Synchrophasor Data Quality Attributes and a Methodology for Examining Data Quality Impacts upon Synchrophasor Applications – Overview

Pacific Northwest National Laboratory and
National Institute of Standards and Technology
Team

March 21, 2016

NASPI North American SynchroPhasor Initiative

Overview

- Phasor Applications Requirements Task Force (PARTF) Background and Expert Team Effort
- Data Quality Framework
- Proposed Test Methodology
- Q&A

PARTF Charter

- PMU data quality management is challenging
  - Application results incorporate issues related to input accuracies and network delivery problems.
  - The breadth and variety of data quality input issues are not widely recognized today.
  - The impacts of these issues on application results are largely unknown, yet power system standards presume accurate results.
- PARTF objective – develop a report that:
  - Clarifies data quality terms to better identify data inaccuracies and data delivery problems
  - Offers a process to understand and identify synchrophasor applications' data quality vulnerabilities

The PARTF Vision

- The synchrophasor community begins using **consistent terms** and definitions for data issues, quality and availability.
- We use the **methodology** to develop a clear understanding of how data issues/filters/data flow issues affect each application and algorithm – and determine the priorities for improving PMUs, data networks, and applications.
- This approach can give grid operators and application users **confidence about the quality and trustworthiness** of the guidance coming out of synchrophasor applications.
- These methodologies get **built into applications** (data quality metric in dashboard), improving application performance, transparency and acceptance.

The PARTF Expert Effort

The complex PARTF scope requires a rigorous methodology and consistent approach to be useful. We can help the synchrophasor community and PARTF volunteers by developing a proposal for review, feedback, and improvement.

- PNNL and NIST have contributed expert resources and funds to develop a proposed methodology and definitions framework
  - Alison Silverstein (NASPI) – framework & readability
  - Laurie Miller (PNNL) – power systems & advanced mathematics
  - Dhananjay Anand (NIST) – applied mathematics & control theory
  - Allen Goldstein (NIST) – electrical engineer & digital signal processing
  - Yuri Makarov (PNNL) – power engineering & advanced mathematics
  - Frank Tuffner (PNNL) – power engineering & PMU applications
- We seek your feedback on these recommendations

Methodology - Definitions

- We need agreed-upon terms to talk about fitness-for-use of PMU data by an application.
  - Most terms describing the fitness-for-use of data for a particular task have multiple meanings.
    - There may be subtle differences in usage of terms among standards, guides, application documentation (latency, gap, quality, “good data”, etc.)
  - When examined in context, many terms eventually prove to have multiple attributes that each need their own definition.
- Used related existing definition sets to inform our discussion
  - Information technology, GIS have good overlap with our problem
  - Our definition set is organized for the PMU applications field

The data quality framework

Data attributes differ according to the type and scope of data

- Data point attributes are mostly about accuracy and metadata
- Data set (a collection of data points) attributes include data coverage (time, topology), consistency (metrology, headers, standards)
- Data stream (a data set in motion) attributes are about the process path and availability

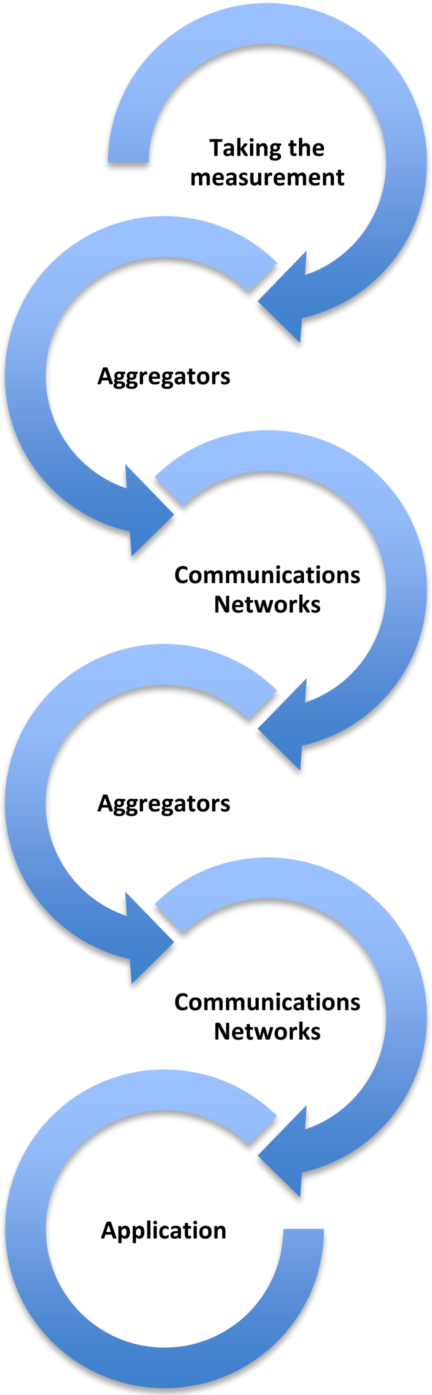
Data Process Path

For each of these three categories, data problems can arise in multiple places along the **data process path** from PMU, through aggregators and communications to the final end-use application.

