

Review Atmospheric geopotentials from ERA5 linked to the daily maximum temperature record-breaking in Spain (1960–2023)

This may be an interesting study, but I found the text a bit difficult to digest due to vague formulations and long and windy descriptions of aspects that I didn't see a clear connection to the core of this topic which this paper is about. I'm not sure if I even quite understood this core topic. Did the authors use AI/ML in assisting them to read this? Also, the paper lacks proper acknowledgement and citation of earlier work related to this study. Finally, I found the paper a bit long.

The abstract is a bit hard to follow. Please consider rephrasing so that it's easier to understand for a little more general readership. E.g. I'm not sure what the meaning is of "Among the tested models, a global model that enhances the base model with geodetic interactions was selected for optimal balance between predictive performance and complexity".

The introduction does not acknowledge the early works that applied record-breaking statistics in climate research. See the overview provided by e.g. DOI:10.1126/sciadv.ado3712. Also, there is no mention of independent and identical distributed (iid) data in the introduction, which the authors seem to label 'stationary climate'. And the claim "... their occurrence follows well-established, distribution-free probabilistic properties" is not quite correct, as the records follow certain patterns as described in previous (uncited) studies. There is also a possibility that autocorrelation may introduce a higher recurrence of record-breaking events. While autocorrelation hardly seems to be an issue for calendar-date record-statistics (but adjacent days may be associated with annual series which are not entirely independent - so it may be more clean to subsample the days - eg every 5 day), it is nevertheless important to discuss the meaning of iid for completeness in the introduction and to convince the reader that they actually are familiar with record-breaking statistics.

As I read the introduction now, it does not provide a convincing argument for the authors having a good knowledge of this topic beyond citing AEMET and NOAA, and that also makes me more doubtful about the rest of their analysis. The neglect to acknowledge previous efforts relevant for the topic can be a slippery slope and cause part of the understanding to fade away when others in turn cite papers with incomplete background information. But to be fair, the concept of iid is mentioned later on in the paper under section 3.1.1, but not properly discussed and it's nevertheless good to introduce it in the introduction.

I expect there will be a connection between geopotential covariates and heatwaves, but it's not so clear for record-breaking events. Heatwaves and record-breaking events have different definitions and different (statistical) behaviors. Record-breaking events have a simple and clear definition, whereas the definition of heatwaves varies from country to country, typically being a number of consecutive days with temperatures above some threshold. A record-breaking event depends on the number of previous measurements and their measured values, and involves a degree of randomness connected to the time and place of the data sampling. This randomness

is expected to make a connection more tricky, while I also am convinced that records tend to fall during blocking highs.

It's a bit annoying that some vague statements about bi-variate analysis or specific downscaling are left in the air and the reader is left to guess or confer with Castillo-Mateo et al. (2025a,b). Please provide a clear and exact explanation so that I don't have to take a break and start reading another paper to understand what the deal is. Furthermore, the cited period 2012–2021 is a fairly short time period (10 years?), and one question is whether 10 data points are sufficient for a proper iid-test - but this is after all a cited study and perhaps not so relevant for this? I'd expect this short time period to be influenced by decadal variability, and would reserve the concept trend to longer time spans. Why this short time period in the first place? Anyway, this part brings me some doubts about whether we have a common understanding of record-breaking statistics, and more the reason for providing a proper account and background to the matter.

I'm not sure that geopotential covariates are well-suited for developing statistical downscaling methods to project the future frequency of record-breaking temperatures under global warming scenarios because it's not clear whether such predictors embed the climate change signal that well. This is the ultimate purpose of this study, is it not? Another alternative could be to downscale the shape of pdfs for the temperatures to infer how their statistics change. A stretching of the tails of the pdfs, after all, provides information about probabilities of record-breaking temperatures. Furthermore, the iid-test was originally inspired by the question whether these distributions were non-stationary. One suboptimal strategy is nevertheless to downscale each data point as in the traditional ways because such an approach would miss the extremes, which I think this paper tries to allude to.

It may be more correct to say that the probability for a new record an iid variable is the reciprocal of the number of observations (n), as the description provided doesn't specify when time t starts. It could be time if t is the number of time steps (t is a discrete number and not a rational number in this case).

Please explain how equation 3 was derived or where it comes from.

In Fig3b, should I not expect the iid case (red curve) to lie over the empirical counts (blue curve) since there are more observed records than for the iid case? The y-axis is after all probability of non-records. I'm confused. Please explain more carefully.

Need a proper reference for the Jaccard index. It's also appropriate to acknowledge others who have contributed with such indices and knowledge.

G300 is not properly introduced and it's not clear. Also "Among the geopotential variables, G300 shows the most similar evolution to T_x , and the trend appears to weaken with increasing height" is confusing. Isn't G300 the geopotential highest up?

For record-breaking statistics, I'd expect to see similar shapes in the cumulated sums of events, as the four panels seem to demonstrate. I cannot see the curve for the iid case in Fig.4, and the figure caption is not very helpful.

It's not helpful to cite figures in the supporting documents, which means that I need to jump between different documents.

It may be useful to consider the model's minimum skilful scale for using the output as a predictor for downscaling. The models' minimum skilful scale is not well defined, but there have been suggestions that it's about 4x4 grid boxes (DOI:0.2151/jmsj.2015-042). One may justify the choice based on that, which is relevant for when this method is applied to the output for global climate models.

The question of how many tests were carried out should also be reflected in the estimation of the so-called 'field significance' (DOI:10.1175/JAM2404.1). When there are many combinations, then I'd expect that one or two will score high based on just chance. Fig. 5 is complex/difficult and not easy to understand.

L291. How was random chance quantified?

Exactly what does "a positive effect on the probability of record" mean?

By page 15, I gave up because it took too much effort to try to understand what was going on and what the point was behind all these exercises. I started to suspect that the text has been drafted with the aid of AI/ML, but don't know if it's the case. At least, I didn't get much out of this. I suggest shortening the paper considerably by moving a fair deal of it to supplementary material as I didn't see the point of having it in the main paper (and several figures before reaching the results section) and exhausting me before I got to the results section.