326/9/10/101002

Yarnal B. 1993. Synoptic climatology in environmental analysis. Belhaven Press: London. https://doi.org/10.1016/0160-4120(93)90293-q

Yarnal B, Comrie AC, Frakes B, Brown DP. 2001. Developments and prospects in synoptic climatology - Review. *International Journal of Climatology* 21: 1923–1950. https://doi.org/10.1002/joc.675

Yosef Y, Aguilar E, Alpert P. 2018. Detecting and adjusting artificial biases in long-term temperature records in Israel. *International Journal of Climatology* 38(8): 3273-3289. https://doi.org/10.1002/joc.5500

Zaitchik BF, Evans JP, Smith RB. 2007. Regional Impact of an Elevated Heat Source: The Zagros Plateau of Iran. *Journal of Climate* 20: 4133-4166. https://doi.org/10.1175/jcli4248.1

Zampieri M, D'Andrea F, Vautard R, Ciais P, de Noblet-Ducoudré N, Yiou P. 2009. Hot European summers and the role of soil moisture in the propagation of Mediterranean drought. *Journal of Climate* **22**: 4747–4758. https://doi.org/10.1175/2009JCLI2568.1

Zittis G, Hadjinicolaou P, Lelieveld J. 2014. Role of soil moisture in the amplification of climate warming in the eastern Mediterranean and the Middle East. *Climate Research* **59**: 27-37. https://doi.org/10.3354/cr01205

Ziv B, Saaroni H, Alpert P. 2004. The factors governing the summer regime of the Eastern Mediterranean. *International Journal of Climatology* 24: 1859–1871. https://doi.org/10.1002/joc.1113

Zschenderlein P, Fink AH, Pfahl S, Wernli H. 2019. Processes determining heat waves across different European climates. *Quarterly Journal of the Royal Meteorological Society* 145: 2973–2989. https://doi.org/10.1002/qj.3599

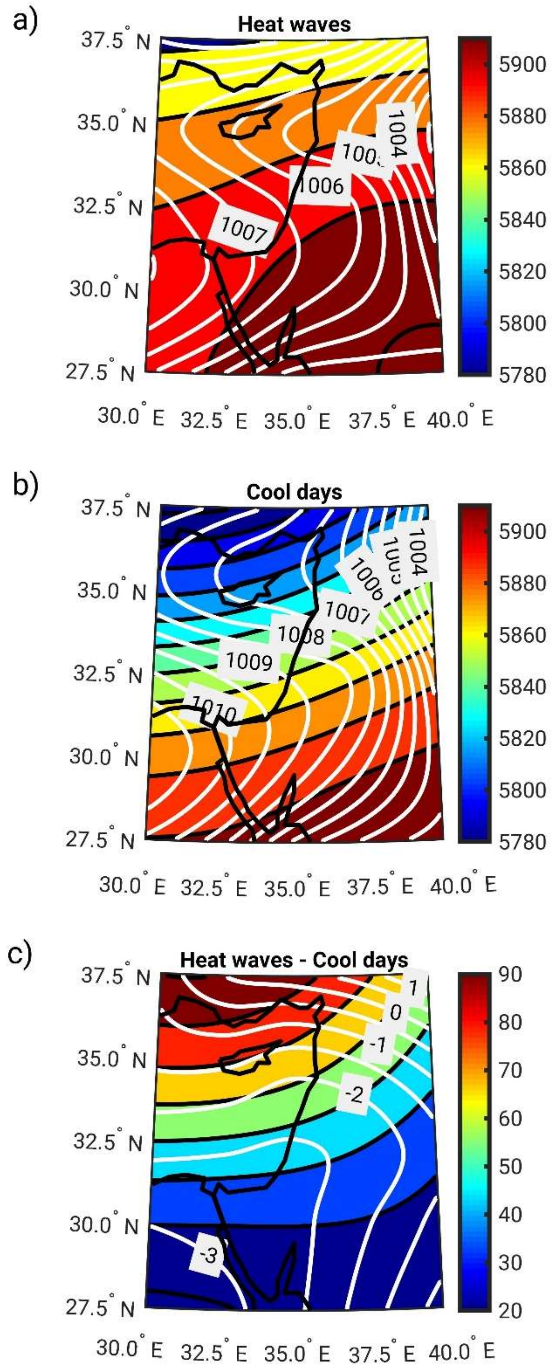


Figure 1 Mean sea level pressure (SLP in hPa, white contours) and 500 hPa geopotential height (Z500 in m, shaded in color) for the 10% of days with the highest (heat waves) and lowest (cool days) Climatic Stress Index (CSI) values. (a) Heat wave days mean composite; (b) cool days mean composite; (c)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