– Measurements are taken at the **PMU** – main point for accuracy

– **Aggregators** refer to any type of historian/database/archive or other storage – affects accuracy and data point availability

– The data usually passes through more than one **communications network** – affects availability and timeliness

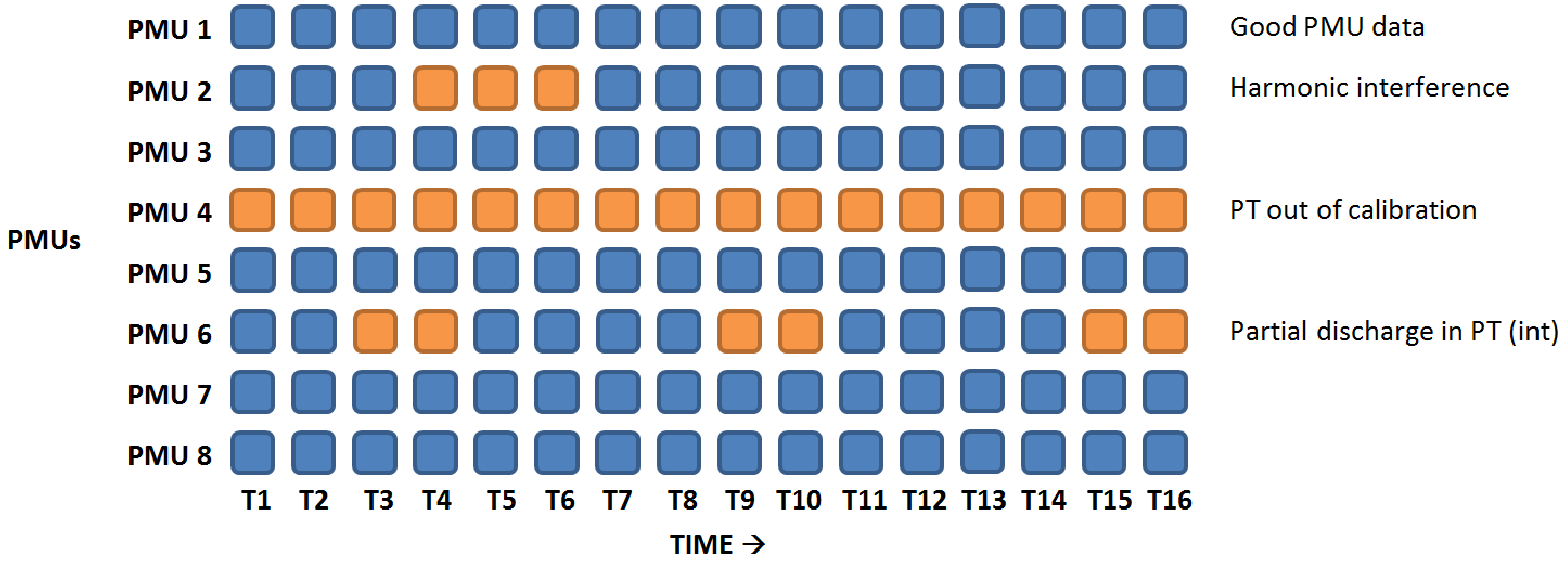


Attributes of single data points

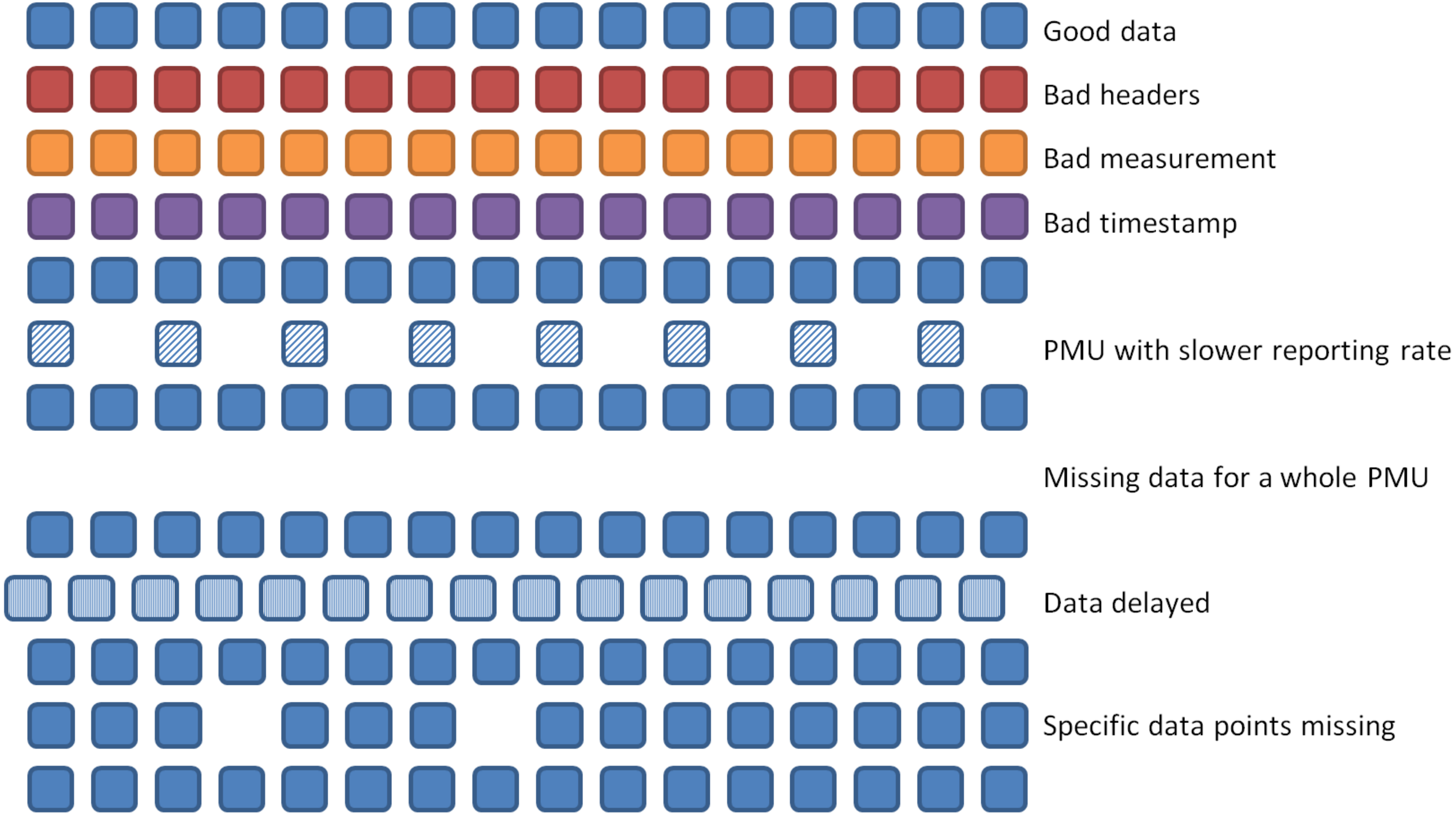
Measurement specifiers		Realized quantity
		Measured result
		Standard units
		Precision
		Confidence interval
Measurement accuracy		Inherent error (source error)
		Introduced error (created by the PMU estimation process)
Attribute accuracy		Temporal accuracy
