**Response to Reviewer 1 Comments**

https://doi.org/10.5194/egusphere-2025-2460-RC1

for

# Understanding European Heatwaves with Variational Autoencoders

https://doi.org/10.5194/egusphere-2025-2460

Submitted to

Earth System Dynamics

by

Aytaç Paçal[1,2], Birgit Hassler[1], Katja Weigel[2,1], Miguel-Ángel Fernández-Torres[3], Gustau Camps-Valls[4], Veronika Eyring[1,2]

[1] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

[2] University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

[3] Department of Signal Theory and Communications, Universidad Carlos III de Madrid (UC3M), Leganés, Madrid, Spain

[4] Image Processing Laboratory (IPL), Universitat de València (UV), Paterna, València, Spain

# Authors' Response to Reviewer 1

**General Comments.** This work uses a non-linear dimensionality reduction method to study heat wave characteristics in western Europe. More specifically, they train a 3D variational autoencoder to reconstruct 11-day windows of multiple atmospheric variables around historical heat wave onset dates. Afterwards, the trained VAE is used to embed heat waves from a test period temporally after the training period. Then, the embeddings are clustered, and a shift in frequency in these clusters between training and testing is observed.
Strengths:

- Non-linear method to analyze heat wave characteristics and their spatio-temporal trends.
- Relevant study area: western europe has recently experienced a series of devastating heat waves, which are assumed to often have been preceded by atmospheric blocking events.
- Detailed description of the methodology, which enables reproducibility
- If valid, the results are interesting: an observed shift of summer time heat wave characteristics in the past 20 years.

We appreciate the reviewer for their thoughtful summary and positive feedback on our work. In the revised manuscript, we will carefully address the reviewer's specific comments to further clarify the methodology, strengthen the interpretation of the results, and enhance the overall robustness and readability of the work.

**Major Comment 1**

1. The VAE may potentially be applied in an OOD scenario: Data from 2001-2022 may be outside of the training distribution (heatwaves during 1941-1990). There is no guarantee that for out of domain samples, neural networks produce meaningful latent representations. This may lead to flawed interpretations on any such latent representations.

2. The previous comment is particularly relevant in the context of one of the main claims of this work: that there is a change in heat wave patterns after 2000. While this may very well be true, it can not be ruled out that this is just an artifact of the way this study was set up.

3. As a way forward I'd suggest to actually not do a temporal splitting of the train/val/test sets, but instead a purely random one. Ultimately you want to extract good representations for the entire time series 1941-2022

**Response:** We appreciate the reviewer's concern about potential out-of-distribution effects. Our primary objective was to assess whether recent heatwaves differ from a historical baseline (1941–1990). We did an additional experiment using random train/val/test splits across the entire 1941–2022 period. This allows us to evaluate whether the clustering and drift patterns persist beyond the temporal split. This approach introduces a smoother latent space representation, as the model was trained on the entire period, as shown in Figure 1. Summer heatwaves still accumulate more closely together. This confirms that the clustering is not simply an artifact of non-stationarity in the time series but reflects consistent structural patterns across periods. We will also revise the text to more clearly articulate why a temporal split was chosen in relation to our scientific question.

