CEA list (C. Gouy-Pailler, cedric.gouy-pailler@cea.fr)

UVSQ – DAVID laboratory (K. Zeitouni & Y. Taher)

Foch Hospital -- UVSQ INSERM – VIMA team (Ph. Aegerter & M. Fischler)

STREAMOPS : OPEN SOURCE PLATFORM FOR RESEARCH AND INTEGRATION OF ALGORITHMS FOR MASSIVE TIME SERIES FLOW ANALYSIS

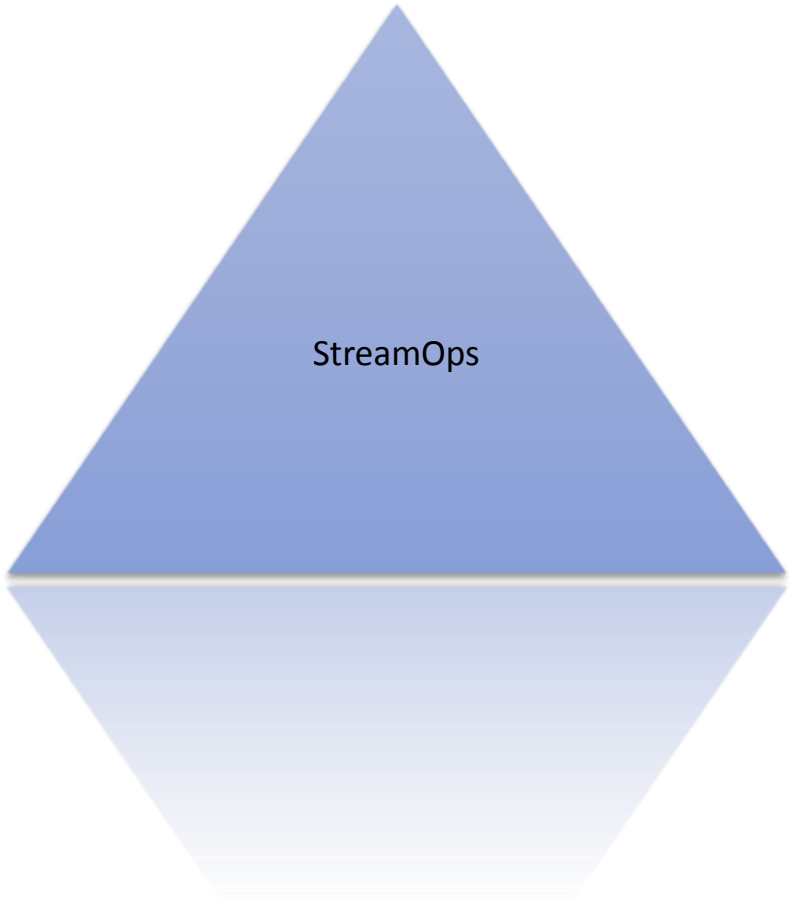
DATAIA-JST International Symposium on Data Science and AI | 11/07/2018

IDENTITY CARD OF THE TEAM

STREAMING APPLICATIONS LANDSCAPE

Innovative algorithms
Specialized approaches
from research community

MOA, research papers



Large-scale and robust
software community

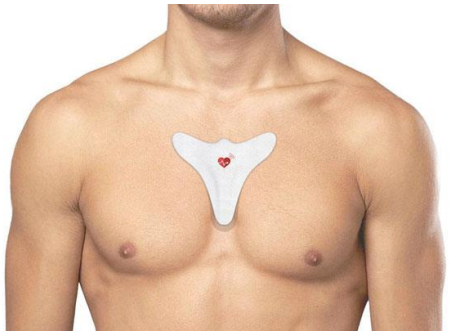
Kafka, Spark, flink, redis,
mongoDB, cassandra, postgreSQL,
Microsoft Azure