		Geospatial accuracy
		Topological accuracy
Data lineage (aka metadata)	Data source	PMU type
		PMU standard followed
		PMU model, firmware version, configuration settings
		PMU-supplied data headers
		Aggregator-supplied data headers
	Data coverage	PMU geographic location
		PMU topological location
	Data transformation methods	Transformations applied to the data at the PMU

Data lineage metadata include specifying information about the data and the PMU, such as the information in a PMU registry.

Quality and usability associated with individual data points
Examples of bad measurements



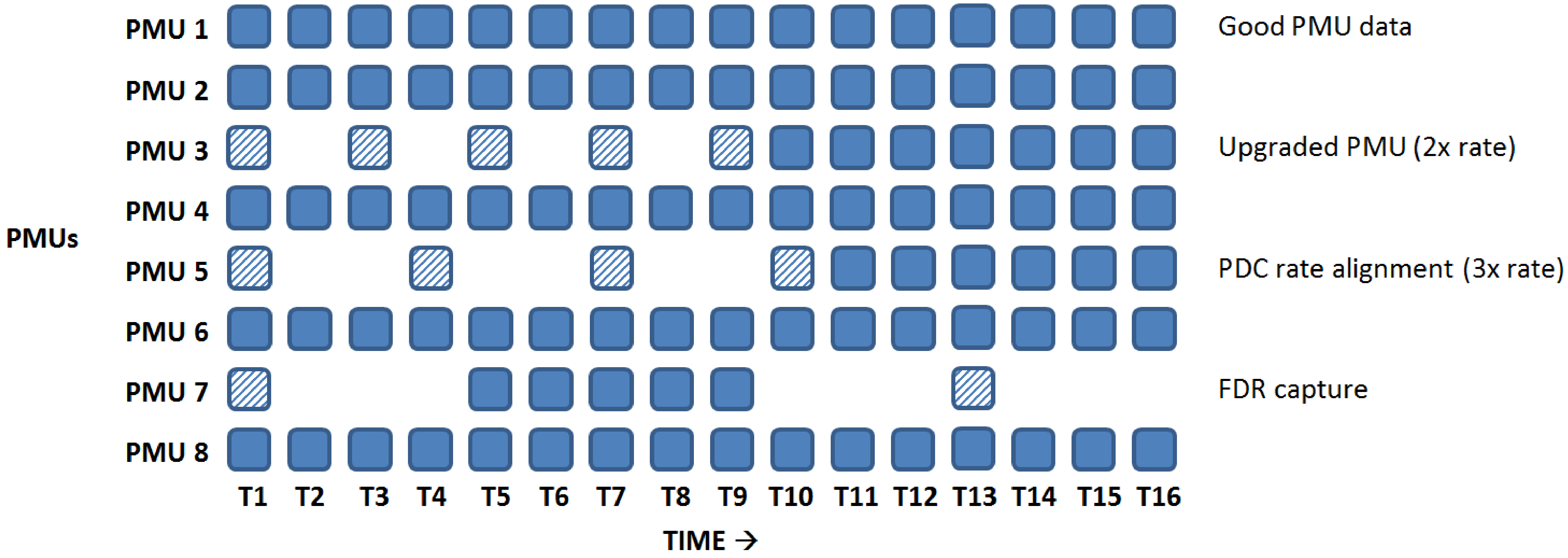
Schema for illustrating data set attributes