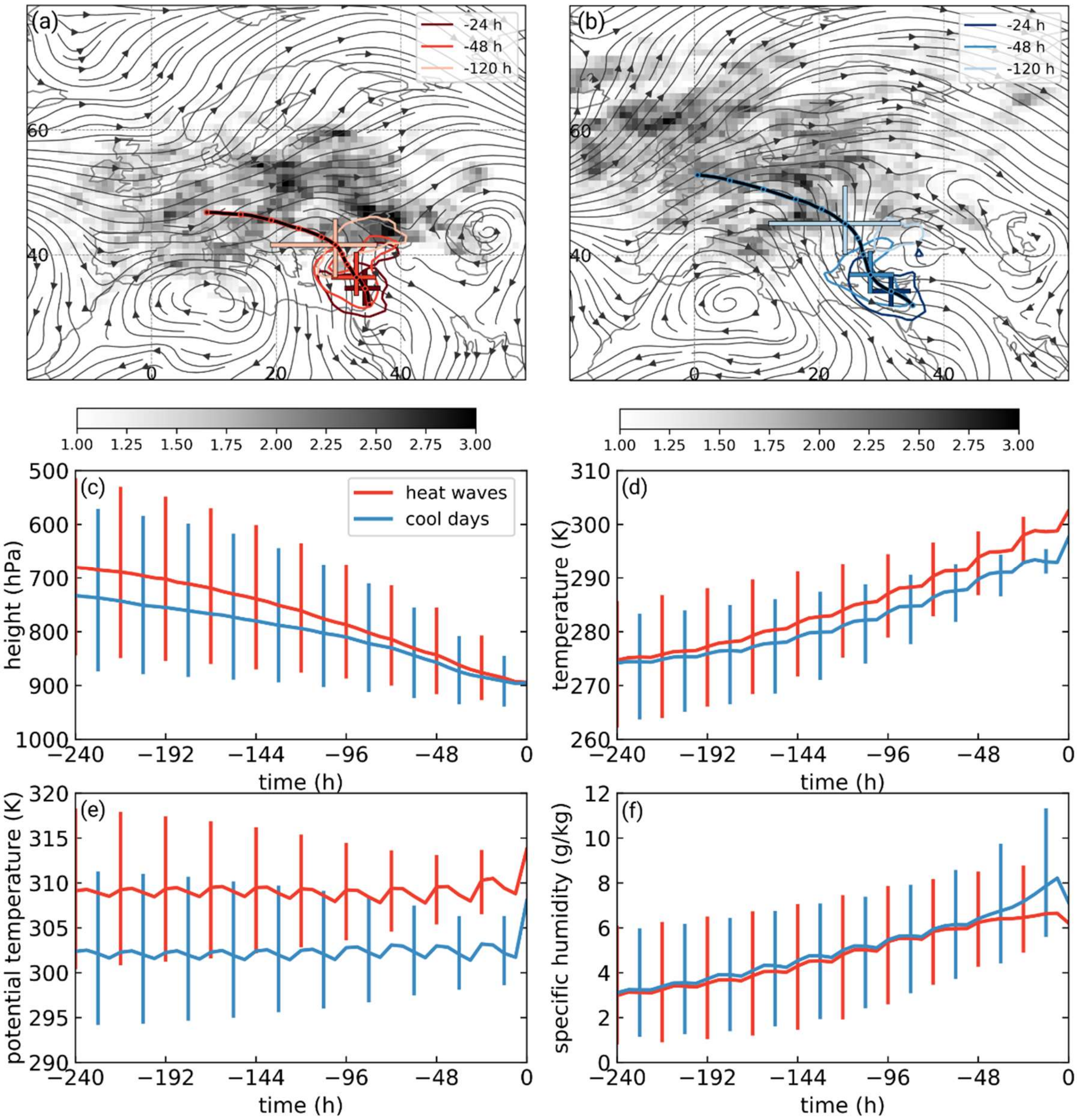


Figure 2 Median backward trajectory for (a) heat waves (upper 10% of CSI) and (b) cool days (lower 10% of CSI), with circles indicating days (from 10 days before onset to onset). Grey shading show trajectory density 10 days before onset (number of trajectories per 1000 km²), while contours show trajectory density for the indicated time lags (5, 2, 1 days before onset, contours denote a density of 20 trajectories per 1000 km²). Streamlines of 800 hPa winds averaged between -5 to -1 days are included. The inter-quartile range of trajectory positions is shown with crosses for the different time lags. Median evolution of (c) pressure (hPa); (d) temperature (K); (e) potential temperature (K); and (f) specific humidity (g kg^{-1}) of air parcels. Heat waves are indicated in red and cool days in blue. The inter-quartile range is plotted for the physical properties of the air parcels. 0 h corresponds to the first day of CSI ≥ 90% or CSI ≤ 10% and at 12UTC.

