



## MOTIVATION

The goal of **CESNET TS-Zoo** is to provide time series datasets with useful tools for preprocessing and reproducibility. The library supports:

- Dataset downloading, configuring, and loading
- Time series preprocessing
- Creation and import of benchmarks
- Creation and import of annotation



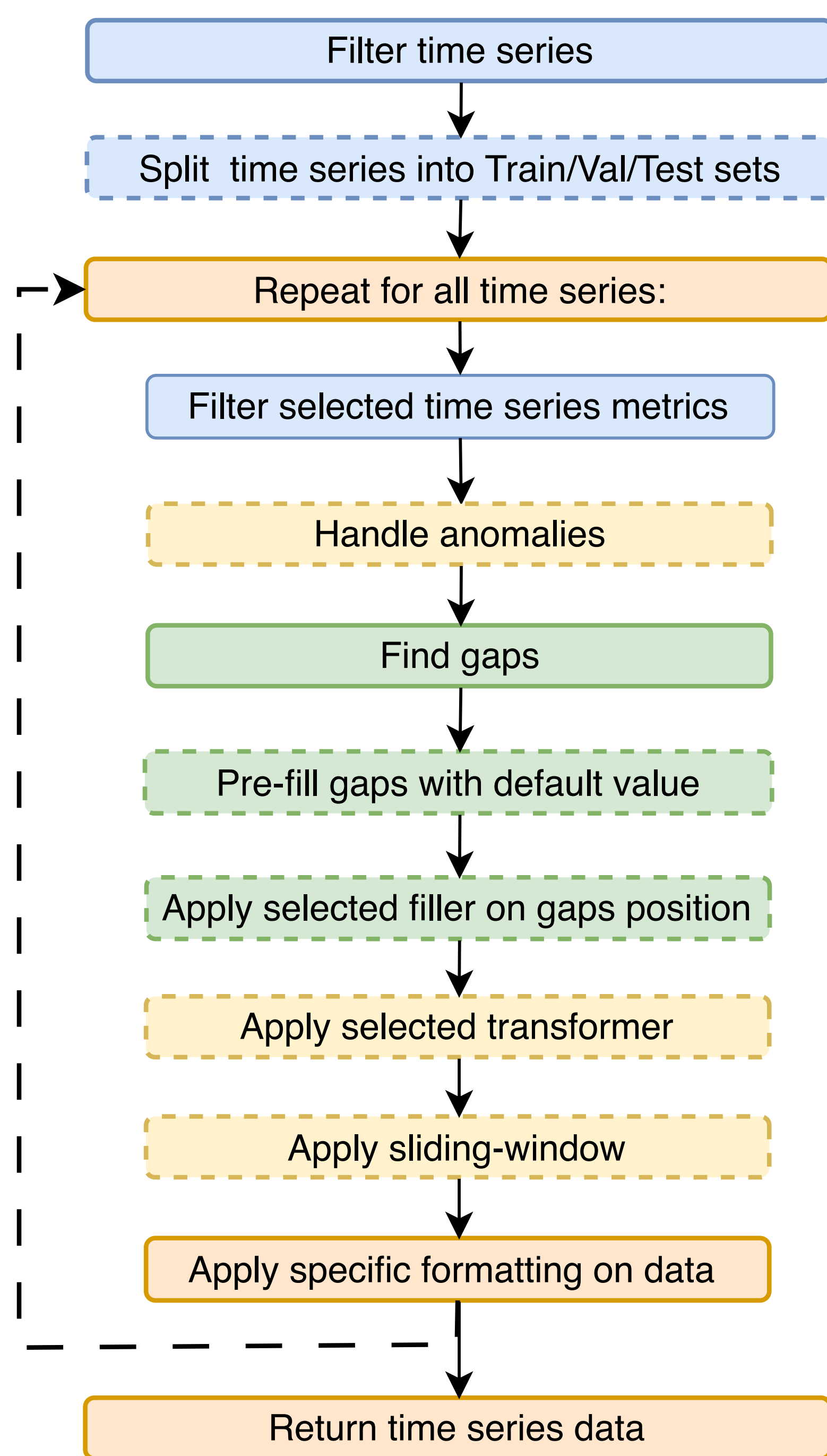
## DATASETS

**CESNET-TimeSeries24 dataset** captures 40 weeks of network traffic from the CESNET ISP network. The dataset offers multivariate time series created through traffic aggregation at three distinct intervals: 10 minutes, 1 hour, and 1 day. Each time series contains 12 different metrics. The dataset contains time series across 283 institutions, 548 institutional subnets, and over 270,000 individual IP addresses.