Data set attributes

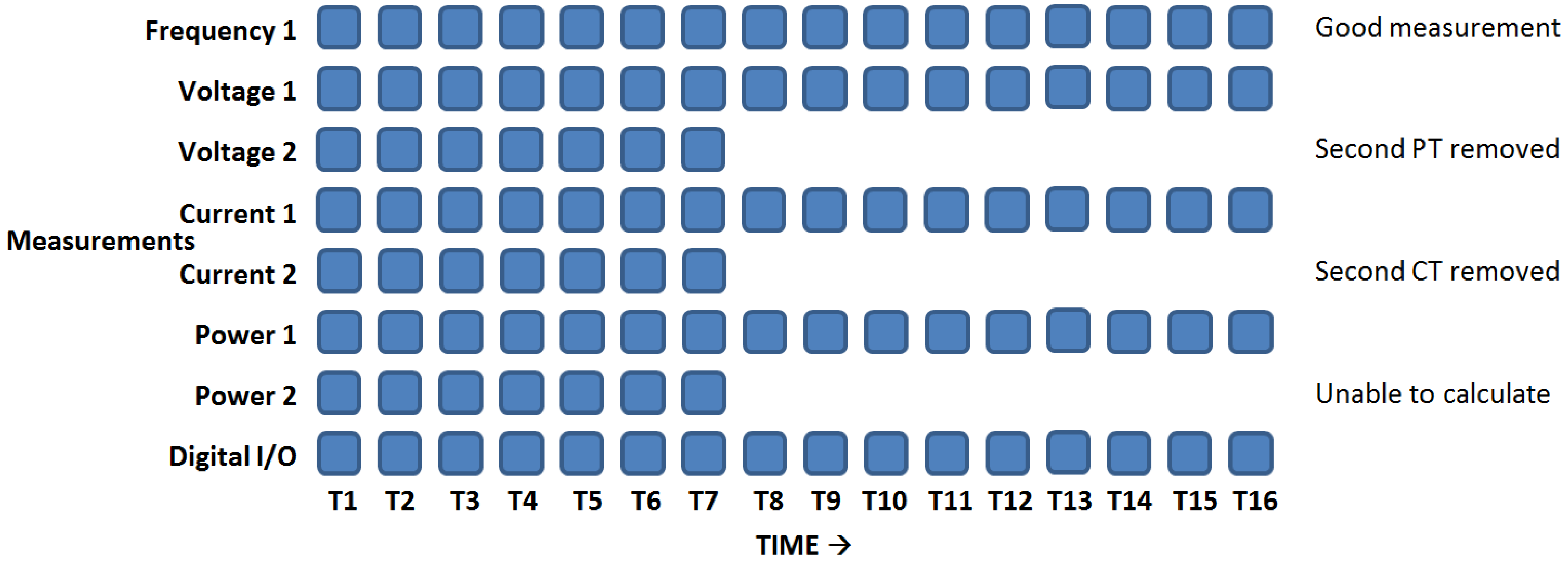
Data lineage (aka metadata)	Data transformation methods *(at an aggregator, during archiving)*
	Data coverage *(PMU temporal coverage, aggregator geographic, topological, and temporal coverage)*
	Data content *(Number of PMUs, channels provided by the PMU and aggregator, PMU reporting rate, aggregator reporting rate)*
Logical consistency	Metrology persistence and consistency
	Header persistence and consistency
	Data frame persistence and consistency
	Standards compliance persistence and consistency
	Reporting rate persistence and consistency
Data Completeness	Gap size *(in data from the PMU and in data from the aggregator)*
	Largest known gap *(in data from the PMU and in data from the aggregator)*
Characteristics of the data process path	Data set manageability
	Data set recoverability
	Data set reliability
	Data set serviceability

