CERIAS

The Center for Education and Research in Information Assurance and Security

Non-Parametric Dimensionality Reduction using Entropy via NEST algorithm

Tyler Lewis, Arvind Sundaram, Hany Abdel-Khalik

Motivation:

Numerous engineering analyses concerned with complex time-dependent data suffer from data scarcity which limits the accuracy and applicability of the wider project, e.g., economics, cybersecurity, physics models, etc., which may introduce high temporal and financial costs to the engineering team responsible for the project. Surrogate data is a well-known solution to this issue in many instances. However, many currently-used methodologies run the risk of biased detrending, which may negate the surrogate data's benefit. This project has produced the NEST algorithm to ensure data are properly detrended with random residuals to bolster the applicability of surrogate data by ensuring capture of all patterns displaying regularity.

Solution: Surrogate Data

What are they?

- Artificial datasets reconstructing all relevant trends while varying background noise

What is their Benefit?

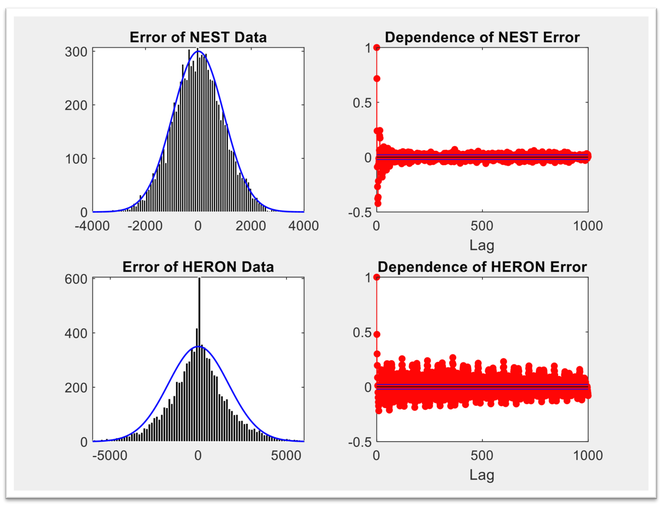
- Uncertainty and vulnerability to noise fall with higher data volume

State-of-the-Art Methods

- HERON is the surrogate algorithm within a resource optimization framework serving as a metric to assess NEST

Results

- Less detrending bias:
  - Normality distributed Residuals
  - Lack of Self-Dependence
- NEST correlations match original data

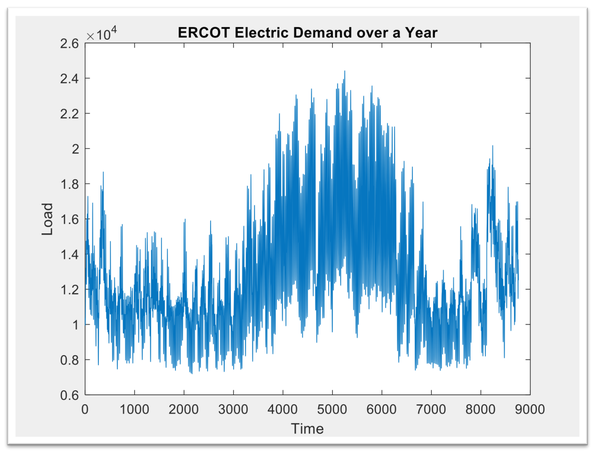


Methodology: NEST Algorithm

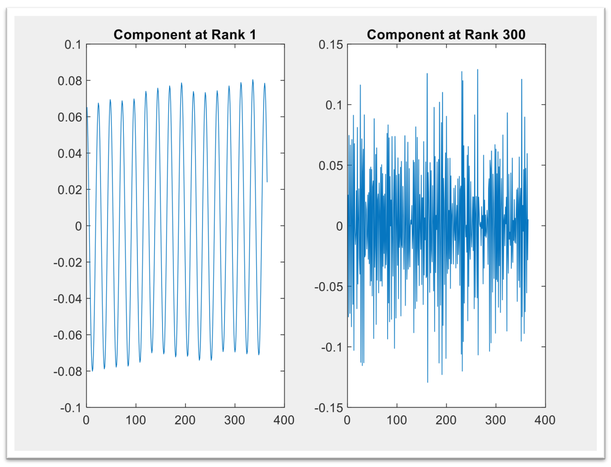
Raw Data ERCOT Electric Demand Data

Timeseries Data

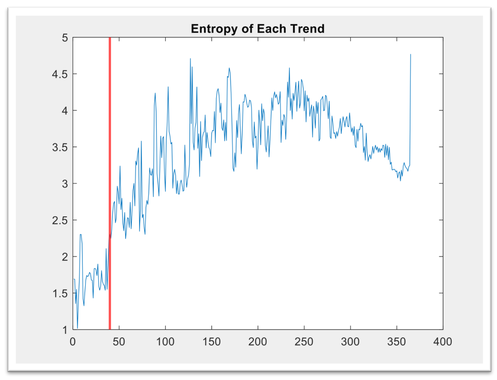
- Complex physical data often varies with time
- This project utilizes the ERCOT electrical load data as a benchmark for which HERON data are available



Data Decomposition: RWD



Rank Selection: SER



Detrending

- HERON employs parametric techniques that fit using external basis functions that may introduce bias into the data
- NEST identifies trends not constrained by shape– a non-parametric technique

Randomized Window Decomposition (RWD)

- Modification of singular value decomposition
- Random windows allow for efficient isolation of fundamental components in the data

Statistical Energy Ratio

- Statistical entropy used to distinguish trend and residual
- Weights each trend via entropy and singular value (by virtue of RWD)

Preserve Trends

Vary Residuals

Form Surrogate Data

PURDUE UNIVERSITY Discovery Park

CERIAS