Knowledge from
applicative side

Predix (GE), Mindsphere
(Siemens), Bosch

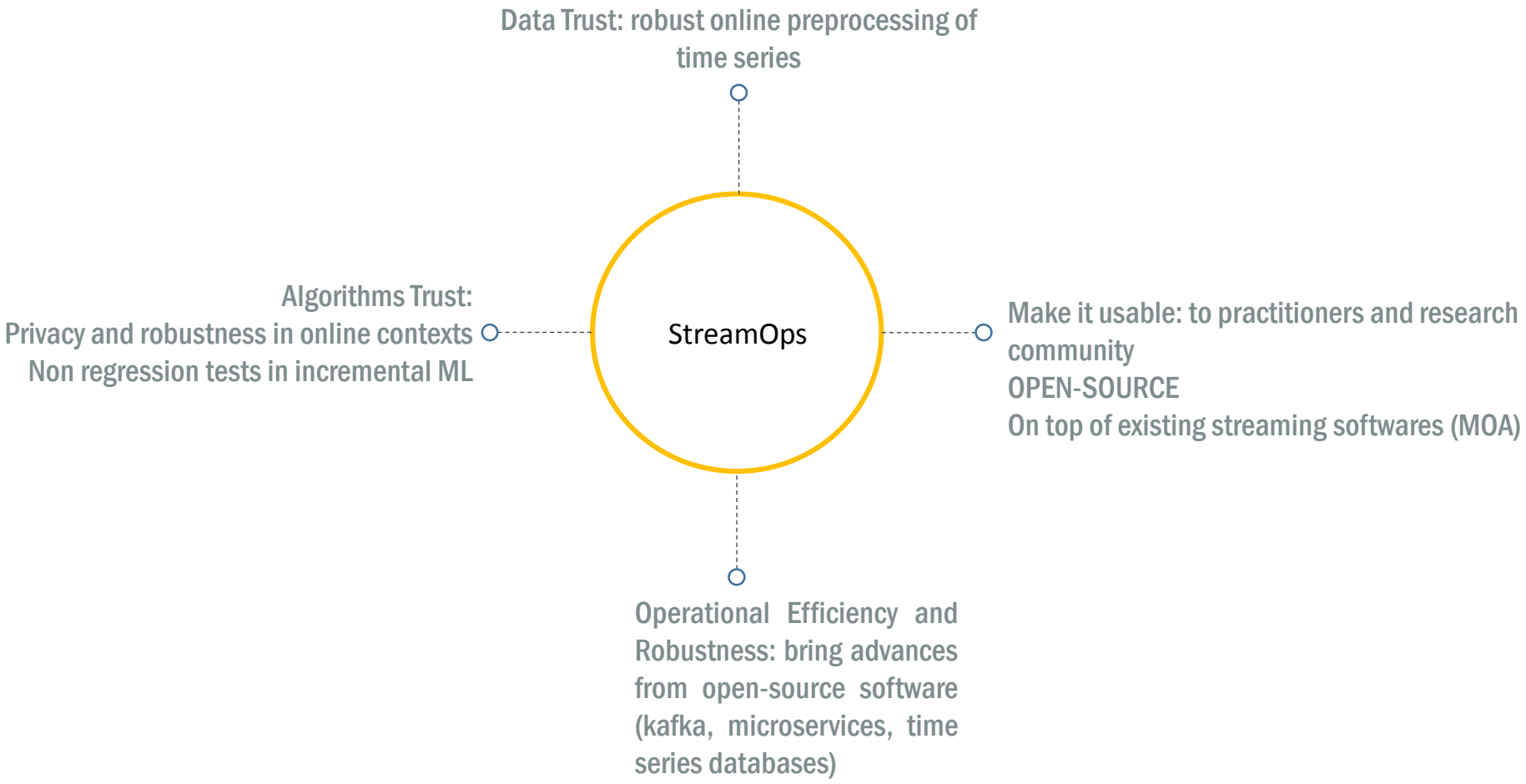
MEDICAL USE CASE EXAMPLE

- To validate data from a connected patch (e.g. to generate alerts) compared to classical devices (e.g. multi-parameters recording from GE)
  - Constrained environment: monitoring during surgeries or post-operative monitoring
  - Long term monitoring: 12h to 24h continuous monitoring