Data set issues
(1) Reporting rate persistence and consistency



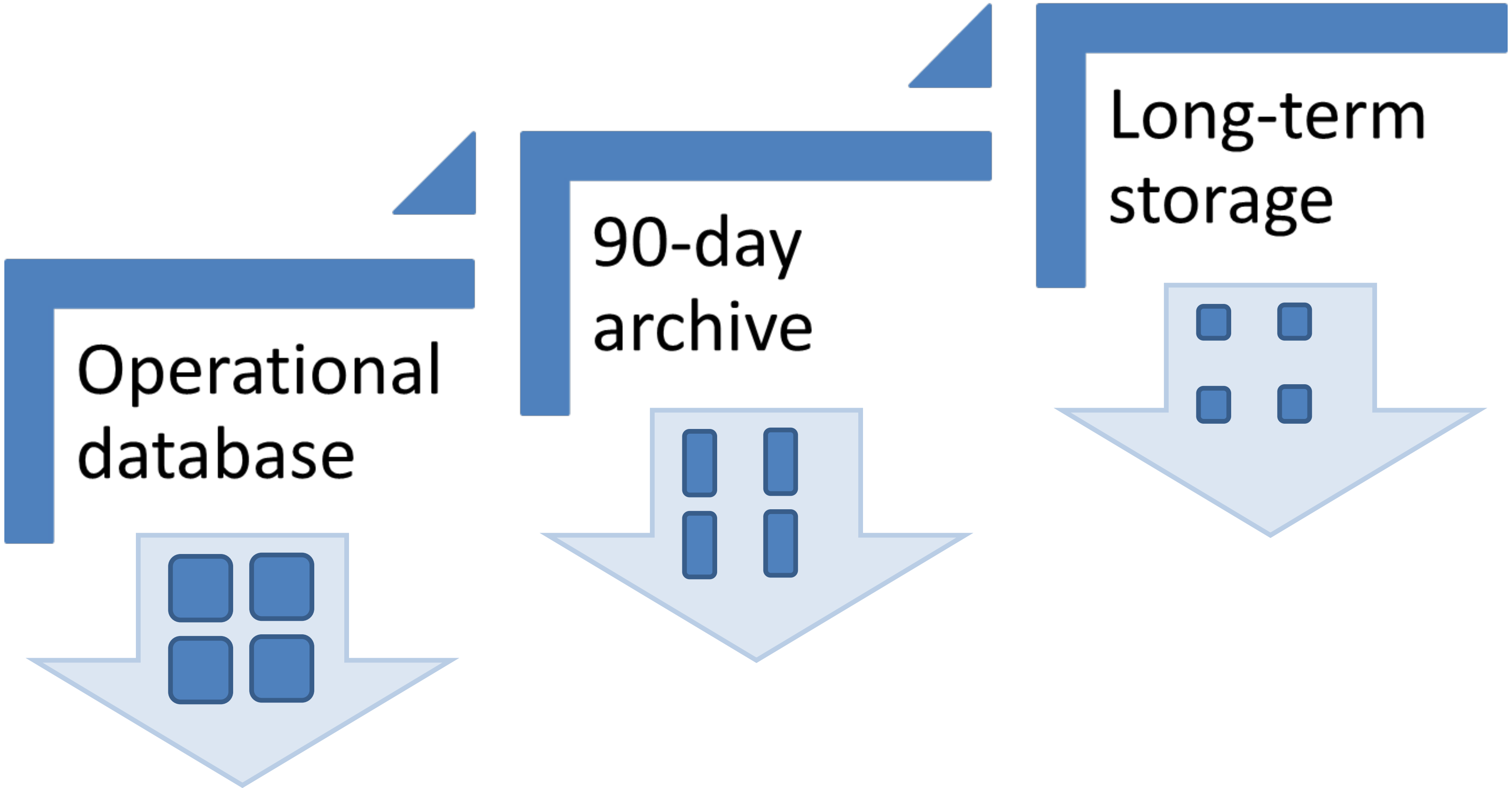
Data set issues
(2) Impact of loss of measurement persistence

Single PMU



Data set issues
(3) Potential impacts of lossy compression in archiving

- Repeated data compression and transformation can irretrievably modify the original PMU data.
- The same data taken from different archive points might no longer be identical.