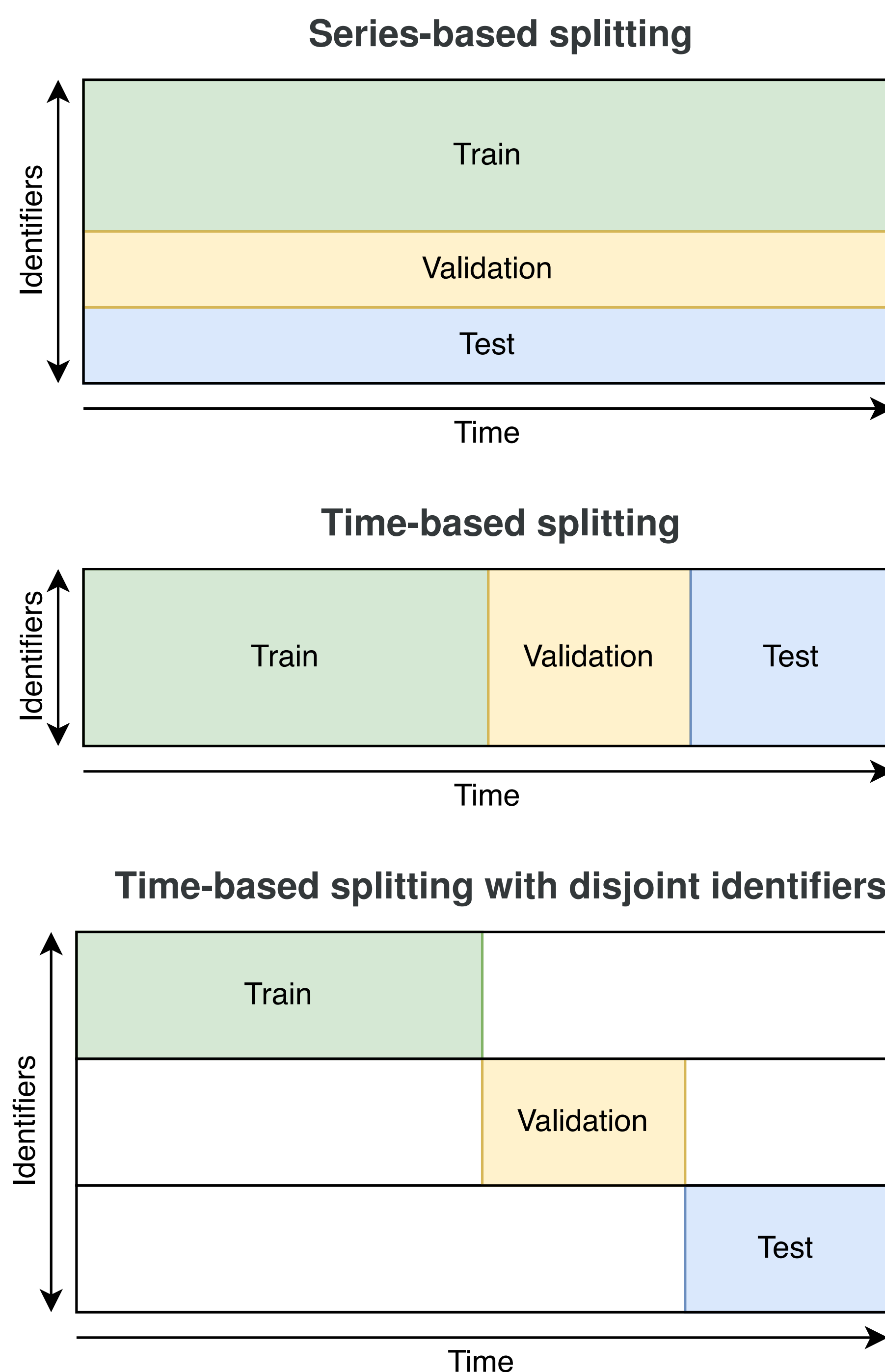
**CESNET-AGG23 dataset** captures two months of network traffic from CESNET network. The dataset offers multivariate time series with 44 metrics.

