The Journal of Open Source Software

TimeHiVE: Hierarchical Moving-Window Time Series Analysis Toolkit in R

Vladimiro Andrea Boselli [1], **Paolo Tagliolato** [1], **and Alessandro Oggioni** [1]

1 Institute for Electromagnetic Sensing of the Environment of the National Research Council of Italy (IREA-CNR), Italy

Abstract

TimeHiVE is an R package designed for hierarchical moving-window statistical analysis of time series data. The package addresses a fundamental challenge in time series analysis: the selection of appropriate window sizes for moving-window calculations. Traditional moving-window approaches require users to specify a fixed window size, which can obscure important patterns if chosen inappropriately. TimeHiVE eliminates this constraint by systematically computing statistics across all possible window sizes, enabling researchers to explore the full spectrum of temporal patterns in their data.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: AHM Mahfuzur Rahman

Reviewers:

- @g-rppl
- @iosefa

Submitted: 02 October 2025
Published: unpublished

Summary

TimeHiVE by Boselli et al. (2025b) is an R package designed for hierarchical moving-window statistical analysis of time series data. The package addresses a fundamental challenge in time series analysis: the selection of appropriate window sizes for moving-window calculations. Traditional moving-window approaches require users to specify a fixed window size, which can obscure important patterns if chosen inappropriately. TimeHiVE eliminates this constraint by systematically computing statistics across all possible window sizes, enabling researchers to explore the full spectrum of temporal patterns in their data.

The package provides implementations for both single time series analysis (including means, trends, and custom statistics) and coupled time series analysis (including Pearson and Mann-Kendall correlations). TimeHiVE features Mann-Kendall statistics algorithms with O(n log n) time complexity, parallel computation capabilities, and customisable visualization tools. These features make it particularly valuable for environmental and climate research, where understanding phenomena across multiple timescales is essential.

Statement of need

Moving-window statistical analysis is a fundamental technique in time series analysis, used to study the evolution of data over time in a dynamic and localized manner. This approach involves calculating statistics on subsets of consecutive data points that progressively move along the series. The technique helps identify localized trends, seasonality, and anomalies while reducing the impact of random fluctuations, highlighting meaningful patterns.

The challenge of selecting an optimal window size is well-documented in time series literature. A window that is too large may obscure important details, while one that is too narrow may be overly sensitive to noise. This problem is particularly relevant in climate science, where researchers need to examine phenomena across multiple timescales. Previous work by Brunetti and colleagues (Brunetti et al., 2006, 2009) demonstrated the value of hierarchical moving-window approaches for climate data analysis, but implementation required custom code.

TimeHiVE makes this analytical approach accessible to a broader research community by providing a well-documented, efficient implementation in R. The package offers several advantages over existing solutions:

1. It eliminates the need for *a priori* window size selection
2. It provides optimized implementations of common statistical tests
3. It supports both single and coupled time series analysis
4. It offers flexible visualization capabilities
5. It allows for custom statistical functions through its extensible architecture

The package is particularly valuable for environmental researchers, climate scientists, and anyone working with temporal data where patterns may manifest across different timescales. By providing a comprehensive toolkit for hierarchical moving-window analysis, TimeHiVE enables more thorough exploratory data analysis and more robust pattern detection in time series data.

TimeHiVE implements two main analytical functions: `TH_single()` for single time series analysis and `TH_coupled()` for analyzing relationships between two time series. Additionally, the `TH_tweak()` function allows users to implement custom statistical functions, making the package extensible to specialized analytical needs.

The package includes comprehensive visualization functions (`TH_plots()`, `TH_plotc()`, and `TH_plott()`) that generate heatmap-style representations of results, with time on the x-axis and window size on the y-axis. This visualization approach, inspired by Brunetti et al. (Brunetti et al., 2006, 2009), enables intuitive interpretation of patterns across timescales. A recent application has been carried out on precipitation time series derived from FAIR datasets provided by the eLTER Research Infrastructure eLTER-RI, offering a first example of their potential to support preliminary data exploration and quality assessment within eLTER (Boselli et al., 2025a).

