B. Maczák, K. P. Komáromi and G. Vadai, "Different Aspects of the Scale-Free Nature of Human Activity – Examination of Its Spectral and Statistical Properties," 2023 International Conference on Noise and Fluctuations (ICNF), Grenoble, France, 2023, pp. 1-4, doi: 10.1109/ICNF57520.2023.10472755.

Different Aspects of the Scale-Free Nature of Human Activity – Examination of Its Spectral and Statistical Properties

Bálint Maczák
Department of Technical Informatics
University of Szeged
Szeged, Hungary
orcid.org/0000-0002-1867-5861

Krisztina Panna Komáromi
Department of Technical Informatics
University of Szeged
Szeged, Hungary
komaromi.krisztina.panna@stud.u-szeged.hu

Gergely Vadai
Department of Technical Informatics
University of Szeged
Szeged, Hungary
orcid.org/0000-0002-2720-3380

***Abstract*—The scale-free nature of human dynamics is commonly identified by fitting different models to distribution functions of various quantities, both in mobility and actigraphy datasets. The spectral scale independence of such recordings is also addressed in the literature in the form of 1/*f* noise, but these examinations have typically been limited to medical purposes by analysing over short time windows. Recently, we have shown that the spectral characteristics of the acceleration measured at the wrist and the activity signals calculated from it in different ways have the same spectral characteristics, including 1/*f* noise. By analysing a set of actigraphic measurements, we show that while the spectral scale-free nature of human activity is general, statistical analyses can lead to different results depending on the various analytical approaches and activity determination methods.**

Keywords—1/f noise, actigraphy, scale-free nature, human dynamics

I. Introduction

Although the scale-free nature of human dynamics is apparent regarding both human mobility and activity, various models emerge from the statistical distributions across different datasets beyond power law, such as the truncated power law, lognormal, exponential, or more complex models from the composition of the above [1].

The locomotor activity of individuals can be measured using the method of actigraphy, which is commonly utilized in medical fields, too. The method quantifies the subject's activity based on the acceleration of its non-dominant wrist. However, the exact procedure widely varies between manufacturers and is commonly unpublished. Therefore, scientists analyse differently produced activity values, typically without giving sufficient details about the methodology [2]. In addition, the statistical analysis of activity data is also diversified: different quantities can be analysed (e.g., activity values, their increment, or active and passive periods) with various statistical analytical methods (e.g., probability density function or complementary cumulative distribution function-based model fitting).

The scale-free nature of human dynamics is also identified in the frequency domain, as 1/*f* noise appears in human mobility and activity signals. We have recently shown that the fluctuations of daily human activity – whether talking about the raw acceleration of the wrist or the activity calculated from it – follow a general spectral characteristic [3], which is very similar to the one we found earlier in the displacement data for human mobility [4]. These spectral properties of human activity are depicted in Fig. 1 subplot a). At low frequencies (below ~10^{-4} Hz), the spectra exhibit white noise with the presence of the peaks at about 24 and 12 hours. At higher frequencies, the spectra contain 1/*f* noise over multiple frequency decades.

In the following, we will investigate if there are differences in the distribution of active and passive periods of human activity depending on how the activity signal is determined from a given measurement and the definition of these active and passive periods – whereas the spectral properties of daily human activity are universal.

II. Determination and Examination of the Activity Signals

While examining the statistical properties of human activity based on an actigraphic dataset, the following aspects can have a major impact on our results in addition to which experimental dataset are we investigating.

A. What kind of actigraphic data do we measure?

Even though actigraphic devices [2] inherently measure the acceleration of the subject's wrist typically along three axes to quantify the locomotor activity, these devices do not usually store the acceleration signal, but the activity values derived from it. However, how activity values are determined is commonly hidden, unstandardised and highly manufacturer or device-dependent [5], therefore, actigraphic recordings from different sources should be treated with caution as they may not be directly comparable. Alongside others, we have already pointed out the dangers of this in one of our previous works [2], while we also mapped, categorised, and compared the most common activity determination methods. Although these methods work differently, we identified the general steps beyond them, along which the activity determination procedures become comparable. The device measures the raw acceleration with a relatively high sampling rate (e.g., 10-100 Hz), then the acceleration data is preprocessed (among other reasons, to eliminate the effect of Earth's gravity, in more detail later). After that, the preprocessed acceleration signal is cut into epochs, which are consecutive timeslots of equal length (e.g., 1 minute). An activity value is then determined for every epoch based on the preprocessed acceleration using an activity metric, which is a set of typically nonlinear operations (numerous metrics will be presented later). As a result of the activity determination procedure, the acceleration signal measured at a relatively high sampling rate is converted into a sparsely sampled activity signal. The heterogeneity of activity determination methods is mainly caused by the diversity in the preprocessing techniques and the usage of different activity metrics. In the following, we will present those we utilise in the current work using the abbreviations defined in our previous work [2].