## PREPROCESSING

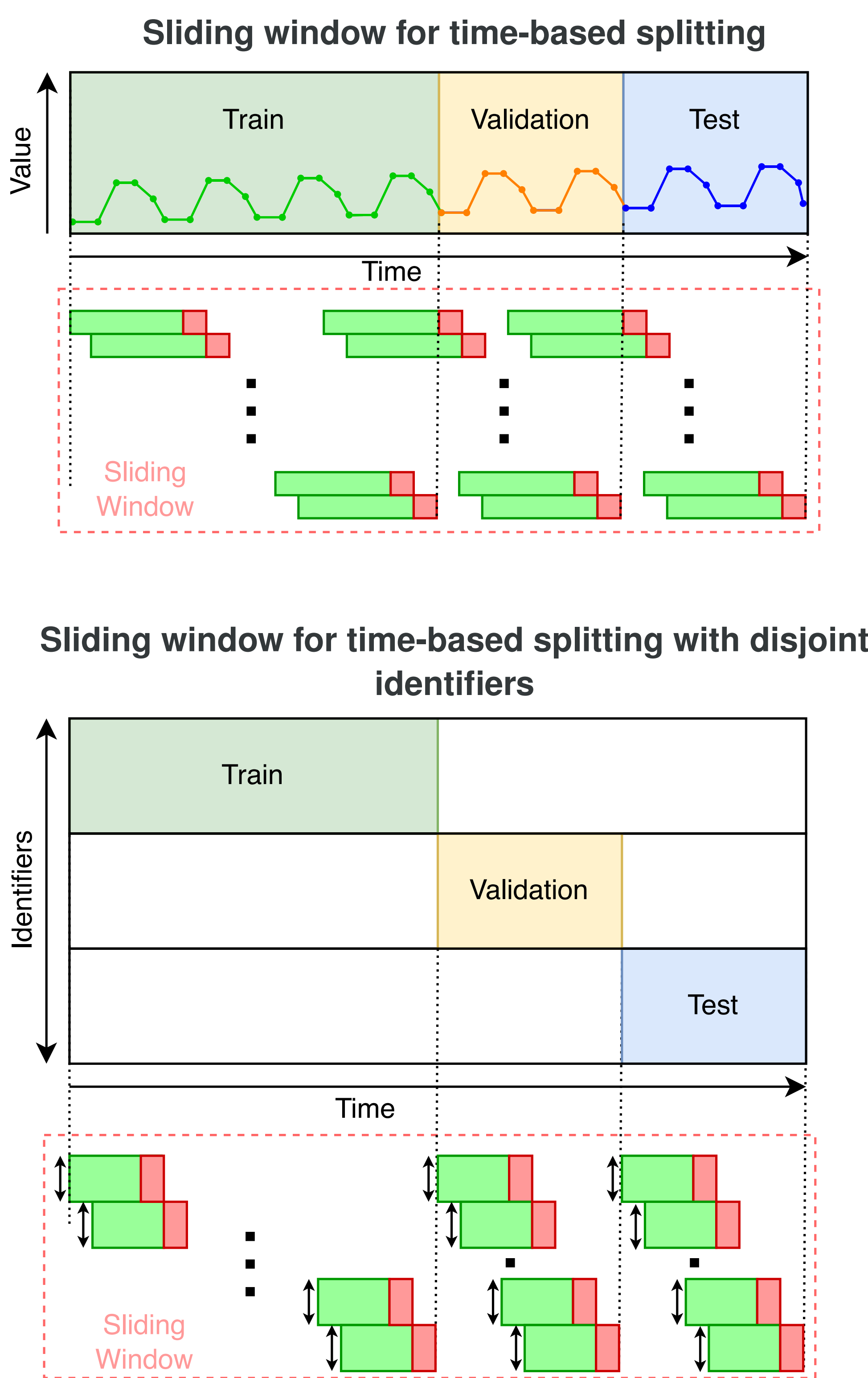
The library supports the preprocessing of time series. The preprocessing includes filtering, handling missing values, handling anomalies, and scaling.



## DATA SPLITTING



## SLIDING WINDOW



## Handlers

The library provides a modular framework for preprocessing and cleaning input data through its data handlers. These handlers ensure data consistency, quality, and readiness for analytical or machine learning workflows. Users can flexibly combine multiple handler types to address various data irregularities and transformations.

- **Handling anomaly values** (Z-score, Interquartile Range, Custom Anomaly Handler)
- **Handling gaps** (Default value, Mean, Forward, Linear interpolation, Custom Filler)
- **Applying transformers** (Min Max Scaler, Standard Scaler, Max Abs Scaler, Log Scaler, L2 Normalizer, Robust Scaler, Power Transformer, Quantile Transformer, Custom Transformer)

## Benchmarks

Reproducible experiments are essential for building trust in scientific results. To address this challenge, we implement functionality for exporting and importing dataset configurations. The library is ready for including benchmarks on forecasting, anomaly detection, classification, and so on.

Exporting the config to a file is possible and can be distributed using, for example, a GitHub repository or the Zenodo platform.

```
dataset.save_config(identifier="<id>")
```

The config file can be imported for reproducing the approach or comparing on the same data.

```
dataset.import_config(identifier="<id>")
```

Moreover, the exporting and importing of dataset configs facilitate the creation of community benchmarks distributed within the library. We implement a method to load a predefined benchmark, as shown below. The library now contains several benchmarks for network traffic forecasting, and more can be added in the future using GitHub Pull Requests.

```
from cesnet_tszoo.benchmarks import load_benchmark
benchmark = load_benchmark(
    "<benchmark_hash>",
    "<path-to-datasets>"
)
dataset = benchmark.get_initialized_dataset()
```

## Release

[GitHub/CESNET/cesnet-tszoo](https://github.com/CESNET/cesnet-tszoo)

`pip install cesnet-tszoo`