Usage examples

In this section we give a visual example for a single time series and a visual example for coupled time series

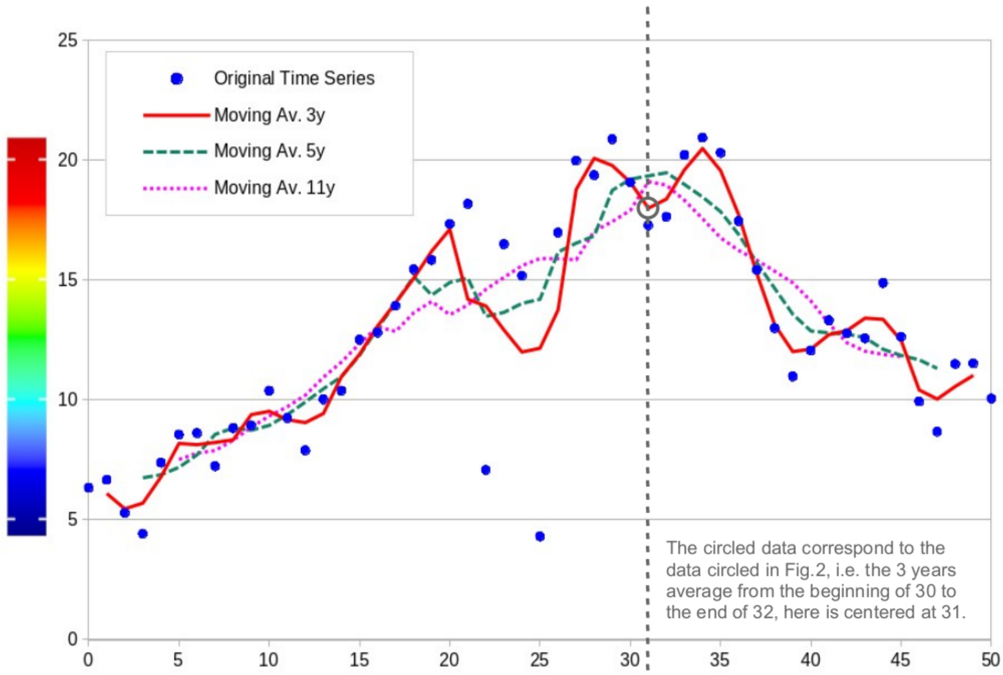


Figure 1: Different Moving Average vs Original Syntetic Time Series. Color scale on y-axis is meant only to match colors in Fig. 2