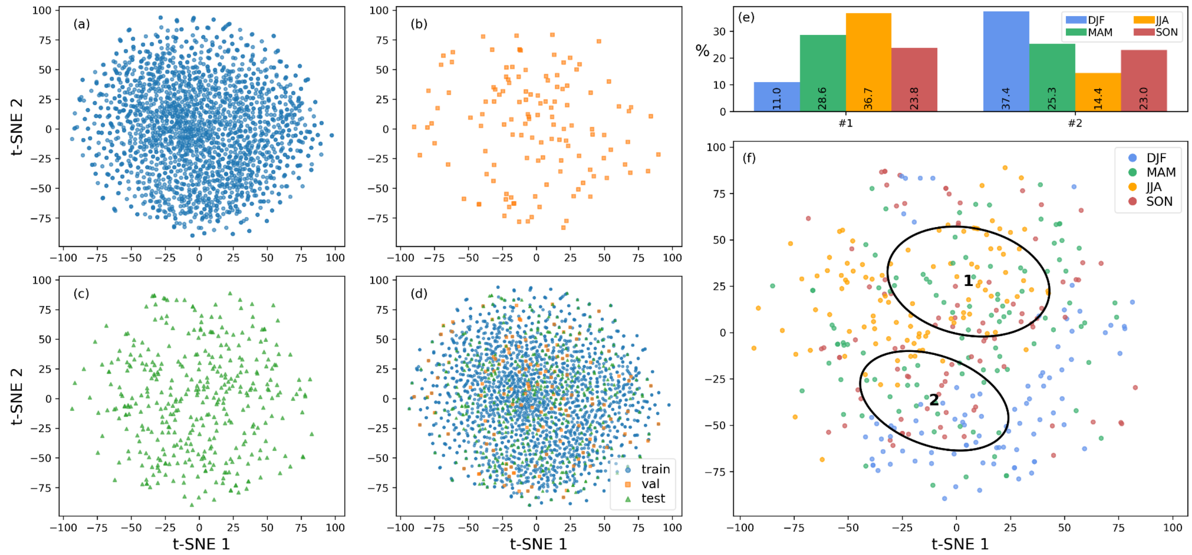


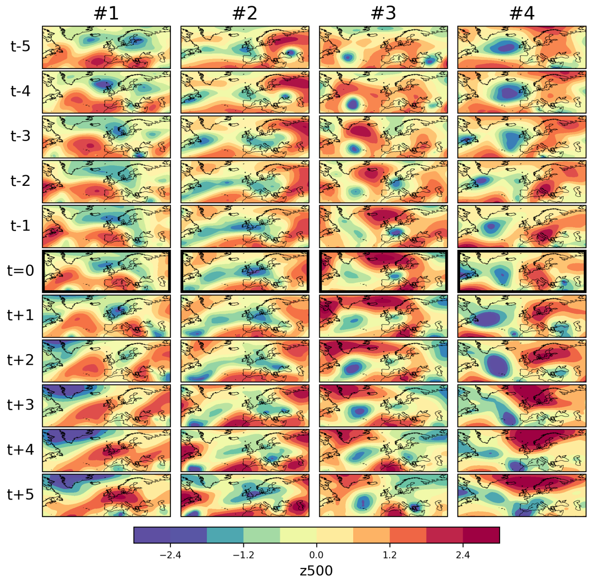

Figure 1: t-SNE representation of the latent space with random-split data (0.8/0.05/0.15). Following our workflow (testing with more than 2 components), a 2-component GMM provided the best fit, as it was closest to a single Gaussian distribution and thus better captured the smooth latent space structure. While this approach provides a smoother latent space, summer heatwaves still accumulate more closely together. This confirms that the clustering is not simply an artifact in the time series but reflects consistent structural patterns across periods.

## Major Comment 2

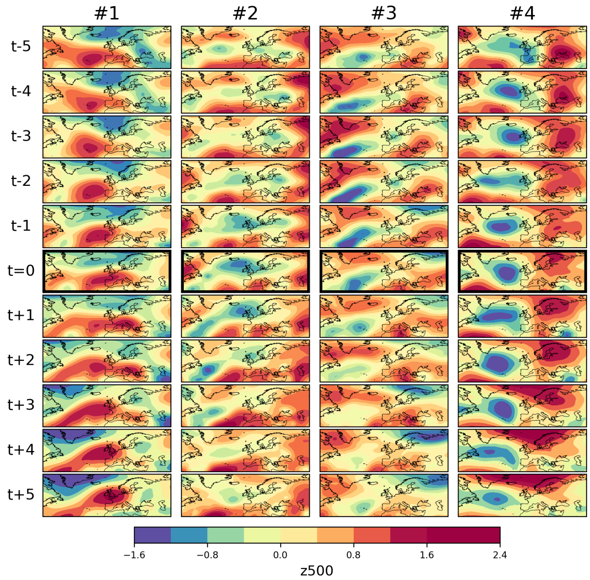
1. Why do you produce composite maps, instead of directly reconstructing the center of each cluster?

2. Could your composite maps not lead to unphysical patterns? For instance, phenomena could cancel out if they are just shifted in space slightly?

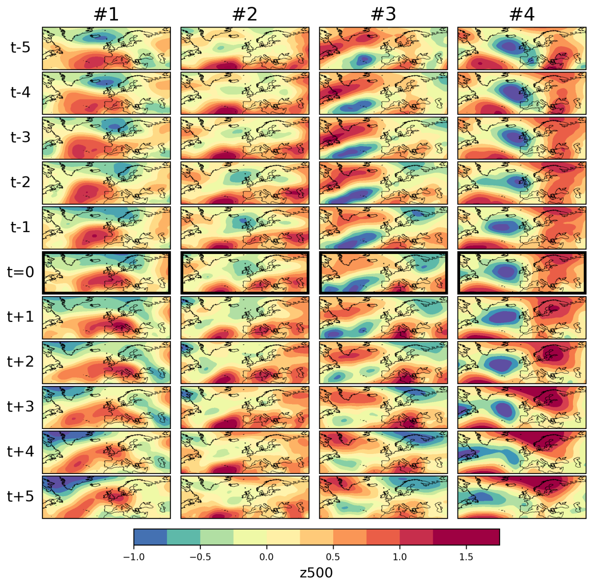
**Response:** The stochastic nature of the method means that the positions of heatwave samples, or cluster centers, in the latent space may not remain stable. Composite maps are therefore more reliable for identifying robust and consistent spatial features, since similar samples are already grouped closely together in the latent space by the VAE. We acknowledge that averaging might cancel signals if patterns are spatially shifted. To address this, we tested composite maps based on different sample sizes (e.g., $N = \{1, 5, 10, 25, 50, 100\}$), which allows us to analyze the stability of the detected patterns and assess potential cancellation effects. While composite maps with more than 5 samples exhibit similar patterns, the 1-sample composite map differs (See Figure 2). This is expected, since the 1-sample map corresponds to cluster centers.