The raw acceleration data can be preprocessed in various ways [2], for example, by calculating the unfiltered magnitude

of acceleration (UFM, i.e., the length of the acceleration vectors) from the unfiltered acceleration values of the x, y, and z axes (UFX, UFY, UFZ, respectively). Then, the effect of Earth's gravity (g) can be eliminated by calculating the absolute distance of the magnitude values from 1 g, which results in the unfiltered normalised magnitude of acceleration (UFNM). The elimination of g can be achieved with a band-pass filter, too, which filters out low-frequency components, and high-frequency noise, such as components related to involuntary movements. The filtering can be executed in two ways. Firstly, the filter can be applied on the magnitude of acceleration data (UFM), resulting in the postfiltered magnitude of acceleration (FMpost). Secondly, the filter can be applied to the per-axis acceleration (UFX, UFY, UFZ) prior to magnitude calculation, resulting in the filtered acceleration values along the x, y, and z axes (FX, FY, FZ, respectively), from which the prefiltered magnitude data can be calculated (FMpre).

These differently preprocessed acceleration data are all suitable to be cut into epochs and to calculate activity from them using an activity metric. However, there are numerous activity metrics in the literature. We have previously [2] collected 7 typical metrics that operate on significantly different principles, as presented. One of the simplest metrics (PIM – Proportional Integration Method) is based on the integration of the acceleration data. Other classical metrics rely on the threshold intersections of the acceleration data: whereas Zero Crossing Mode (ZCM) counts the threshold crosses, Time Above Threshold (TAT) measures the time when the data exceeds the threshold. In addition, there are metrics that averages the acceleration data: while the Euclidean Norm Minus One (ENMO) utilizes a special averaging rule, the High-pass Filtered Euclidean Norm (HFEN) requires specially preprocessed acceleration data. Beyond these, metrics relying on the standard deviation (MAD – Mean Amplitude Deviation) of the magnitude of acceleration or variance (AI – Activity Index) of the per-axis data also exist. Thus, there are certain metrics that require acceleration data preprocessed in a strict way, while others can also be applied to differently preprocessed acceleration signals. The details about the limitations, compatibility, and proper combinations of preprocessing techniques and activity metrics can be found in our previous work [2]. To identify activity signals determined in a given way, we will use the activity metric as an operation, and the preprocessing method as its argument (e.g., PIM(UNFM)).

Nowadays actigraphs that can store the raw and frequently sampled acceleration data are becoming more and more popular, and in this case, any kind of activity data can be determined after the measurement using the definitions above. Previously [2], we also compared these methods and found out that there could be considerable differences – in the sense of linear relationship – between activity signals calculated from differently preprocessed acceleration data. This suggests that the way activity signals are determined may have a substantial impact on the statistical analysis of such data.

B. *What kind of activity-related quantity are we examining?*

When analysing the properties of activity signals, numerous quantities can be considered, such as the distribution of activity data, or the distribution of the active and passive periods of the motion separately. Even if we limit ourselves to the latter and more common method, i.e., the analysis of active (AP) and passive periods (PP) of activity signals $a(t)$, there are considerable differences in the related literature. Various threshold rules are prevalent to define the active and passive activity values, such as the overall average of the activity data (e.g., mean(a)) [6], 70% of the mean(a) [7], or the average of the non-zero activity values (e.g., mean($a > 0$)) [8]. A different theoretical approach is to consider consecutive non-zero activity values as events of a given duration and the time elapsed between them as waiting times; however, similar models could be fitted to the distributions in this sense, too [9]. Therefore, a technical analogy can be instituted between this and the former approach if the threshold level is chosen to be negligibly small (i.e., 1% of the mean(a)).