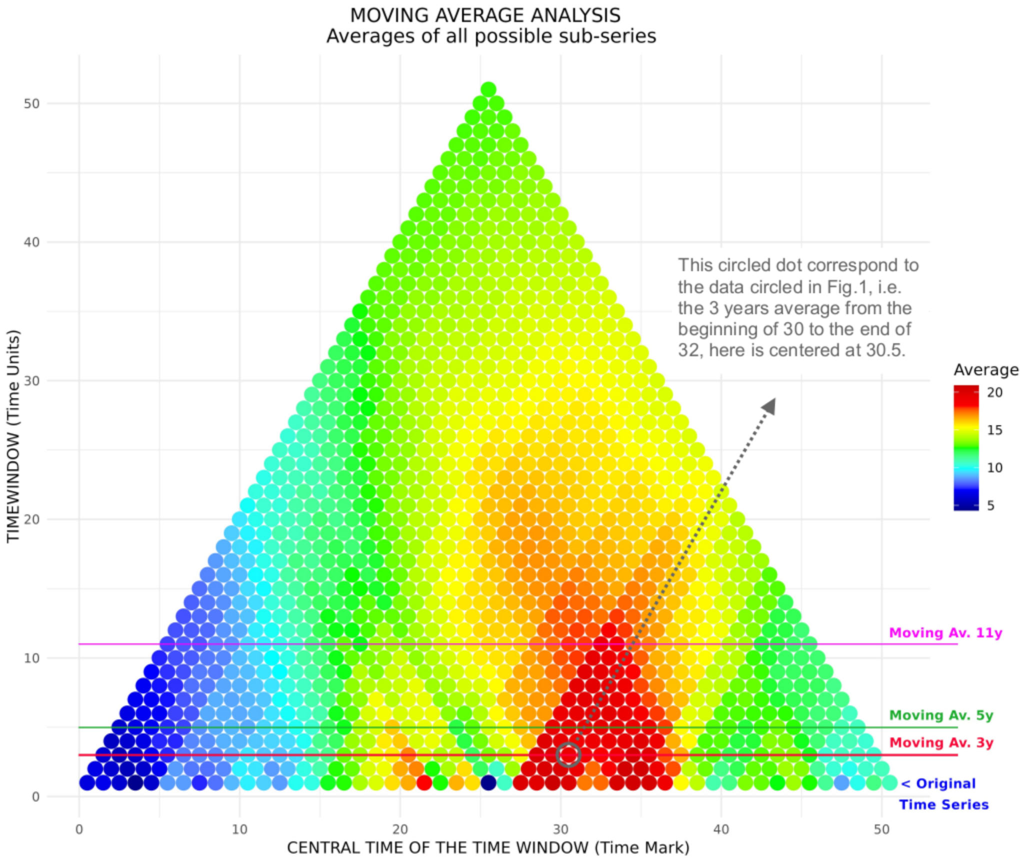
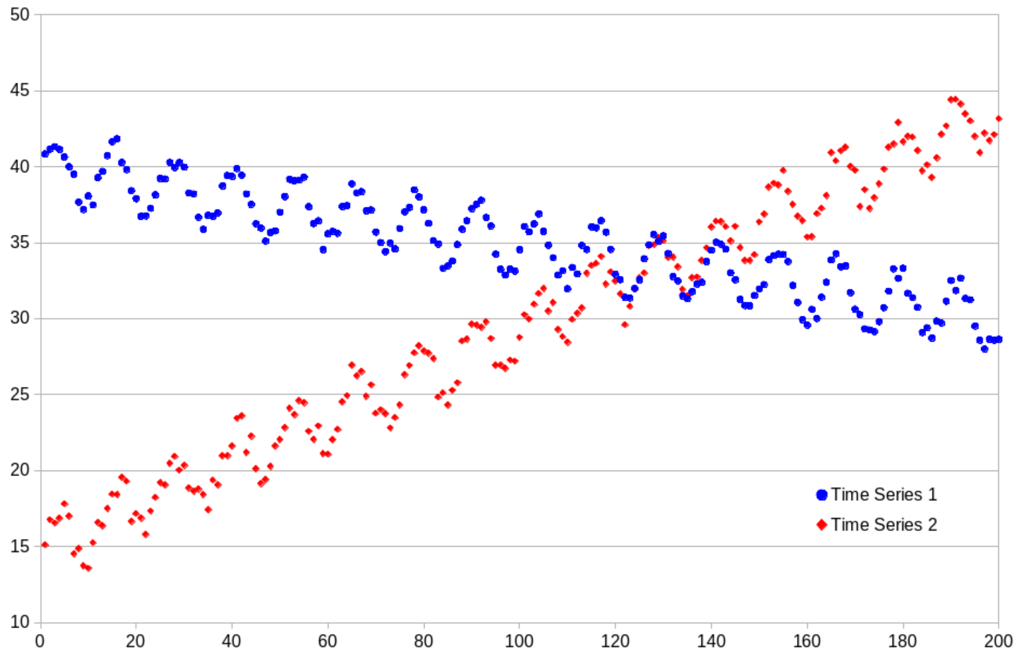


Figure 2: Representation of all the possible moving averages for the Original Syntetic Time Series, moving average analyses represented in Fig. 1 are highlighted with comments on the results.



The Journal of Open Source Software

Figure 3: Representation of two Time Series positively correlated for short periods but negatively correlated for long periods. The series are built as: `TS1 = 40 + 2*sin(t/2) - t/20 - rand(-2/3, 2/3)` *and* `TS2 = 15 + 2*sin(t/2) + t/7 - rand(-2/3, 2/3)`*. The sine component creates short-term positive correlation, while the* `t/n` *terms drive long-term negative correlation.*

