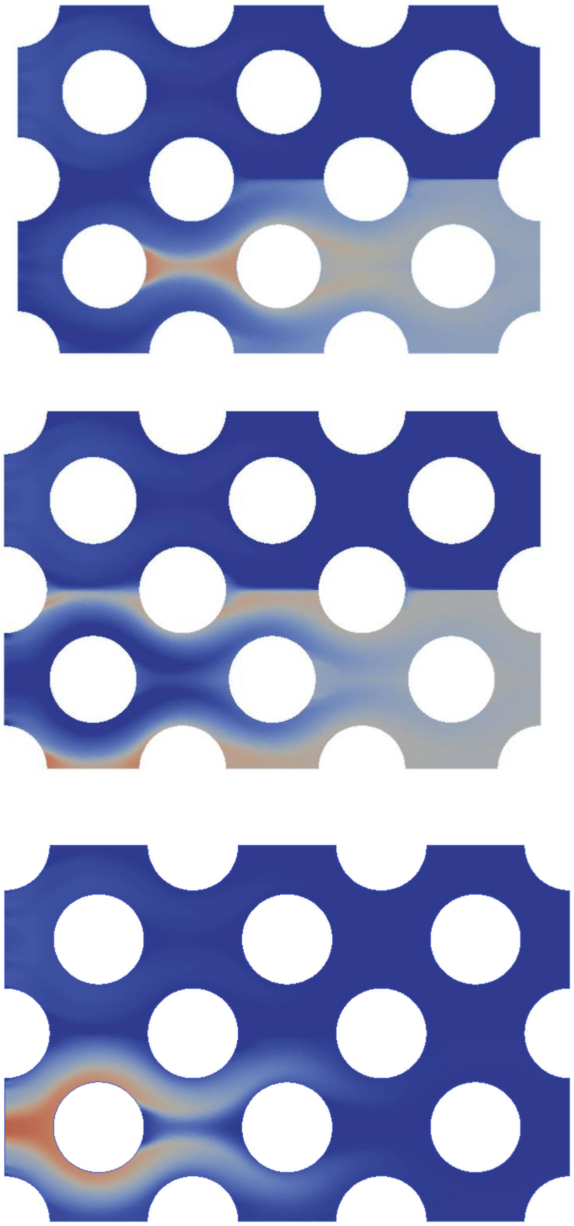
MELISSA : Challenges computationnels de l’étude de l’incertitude dans les simulations numériques

Alejandro Ribés
EDF R&D Paris-Saclay

ETICS
17 septembre 2021



THE HEROES OF ETICS 2021: BAIN DE MINUIT

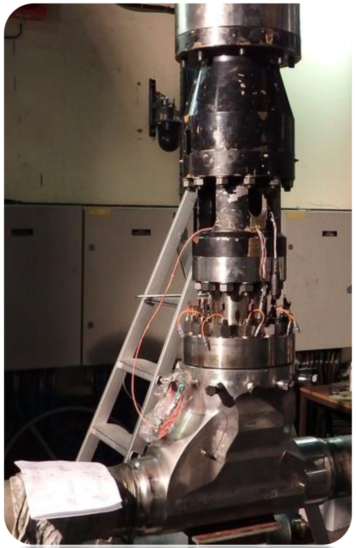
- **Claire**
- **Thibault**
- **Julien**
- **Bruno**
- **Clément**
- **Batiste**

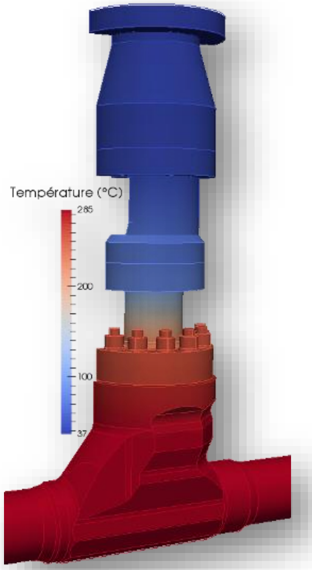
EDF – ELECTRICITÉ DE FRANCE

- **Electric utility company**
- **58 active nuclear reactors in France (all PWRs)**
- **EDF Energy in UK**
  - 8 nuclear power stations (7 AGR)
- **EDF R&D**
  - About 2,000 researchers
  - Saclay →
  - Several top500 supercomputers
    - Currently 2 clusters
  - Extensive use of numerical simulation