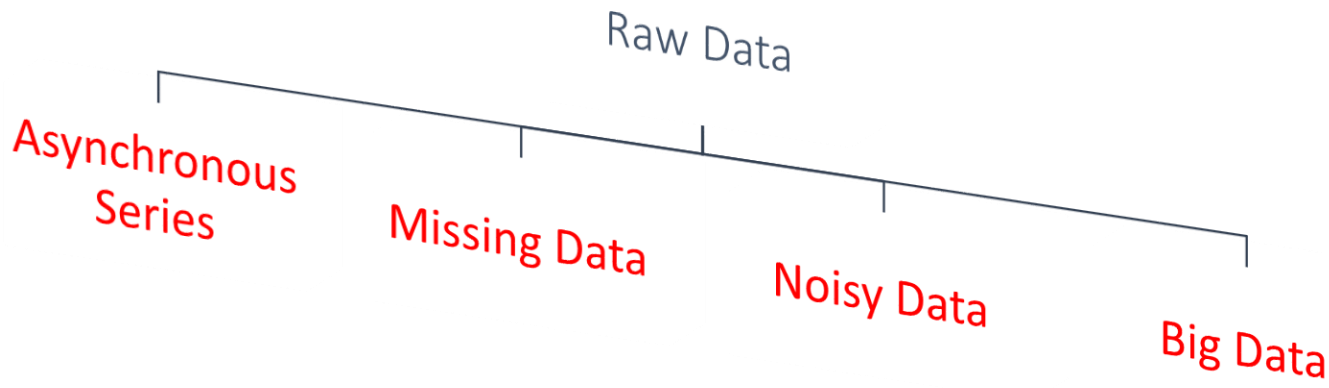


=?