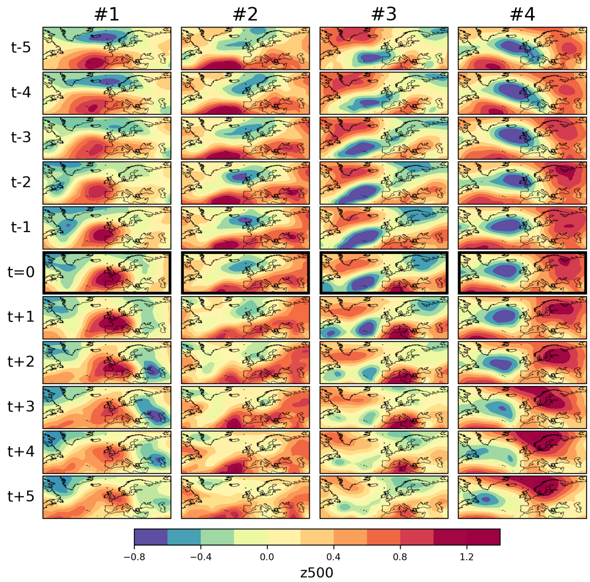

(a) 1-sample composite map



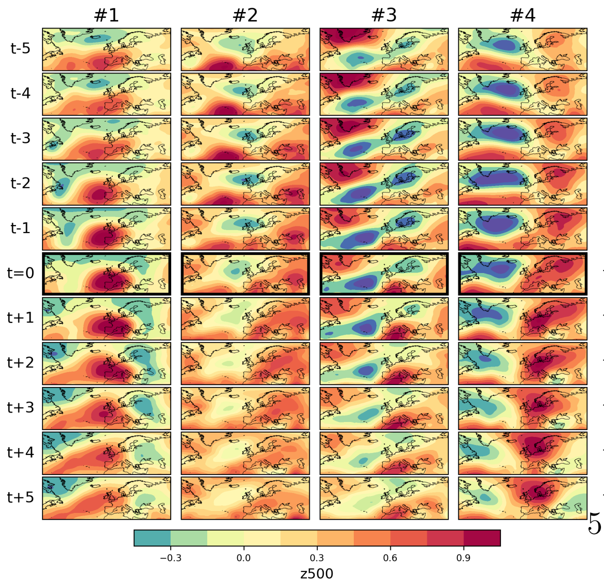

(b) 5-sample composite map



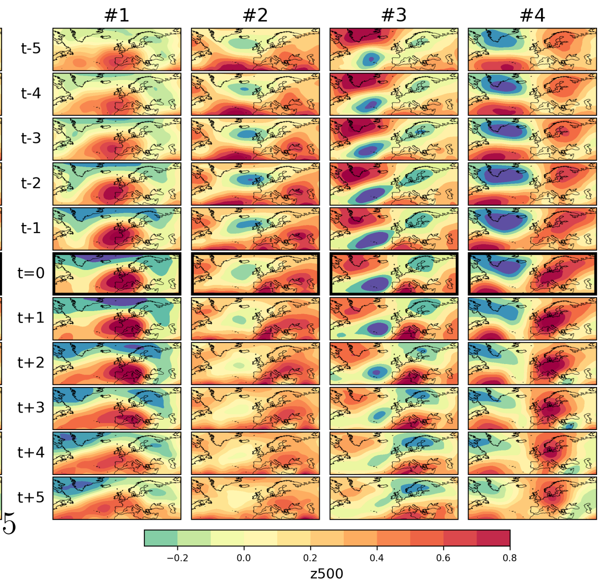

(c) 10-sample composite map




(d) 20-sample composite map




(e) 50-sample composite map




(f) 100-sample composite map

Figure 2: Composite maps for geopotential height at 500 hPa (*z500*).

## Major Comment 3

In this study you introduce atmospheric patterns that have been observed prior to heat waves. One open question remains, if these patterns exclusively arise prior to heatwaves - or if they are not always succeeded by extreme temperature anomalies. This would be very valuable to understand the physical mechanisms but also assess potential predictive skill.

**Response:** We agree that this is an important point. In this study, we focused on identifying and characterizing the atmospheric patterns that are commonly observed before heatwaves. However, these patterns may also occur without being followed by extreme temperature anomalies, and distinguishing such cases is indeed crucial for both understanding the underlying physical mechanisms and assessing predictive skill. While a complete predictive analysis is beyond the scope of the present paper, we computed the mean bias between the summer cluster center and 11-day running window samples from the summer months of 2015–2021 for each variable, as shown in Figure 3. These biases did not reveal a strong predictive signal for heatwaves. We will clarify this limitation in the revised manuscript and highlight it as an important direction for future work.