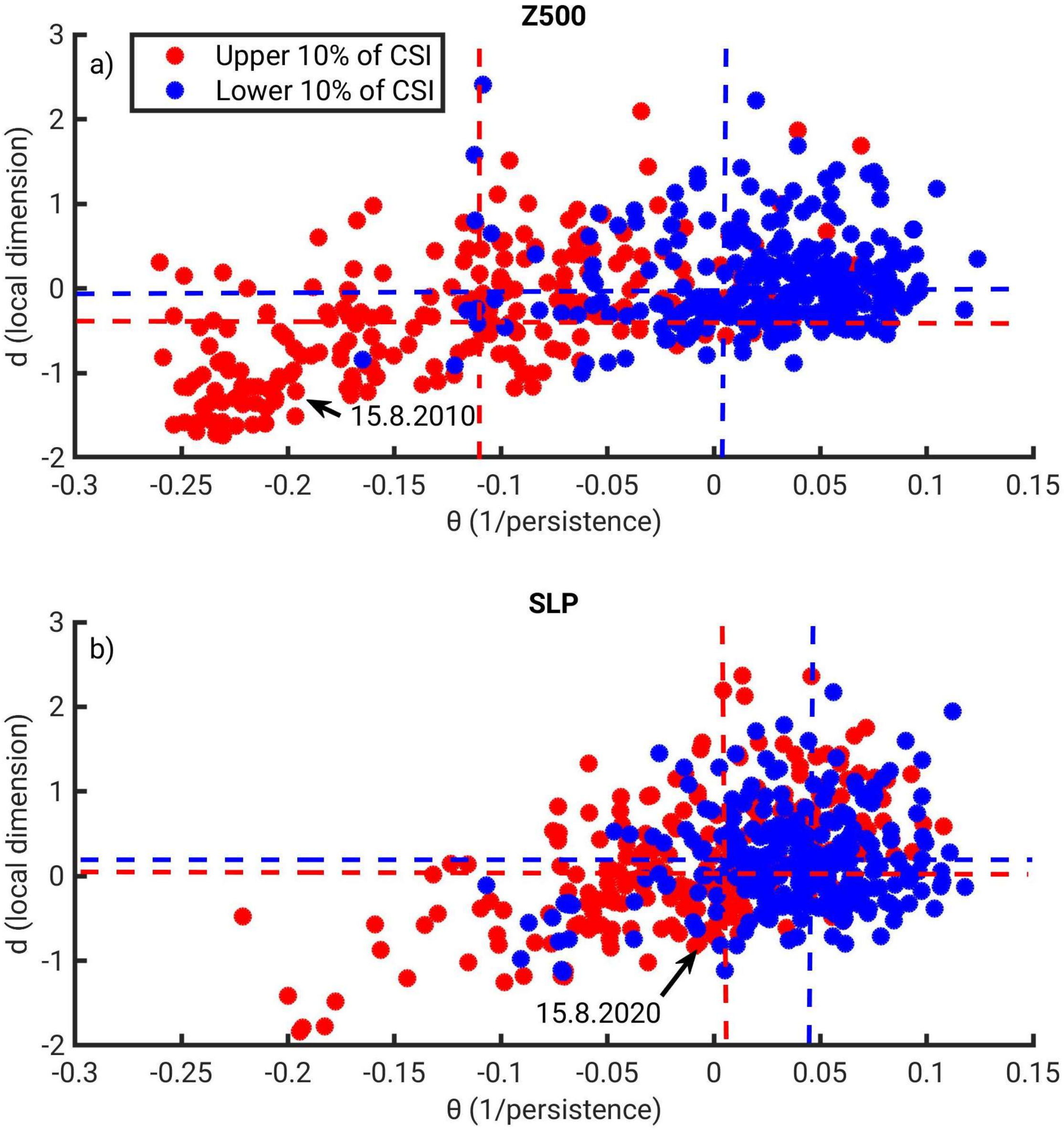


Figure 3 A phase-plane diagram for the upper and lower 10% of CSI daily values (heat waves in red and cool days in blue). The de-seasonalized dynamical systems metrics (d and θ) were computed for: (a) Z500 and (b) SLP. Dashed lines represent the median values of d and θ. The 15.8.2010 is indicated by the black arrows.