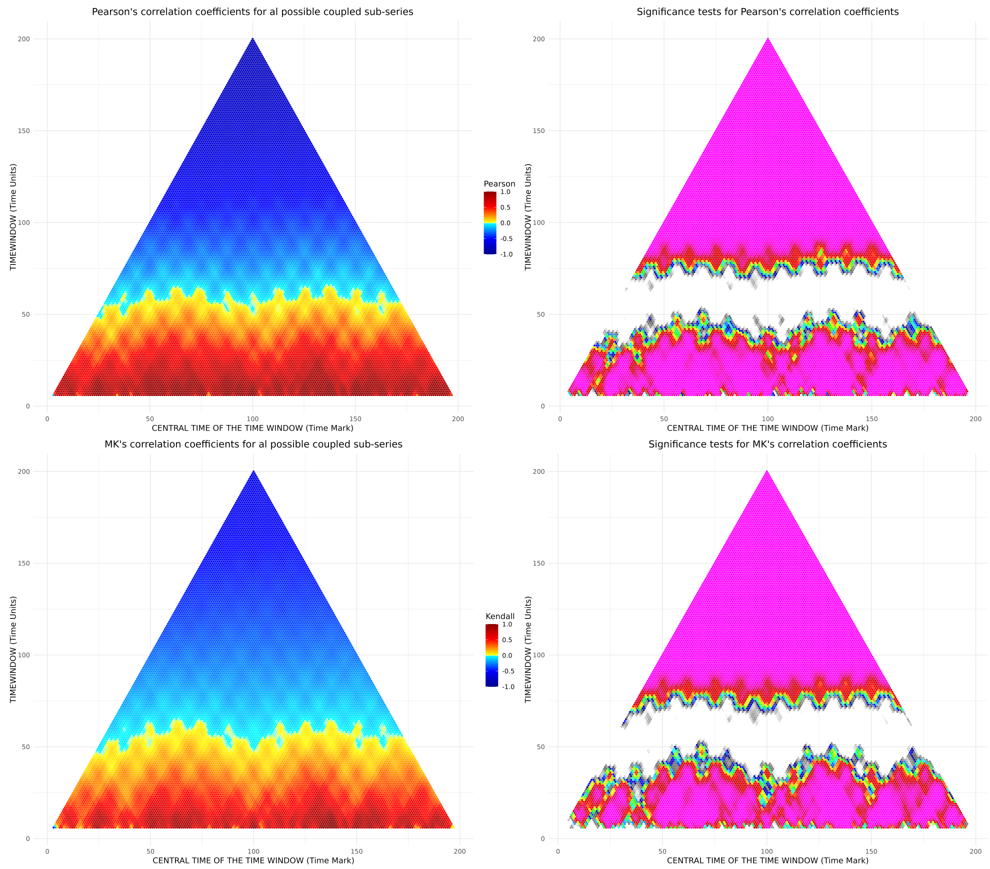


Figure 4: Here we show the Moving Correlation Analysis for the coupled Time Series represented in Fig. 4, the first row shows the results for Pearson's correlation coefficient (Top Left) and relative p-values (Top Right), the second row shows the same analysis with MK's correlation coefficients (Bottom Left) and relative p-values (Bottom Right). The inversion of the correlation between short and long period of analysis is quite clear.

Performance and implementation

TimeHiVE is implemented in R and utilizes several optimization strategies to ensure computational efficiency. The package employs parallel computation where appropriate, leveraging the `parallel` package to distribute calculations across available CPU cores. Key algorithms for Mann-Kendall Trend Test `TH_MK_Trend()` and Kendall Rank Correlation Coefficient `TH_MK_Corr()` are optimized to run O(n log n) time complexity, making them feasible for typical time series lengths encountered in environmental research (Christensen, 2005; Knight, 1966; Sepulveda, 2025).

The package implementation follows modern R package development standards, including comprehensive documentation, unit tests, and examples. TimeHiVE is available on GitHub under the GPL-3.0 license, encouraging community contributions and extensions.

Acknowledgements

This work has been partially funded from the Horizon Europe eLTER EnRich project Grant Agreement No. 101131751 (DOI: 10.3030/101131751)

References

Boselli, V. A., Tagliolato, P., & Oggioni, A. (2025a). A software package for exploratory time series analysis of environmental data. *ARPHA Conference Abstracts*, *8*, e152121. https://doi.org/10.3897/aca.8.e152121

Boselli, V. A., Tagliolato, P., & Oggioni, A. (2025b). *Time-series hierarchical visual explorer*. https://doi.org/10.5281/zenodo.15697261

Brunetti, M., Lentini, G., Maugeri, M., Nanni, T., Auer, I., Böhm, R., & Schöner, W. (2009). Climate variability and change in the greater alpine region over the last two centuries based on multi-variable analysis. *International Journal of Climatology*, *29*(15), 2197–2225. https://doi.org/10.1002/joc.1857

Brunetti, M., Maugeri, M., Nanni, T., Auer, I., Böhm, R., & Schöner, W. (2006). Precipitation variability and changes in the greater alpine region over the 1800–2003 period. *Journal of Geophysical Research: Atmospheres*, *111*(D11). https://doi.org/10.1029/2005JD006674

Christensen, D. (2005). Fast algorithms for the calculation of kendall's τ. *Computational Statistics*, *20*(1), 51–62. https://doi.org/10.1007/bf02736122

Knight, W. R. (1966). A computer method for calculating kendall's tau with ungrouped data. *Journal of the American Statistical Association*, *61*(314), 436–439. https://doi.org/10.1080/01621459.1966.10480879

Sepulveda, M. V. (2025). Kendallknight: An r package for efficient implementation of kendall's correlation coefficient computation. *PLoS ONE*, *20*(6), e0326090. https://doi.org/10.1371/journal.pone.0326090