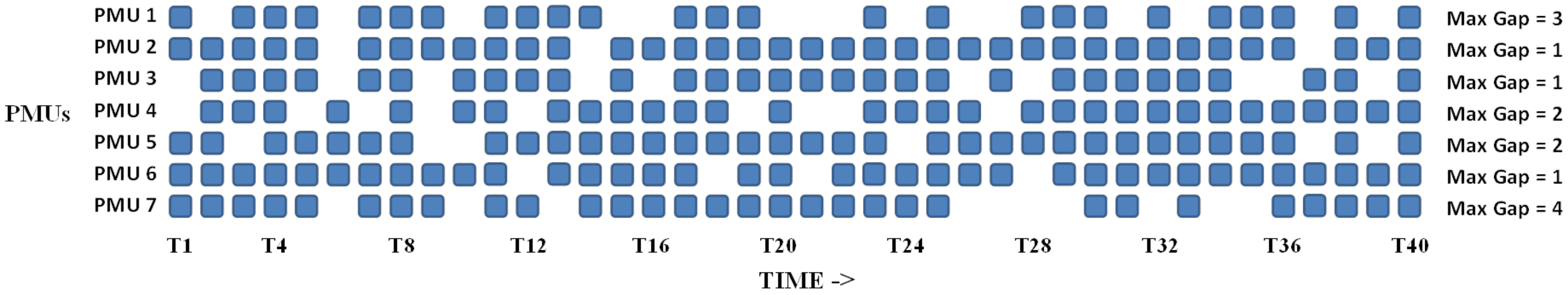
Data stream attributes

Characteristics of the data process path *	Data stream manageability
	Data stream recoverability
	Data stream reliability
	Data stream serviceability
Data stream availability *	Message rate
	Message rate type
	Message delivery time
	Expected dropout rate †
	Expected dropout size †
	Message continuity period †

* Assess regularly -- this attribute does not need to accompany every measurement, but should be weighed when determining the suitability of data for an application or study or when troubleshooting problems with application results.

† There are a number of statistics that could be used to characterize a data set or a live PMU data stream; this attribute is provided as one example. Which statistics should be chosen to consider the fitness-for-use of a data set or a live PMU data stream by an application must be determined by the user for the application and specific study to be performed.

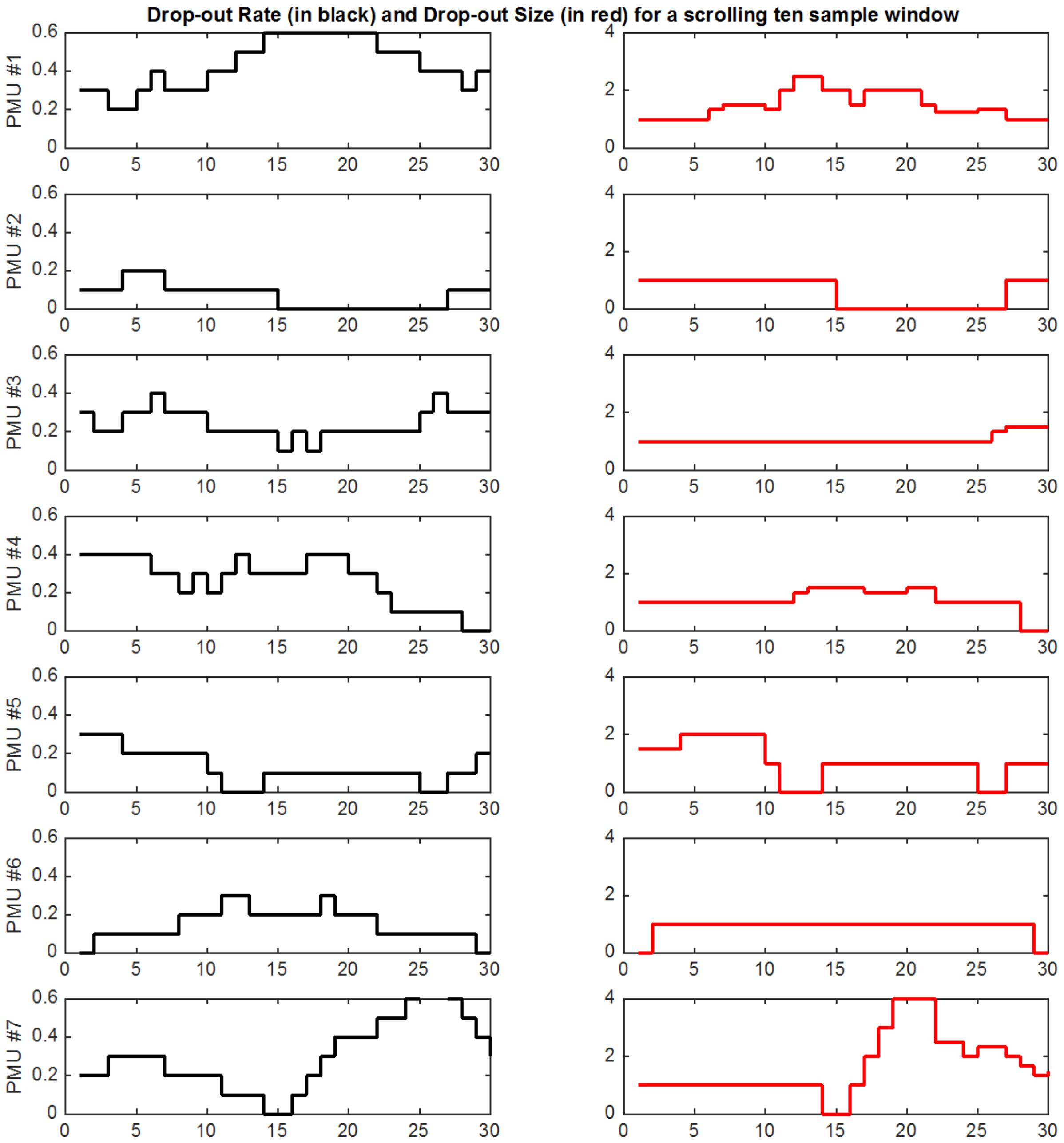
Data stream issues
(1) Data stream dropouts with random network losses