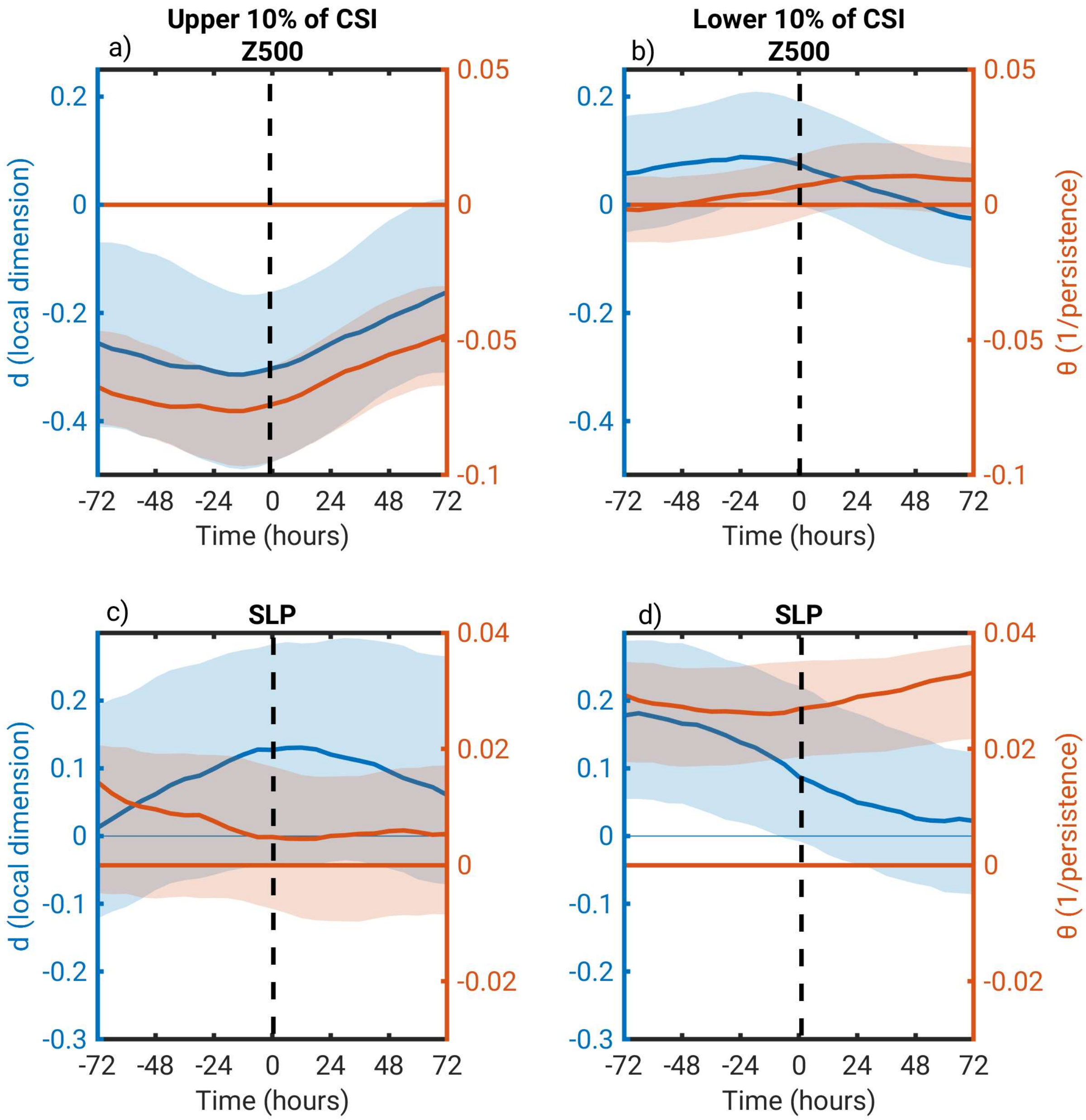


Figure 4 The average temporal evolution of the dynamical systems metrics (d and θ) for heat waves (upper 10% of CSI) and cool days (lower 10% of CSI) events. The dynamical systems metrics were computed for: (a, b) Z500 and (c, d) SLP. The events are centered (0 h) on the first day of CSI ≥ 90% or CSI ≤ 10% and at 12UTC. A 95% bootstrap confidence interval is plotted in shading.