C. *What kind of numerical method do we use for statistical analysis?*

Various numerical methods are available to examine the distribution of the active/passive periods, such as the estimation of their Probability Density Function (PDF) [10]. To calculate PDF, the range of the samples (e.g., passive periods) must be split into bins, but choosing the number of bins can be problematic. If it is insufficiently low, the PDF will have a low resolution; at the other extreme, the PDF will have a lot of bins with a negligible number of samples. The use of PDF has additional considerations if examining heavy-tailed distributions on log-log scales (e.g., power-law distribution), which – as we will see – describes the active/passive periods. The binning can happen either linearly (constant bin widths) or logarithmically; the latter approach reduces noise, especially in the tail, as it will be covered by fewer bins containing more samples.

However, the binning is avoidable if calculating the Complementary Cumulative Distribution Function (CCDF or survival function, i.e., 1-CDF) [11] instead of PDF, as follows. For each unique sample (e.g., passive periods) sorted in ascending order, the number of samples that are greater than or equal to it must be divided by the total number of samples. Therefore, the CCDF defines the probability that the active/passive period is longer than or equal to a given length. The so-called rescaling of CCDFs is also prevalent in the literature [8, 12]: before CCDF calculation, the samples are divided by their mean. Rescaling is relevant when comparing CCDFs whose underlying data may cover different ranges, but it also raises further questions of when to use it. Although CCDF is the widespread approach to statistically examine the active/passive periods in actigraphic data [1, 6–9, 12, 13], it has limitations, too. Even if the distribution of the observed quantity theoretically follows pure power-law scaling, the finite-size of the empirical data can lead to the truncation (i.e., downward bending from the straight line at larger values on log-log scales) of the power-law form [14–16], whereas PDFs are less sensitive to such distortions. Nonetheless, if the CCDF follows power-law scaling with exponent α, then so is the PDF but with exponent of $\alpha + 1$.

Thus, the distribution of active/passive periods of human activity can be examined either through PDFs or CCDFs, but the decision between these methods is not obvious, as both have various limitations and advantages and could lead to different interpretations if analysing heavy-tailed – especially power-law – distributions [17], which is a common perception

for the passive periods of human activity [8, 12]. Additional aspects may also influence these statistical analyses, such as the following. If we have data from multiple subjects, the question is whether we calculate the distributions per subject [6], the average of these distributions [8], or determine the distribution of the aggregated activity data [7]. Beyond visual interpretation, the evaluation of the distributions can be done by fitting various models (e.g., power-law, exponential, or lognormal) using different fitting methods (e.g., least squares, or maximum likelihood method), whose correct choice raises further questions. [1, 11, 17].

III. Results

To investigate the impact of methodological differences in human activity's statistical analysis presented above, we have reproduced the different steps of several important studies of the last decade [6–9]. For this, we used the same 10-day-long raw acceleration recordings of 42 healthy, free-living individuals that we also used for our previous works [2, 3]. These signals were measured at the sampling rate of 10 Hz in the ±8 g range with such an actigraphic device [2] that was specially developed to store raw triaxial acceleration data. From the raw acceleration recording of each subject, were able to calculate the different kinds of activity signals used by the 4 relevant works for the same movement, and then we applied the different threshold rules related to these works to analyse the aggregated distribution (i.e., distribution of the measured subjects' pooled data) of active and passive periods. Although it is not always known what kind of preprocessing has been used to determine activity values in these related works, and in some cases, the activity metrics differ from those we have previously standardised [2], we have intended to choose the closest method during reproduction. As it can be seen in Fig. 1, the PSD of the 3 differently computed activity signals – which are related to the works to be reproduced – matches and exhibits $1/f^{\alpha}$ scaling ($\alpha = 1$). In contrast, the non-rescaled CCDF of their passive periods follow power-law scaling (i.e., exponent of 2 for the PDF) in different degrees depending on the methodological differences.