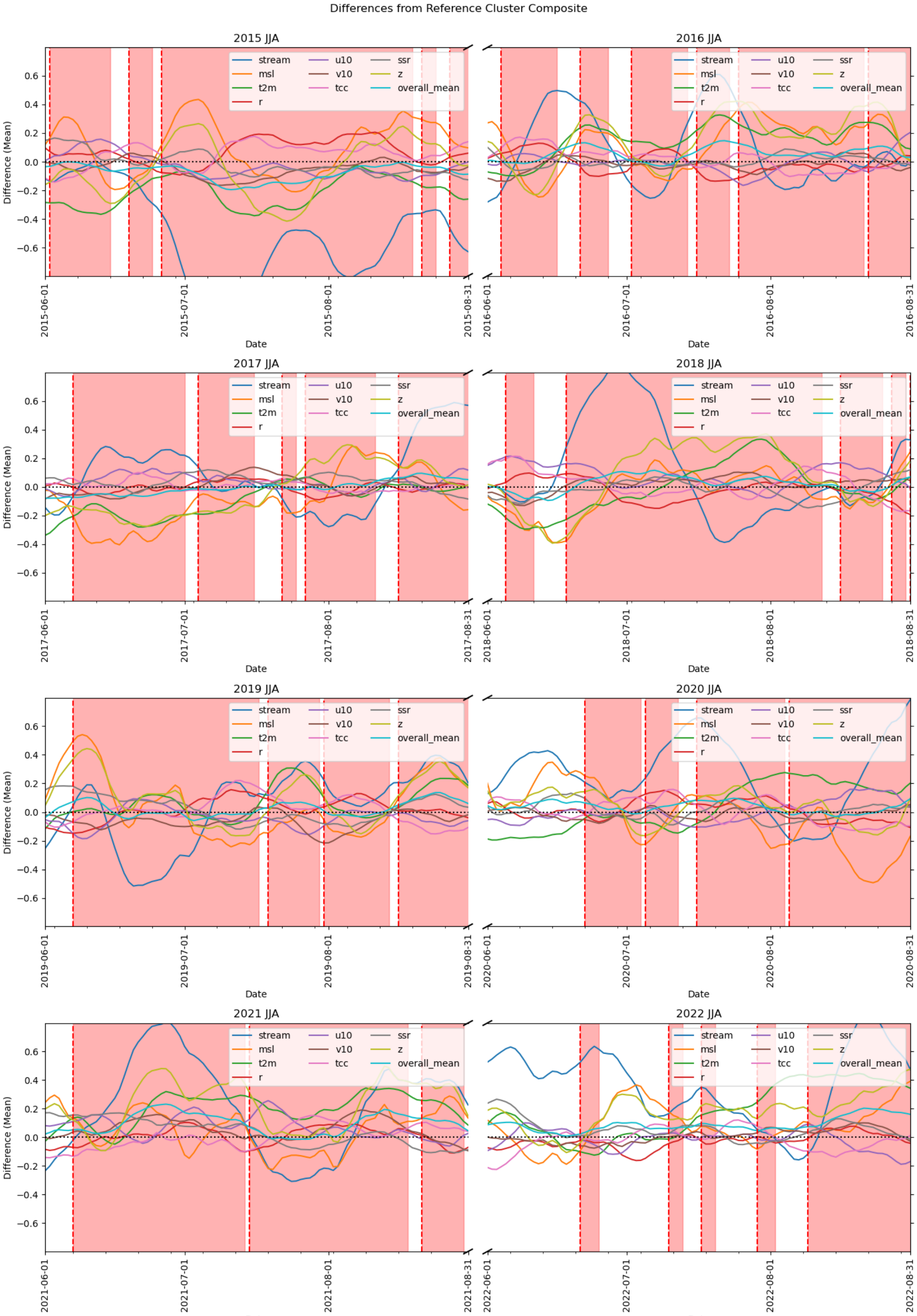



Figure 3: The mean bias between the center sample from summer cluster and the 11-day running window samples from summer months. x-axis is the central date for 11-day window. The different lines correspond to individual variables. The red shaded regions next to the dashed line indicate when heatwaves start and how long they last. To keep the plot simple, heatwaves that happen at the same time in different locations are merged into one representation.

## Major Comment 4

Why first t-SNE and then cluster? Why not directly cluster on the embeddings?

**Response:**

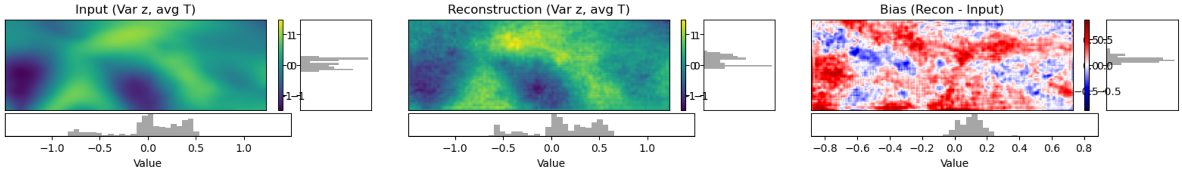
The pre-reduction step was recommended in the t-SNE paper (Maaten and Hinton 2008) and the documentation for the scikit-learn function (Pedregosa et al. 2011), which suggests applying a preliminary reduction step such as PCA before using t-SNE when working with latent spaces of higher dimension (above 50). In our case, the latent space had 128 dimensions. Applying PCA first was faster while producing similar patterns to using t-SNE alone, except for the inherent randomness in each run. We therefore adopted this approach to ensure both efficiency and robustness. To clarify this, we will modify the manuscript as follows:

> As an intermediate processing step, we first used Principal Component Analysis (PCA) to reduce the dimensionality of all heatwave samples from 1941-2022 to 50 components. Then, we applied the t-SNE algorithm to all heatwave samples to reduce them to 2 dimensions to visualize the latent space. This two-step approach is recommended for the analysis of a high-dimensional latent space to reduce the number of dimensions (see Maaten and Hinton (2008) and Pedregosa et al. (2011)) and helps to ensure efficiency and robustness. Because these steps involve stochasticity, we fixed the random seed to 42 (Adams 1979) for PCA, t-SNE, and GMM to ensure consistency across visualization runs.
>
> Adams, Douglas (1979). *The Hitch Hiker's guide to the Galaxy*. eng. Pan original. London: Pan Books.
>
> Maaten, Laurens van der and Geoffrey Hinton (2008). "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9.86, pp. 2579–2605.
>
> Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