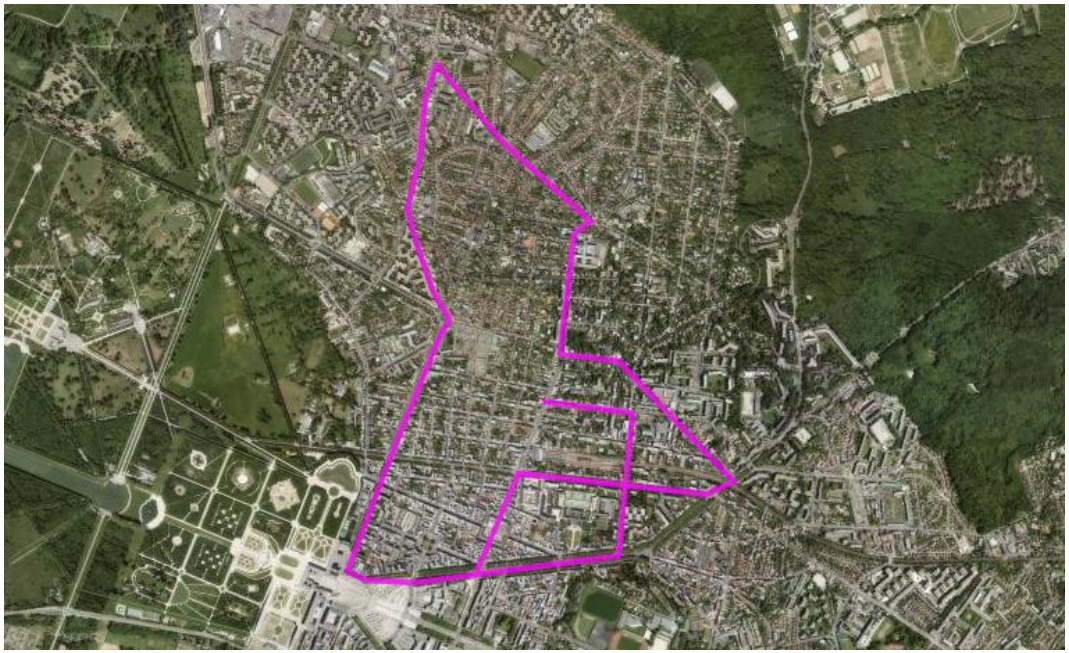
POSITION OF THE STREAMOPS PROJECT – CHALLENGES TRUST, USABILITY AND ROBUSTNESS



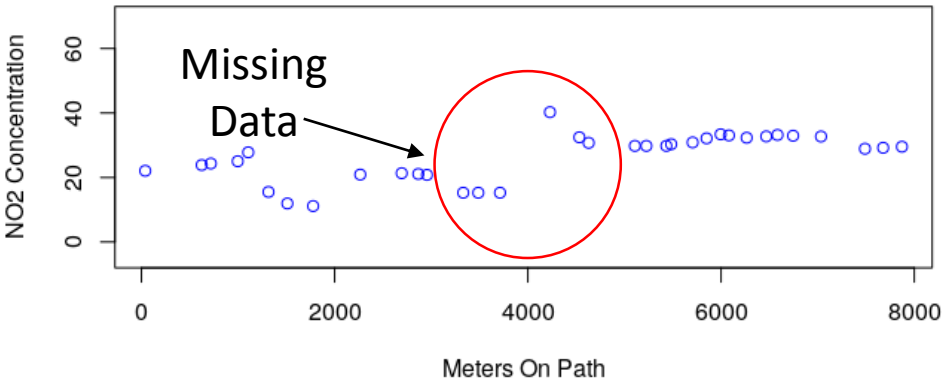
SENSOR RAW DATA PROBLEMS



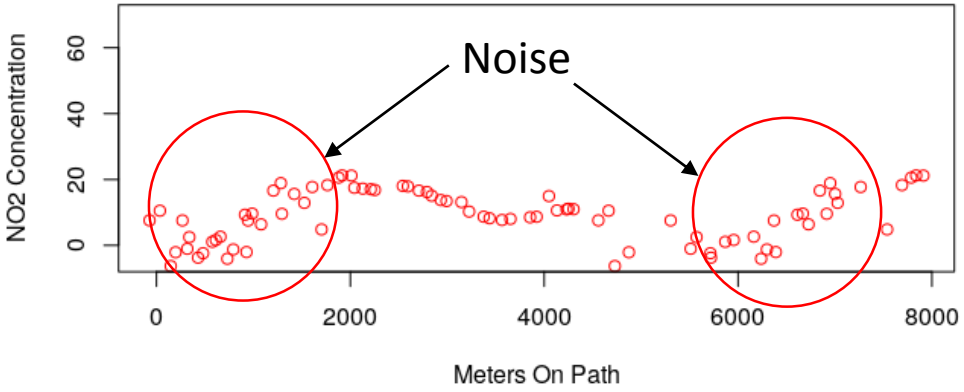
SENSOR RAW DATA PROBLEMS



Data series from Sensor 1



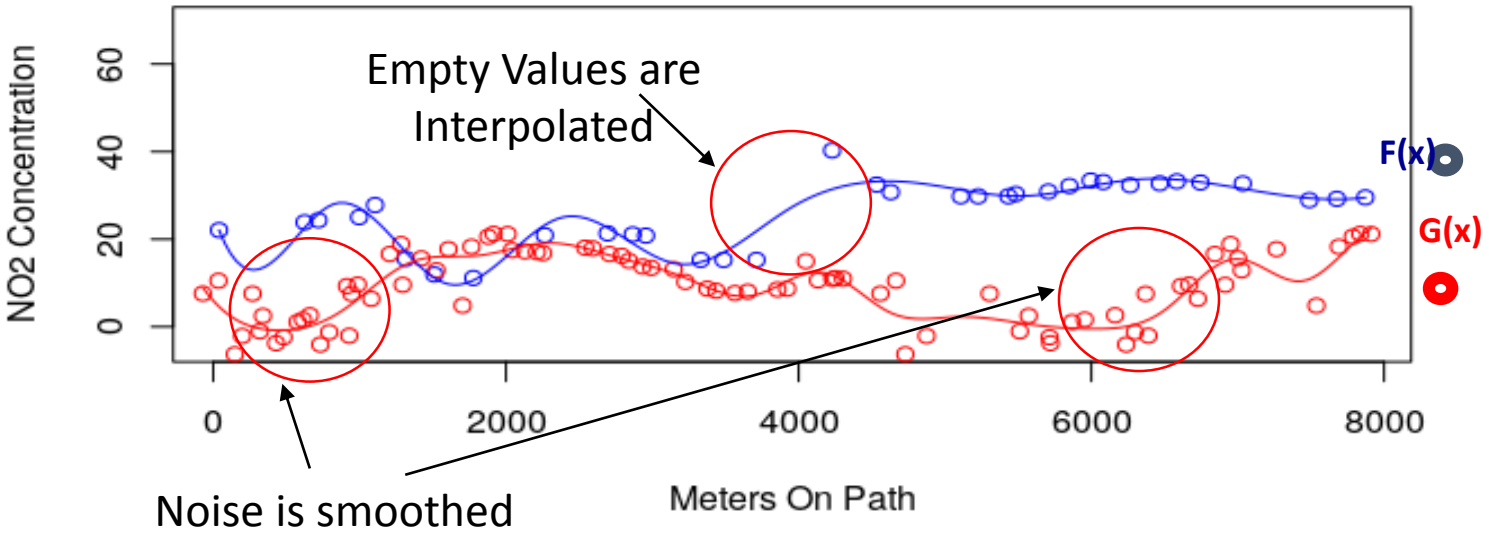
Date Series from Sensor 2



SENSOR RAW DATA PROBLEMS

We adopt Basis Function Expansion F(x) represented by a linear aggregation of basis functions :

$$F(x) = \sum_{1}^{m} c_i B_i(x) = c_1 B_1(x) + c_2 B_2(x) + \ldots + c_n B_m(x)$$



PRIVACY, ROBUSTNESS AND NON-REGRESSION IN INCREMENTAL LEARNING

- **The need for privacy**
  - When the ML model potentially brings information about individuals in the training set
  - Incremental updates of the model by a milicious attacker could bring some leaks in these data
  - Homomorphic encryption does not fit all use cases
  - Research directions
    - Differential privacy in ML online applications
- **Robusness and privacy**
  - Robustness in adversarial contexts or non-adversarial contexts (analyzed in the statistical query framework in deterministic algorithms)
  - In randomized online algorithms (necessary in private contexts) → define and analyze general definition of robustness (with attacks and defenses)
  - Research directions
    - Robustness in private online algorithms
- **Non-regression in incremental learning**
