ORBITS: Online Recovery of Missing Values in Multiple Time Series Streams

UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

Mourad Khayati, Ines Arous, Zakhar Tymchenko, and Philippe Cudré-Mauroux
{firstname.lastname}@unifr.ch

Goal and Contribution

Motivation: Missing values are ubiquitous in time series data. They do not only affect real-time monitoring but also compromise the quality of online data analyses.

Goal: An efficient online recovery of missing blocks for multiple time series streams.

Contributions:

1. ORBITS: Anticipatory algorithm to recover missing blocks in a streaming fashion.
2. It leverages the inter and intra series' correlation.
3. It produces a shape-preservation recovery.

Applications

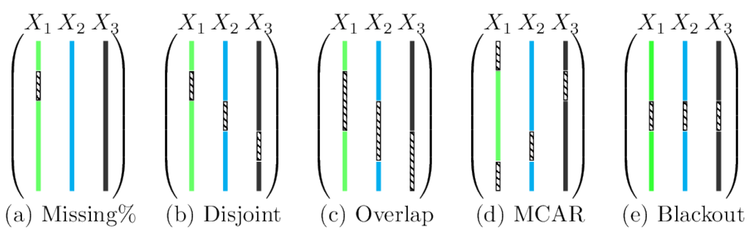
Sports Analytics (Soccer)

- Analyze the team behavior in real-time.
- Sensors often detach from players' equipment yielding missing blocks.
- On-the-fly recovery allows coaches to immediately adjust their tactics.

Other applications: Water discharge monitoring, smart grids, pollution tracking, etc.

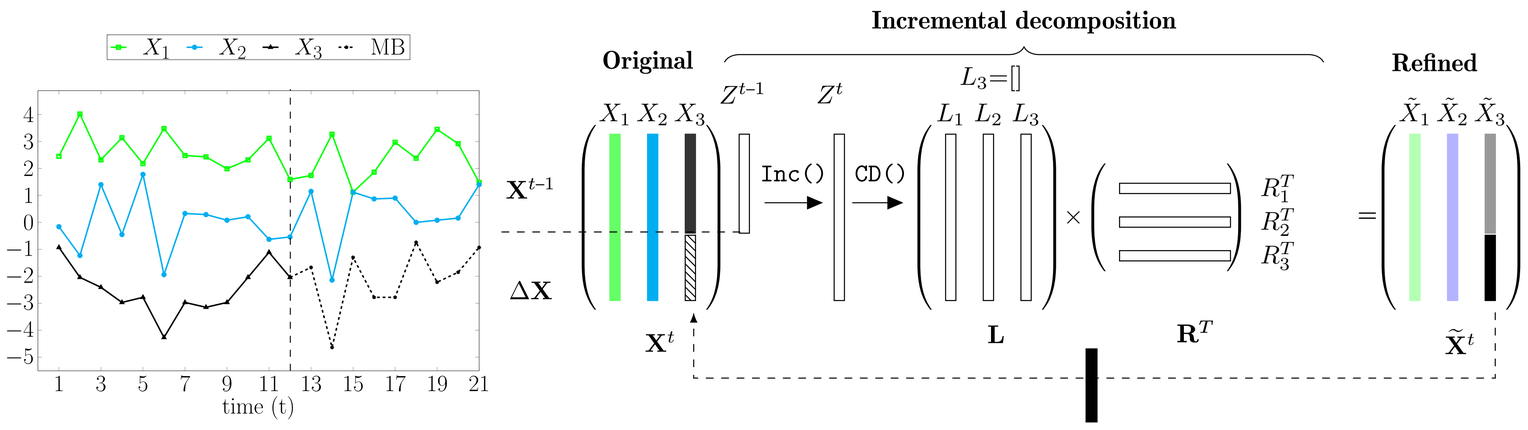
Challenges

- Large missing blocks in multiple series.
- High-frequency data streams (e.g., 200Hz in sports data).
- Different missing patterns:



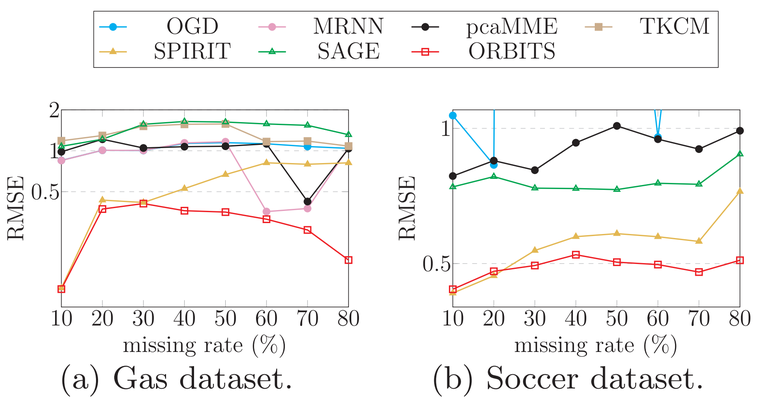
(a) Missing% (b) Disjoint (c) Overlap (d) MCAR (e) Blackout

ORBITS in Action



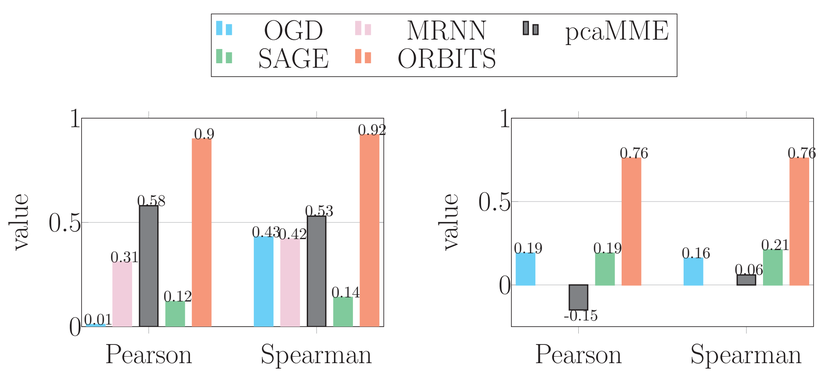
Experiments

- ORBITS is on average 25% more accurate than the state-of-the-art.



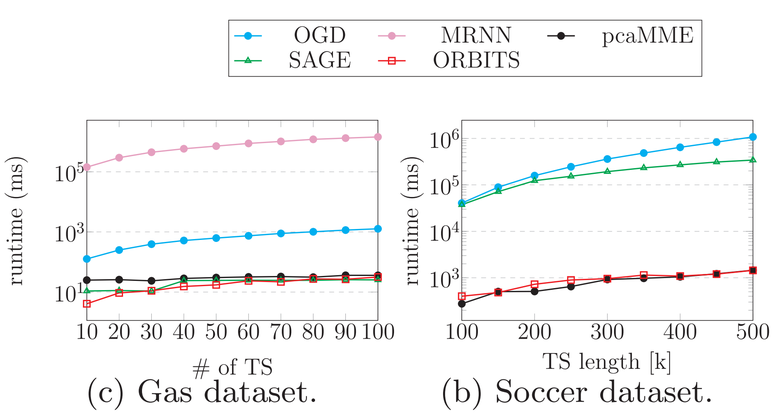
(a) Gas dataset. (b) Soccer dataset.

- ORBITS preserves the linear and nonlinear correlation across time series.



(a) Gas dataset. (b) Soccer dataset.

- ORBITS takes 1.4 sec to recover 10 time series each with 500k.



(c) Gas dataset. (b) Soccer dataset.

Conclusions

- ORBITS is a fast incremental algorithm to recover missing blocks in time series streams.
- Our anticipatory termination can be extended to speedup other matrix decomposition techniques (e.g., SVD, NMF, QR, etc.).
- Future work: Apply missing values recovery to repair anomalies.

Acknowledgments

- Swiss National Foundation (SNF).
- European Research Council (ERC).
- Swiss Federal Office for the Environment (BundesAmt Für Umwelt – BAFU).

Additional Info

- **Github**: `https://github.com/eXascaleInfolab/orbits`
- **GUI**: `http://revival.exascale.info/streaming/datastream.php`
- **Related work**: Mourad Khayati, Alberto Lerner, Zakhar Tymchenko, and Philippe Cudré-Mauroux: "Mind the Gap: An Experimental Evaluation of Imputation of Missing Values Techniques in Time Series", PVLDB 2020.

Published in the *Proceedings of the 47th International Conference on Very Large Databases (VLDB)*, Copenhagen, Denmark, August 2021.