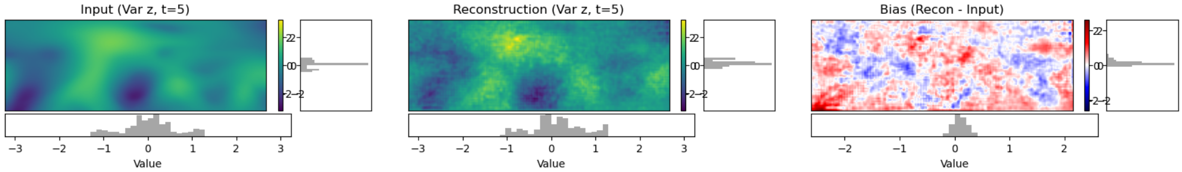
### Major Comment 5

Since your reconstruction accuracy is far from perfect (table 2), I wonder if you have looked into the model residuals. What are the errors the model makes? Does it miss any significant patterns related to heat waves? Is it unbiased? How well does it capture the extremes (i.e. the grid cells with heat waves)? This may be especially tricky since you seem to train on MSE.

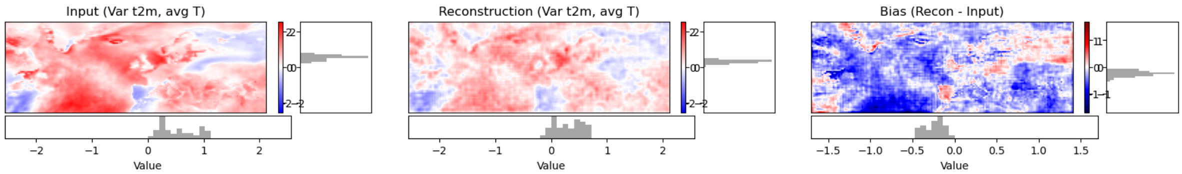
**Response:** We thank the reviewer for this suggestion. The reconstructions reproduce large-scale anomaly structures well but are smoother than the inputs, leading to a slight underestimation of extreme intensities. This outcome is consistent with the use of MSE loss, which favors average solutions. Importantly, the model retains the spatial location and amplitude of large-scale features, indicating that the essential heatwave-related circulation patterns are preserved despite reduced intensity of extremes (see Figure 4).



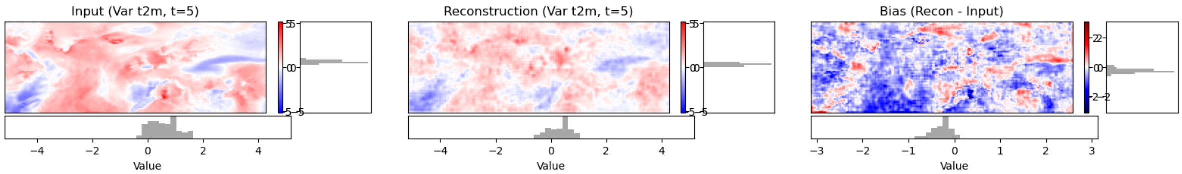

(a) Average geopotential height at 500 hPa bias as time mean




(b) Average geopotential height at 500 hPa bias on heatwave onset




(c) Average temperature bias as time mean




(d) Average temperature bias on heatwave onset

Figure 4: Input, reconstruction, and bias maps of geopotential height and temperature, shown as the time mean and at the heatwave onset date.

### Major Comment 6

Figure 5 would greatly benefit from actually including the maps of z500 from Lhotka & Kysely (2022) and plotting the differences.

**Response:** We thank the reviewer for this constructive suggestion. Since our focus here was on exploring the capability of the VAE, we chose to compare our results qualitatively with the published maps rather than repeating their analysis. We will expand the discussion around Figure 5 to further clarify the comparison to Lhotka and Kyselý (2022).

Lhotka, Ondřej and Jan Kyselý (2022). "The 2021 European Heat Wave in the Context of Past Major Heat Waves". In: *Earth and Space Science* 9.11, e2022EA002567.

DOI: https://doi.org/10.1029/2022EA002567. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022EA002567.

### Major Comment 7

L. 342ff can you elaborate a bit more how this work improves over the previous work from Happé et al 2024? Also, as is I assume the "in their" should be replaced by "to their"?

**Response:** We appreciate this important point. Our study builds directly on the framework of Happé et al. (2024), but extends it in several significant ways. While Happé et al. (2024) focused on summer heatwaves in the KNMI-LENTIS model ensemble using two atmospheric circulation variables and shorter temporal windows, our work applies the approach to ERA5 reanalysis, incorporates nine variables, and examines heatwaves year-round with longer pre-onset windows. Crucially, we also analyze temporal shifts by comparing historical (1941–1990) with recent (2001–2022) periods, revealing climate-change–related changes in heatwave dynamics. We will revise the Abstract, Introduction, and Discussion to make this novelty explicit and clearly separate where our study confirms previous results versus where it extends them.

### Major Comment 8

L 386ff - the current limitations read somewhat superficial. This approach builds on top of many assumptions. These should be clearly stated, please extend the section on limitations.

**Response:** We thank the reviewer for this important point. We extended the discussion of limitations to clarify the assumptions on which the methodology is based as follows:

> Despite these promising results for identifying heatwaves and their underlying atmospheric conditions, several limitations of the current approach must be acknowledged. First, we adopt a percentile based threshold for detecting heatwave sample onset dates using 2-meter temperature anomalies over Western Europe.