  - Anology with software development: how do you ensure that your model updates still satisfy some constraints?
  - Unit testing and non-regression tests for ML models
    - Optimal training set sampling

SOME REFERENCES

T. L. Coelho da Silva, K. Zeitouni, J. A. F. de Macêdo, et M. A. Casanova, « *CUTiS: Optimized Online ClUstering of Trajectory Data Stream* », in IDEAS 2016

Morvan, K. Choromanski, C. Gouy-Pailler, et J. Atif, « *Graph sketching-based Massive Data Clustering* », SIAM Int. Conf. Data Min. SDM 2018

Morvan, A. Souloumiac, C. Gouy-Pailler, et J. Atif, « *Streaming Binary Sketching based on Subspace Tracking and Diagonal Uniformization* », ICASSP 2018.

R. Pinot, « *Minimum spanning tree release under differential privacy constraints* », ArXiv180106423 Cs Math Stat, janv. 2018.

R. Mousheimish, Y. Taher, et K. Zeitouni, « *Automatic Learning of Predictive CEP Rules: Bridging the Gap Between Data Mining and Complex Event Processing* », in ACM DEBS 2017

Sandu Popa, K. Zeitouni, V. Oria, D. Barth, et S. Vial, « *Indexing in-network trajectory flows* », VLDB J., vol. 20, no 5, p. 643, oct. 2011.

Rafael Pinot, Anne Morvan, Florian Yger, Cedric Gouy-Pailler, Jamal Atif. « *Graph-based Clustering under Differential Privacy*". To appear in UAI 2018, Monterey, USA.

Currently in Kyoto, JSPS summer program (2 months). Hisashi Kashima's lab.