Toward Context-Aware Imputation and Modeling of Snow Pillow Time Series: Foundations and Future Directions

Abstract:

Snow pillow data gaps can significantly undermine flood forecasting and water resource planning, particularly at high-elevation stations with erratic melt dynamics. Our initial analysis of datasets from the Tuolumne and San Joaquin basins reveals that imputation performance is not governed merely by data volume or missingness percentage, but instead by a station's contextual characteristics. This includes elevation, climatic behavior, temporal regularity, and sensor stability. Merged datasets further complicate the problem: shared stations introduce structural overlap, and inconsistent date ranges and index resolutions widen both visual and analytical gaps. These observations demonstrate the inadequacy of one-size-fits-all imputation strategies.

We propose a new class of adaptive, context-aware imputation pipelines that respond dynamically to metadata and localized data behavior. Rather than pre-selecting models like LSTM or Gaussian Process Regression (GPR), we will cluster stations based on shared spatiotemporal traits and develop flexible frameworks capable of switching between multiple imputation modes (e.g., temporal linear interpolation, seasonal decomposition, decision trees, or deep learning) based on gap structure, variance, and temporal depth. Particular emphasis will be placed on preserving real zero values and reconstructing long, structured gaps while avoiding model overfitting.

Model evaluation will rely on temporally structured k-fold cross-validation strategies that avoid random shuffling and simulate deployment conditions. These include varying the test window length and validating across climate zones to assess robustness and generalizability.

This work aims to establish a foundation for intelligent, real-time snow data platforms. Its broader implications extend to climate informatics and environmental modeling, where nuanced, spatially informed imputation is essential for trustworthy decision-making in the presence of structured missingness.

Toward Context-Aware Imputation and Modeling of Snow Pillow Time Series: Symmetry, Subgroups, and Structural Optimality

Abstract:

Missing data poses a persistent challenge in environmental monitoring, particularly in snow pillow sensor networks used to estimate snowpack and inform water resource planning. Traditional clustering and imputation techniques often fall short because they rely on spatial or Euclidean proximity, neglecting latent symmetries in snow accumulation and melt cycles.

Patterns are frequently shaped by elevation, climate zone, and hydrologic regime. Building on prior work in time series imputation, we propose a mathematically principled framework that integrates finite group theory, representation theory, and high-dimensional lattice geometry to address structured missingness in environmental time series.

We model snow pillow stations as elements of a finite group, where Sylow subgroups capture elevation-driven melt regimes. Inspired by the McKay Conjecture, we analyze normalizer subgroups to derive imputation rules that preserve local symmetry within dynamically similar subgroups while maintaining coherence at the system-wide level. Further, we draw on Viazovska's breakthrough in E_8 sphere packing and Leech lattice structures to treat multivariate gaps as high-dimensional voids. These lattice packings offer a foundation for geometric regularization within machine learning models—such as Gaussian Process Regression or LSTM—guiding imputation toward solutions that minimize spatial-temporal distortion while retaining structural integrity.

We will validate this framework using synthetic datasets embedded with E_8-symmetric structure to assess gains in imputation stability, interpretability, and generalization relative to standard Euclidean approaches. These insights build on our prior benchmarking of over two dozen imputation methods across varying spatial configurations, including basin-specific, targeted, and merged datasets, demonstrating the limits of traditional models in topographically and climatically complex settings. These could result in substantial improvements, particularly in high-missingness or topographically complex scenarios.

By unifying group-theoretic insights with high-dimensional optimization, this project advances a new class of robust, symmetry-aware imputation strategies. While developed in the context of hydrology, the framework extends naturally to any domain characterized by structured, subgrouped, and symmetry-governed missing data. This includes sensor networks, genomics, and remote sensing, which offers broad relevance for intelligent data modeling in complex systems.