However, there is no universally accepted definition of heatwaves (Barriopedro et al. 2023; Boni et al. 2023). Alternative heatwave indices could yield different heatwave samples and thus different atmospheric patterns. Second, while we used variables that are related to atmospheric heatwaves, other potentially relevant drivers of heatwave dynamics might not be represented. As a result, the clusters we identify should be interpreted as emerging from the chosen subset of variables rather than from the full spectrum of processes affecting heatwaves. In addition, ERA5 itself has known uncertainties (Soci et al. 2024). Future work could improve this by also analyzing other variables to capture additional feedback processes. Third, the VAE is trained on historical heatwaves (1941–1990) and tested with recent heatwaves. While this was a deliberate choice to investigate how recent heatwaves differ from historical heatwaves, this assumes that the latent representation learned from the historical period can adequately model unprecedented patterns arising due to climate change. Furthermore, the VAE approach introduces some uncertainties due to the choices in model architecture. Key hyperparameters (such as the dimensionality of the latent space, learning rate, batch size, and the number of layers) can significantly influence training outcomes and the structure of the latent space. In addition to this, dimensionality reduction using PCA and t-SNE, followed by clustering with GMM, further affects how heatwave events are grouped in latent space, as all of these steps involve stochastic elements. Finally, the VAE reconstructions tend to smooth extremes, and the use of composite maps averaging 100 nearest latent samples, while reducing randomness and providing a more stable representation of atmospheric conditions, may mask variability within clusters. These limitations highlight that the VAE clustering approach is one possible data-driven analysis of atmospheric dynamics associated with heatwaves, rather than a definitive classification.

Barriopedro, D. et al. (2023). “Heat Waves: Physical Understanding and Scientific Challenges”. In: *Reviews of Geophysics* 61.2, e2022RG000780. DOI: `https://doi.org/10.1029/2022RG000780`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022RG000780`.

Boni, Zofia et al. (2023). “What is a heat(wave)? An interdisciplinary perspective”. en. In: *Climatic Change* 176.9, p. 129. DOI: `10.1007/s10584-023-03592-3`.

Soci, Cornel et al. (2024). “The ERA5 global reanalysis from 1940 to 2022”. In: *Quarterly Journal of the Royal Meteorological Society*. DOI: `10.1002/qj.4803`.

### Major Comment 9

Overall the storyline of the discussion needs to be streamlined. It jumps between results, limitations and outlooks. For instance, why is L 395ff after the limitations section?

**Response:** We thank the reviewer for bringing this to our attention. In the revised version, we will restructure the discussion to remove the current back-and-forth and improve the flow. For example, we improved the text on limitation and moved some parts to methodology to have a better flow as shown in Comment 4 and 8.

### Major Comment 10

The choice of splitting the dataset temporally also potentially influences attribution of the breakpoint you claim to find in heat wave characteristics. How robust is this finding if you change the periods over which you aggregate the latent space samples (in figure 4)?

**Response:** We thank the reviewer for bringing this point to our attention. Our goal was to assess whether recent heatwaves (2001–2022) differ from those in a historical baseline (1941–1990), which requires a temporal split by design. Importantly, the strong warming trend in Europe in terms of heatwave occurrence and intensity after the 1980s is well-documented in the literature (IPCC 2021; Reid et al. 2016; Lo and Hsu 2010), and our findings are consistent with this broader evidence. The breakpoint signal would not be expected in a random split, where events from both periods are mixed. We will clarify this in the revised manuscript and emphasize that the temporal split is an intentional feature of the study, aligned with our research question.

IPCC (2021). *Climate Change 2021: The Physical Science Basis.* Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Edited by V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou. Cambridge University Press (In Press).

Lo, Tzu-Ting and Huang-Hsiung Hsu (2010). “Change in the dominant decadal patterns and the late 1980s abrupt warming in the extratropical Northern Hemisphere”. In: *Atmospheric Science Letters* 11.3, pp. 210–215. DOI: https://doi.org/10.1002/asl.275. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/asl.275.

Reid, Philip C. et al. (2016). “Global impacts of the 1980s regime shift”. In: *Global Change Biology* 22.2, pp. 682–703. DOI: https://doi.org/10.1111/gcb.13106. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.13106.

**Minor Comment 1**

L 1-2 & L. 42f not sure I can follow the logic here. To me these are two separate aspects: one is defining heatwaves, which should be done using temperatures (or if you like, including also humidity) - and the other one is understanding their dynamics, for which other variables are also important.

**Response:** Thank you very much for bringing this issue to our attention. We will rewrite the abstract opening to state that definitions are typically based on temperature thresholds, while the study of underlying dynamics requires a broader multivariate perspective. We also adapt the introduction to explicitly separate these two aspects.

**Minor Comment 2**

L. 64 reads misleading, i would say the predominant use of VAEs is generative modeling - and anomaly detection is just one application.

**Response:** Thank you very much for your comment. We will modify the revised manuscript as follows:

> ... VAEs are primarily developed as generative models, able to learn complex data distributions and generate realistic synthetic samples (Kingma and Welling 2019). Beyond generative tasks, VAEs have also been widely used in anomaly detection in domains such as network security, risk management, health monitoring, and

computer vision (Pang et al. 2021; Nassif et al. 2021; Albuquerque Filho et al. 2022).

Albuquerque Filho, José Edson De et al. (2022). "A Review of Neural Networks for Anomaly Detection". In: *IEEE Access* 10, pp. 112342–112367. DOI: 10.1109/ACCESS.2022.3216007.