Beyond the way of activity values are determined and the selection of the active/passive threshold rule, there are further methodological considerations about the statistical analytical method, for example, whether examining the distributions through CCDFs or PDFs, and if the samples are preliminarily rescaled. As presented in Fig. 2, the rescaling of the passive

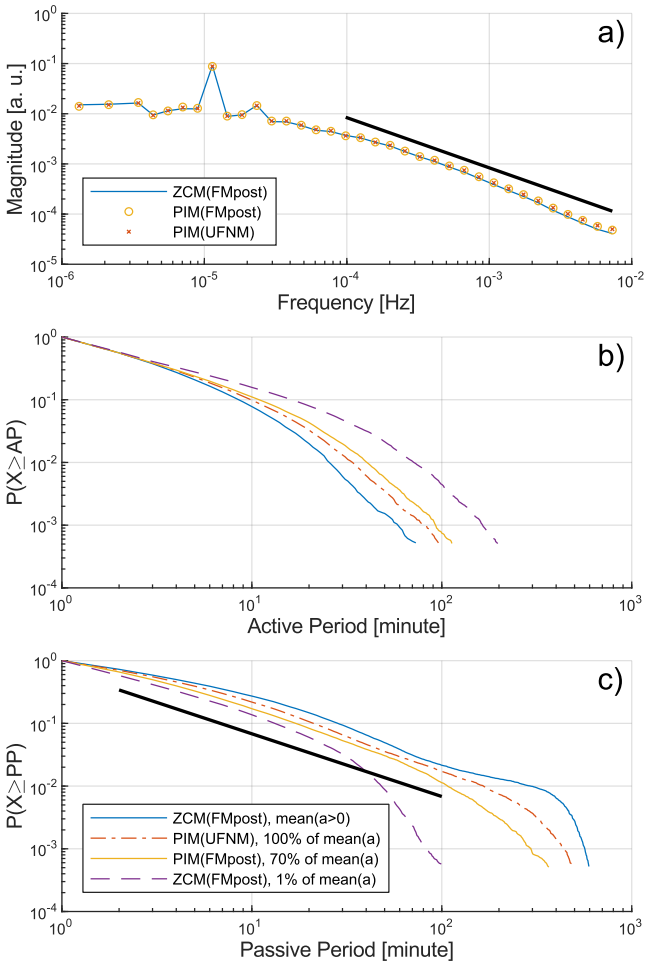

Fig. 1. a) The PSDs of the examined activity signals. b-c) The aggregated, non-rescaled CCDF of active and passive periods of these activity signals, using various threshold levels. The power-law scaling with an exponent of 1 is illustrated as a black trendline.

periods may help to visually compare distributions originating from different datasets, in our case, different types of activity signals. By comparing the PDFs (Fig. 2 subplot c) calculated using 100 logarithmically spaced bins and the CCDFs (Fig. 2 subplot a) of the non-rescaled passive periods, the advantages, and disadvantages of the two approaches can be observed. The power-law scaling is more prominent for the PDFs, however, the resolution, as well as the noise in the tail is heavily depending on the number of bins. We can avoid the necessity