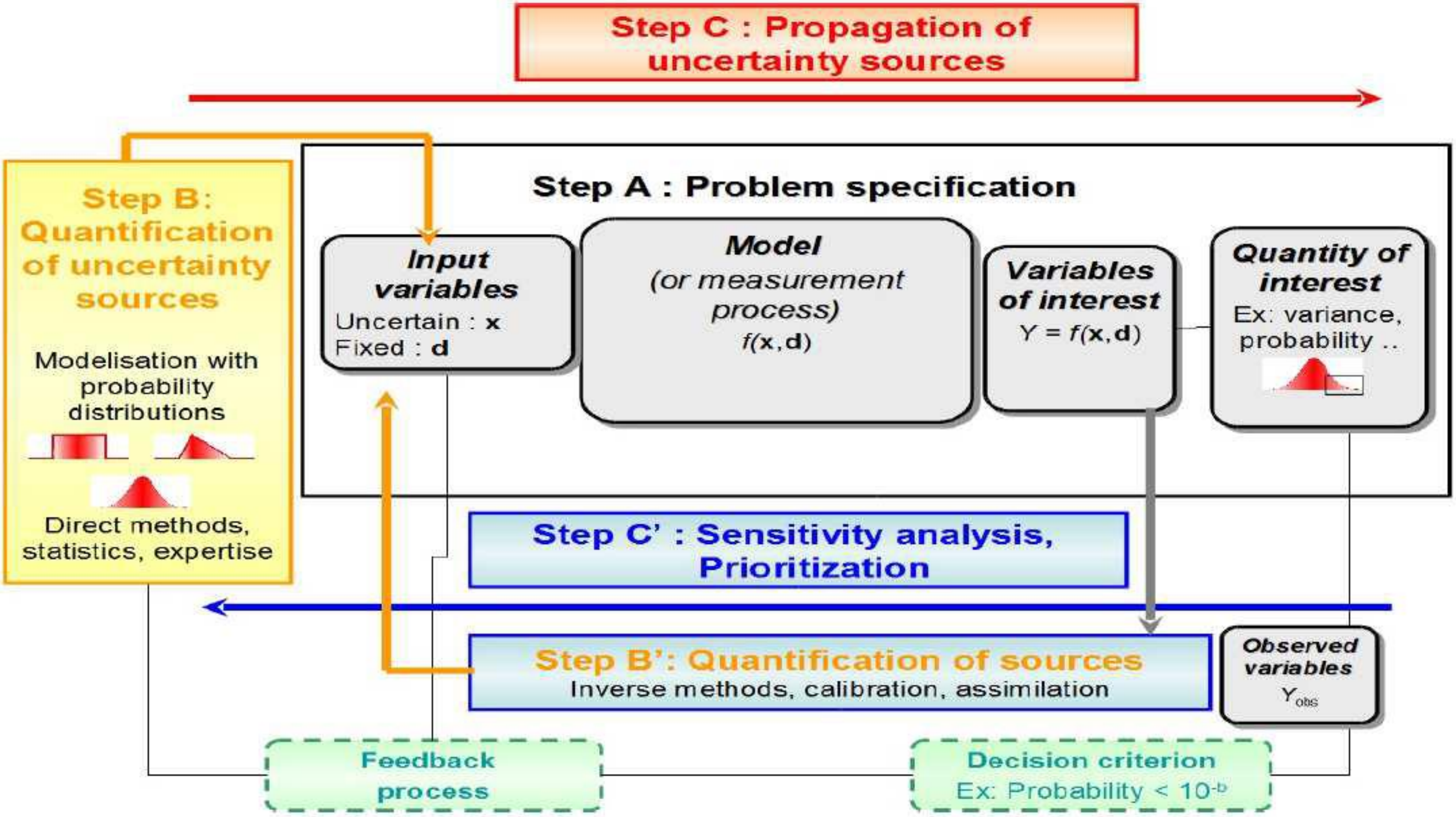
SUMMARY

1. General context
2. **The problem**
3. Ubiquitous statistics
4. Melissa
5. Conclusion



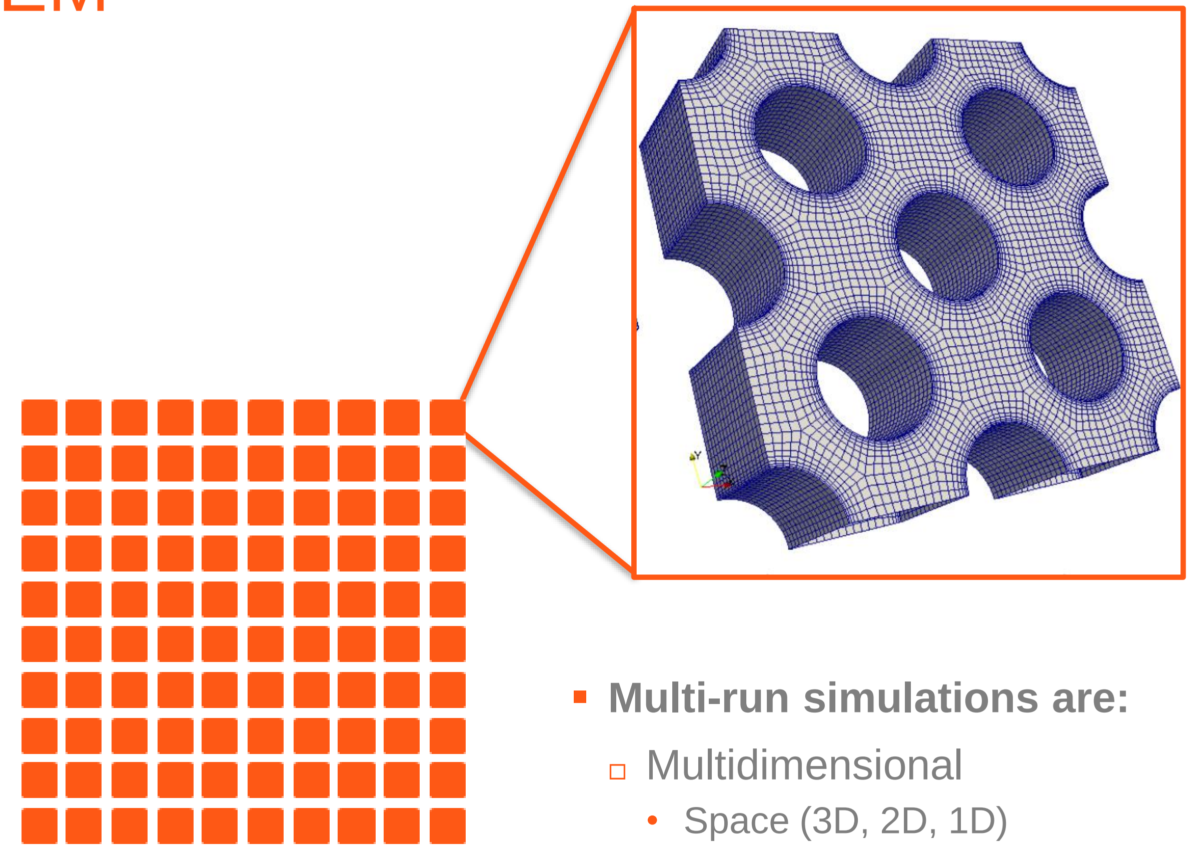


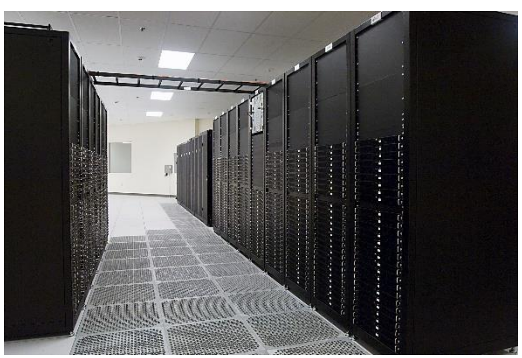
THE PROBLEM



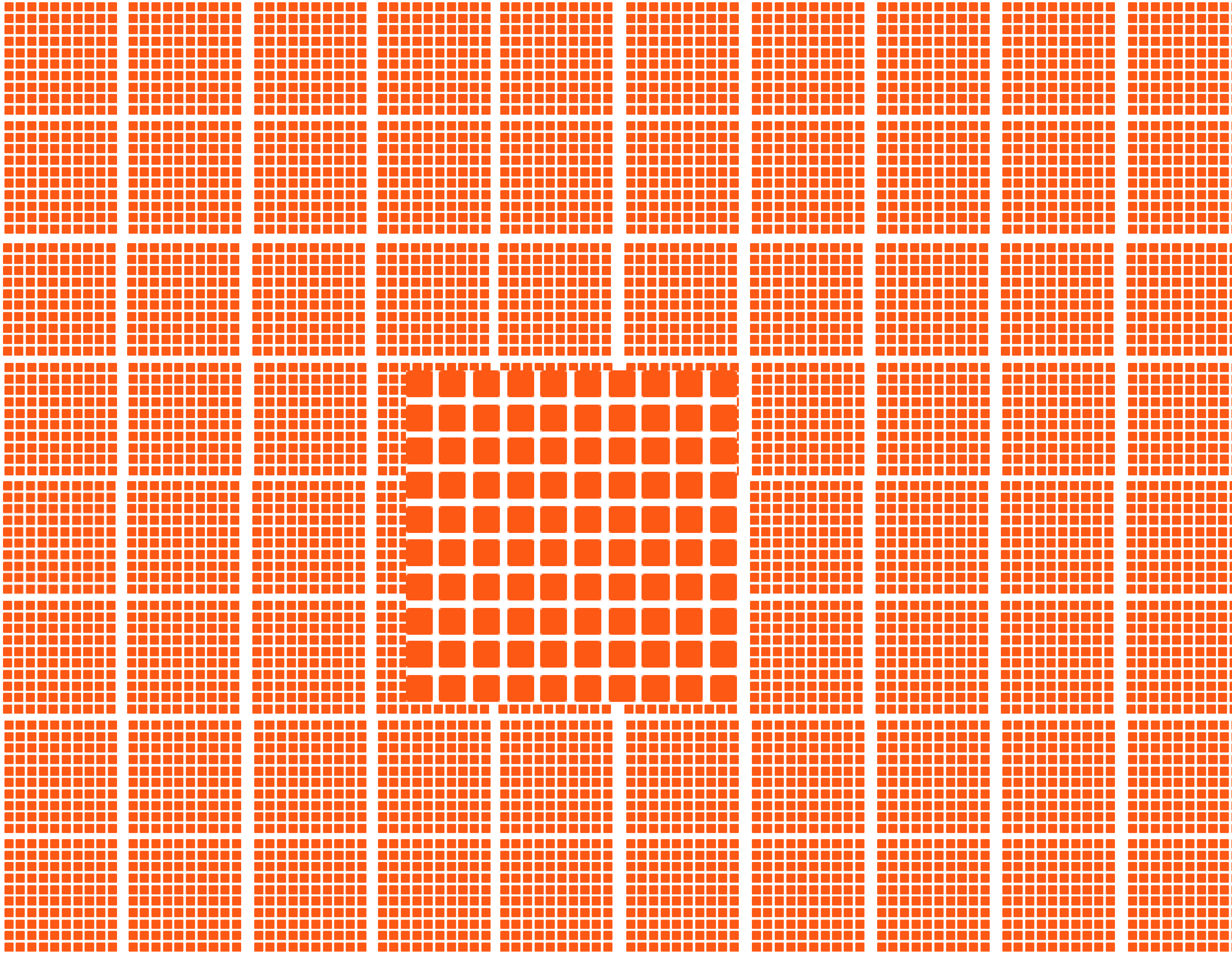
Step C : Propagation of uncertainty sources
Step B: Quantification of uncertainty sources
Modelisation with probability distributions
Direct methods, statistics, expertise
Step A : Problem specification
Input variables
Uncertain : x
Fixed : d
Model
(or measurement process)
f(x,d)
Variables of interest
Y = f(x,d)
Quantity of interest
Ex: variance, probability ..
Step C' : Sensitivity analysis, Prioritization
Step B': Quantification of sources
Inverse methods, calibration, assimilation
Observed variables
Yobs
Feedback process
Decision criterion
Ex: Probability < 10-b

THE PROBLEM

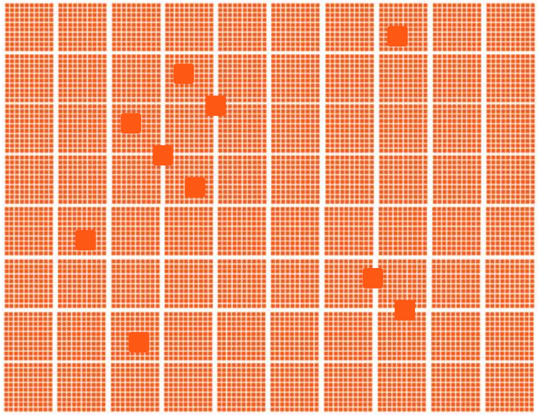




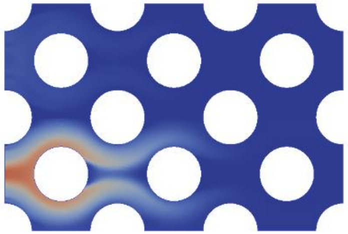
- **Multi-run simulations are:**
  - Multidimensional
    - Space (3D, 2D, 1D)
    - Time
  - Multivalued (temperature, pression, height, etc)
  - Multivariate (1,000 or 100,000?)



THE PROBLEM

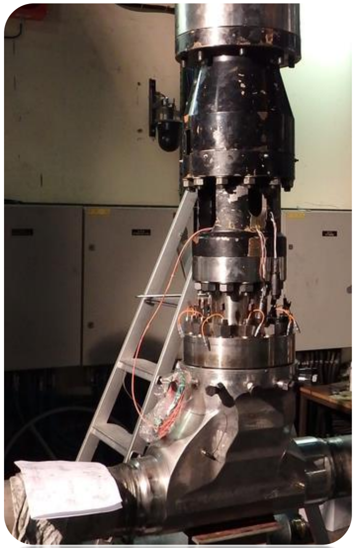


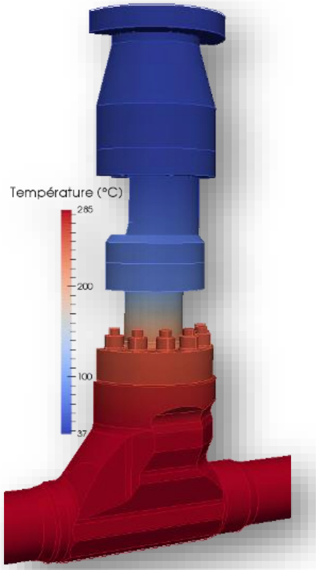
Post-treatment



SUMMARY

1. An overview of SALOME platform
2. The problem
3. **Ubiquitous statistics**
4. Melissa
5. Conclusion



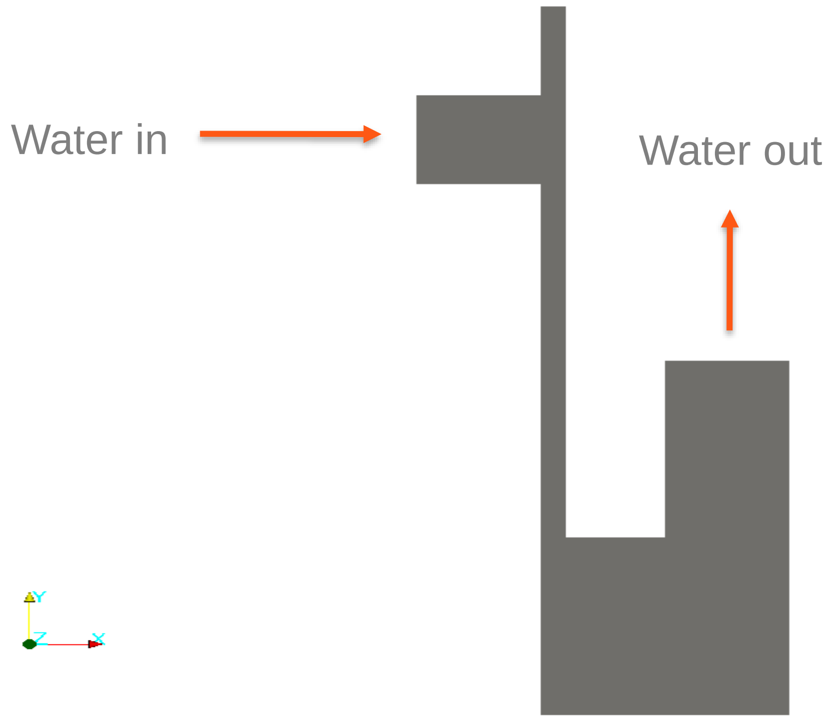


THE SOLUTION: UBIQUITOUS STATISTICS

- **UBIQUITOUS STATISTICS are:**
  - Multidimensional
    - Space (3D, 2D, 1D)
    - Time
  - Multivalued (temperature, pression, height, etc)

THE SOLUTION: UBIQUITOUS STATISTICS

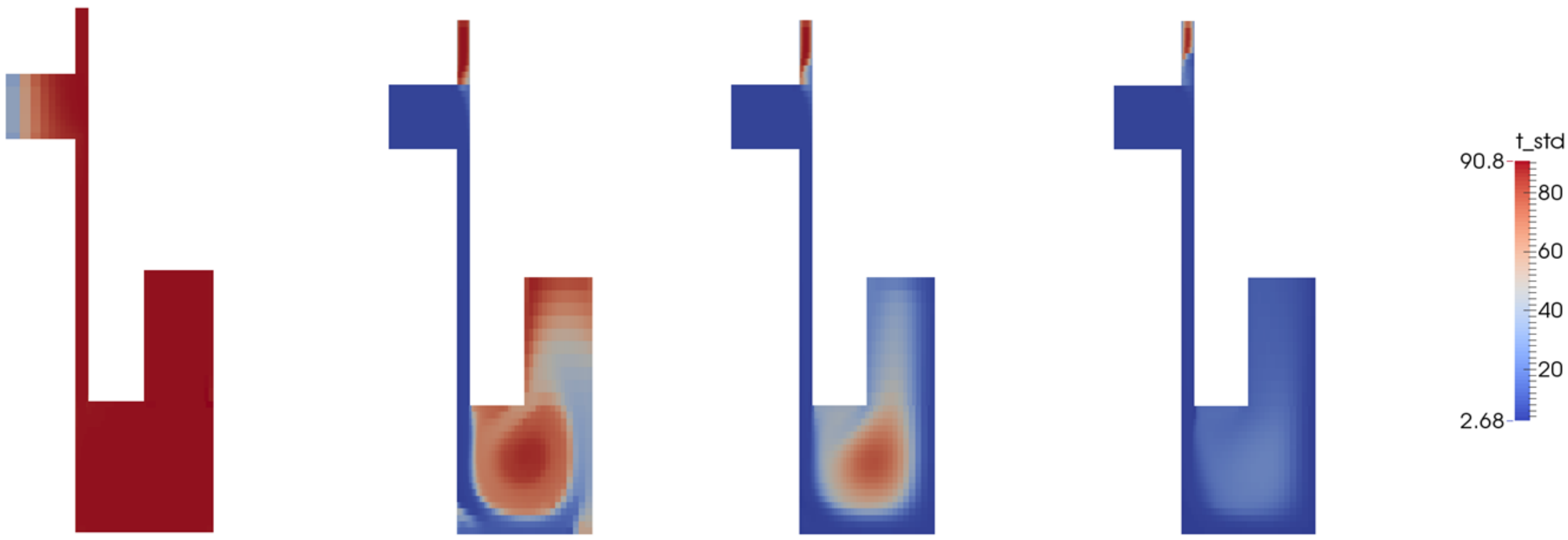
- **Purge of a vessel**
  - Temperature inside de vessel: from 200 to 490
  - 30 simulations (10 C intervals)