- **Gaps** in data sent by PMUs/aggregators
- **Dropouts** introduced by communications networks
- Gap rate primarily deals with whether the information exists in any form and is obtainable, whereas dropout rate is associated with timely or successful receipt of streamed data.

Data stream issues (2) Comparing scrolling window dropout rate and dropout size

- Application requires a dropout rate < 0.5 (e.g. < 5 per 10 timestamps)
  - PMU #1 out of spec at T13, back in spec at T25
  - PMU #7 out of spec at T22, back in spec at T28
- Application requires a maximum dropout size < 4 timestamps in a row
  - PMU #7 out of spec at T19, back in spec at T22

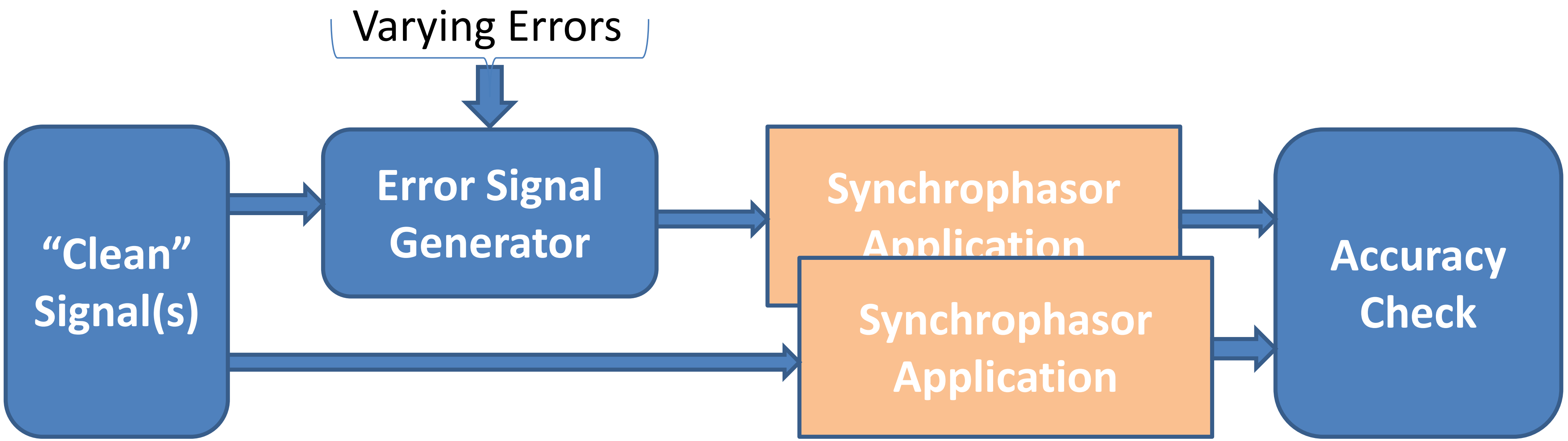


NASPI North American SynchroPhasor Initiative

Proposed method for determining the impact of data problems on application performance

