MathMOuse: A Mathematical Models Warehouse to handle both Theoretical and Numerical Data

**Context**

**Time Series in Experimental Sciences**
- Numerical technologies evolution → increase in the volume of data to process.
- In automatic control: time series processing → differential equations.
- Increasing volume of data → increase of derived differential equations.

**Database as service for data scientists?**
- Time series in the context of databases are stored as raw numerical data → The models resulting from the scientists work are not stored.
- Storing and organizing models to ease their management and retrieval became essential.
- A storage of the models would have a better interest → How to exploit the models database to aid the retrieval of models from time series?

**Proposal**

- MathMOuse: an enriched database storing differential equations
- MathMOuse has a « Query by data » system: time series are used to query the database and retrieve relevant equations.
- It contains a GUI to allow user to load models, navigate through them, visualize their data and perform queries, etc.

**Example of a time series**

<table>
<thead>
<tr>
<th>Time (ms)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00000000+00</td>
<td>0.00000000+00</td>
</tr>
<tr>
<td>1.00000000+00</td>
<td>0.41053650-02</td>
</tr>
<tr>
<td>2.00000000+00</td>
<td>1.81202077-01</td>
</tr>
<tr>
<td>3.00000000+00</td>
<td>2.71290651-01</td>
</tr>
<tr>
<td>4.00000000+00</td>
<td>3.58501601-01</td>
</tr>
<tr>
<td>5.00000000+00</td>
<td>4.35010190-01</td>
</tr>
<tr>
<td>6.00000000+00</td>
<td>5.12093310-01</td>
</tr>
<tr>
<td>7.00000000+00</td>
<td>5.86119201-01</td>
</tr>
<tr>
<td>8.00000000+00</td>
<td>6.57242311-01</td>
</tr>
<tr>
<td>9.00000000+00</td>
<td>7.25756212-01</td>
</tr>
</tbody>
</table>

**Numerical tools**
- Equation solver (Runge-Kutta)
- Time Series comparison
  - Time Series contains continuous values ⇒ comparison is subject to noise, approximation,…
  - Strict equality is not relevant
  - Comparison: compute criteria (average error, standard deviation of the error), and threshold values (minimum value for acceptance, maximum value for rejection).

**Models**
Models are:
- Differential equations
- Ordinary
- Linear (storage structure allow non-linearity, but numerical solving is more complex)

**Database conversion**
Conversion XML to tuples

**Implementation**

**Technical Setup**
- Docker, Docker-compose
- PostgreSQL
- Microservices developed in Java
- Docker rabbitmq
- https://forge.lias-lab.fr/projects/mathmouse

**Participants**
- Cyrille PONCHATEAU
  PhD Student
  ponchat@ensma.fr
- Ladjel BELLATRECHE
  Prof. at l’ENSMA
  bellatreche@ensma.fr
- Carlos ORDONEZ
  Prof. at University of Houston
  ordonez@cs.uh.edu
- Mickael BARON
  Research associate
  baron@ensma.fr

**Tests**

**Set of data used**
- 100 randomly generated differential equations (order up to 5)
- A set of Time Series, derived from the above random equations, with up to 5000 elements

**Parallel computing**
- Microservices → comparison and generation can be parallelized
- Prototype tested with different numbers of comparison and generation parallel processes

**To know more about it…**

**Related Publications**
- Cyrille PONCHATEAU, Ladjel BELLATRECHE, Mickael BARON, Entrepôts de Données dans l’air Data Science : De la Données au Modèle, (EDA 2016), June, 2016
- Cyrille PONCHATEAU, Ladjel BELLATRECHE, Carlos ORDONEZ, Mickael BARON, A Database Model for Time Series : From a traditional Data Warehouse to a Mathematical Models Warehouse, 32e journées Bases de Données Avancées (BDA), Potters, 2016