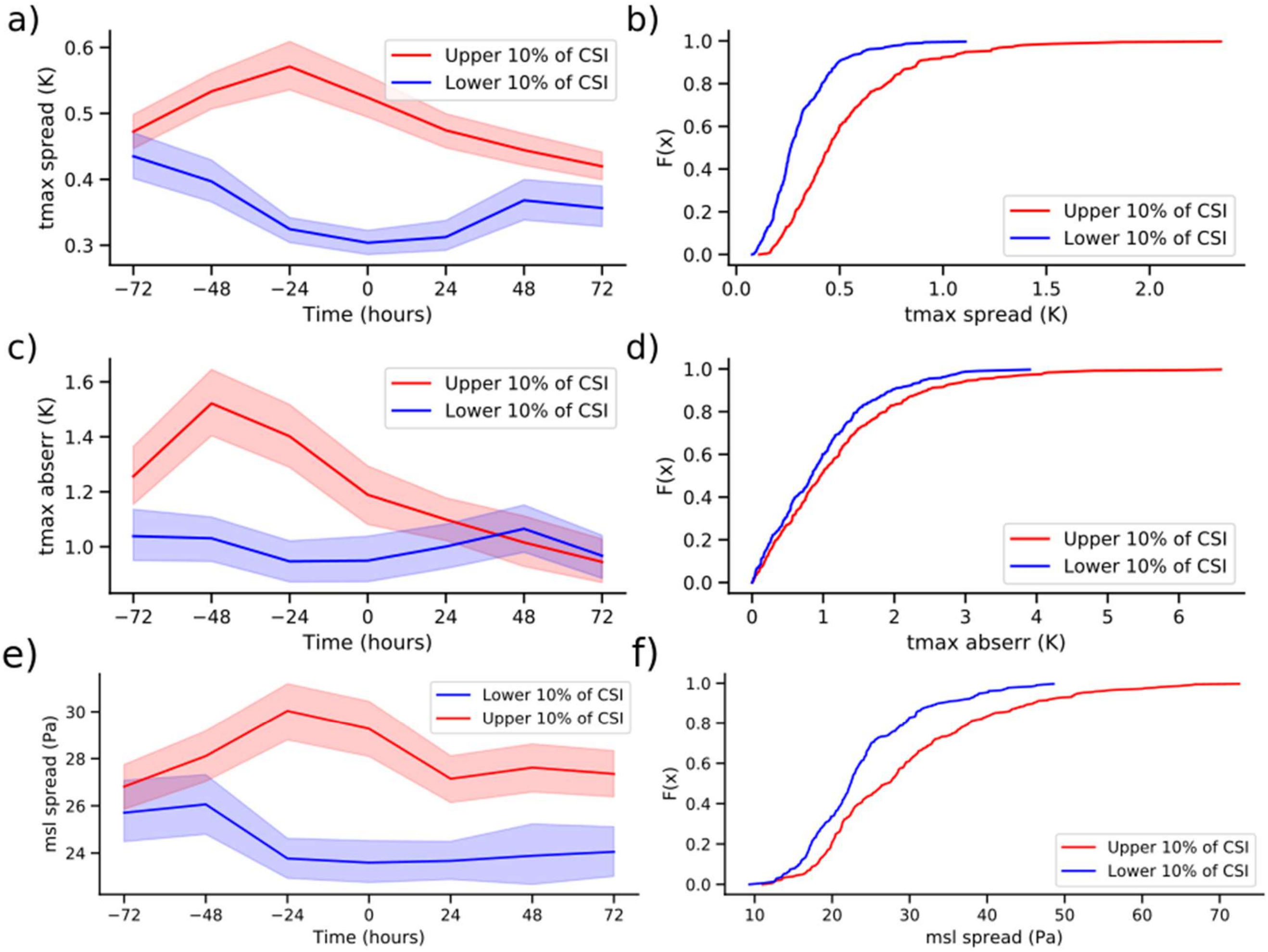
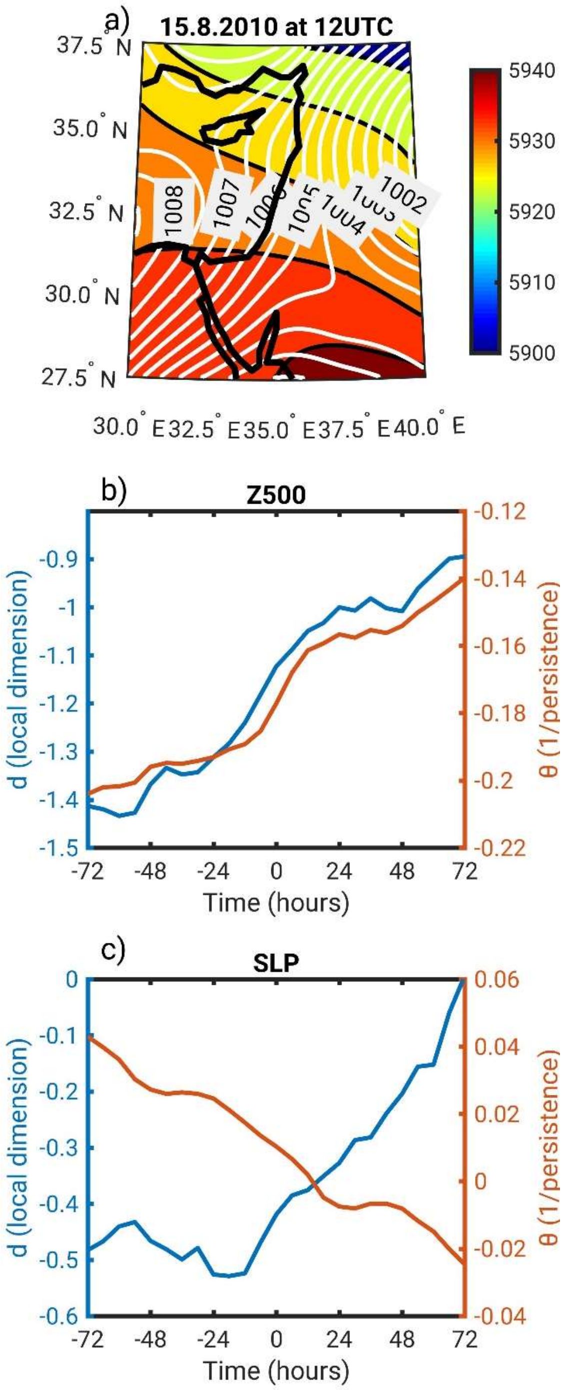