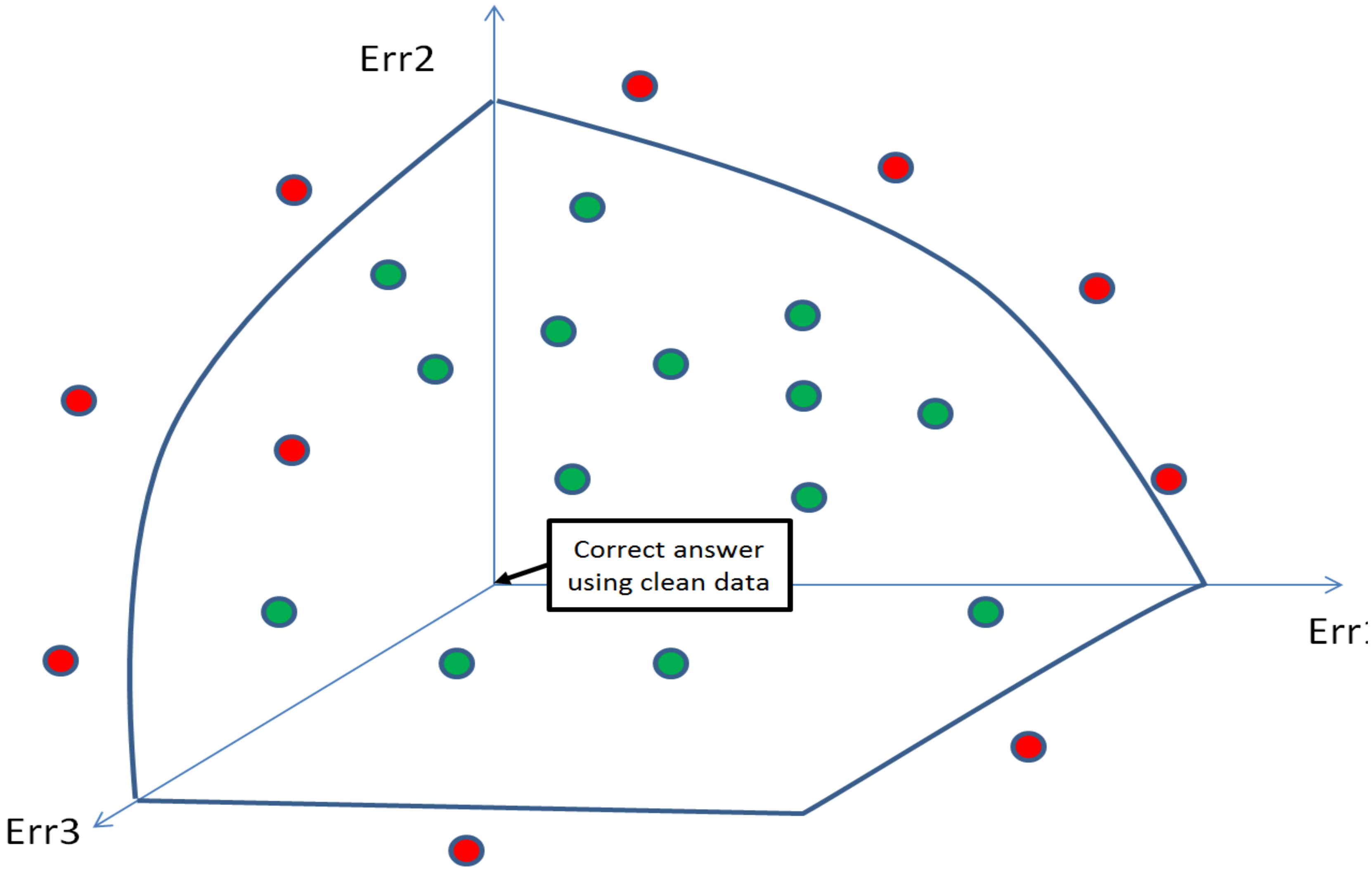
- Test an application many times using many data sets with known errors relative to a clean data set(s) to determine how the different errors affect the application's performance (i.e., multi-dimensional PMU data error analysis).
- Develop a performance envelope that shows which combinations and magnitudes of data errors provide sufficiently accurate output (based on users' or developers' acceptability requirements).
- This test process should reveal which data problems (accuracy and availability) are most influential on application output accuracy and trustworthiness.
  - Distinguish whether improvement needed from PMUs, networks or aggregators.
  - Develop on-line application evaluator that warns when incoming data stream makes the application results untrustworthy.
  - Help make applications less vulnerable to various types of data problems.

Application Performance Envelope Methodology Outline



NIST is developing clean synchrophasor data sets and data set modification methods that we can use to generate synthetic, varying data sets for application testing.

Illustrative Application Performance Envelope



Next Steps

- Get comments back from PARTF, NASPI community and others on initial white paper
  - Find the PARTF proposal at https://www.naspi.org/File.aspx?fileID=1689
    - "NASPI-2016-TR-002 Synchrophasor Data Quality Attributes and a Methodology for Examining Data Quality Impacts upon Synchrophasor Applications"
  - Please send feedback to PARTF at naspi@pnnl.gov by April 18.
- Refine the documents and methodology into a "Version 2" document to share with NASPI community and other stakeholders
- Look at how to use these ideas within on-going data quality and application development activities
  - Incorporate into DNMTT and SMS guidance?
  - Encourage further national lab work on data set development and application test methods
  - Encourage vendors to implement these ideas into their tools and applications

QUESTIONS