Wave Heights Long-Term Trend Analysis

Overall Description

This program analyzes long-term trends in significant wave heights (SWH) using data from a CSV file (`input.csv`). To reduce methodological biases, advanced statistical techniques are applied, including adjustments for autocorrelation and seasonal variability. In addition to decadal trend analyses, the program performs a complete Seasonal Mann–Kendall test by splitting the deseasonalized data into 12 monthly series and then conducts a detailed monthly decadal analysis—grouping each month's data by decade and testing for differences across decades.

Key Processing Steps

The key steps include:

1. **CSV Data Reading and Parsing**:
    - Reads the CSV file containing time series data of wave heights.
    - Extracts date/time and corresponding SWH values.
    - Discards invalid or improperly formatted rows.
2. **Data Sorting and Chronological Ordering**:
    - Sorts the parsed data in ascending order by year and month.
3. **Seasonal Effect Removal (Deseasonalization)**:
    - Computes the average SWH for each calendar month over the entire record.
    - Subtracts the monthly average from each measurement to remove seasonal cycles, isolating the long-term trend.
4. **Grouping into Full Decades**:
    - Determines the earliest decade from the data and the last complete decade.
    - Groups the deseasonalized data into complete decades (only decades with 10 full years are used).
5. **Advanced Statistical Trend Analysis**:
    - **Modified Mann–Kendall Test**:
        - Computes the trend statistic (S) using inversion counts.
        - Adjusts the variance for lag-1 autocorrelation.

        - Outputs a standardized Z value indicating trend significance.
    - **Sen's Slope Estimator**:
        - Computes the median of all pairwise slopes as a robust trend estimate.
        - For very large datasets, a random sample of pairs is used.
    - **Seasonal Mann–Kendall Test**:
        - Splits the deseasonalized data into 12 monthly series.
        - Applies the Mann–Kendall test to each monthly series.
        - Combines the monthly S statistics and variances to assess the overall seasonal trend.
    - **Monthly Decadal Analysis**:
        - For each calendar month, groups the deseasonalized data by decade.
        - Computes basic statistics (count, mean, standard deviation) for each month–decade group.
        - Performs one-way ANOVA (and Tukey HSD post-hoc tests if applicable) to test if a month's SWH evolution differs significantly across decades.
        - Ranks decades for each month by average deseasonalized SWH.
    - **One-Way ANOVA with Tukey HSD Post-hoc Test (Decadal Analysis)**:
        - Tests whether the mean deseasonalized SWH differs among full decades.
        - If the overall F-statistic exceeds a rough threshold (F > 2), pairwise comparisons are performed.

6. **Ranking**:

    - Computes the average deseasonalized SWH for each full decade and for each month (across decades).
    - Ranks decades (and months) from highest to lowest average SWH.

7. **Report Generation**:

    - Compiles all processing details, statistical test results, and final conclusions into an extremely detailed report (`report.txt`) written in plain language for non-technical readers.

In summary, by removing seasonal effects, adjusting for autocorrelation, and analyzing both decadal and monthly trends, the program minimizes biases and produces robust, evidence-based conclusions regarding long-term changes in wave heights.

Compile with:

```
g++ -O3 -fopenmp -Wall wave_height_trends.cpp -o wave_height_trends -static -static-libgcc -static-libstdc++
```

Compilation Details:

- `-O3`: High-level optimizations.
- `-fopenmp`: Enable parallel processing with OpenMP.
- `-Wall`: Enable all warnings.
- `-static, -static-libgcc, -static-libstdc++`: Produce a fully statically linked executable.

Technical References

- Hirsch, R.M., Slack, J.R., & Smith, R.A. (1982). Techniques of trend analysis for monthly water quality data. Water Resources Research.
- Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association.
- Tukey, J.W. (1949). Comparing Individual Means in the Analysis of Variance. Biometrics.
- Box, G.E.P. & Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control.
- Cochrane, D., & Orcutt, G.H. (1949). Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms.