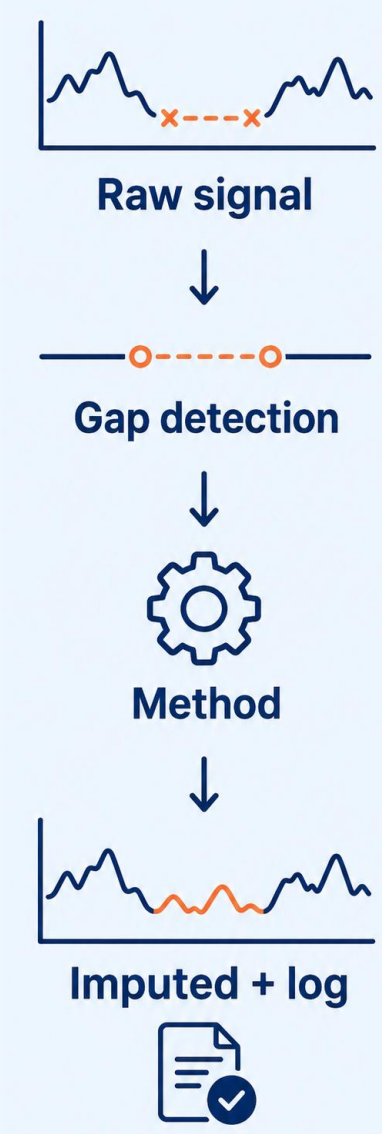


1 Motivation

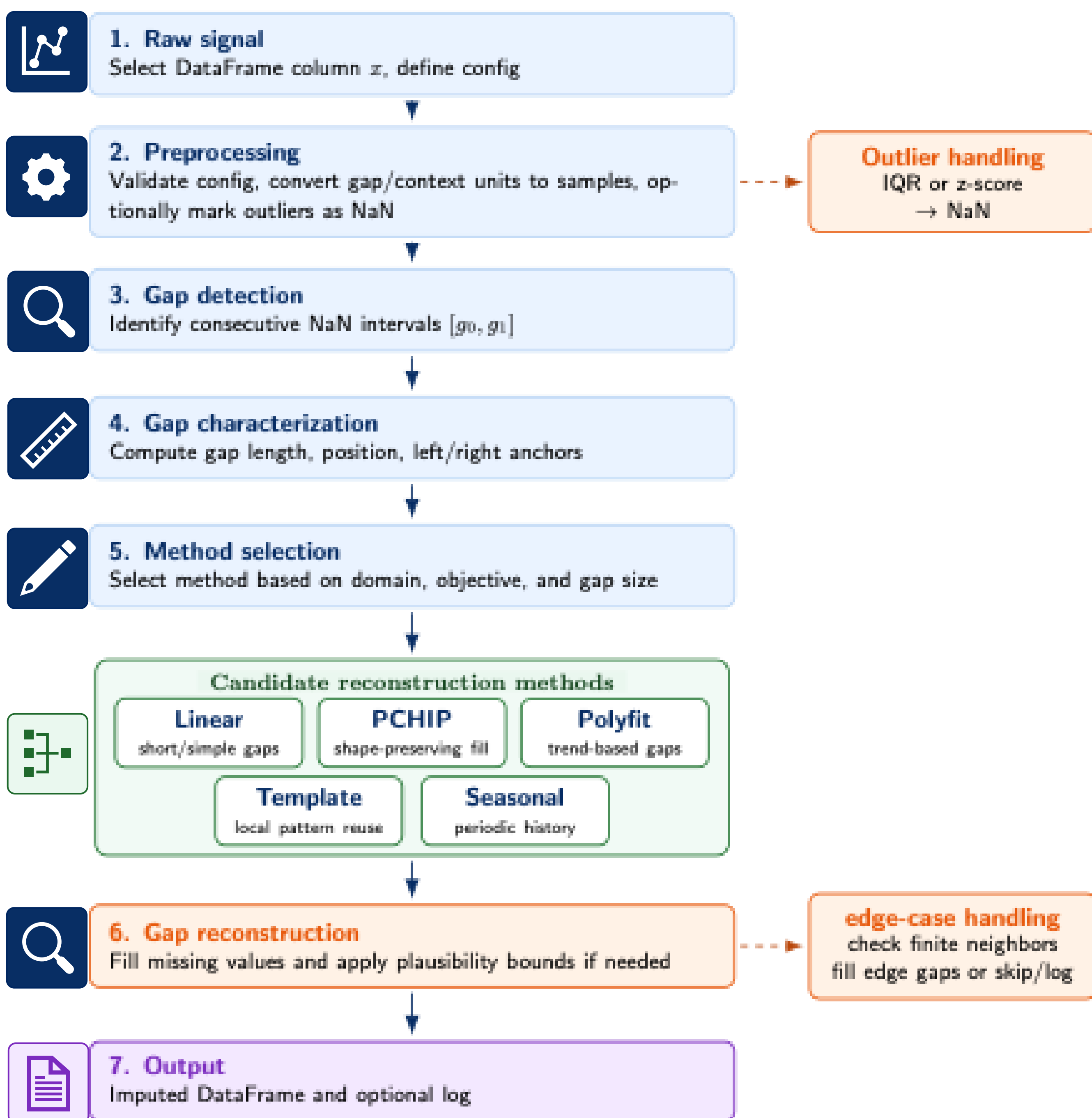
Missing values are inevitable in real-world time series. GapTrace provides transparent, domain-aware decision support to select and apply appropriate gap imputation methods. Preprocessing choices, detected gaps, selected methods and outcomes are logged to improve reproducibility and comparability.



2 Core idea & contributions

- Domain-aware gap imputation
- Automatic gap profiling and method recommendation
- Provenance logging of preprocessing decisions
- Objective-specific strategies for accurate imputation and visual reconstruction

3 Simplified pipeline



4 Domain-aware method selection

Objective: Accurate imputation				Objective: Visual reconstruction		
Domain	Small gap	Medium gap	Large gap	Small gap	Medium gap	Large gap
Eye tracking	PCHIP	PCHIP	Linear	PCHIP	Template	Template
Medical	Linear	Polyfit	Linear	Linear	Template	Template
Traffic	Linear	PCHIP	Seasonal	Linear	Template	Seasonal
Weather	PCHIP	Polyfit	Template	PCHIP	Template	Template

5 Provenance

```

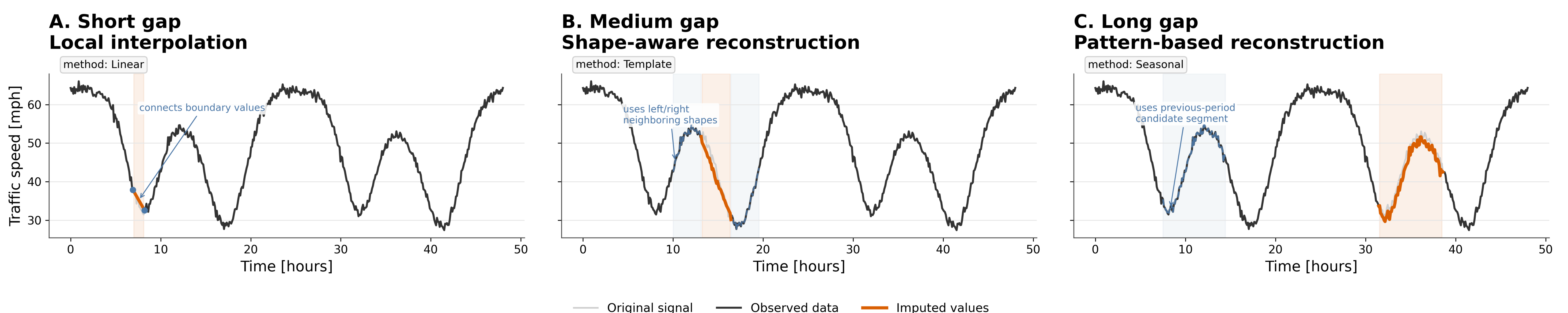
ProvCard
source_id: preprocessing_dataset.csv
type: entity
description: dataset after outlier detection
gaps:
  - {gap_start: 3, gap_end: 4, method: linear, ...}
  - {gap_start: 579, ...}
input:
  - {type: entity, source_id: original_dataset.csv, ...}
  - {type: entity, source_id: config.txt, ...}
  - {type: activity, source_id: outlier_detection.ipynb, ...}
entity: processed_data
  
```

Provenance describes information about how an end product was created, whether it is data or a physical object. Provenance is collected while the workflow is running.

A ProvCard is a machine-readable JSON metadata element that summarizes the provenance of a specific workflow component.

It stores key details about **entities** such as datasets or code, **activities** such as processing or gap imputation steps, and **agents** such as users, tools, or systems.

6 Demonstration scenarios: Traffic time series – visual reconstruction



7 Key takeaway

Different gap types and domains require different reconstruction strategies.



GapTrace selects objective- and domain-aware imputation methods and logs every preprocessing and imputation decision for reproducible analysis.

8 Evaluation summary (Average RMSE, lower is better)

Domain	Accurate	Visual	Linear	PCHIP	Fill/BFill	Mean
Eye tracking	62.992	69.039	61.346	60.930	93.232	193.132
Medical	7.215	7.349	7.223	7.522	9.200	18.726
Traffic	5.618	5.646	5.648	5.791	6.781	8.131
Weather	2.767	2.697	2.924	2.933	3.747	8.126