Figure 5 Forecast spread/skill for heat waves (upper 10% of CSI) and cool days (lower 10% of CSI). The lines show the mean temporal evolution of the ensemble model spread for Tmax (a), SLP (e) and absolute error for Tmax (c) of forecasts with lead-time 69h, initialized at different time lags with respect to the events, calculated every 24 hours. The events are centered (0 h) on the first day of CSI ≥ 90% or CSI ≤ lower 10% and at 12UTC. The CDFs of the mean ensemble forecast model spread for Tmax (b), SLP (f) and absolute error of Tmax (d) for the forecasts with lead-time 69h initialised at 00UTC. A 95% bootstrap confidence interval is shown in shading for the temporal evolution plots (a, c, e). The plots are displayed for the time of forecast initialization (see Sect. 2.4).



780 **Figure 6** A dynamical systems analysis for the mid-august 2010 heat wave. (a) SLP (white contours in hPa) and Z500 (shading in m) on 15.8.2010 at 12UTC. The dynamical systems metrics' (d and θ) temporal evolution centered on 15.8.2010 at 12UTC (0 h) computed on (b) Z500 and (c) SLP.

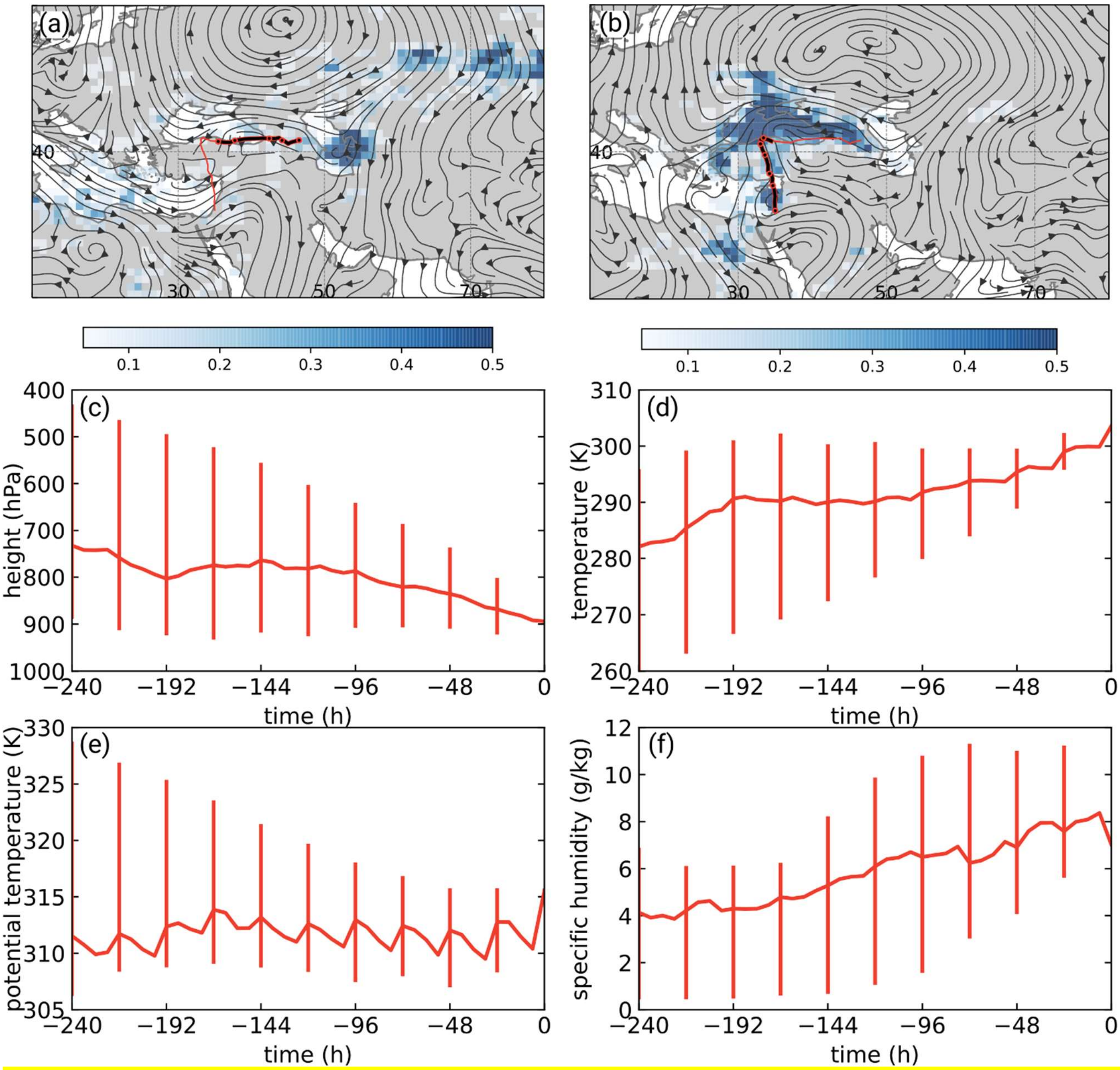


Figure 7 Backward trajectory air parcel tracking for the mid-August 2010 heat wave initialized on 15.8.2010 at 12UTC with (a) circles indicating days (from -10d to -6d before 15.8.2010 at 12UTC), grey shading indicating trajectory density 10 days before onset (number of trajectories per 1000 km²), and stream lines of 800-hPa wind (averaged between -10d to -6d before 15.8.2010 at 12UTC). (b) as in (a), but for -5d to -1d and trajectory density 5 days before onset. Median evolution of (c) height (hPa); (d) temperature (K); (e) potential temperature (K); and (f) specific humidity (g kg^{-1}) of the tracked air parcels. The inter-quartile range is plotted for the physical properties of the air parcels. 0 h corresponds to 15.8.2010 at 12UTC.

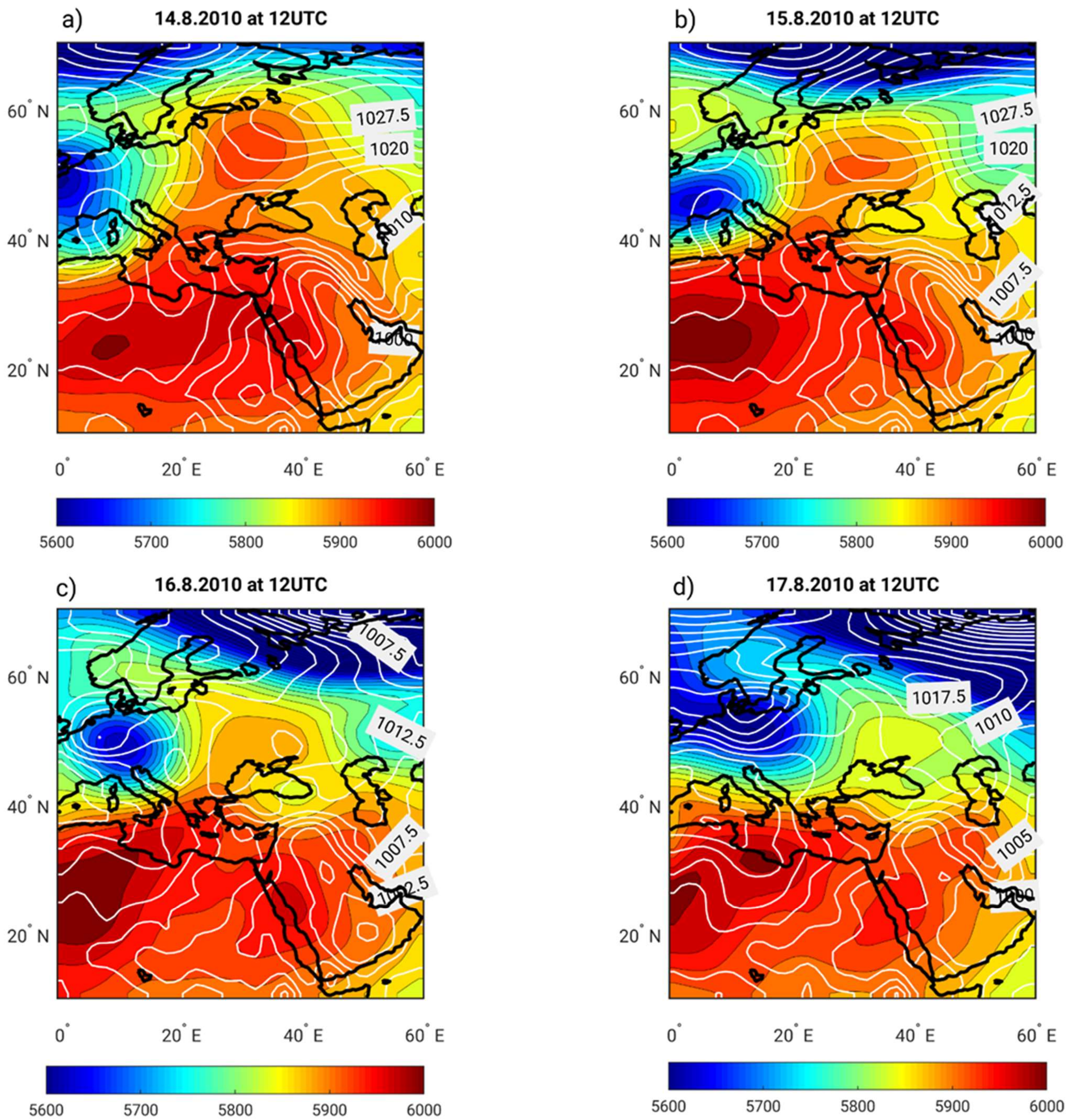


Figure 8 The large-scale evolution of SLP (white contours in hPa) and Z500 (shaded color in m) for the mid-August 2010 heat wave. a) 14.8.2010 at 12UTC; b) 15.8.2010 at 12UTC; c) 16.8.2010 at 12UTC; and d) 17.8.2010 at 12UTC.