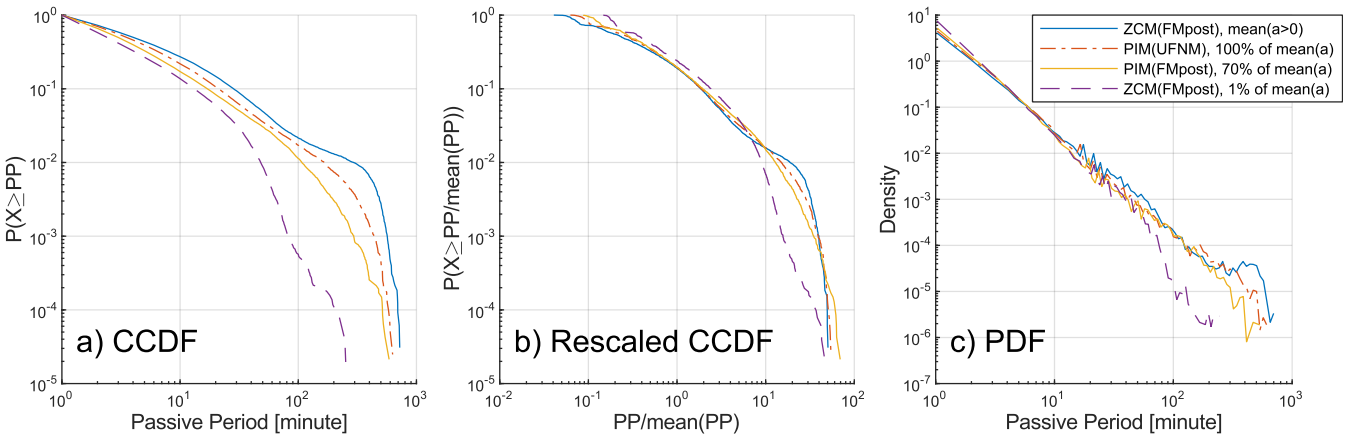

Fig. 2. a) Similar to Fig. 1 c), but the CCDFs were visualized over their entire range, without omitting passive periods whose P(X≥PP) is less than $10^{-3.5}$. b) The rescaled CCDFs of a), i.e., the samples were divided by their mean prior to CCDF calculation. c) The PDFs (calculated using logarithmic binning for 100 bins) of the same samples as for a).

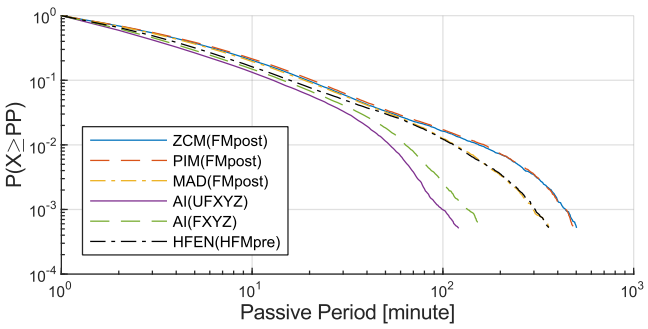

Fig. 3. The aggregated, non-rescaled CCDF of passive periods for different types of activity signals using a threshold value of mean(a). The curves of the unplotted activity signal types coincide with those of ZCM and PIM.

of binning using CCDFs, but at longer periods, the CCDF deviates from the ideal line of the power-law scaling that would otherwise be assumed based on the shape of the PDF. Such truncation for the CCDF of power-law distribution may happen due to the finite-size of the empirical data, which is a known phenomenon [14–16]. Compared to Fig. 1 c), in Fig. 2, we have plotted the CCDFs over their entire range, down to much lower probability values. However, these small probabilities are substantially influenced by the heterogeneity of the measured subjects as we analysed their aggregated samples. Parallelly, the heavy tails of the PDFs also become especially noisy, as negligible number of samples falls in the bins at the tail. Thus, another question is the range of the distributions that we want to analyse and fit model to it.

To observe the effect of the activity determination, we calculated activity signals using the different metrics [2] while the preprocessing method (where it was technically possible) and the threshold rule of separating active/passive periods were fixed (detailed in Fig. 3). As we presented previously, the spectral nature of these activity signals is generally independent of the way these signals are determined. On the contrary, Fig. 3 shows that the distribution of the passive periods differs for several activity metrics and follow power-law scaling to various extent, which indicates that the way of activity determination influences the results. The fact that the passive distributions of the ZCM and PIM activity signals are identical for a same threshold rule in Fig. 3 confirms which we have observed in Fig. 1: the choice of the threshold level has a large impact on the results. Yet, Fig. 3 shows that the use of several other metrics can cause major differences in the results, too.

IV. Conclusion

We have shown that the distribution of active and passive periods derived from the same measurements using different methodologies (e.g., activity calculation, active/passive threshold level) prevalent in the literature widely differ, whereas the underlying activity signals generally follow the same spectral characteristics – including $1/f$ noise. Consequently, our results suggest that the scale-free nature of human activity is more directly addressable in the frequency domain compared to the statistical approach, whose methodology-dependence may affect the model fitting and further interpretation.

Acknowledgement

The authors thank Zoltán Gingl for his valuable insights and support in interpreting the results of this study.

This research was supported by project TKP2021-NVA-09. Project no. TKP2021-NVA-09 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme.

References

[1] J. J. Chapman, J. A. Roberts, V. T. Nguyen, and M. Breakspear, “Quantification of free-living activity patterns using accelerometry in adults with mental illness,” *Sci Rep*, vol. 7, no. 1, Art. no. 1, Mar. 2017, doi: 10.1038/srep43174.