Kingma, Diederik P. and Max Welling (2019). "An Introduction to Variational Autoencoders". In: *Foundations and Trends® in Machine Learning* 12.4, pp. 307–392. DOI: 10.1561/2200000056.

Nassif, Ali Bou et al. (2021). "Machine Learning for Anomaly Detection: A Systematic Review". In: *IEEE Access* 9, pp. 78658–78700. DOI: 10.1109/ACCESS.2021.3083060.

Pang, Guansong et al. (2021). "Deep Learning for Anomaly Detection: A Review". In: *ACM Computing Surveys* 54.2, pp. 1–38. DOI: 10.1145/3439950.

## Minor Comment 3

L. 87, just to be sure, you compute the climatology, after having applied the temporal operators from table 1, correct?

**Response:** We thank the reviewer for bringing this issue to our attention, and yes, that is correct. We first apply the temporal operators listed in Table 1 to the hourly ERA5 data, and then compute the climatology from these aggregated fields. We will clarify this in the manuscript to avoid any ambiguity as follows:

> ... We first applied the temporal aggregation operators listed in Table 1 to convert the hourly ERA5 fields to daily values. From these daily fields, we then computed the 1941–1980 climatological mean and standard deviation using a 15-day moving window centered on each calendar day, which we used to standardize the variables and remove seasonal variability, since this period precedes the stronger warming trends after 1980 in Europe (Elguindi, Rauscher, and Giorgi 2013; Reid et al. 2016).

Elguindi, N., S. A. Rauscher, and F. Giorgi (2013). "Historical and future changes in maximum and minimum temperature records over Europe". en. In: *Climatic Change* 117.1, pp. 415–431. DOI: 10.1007/s10584-012-0528-z.

Reid, Philip C. et al. (2016). "Global impacts of the 1980s regime shift". In: *Global Change Biology* 22.2, pp. 682–703. DOI: https://doi.org/10.1111/gcb.13106. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.13106.

### Minor Comment 4

Have you tried a more standard 2D VAE (e.g. using an Imagenet pretrained ConvNext backbone)? In many applications the additional inductive bias on the temporal structure of the data may not translate in actual better performance, so it could be interesting to see if this is one case where it does.

**Response:** We thank the reviewer for this suggestion. We agree that comparing our approach with a more standard 2D VAE architecture would be an interesting exercise. In fact, we also experimented with a VAE with 2D convolutional encoder (decoder) layers to capture spatial features followed (preceded) by LSTM layers to capture temporal information. However, this setup did not yield stable or satisfactory reconstructions, likely because the LSTM struggled to capture the complex spatiotemporal dynamics of heatwave events when applied after a spatially compressed representation. By contrast, the 3D convolutional VAE provided more robust and coherent representations. Since our focus is on demonstrating the feasibility of incorporating temporal information into the VAE for heatwaves, we therefore focused on the 3D VAE for this study.

### Minor Comment 5

L. 191 not sure "temporal change" is the best term here, i read it first as if the temporal characteristics of heatwaves changed

**Response:** We thank the reviewer for pointing this out. Our intention was not to suggest that the temporal characteristics of individual heatwaves changed, but rather that the distribution of heatwave types shifted over time in the latent space. To avoid ambiguity, we will rephrase this part.

**Minor Comment 6**

L. 200 not sure it proves that the patterns are necessarily multivariate - probably you could learn this from a single variable like air temperature also.

**Response:** We thank the reviewer for this helpful observation. While certain patterns could in principle be learned from a single variable, our goal was to demonstrate that a multivariate approach provides a more detailed interpretation of anomaly patterns associated with heatwaves. We will revise the wording accordingly to avoid implying that the VAE model captures the multivariate nature, but instead emphasize that the added variables improve interpretability.

**Minor Comment 7**

L. 210 - cluster 1 & 2 are clearly dominated by one season, but 3&4 not sure i agree with your assignment, e.g. cluster 4 also has 29% DJF

**Response:** We thank the reviewer for this careful observation. Our assignment of Clusters 3 and 4 to transitional seasons was based on their modal distributions, which are dominated by SON and MAM events, respectively. We acknowledge that Cluster 4 also contains a notable fraction of DJF events (29%), which indicates that the separation between seasonal and transitional clusters is not absolute. We also examined the mean anomaly values of each cluster as shown in Figure 5. For example, Cluster 1 represents winter heatwaves with weaker t2m anomalies but warm advection (positive mean v10 anomalies) and lower solar input; Cluster 2 corresponds to summer heatwaves with the highest mean t2m, humidity, tcc and z500 anomalies; and Clusters 3 and 4 capture transition season events, one with high solar radiation and the other with strong pressure anomalies, respectively. In the revised manuscript, we will clarify that Clusters 3 and 4 capture mixed-season dynamics, with SON/MAM as the dominant but not exclusive contributions.