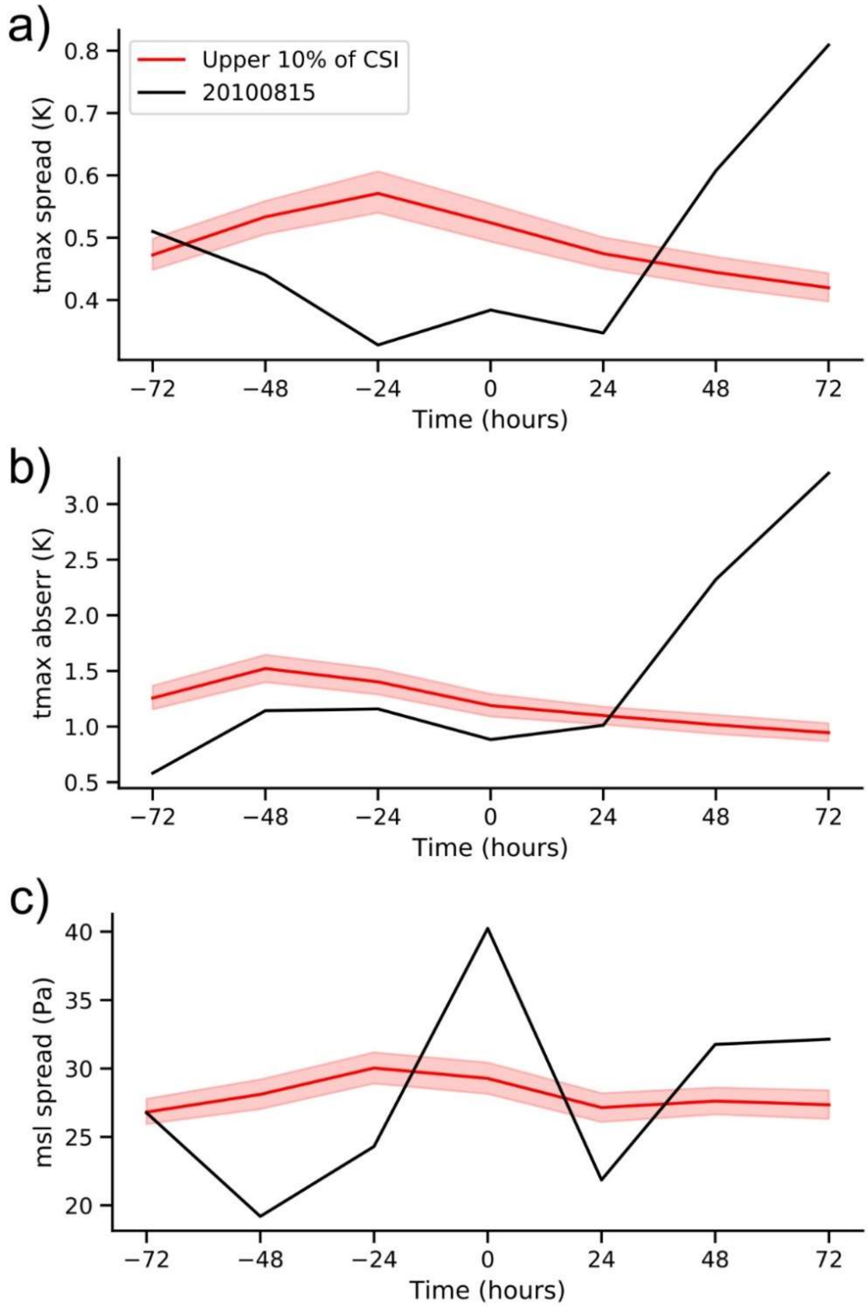


Figure 9 Forecast spread/skill for the mid-August 2010 heat wave, centered (0 h) on 15.8.2010 at 12UTC (black line). The mean temporal evolution of the ensemble model spread for Tmax (a), SLP (c) and absolute error for Tmax (b) of forecasts with lead-time 69h, initialized at different time lags with respect to the event, computed every 24 hours. The heat waves (upper 10% of CSI - red lines) are displayed for reference. A 95% bootstrap confidence interval for all heatwaves is displayed in shading.