THE SOLUTION: UBIQUITOUS STATISTICS

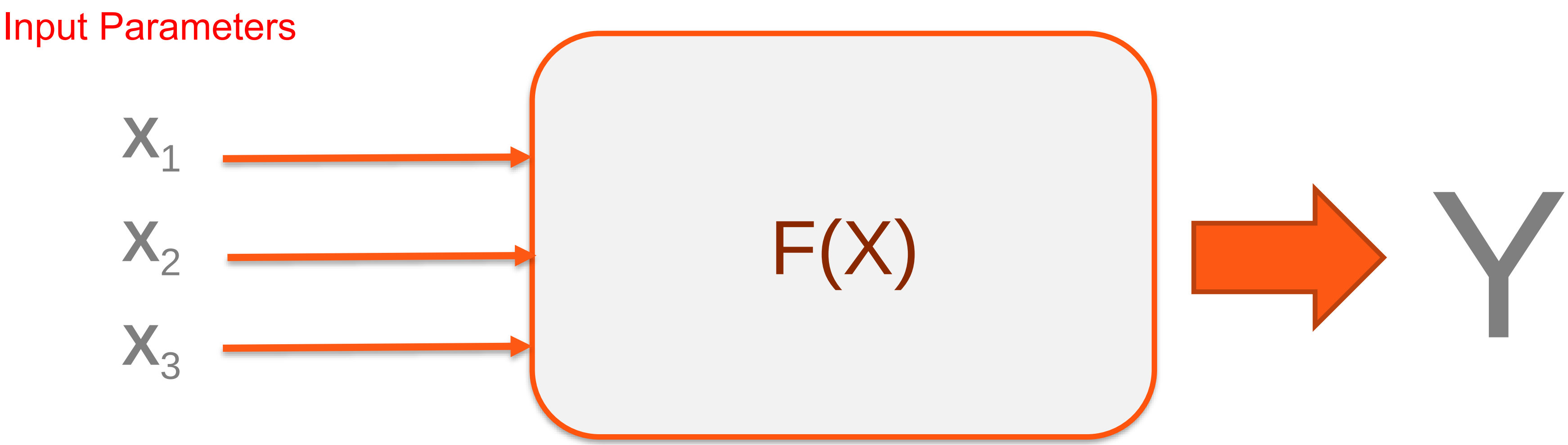
- **Purge of a vessel**
  - Temperature inside de vessel: from 200 to 490
  - 30 simulations (10 C intervals)

Standard deviation



UBIQUITOUS STATISTICS: SOBOL INDICES

- **Sobol Indices**
  - Part of variance $S_i = Var(E[Y|X_i]) / Var(Y)$



UBIQUITOUS STATISTICS: SOBOL INDICES

- **Sobol Indices**
  - Part of variance

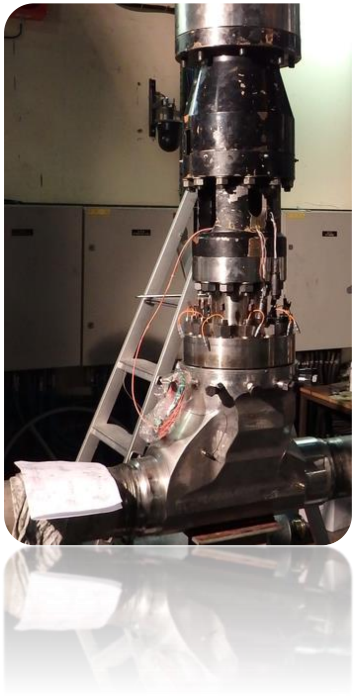
$$\mathbf{S}_i = \mathrm{Var}(\mathbf{E}[\mathbf{Y}\ |\mathbf{X}_i\ ])\ /\ \mathrm{Var}(\mathbf{Y}\ )$$

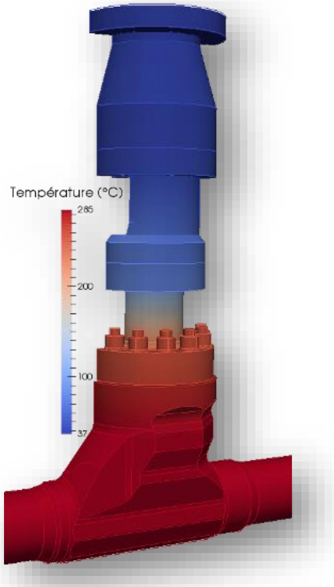
- **UBIQUITOUS Y:**
  - Multidimensional
    - Space (3D, 2D, 1D)
    - Time
  - Multivalued (temperature, pression, height, etc)
  - Multivariate

Input Parameters

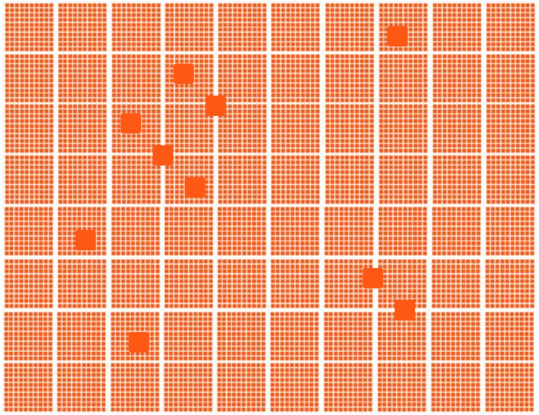
SUMMARY

1. An overview of SALOME platform
2. The problem
3. What people do?
4. Ubiquitous statistics
5. **Melissa**
6. Conclusion