[2] B. Maczák, G. Vadai, A. Dér, I. Szendi, and Z. Gingl, “Detailed analysis and comparison of different activity metrics,” *PLOS ONE*, vol. 16, no. 12, p. e0261718, Dec. 2021, doi: 10.1371/journal.pone.0261718.

[3] B. Maczák, Z. Gingl, and G. Vadai, “General spectral characteristics of human activity and its inherent scale-free fluctuation.” 2023. doi: 10.21203/rs.3.rs-2539448/v1.

[4] G. Vadai, A. Antal, and Z. Gingl, “Spectral Analysis of Fluctuations in Humans’ Daily Motion Using Location Data,” *Fluct. Noise Lett.*, vol. 18, no. 02, p. 1940010, Jun. 2019, doi: 10.1142/S0219477519400108.

[5] F. Richert, “movisens GmbH - Activity Monitoring and Existing Algorithms Regarding Energy Expenditure Estimation: A Short Review.” Accessed: Jun. 24, 2021. [Online]. Available: https://www.movisens.com/wp-content/downloads/WPMove3_Activity_Algorithms.pdf

[6] S. E. Huber *et al.*, “Assessment of Fractal Characteristics of Locomotor Activity of Geriatric In-Patients With Alzheimer’s Dementia,” *Front Aging Neurosci*, vol. 11, p. 272, 2019, doi: 10.3389/fnagi.2019.00272.

[7] J.-H. Lee *et al.*, “Statistical properties of human activity and criticality in active behavior,” *EPL*, vol. 126, no. 6, p. 68001, Jul. 2019, doi: 10.1209/0295-5075/126/68001.

[8] T. Nakamura *et al.*, “Of Mice and Men — Universality and Breakdown of Behavioral Organization,” *PLOS ONE*, vol. 3, no. 4, p. e2050, Apr. 2008, doi: 10.1371/journal.pone.0002050.

[9] D. Chialvo *et al.*, “How we move is universal: Scaling in the average shape of human activity,” *Papers in Physics*, vol. 7, pp. 070017–070017, Nov. 2015, doi: 10.4279/pip.070017.

[10] E. P. White, B. J. Enquist, and J. L. Green, “On Estimating the Exponent of Power-Law Frequency Distributions,” *Ecology*, vol. 89, no. 4, pp. 905–912, 2008, doi: 10.1890/07-1288.1.

[11] A. Clauset, C. R. Shalizi, and M. E. J. Newman, “Power-Law Distributions in Empirical Data,” *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, Nov. 2009, doi: 10/dd7xhj.

[12] J. K. Ochab *et al.*, “Scale-free fluctuations in behavioral performance: Delineating changes in spontaneous behavior of humans with induced sleep deficiency,” *PLoS ONE*, vol. 9, no. 9, 2014, doi: 10.1371/journal.pone.0107542.

[13] T. Nakamura, K. Kiyono, K. Yoshiuchi, R. Nakahara, Z. R. Struzik, and Y. Yamamoto, “Universal Scaling Law in Human Behavioral Organization,” *Phys. Rev. Lett.*, vol. 99, no. 13, p. 138103, Sep. 2007, doi: 10.1103/physrevlett.99.138103.

[14] N. Marshall, N. M. Timme, N. Bennett, M. Ripp, E. Lautzenhiser, and J. M. Beggs, “Analysis of Power Laws, Shape Collapses, and Neural Complexity: New Techniques and MATLAB Support via the NCC Toolbox,” *Frontiers in Physiology*, vol. 7, 2016, Accessed: Sep. 21, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphys.2016.00250

[15] S. M. Burroughs and S. F. Tebbens, “Upper-truncated Power Laws in Natural Systems,” *Pure appl. geophys.*, vol. 158, no. 4, pp. 741–757, Apr. 2001, doi: 10.1007/PL00001202.

[16] J. Alstott, E. Bullmore, and D. Plenz, “powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions,” *PLOS ONE*, vol. 9, no. 1, p. e85777, Jan. 2014, doi: 10.1371/journal.pone.0085777.

[17] J. Savre and G. Craig, “Fitting Cumulus Cloud Size Distributions From Idealized Cloud Resolving Model Simulations,” *Journal of Advances in Modeling Earth Systems*, vol. 15, no. 5, p. e2022MS003360, 2023, doi: 10.1029/2022MS003360.