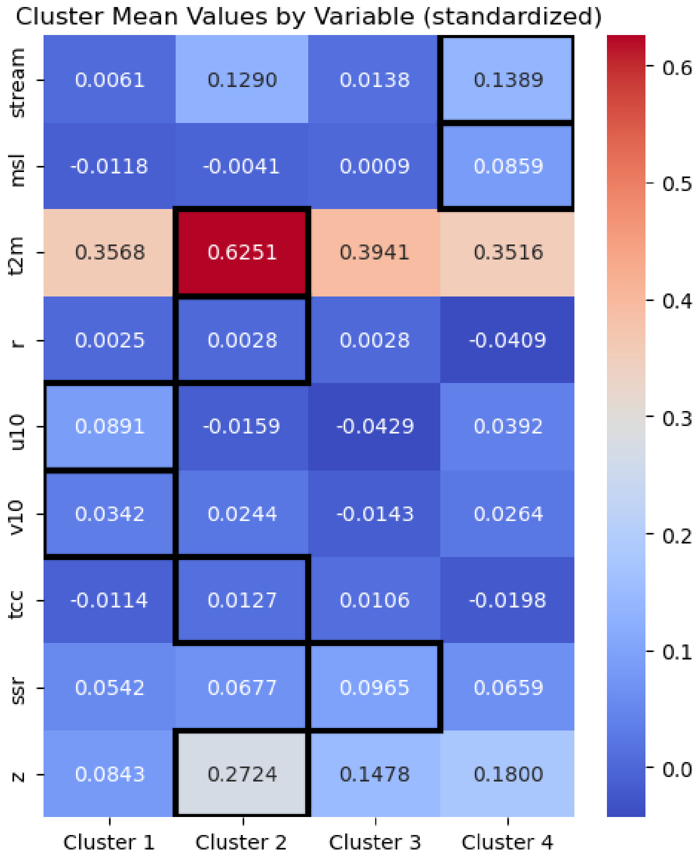



Figure 5: Mean cluster anomaly values for each variable.

**Minor Comment 8**

L 356-363 this reads like methods and not discussion & conclusion

**Response:** Thank you for your comment. We moved this part to Variational Autoencoder subsection:

> . . . This low-dimensional latent space compresses the input data and allows the discovery of hidden patterns within the data. It can be further analyzed using visualization techniques and clustering algorithms to identify patterns and regimes in the data (Happé et al. 2024; Lindhe, Ringqvist, and Hult 2021). The reconstruction produced by the VAE can be interpreted as a learned climatological baseline or ‘normal state,’ against which deviations can be studied. We focused on using latent space to understand the atmospheric patterns underlying different heatwave types. Similar latent space approaches have been successfully applied in climate

research (Behrens et al. 2022; Oliveira et al. 2022; Shamekh et al. 2023; Mooers et al. 2023; Camps-Valls et al. 2025).


Behrens, Gunnar et al. (2022). "Non-Linear Dimensionality Reduction With a Variational Encoder Decoder to Understand Convective Processes in Climate Models". In: *Journal of Advances in Modeling Earth Systems* 14.8, e2022MS003130. DOI: `https://doi.org/10.1029/2022MS003130`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022MS003130`.

Camps-Valls, Gustau et al. (2025). "Artificial intelligence for modeling and understanding extreme weather and climate events". en. In: *Nature Communications* 16.1, p. 1919. DOI: `10.1038/s41467-025-56573-8`.

Happé, T. et al. (2024). "Detecting Spatiotemporal Dynamics of Western European Heatwaves Using Deep Learning". In: *Artificial Intelligence for the Earth Systems* 3.4, e230107. DOI: `10.1175/AIES-D-23-0107.1`.

Lindhe, Adam, Carl Ringqvist, and Henrik Hult (2021). *Variational Auto Encoder Gradient Clustering*. DOI: `10.48550/arXiv.2105.06246`.

Mooers, Griffin et al. (2023). "Comparing storm resolving models and climates via unsupervised machine learning". In: *Scientific Reports* 13.1, p. 22365. DOI: `10.1038/s41598-023-49455-w`.

Oliveira, Dario A. B. et al. (2022). "Controlling Weather Field Synthesis Using Variational Autoencoders". In: *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5027–5030. DOI: `10.1109/IGARSS46834.2022.9884668`.

Shamekh, Sara et al. (2023). "Implicit learning of convective organization explains precipitation stochasticity". In: *Proceedings of the National Academy of Sciences* 120.20, e2216158120. DOI: `10.1073/pnas.2216158120`. eprint: `https://www.pnas.org/doi/pdf/10.1073/pnas.2216158120`.


**Minor Comment 9**

Could the latent space clustering primarily reflect heatwave intensity rather than pattern type? How important is the non-linear/multimodal aspect of your method for separating these types?

**Response:** We thank the reviewer for this insightful comment. We agree that clustering in the latent space could be dominated by heatwave intensity. To address this issue, we examined the mean anomaly values of each cluster as shown in Figure 5. While the heatwave intensity does play a role in the clustering (e.g., Cluster 2 shows the strongest positive t2m anomalies), clusters differ in other variables, indicating that the separation is not only based on intensity. For example, Cluster 1 represents winter heatwaves with weaker t2m anomalies but warm advection (positive mean v10 anomalies) and lower solar input; Cluster 2 corresponds to summer heatwaves with the highest mean t2m, humidity, and z500 anomalies; and Clusters 3 and 4 capture transition season events, one with high solar radiation and the other with strong pressure anomalies, respectively. We will add this analysis to the manuscript.

### Minor Comment 10

A stylistic note: The introduction might come across as overly alarmist. Consider softening the tone slightly.

**Response:** We thank the reviewer for this remark. We will revise the introduction to soften the tone and maintain the focus on the scientific background. We will review the parts related to the impact of climate change on global health.