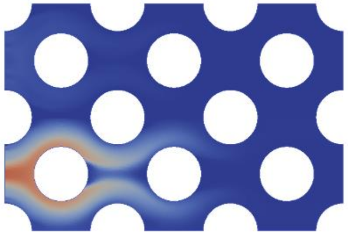




THE PROBLEM



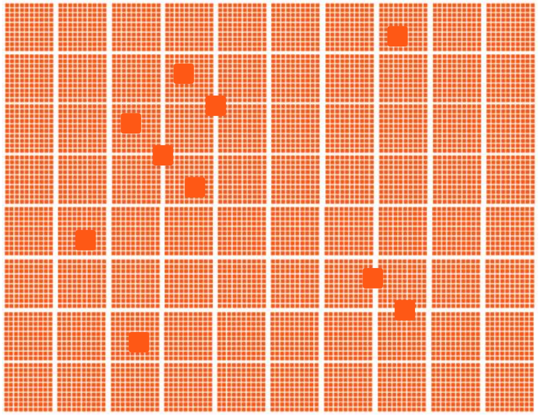
Post-treatment



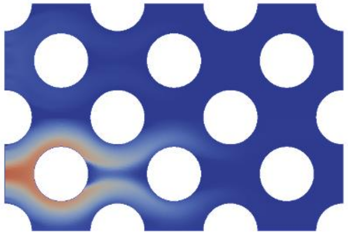
48TB

Some MB

IN-TRANSIT STATISTICAL ANALYSIS



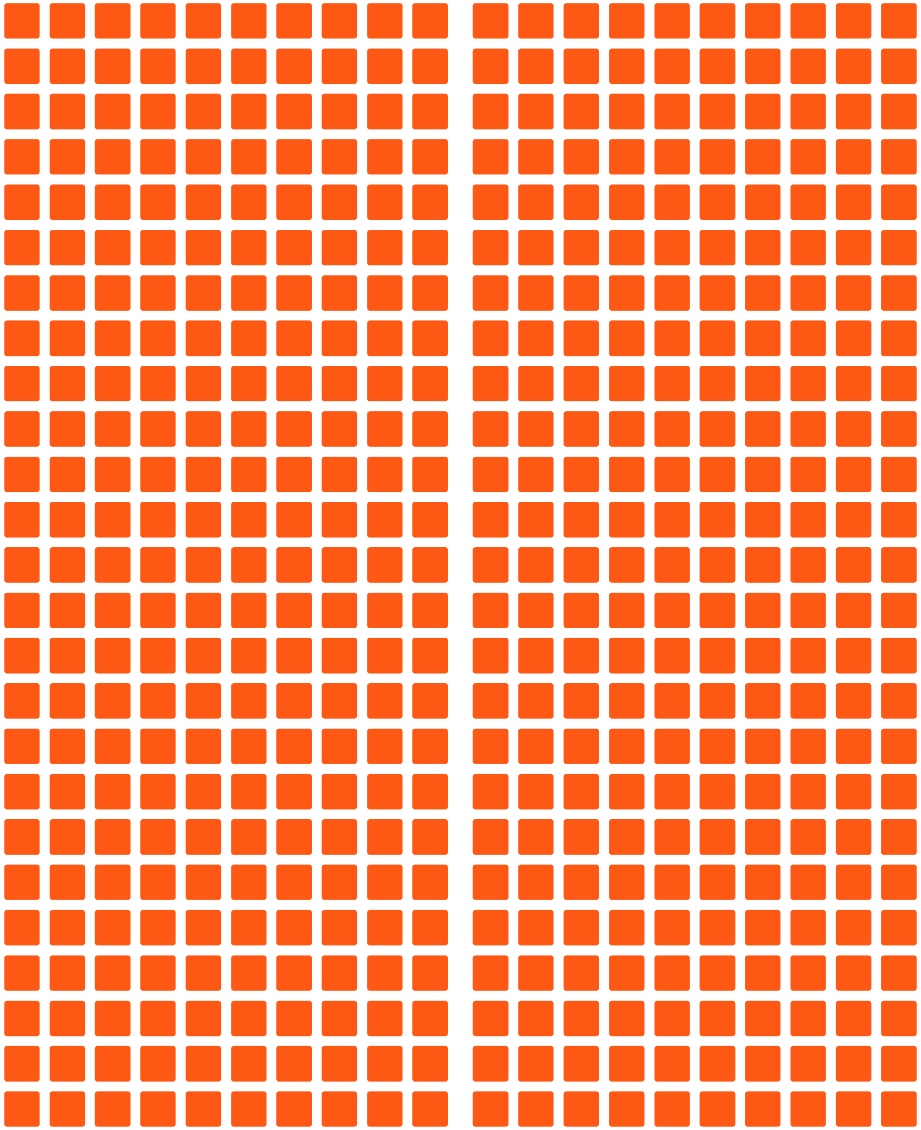
MELISSA

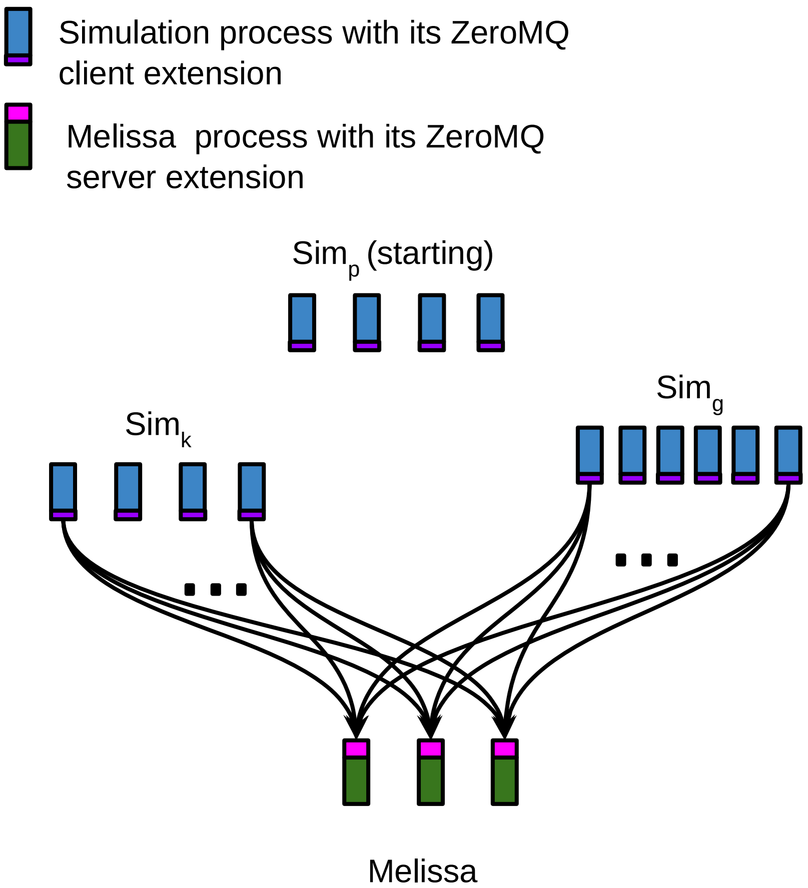


48TB

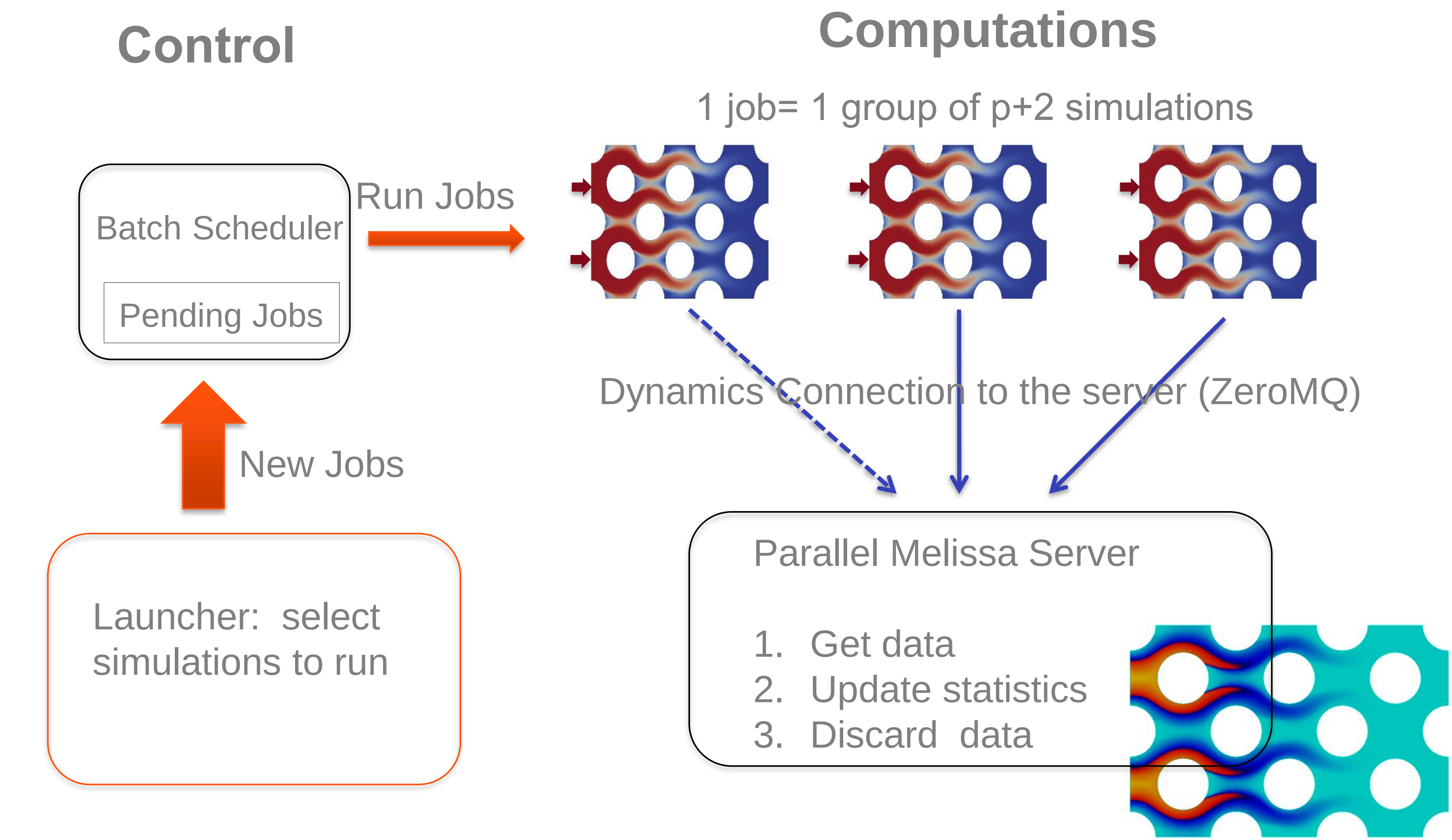
Some MB

MELISSA: UBIQUITOUS AND ASYNCHRONOUS





MELISSA FRAMEWORK



MELISSA: ITERATIVE STATISTICS

MELISSA

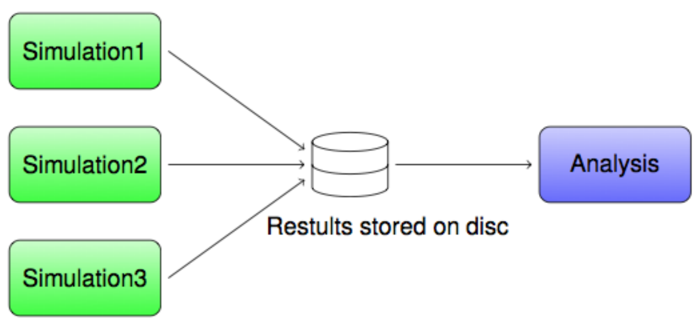
Variable

Average of variable

Spatial coordinates

$$\mu(X,t) = \frac{\sum_{i=1,N} u(i,X,t)}{N}$$

Time step



MELISSA: ITERATIVE STATISTICS

MELISSA

Zero intermediate files thanks to iterative statistics

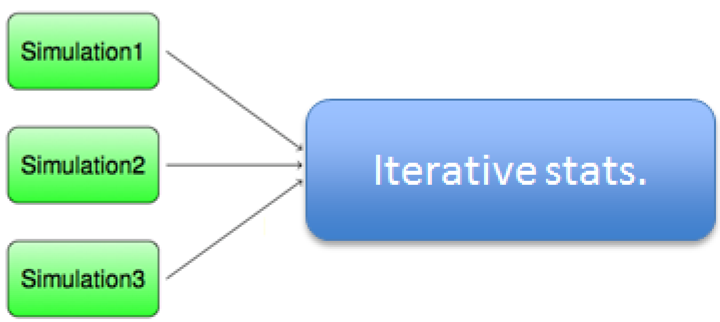
Iterative average (i^{th} update):

$$\mu_i(X,t) = \mu_{i-1}(X,t) + \frac{1}{i}(u(i,X,t) - \mu_{i-1}(X,t))$$

Iterative standard deviation (i^{th} update):

$$V_i(X,t) = V_{i-1}(X,t) + (u(i,X,t) - \mu_{i-1}(X,t))(u(i,X,t) - \mu_i(X,t))$$

Iterative Sobol' Indices, …



SOBOL' INDEX ESTIMATION: PICK FREEZE METHOD

parameters

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,p} \\ \vdots & \ddots & \\ a_{n,1} & & a_{n,p} \end{pmatrix}; B = \begin{pmatrix} b_{1,1} & \cdots & b_{1,p} \\ \vdots & \ddots & \\ b_{n,1} & & b_{n,p} \end{pmatrix}$$

A and B are random matrices

$$C^k = \begin{pmatrix} a_{1,1} & \cdots & a_{1,k-1} & b_{1,k} & a_{1,k+1} & \cdots & a_{1,p} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i,1} & \cdots & a_{i,k-1} & b_{i,k} & a_{n,k+1} & \cdots & a_{i,p} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,k-1} & b_{n,k} & a_{n,k+1} & \cdots & a_{n,p} \end{pmatrix}$$

Martinez estimator for first order and total Sobol' Indices [Baudin 2016]

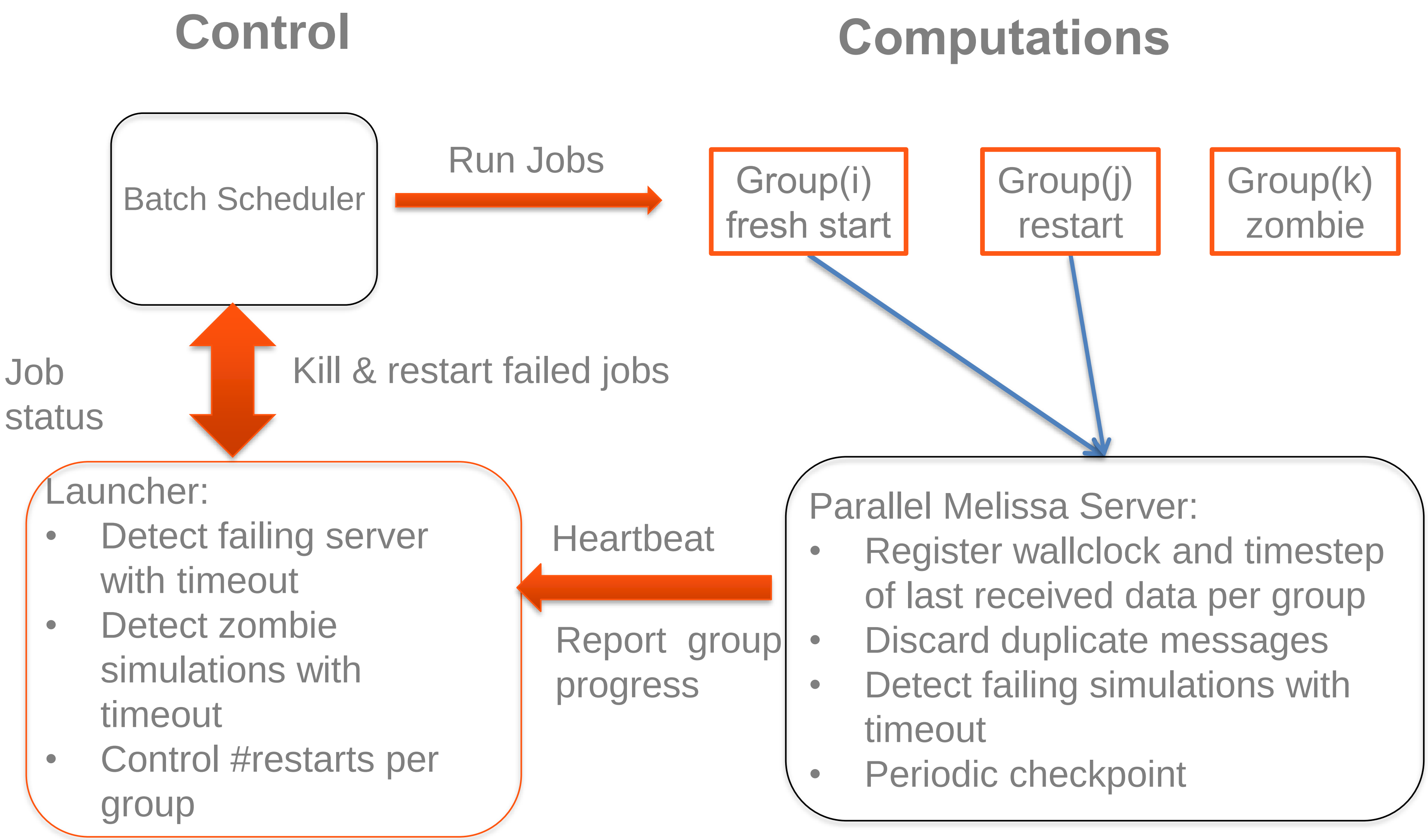
C^k built from A and B

$$S_k(f, A, B) = \frac{Cov(Y^B, Y^{C^k})}{\sqrt{\mathbb{V}(Y^B)}\sqrt{\mathbb{V}(Y^{C^k})}},$$

$$ST_k(f, A, B) = 1 - \frac{Cov(Y^A, Y^{C^k})}{\sqrt{\mathbb{V}(Y^A)}\sqrt{\mathbb{V}(Y^{C^k})}}$$

Require running n x (p+2) simulations, with parameters given by each raw of A, B, C^k (k=1..p)

FAULT TOLERANCE



Checkpointing: 491GB, +0,5% exec time

EXPERIMENTS

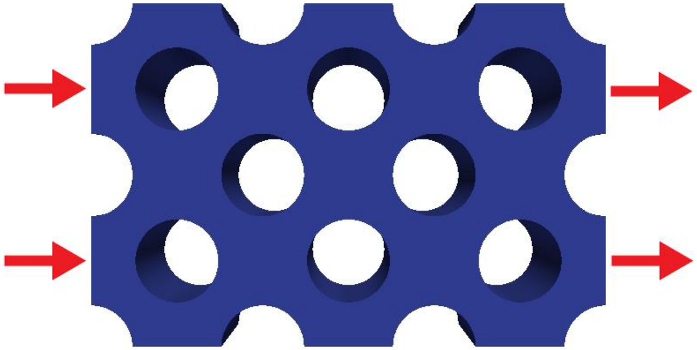
Fluid simulation with Code_Saturne [EDF]

Curie Machine (80K cores)

9M hexahedral mesh – 100 timesteps

6 parameters, 3 per injector:

- Dye concentration
- Injection width
- Injection duration



Ubiquitous Sobol' indices: 9x100X2=1800M indices (dye concentration)

8 simulations per group, 1000 groups, each one running on 512 cores

Generate 48TB of intermediate results

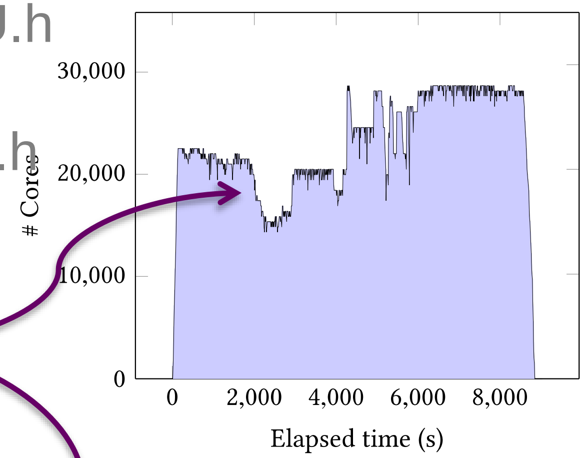
Server size: enough nodes to work in memory (491GB)

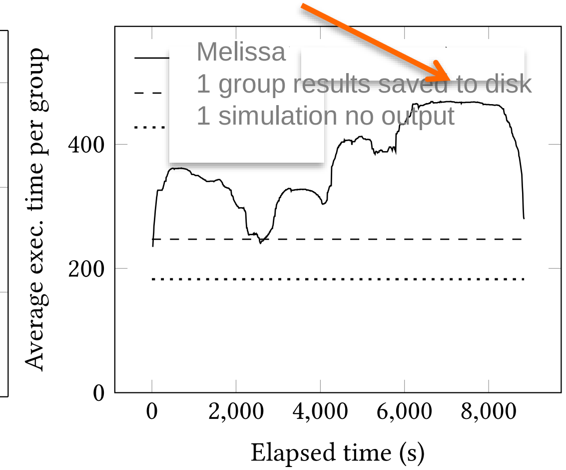
EXPERIMENTS

Bottleneck on the server side

15 server nodes: 600 CPU.h (1%)

Simulations: 56500 CPU.h

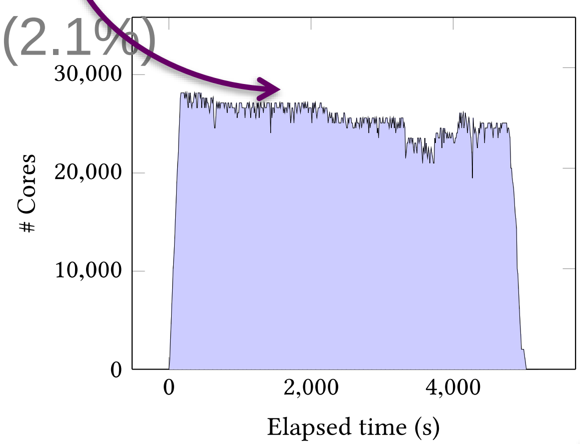


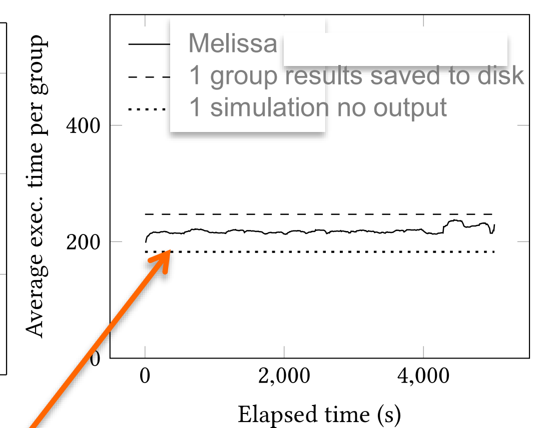


Elastic execution up to 28K cores

32 server nodes: 740 CPU.h (2.1%)

Simulations: 34000 CPU.h





From 15 to 32 server nodes:

- 1% more cores
- 40% less CPU.h

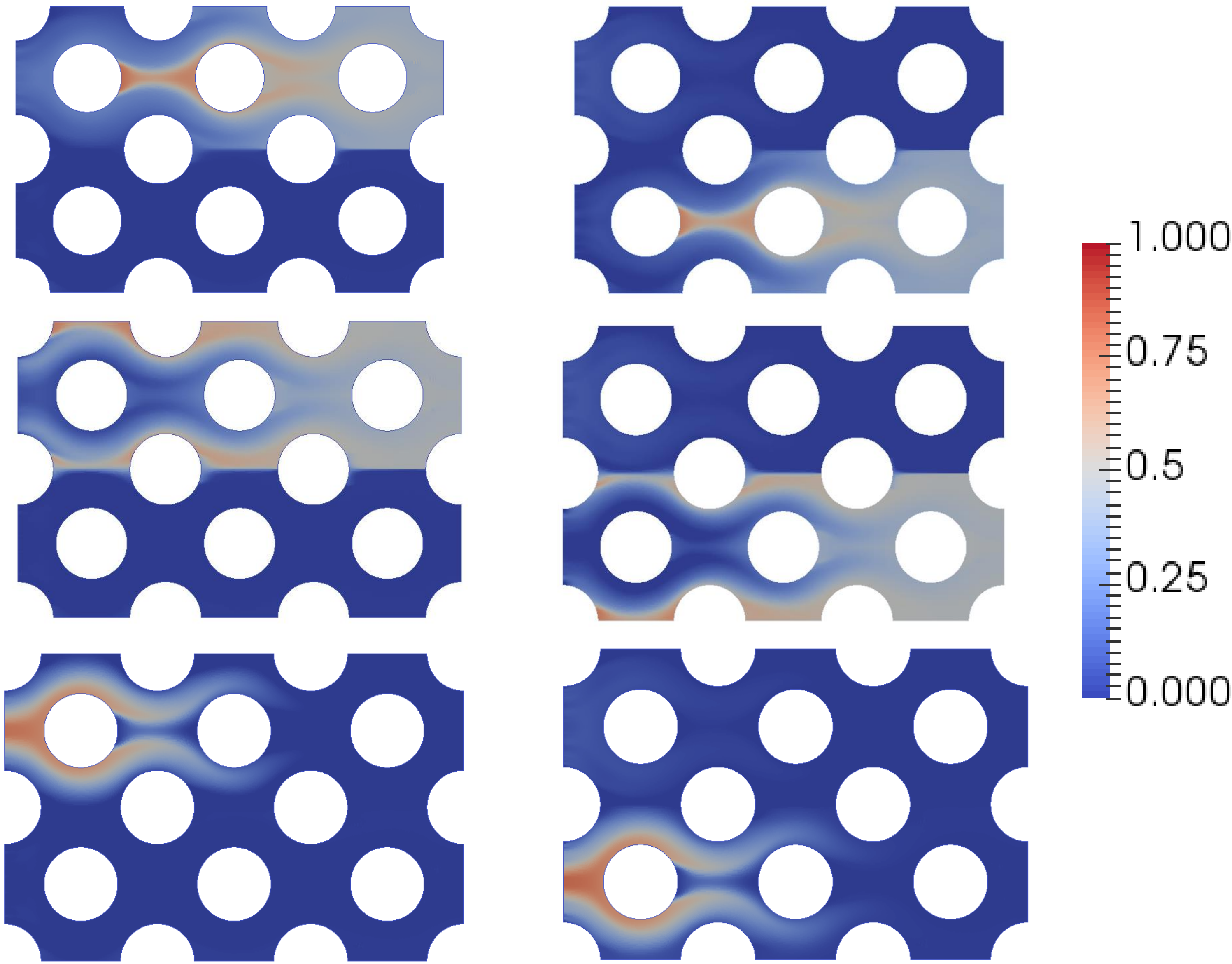
Groups run 13% faster on average than when writing to disk

Injector 1 Injector 2

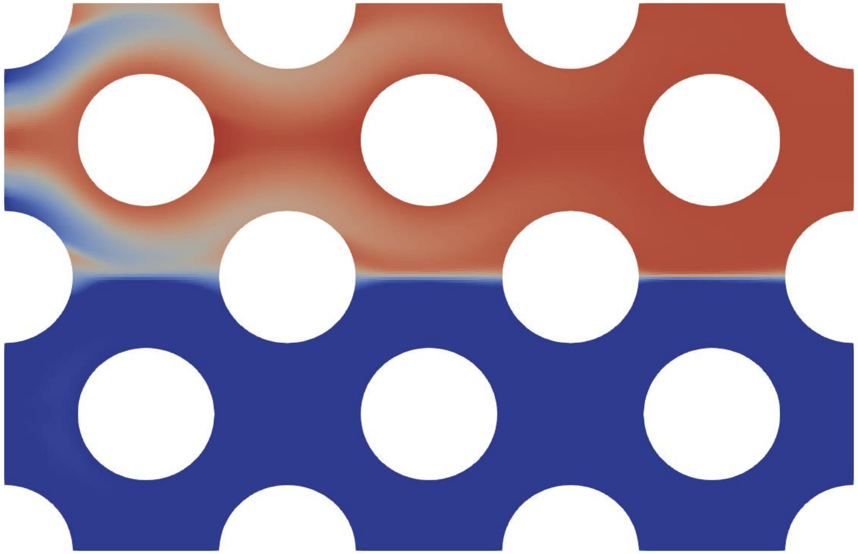
Die concentration

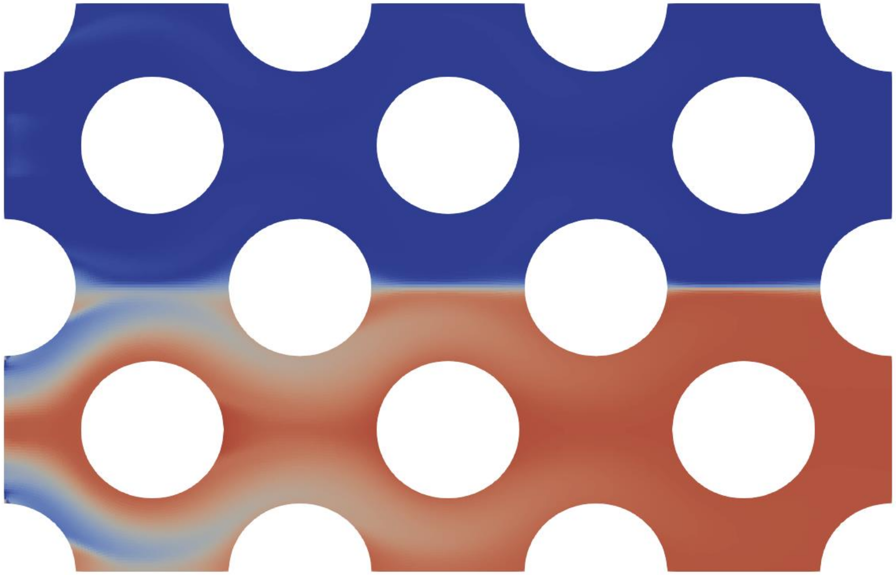
Width

Duration



- Composition d'indices de Sobol



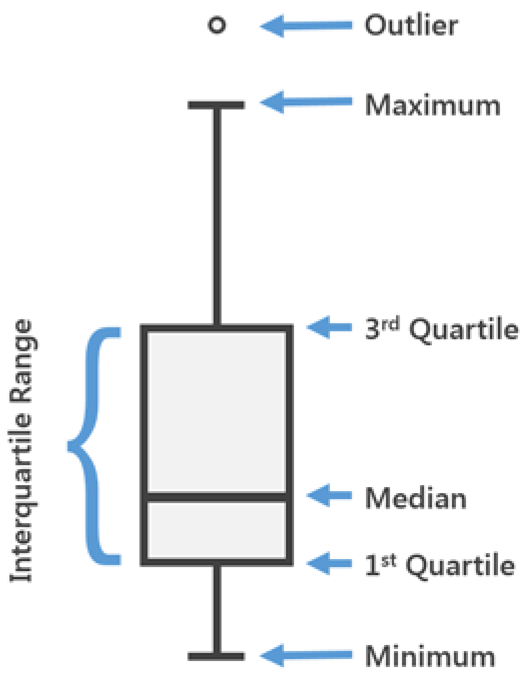


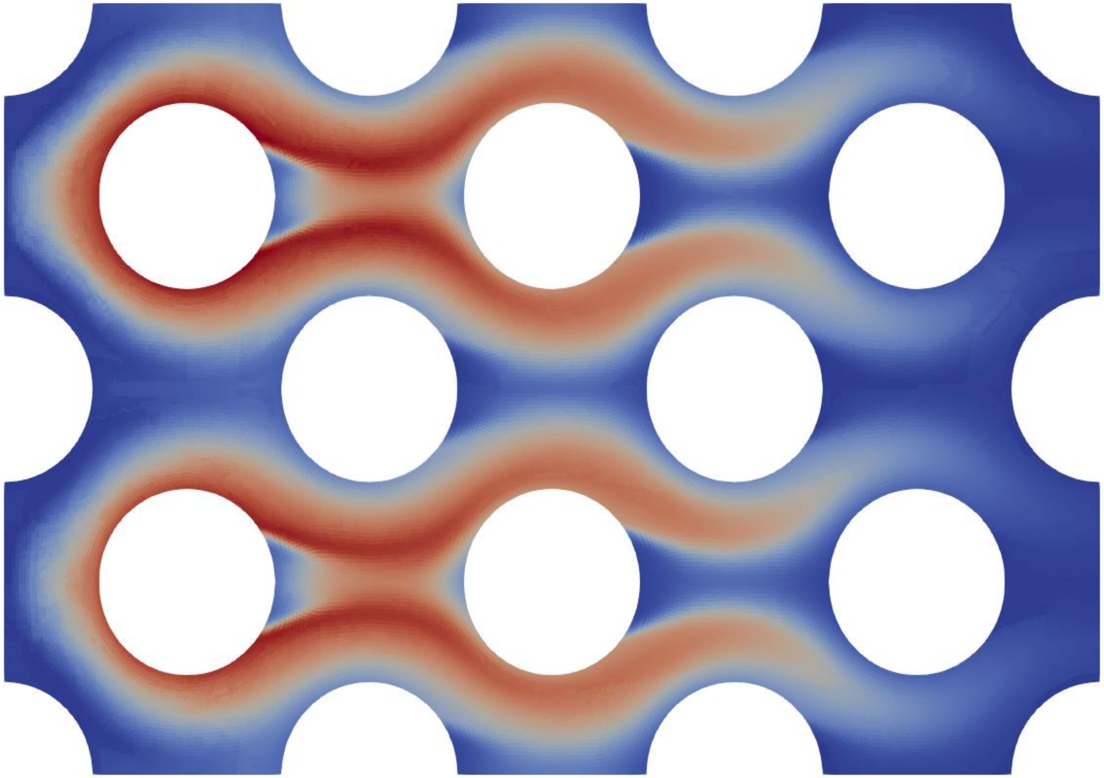
Statistiques d'ordre : QUANTILES

- Quantiles
- Image : bande inter-percentile à 90%

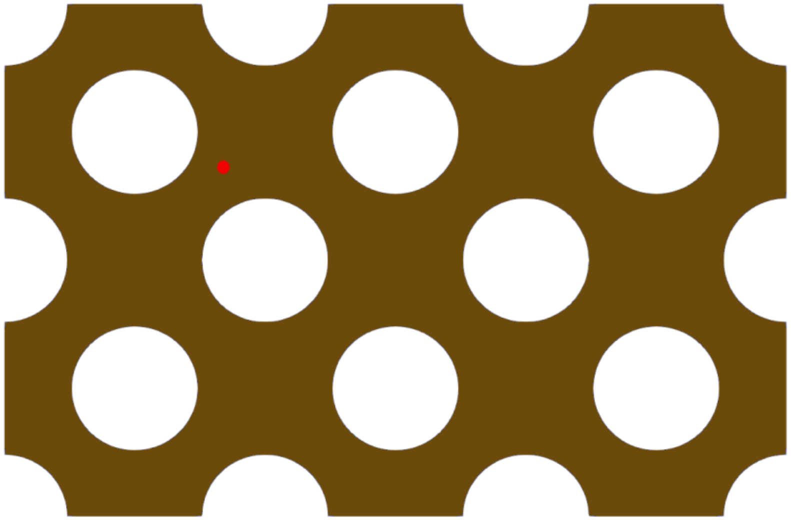
$$q_{n+1} = q_n - \frac{1}{n^\gamma}\left(1_{Y_{n+1} \leq q_n} - \alpha\right)$$

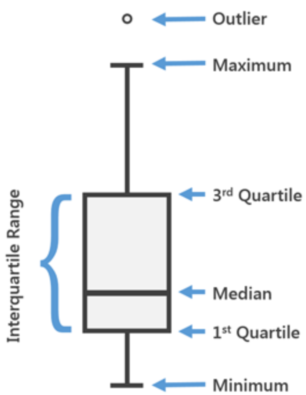
Robbins-Monro estimator

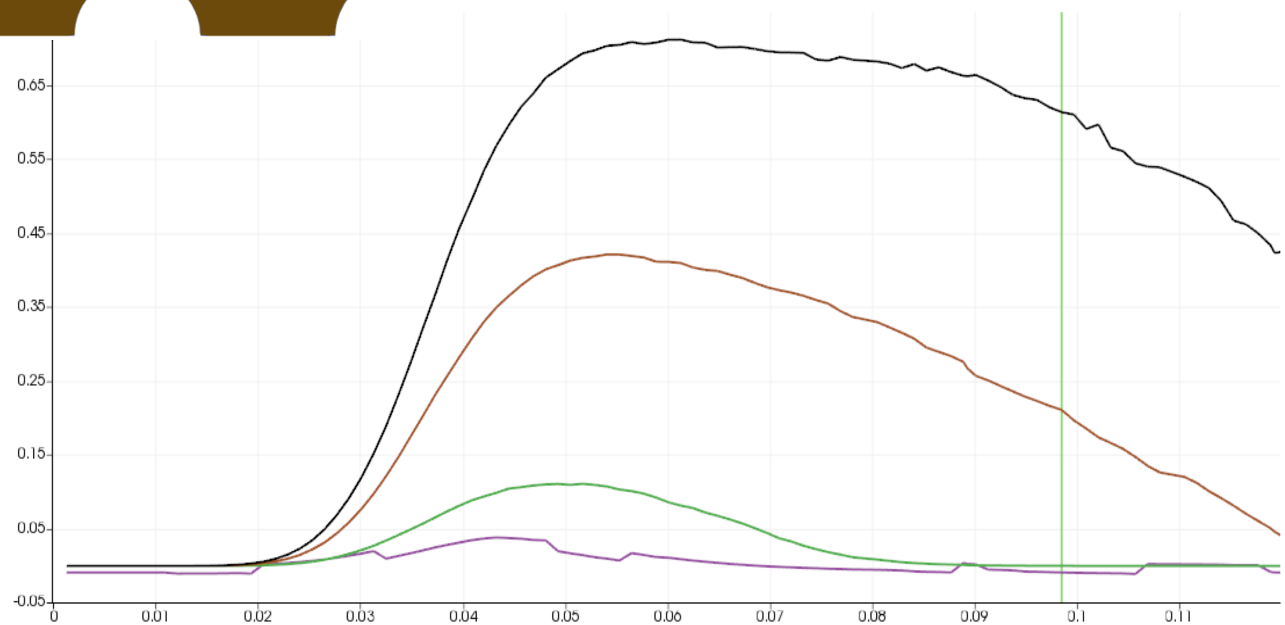




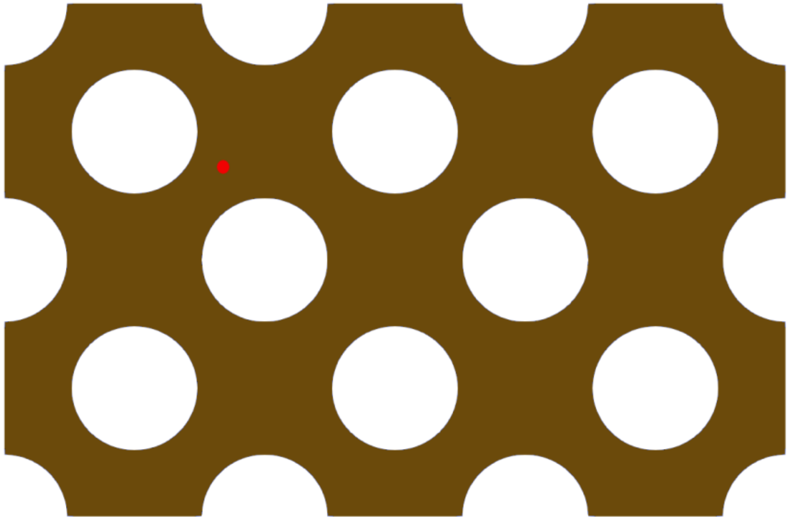
Evolution temporelle des quantiles

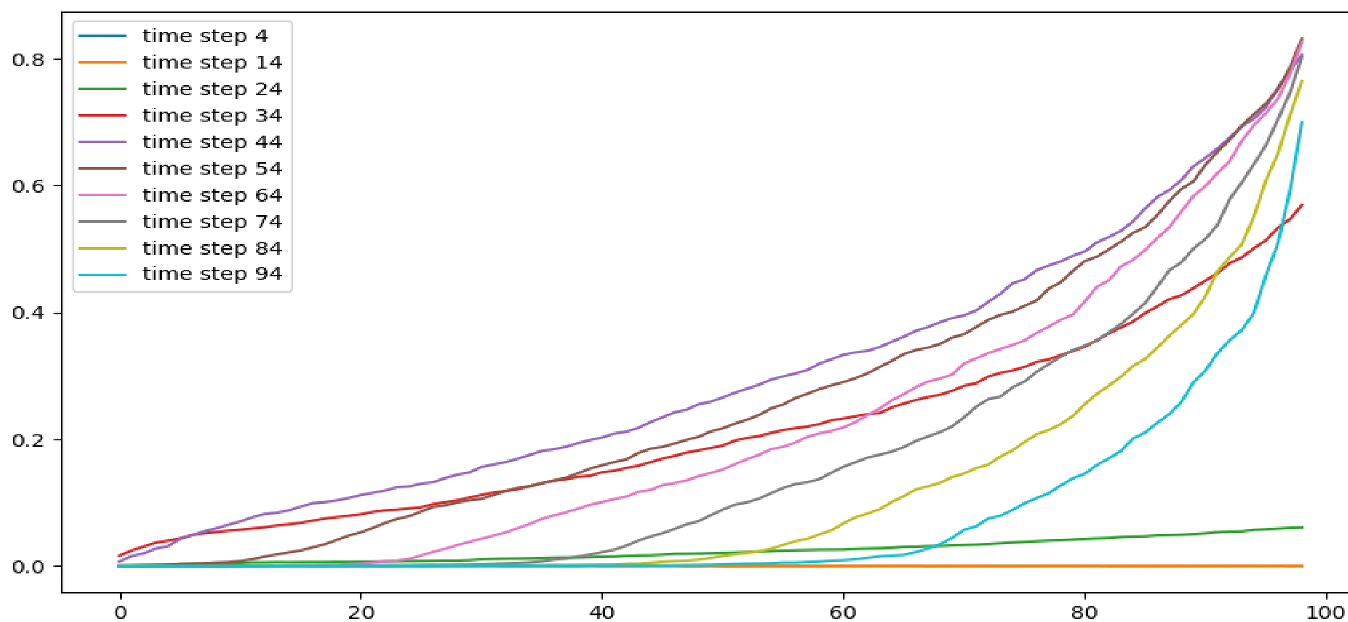






Evolution temporelle des fonctions percentiles

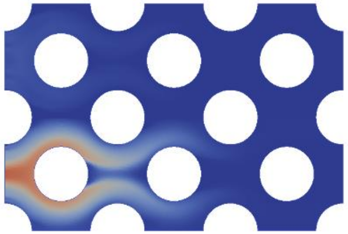




CONCLUSION

- **In-situ analysis of multi-run simulations**
- **Problems:**
  - Size
  - Complexity
- **Ubiquitous statistics**
- **Melissa**
  - In-transit
  - Iterative statistics
  - Fault –tolerant
  - Open-source : **https://melissa-sa.github.io**

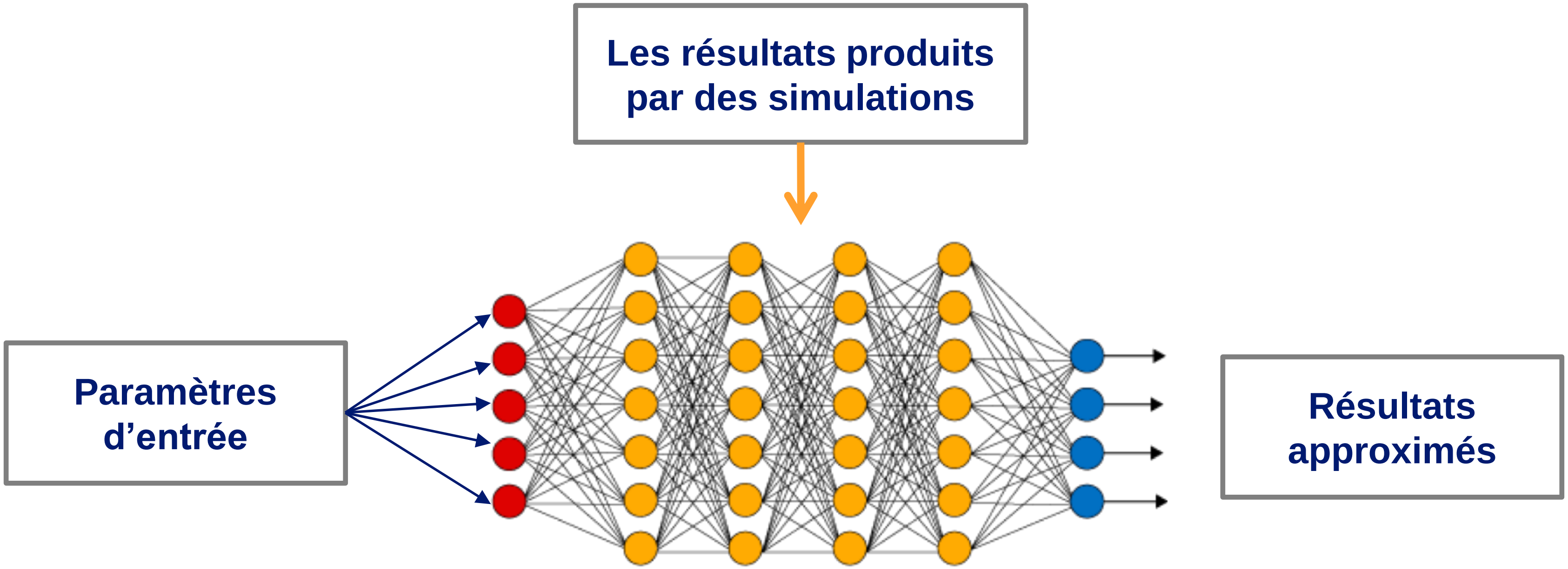
MELISSA



CONCLUSION

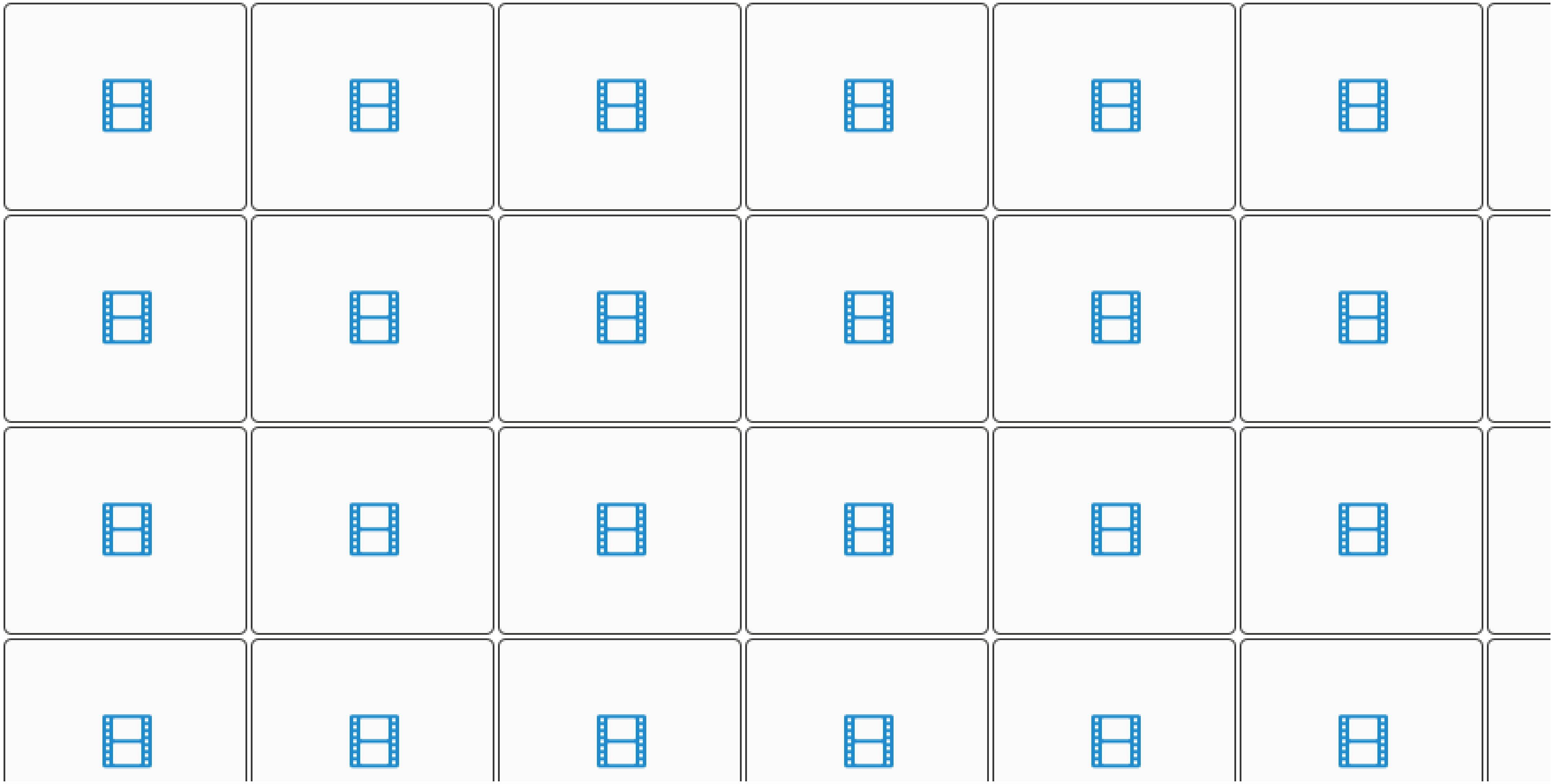
- **In-situ analysis of multi-run simulations: AVIDO**
  - A. Ribes, T. Terraz, Y. Fournier, B. Iooss, and B. Raffin. Large Scale Computation of Quantiles using Melissa. In Proceedings of Super Computing conference, Dallas, Texas USA, November 2018 (SC'18).
  - **T. Terraz, A. Ribes, Y. Fournier, B. Iooss, and B. Raffin. 2017. Melissa: Large Scale In Transit Sensitivity Analysis of Model Outputs Avoiding Intermediate Files. In Proceedings of Super Computing conference, Denver, Colorado USA, November 2017 (SC'17).**
  - A. Ribes "Computing Ubiquitous Statistics: Computational Challenges", Keynote at ISAV (In Situ Infrastructures for Enabling Extreme-scale Analysis and Visualization), November, 12th 2017. Denver, USA.
  - T. Terraz, B. Raffin, A. Ribes, Yvan Fournier. "In Situ Statistical Analysis for Parametric Studies". Proceedings of the In Situ Infrastructures for Enabling Extreme-scale Analysis and Visualization (ISAV), Salt Lake City, USA, November 2016.
- **Integration of Catalyst into SALOME-YACS**
  - A. Ribes, O. Mircescu, A. Geay and Y. Fournier. "In-situ Visualization for Computation Workflows", ISC workshop "Visualization at Scale". Frankfurt, Germany. June 2017.
  - A. Ribes, A. Bruneton. "Visualizing results in the SALOME platform for large numerical simulations: an integration of ParaView". IEEE symposium on Large Scale Data Analysis and Visualization (LDAV), Paris, France, November 2014.
- **Integration of Catalyst into Code_Saturne:**
  - A. Ribes, B. Lorendeau, J. Jomier, Y. Fournier. "In-Situ Visualization in Fluid Mechanics using Open-Source tools: integration of Catalyst into Code_Saturne". Book Chapter in Topological and Statistical Methods for Complex Data -- Tackling Large-Scale, High-Dimensional, and Multivariate Data Sets. Springer. Pages 21-37. 2015.
  - B. Lorendeau, Y. Fournier, A. Ribes. "In-Situ visualization in fluid mechanics using Catalyst: a case study for Code Saturne". IEEE symposium on Large Scale Data Analysis and Visualization (LDAV), Atlanta, USA, October 2013.

DEEP LEARNING




```
model = Sequential()
model.add(Dense(N*2, input_dim=3,activation='relu'))
model.add(Dense(N*3, kernel_initializer='normal', activation='relu')
model.add(Dense(N*4, kernel_initializer='normal', activation='relu')
model.add(Dense(N*3, kernel_initializer='normal', activation='relu')
model.add(Dense(N*2, kernel_initializer='normal', activation='relu')
model.add(Dense(N, kernel_initializer='normal'))
model.compile(optimizer='adam',loss='mse', metrics=['mse'])
```

MÉTA-MODÈLE D’UNE SIMULATION DE CODE_SATURNE



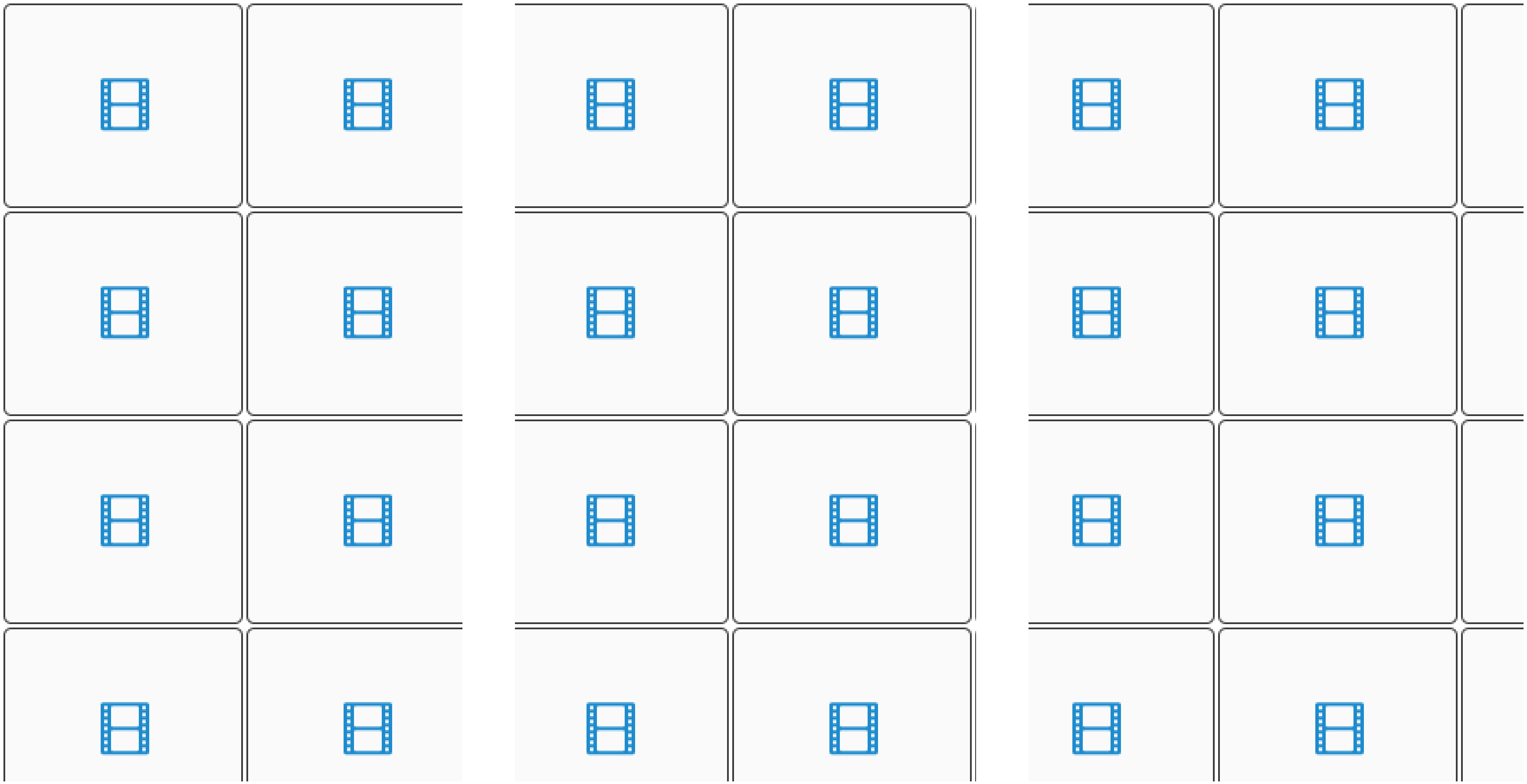
Résultat d’un simulation de Code_Saturne

Résultat donné par réseau des neurones entrainé avec 200 simulations

Résultat donné par réseau des neurones entrainé avec 10000 simulations

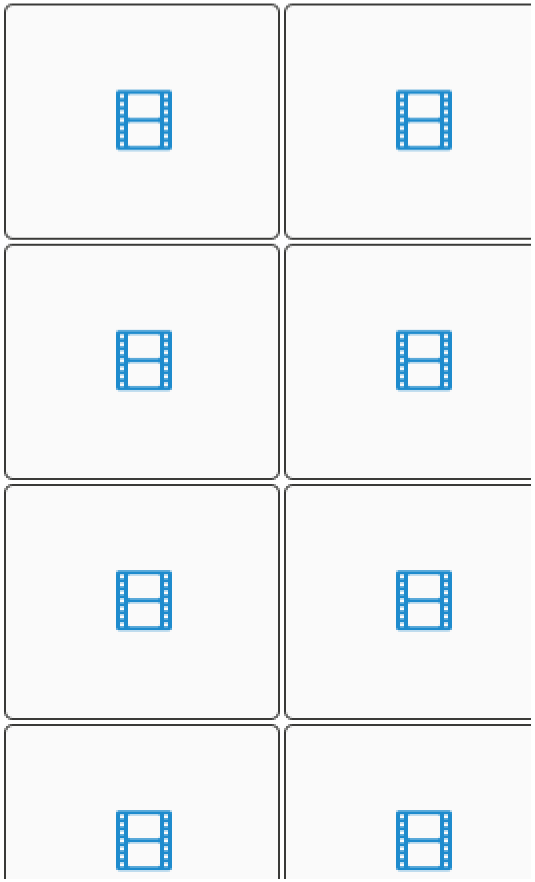
Résultat donné par réseau des neurones entrainé avec 50000 simulations

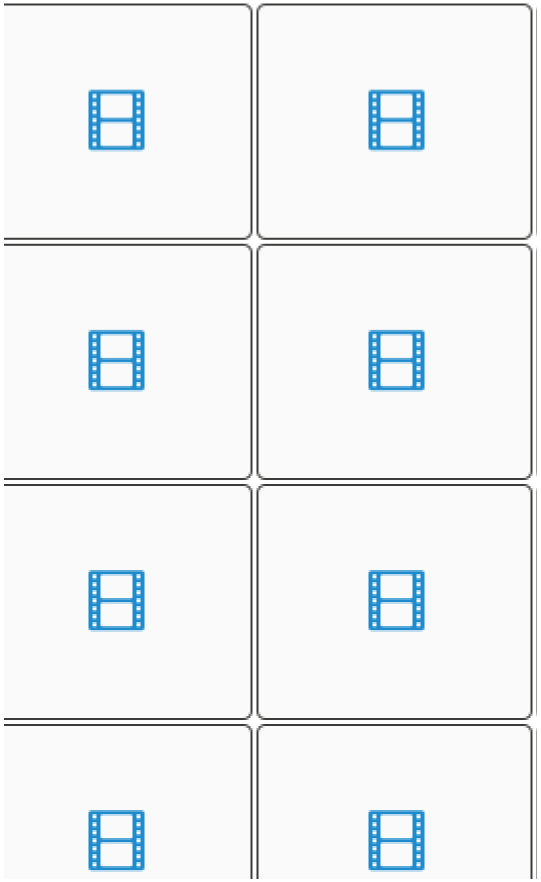
MÉTA-MODÈLE D’UNE SIMULATION DE CODE_SATURNE

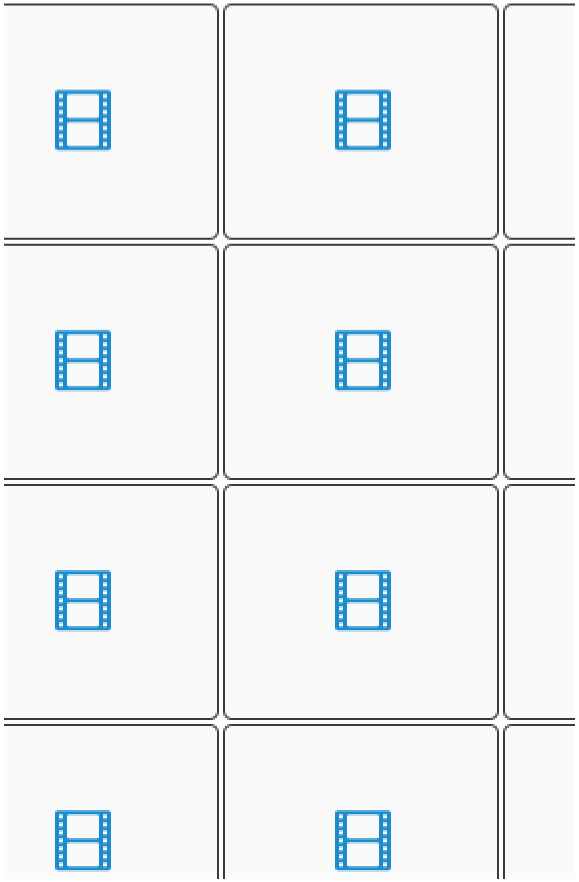


La différence en valeur absolue entre les résultats donnés par la simulation et les méta-modèles

MÉTA-MODÈLE D'UNE SIMULATION DE CODE_SATURNE



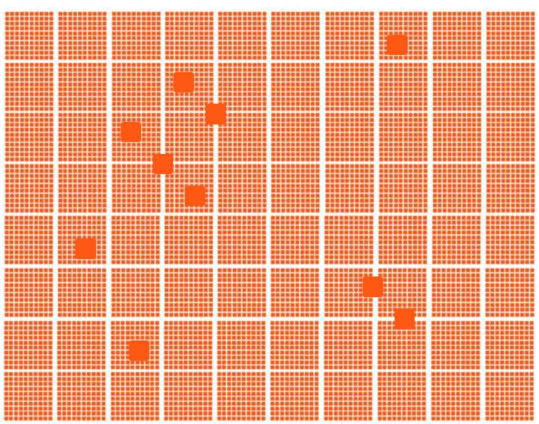




La moyenne des différences des résultats en valeur absolue entre les simulations de Code_Saturne et les méta-modèles avec 1000 sets de paramètres différents.

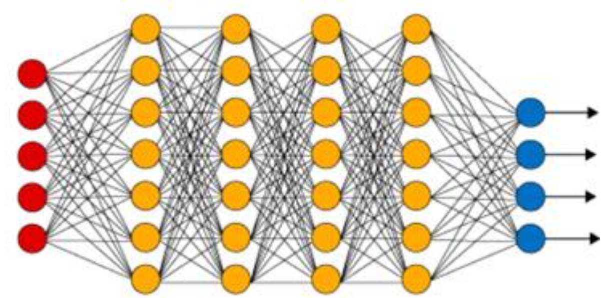
DEEP-MELISSA

- **Objectifs**
  - Méta-modélisation
  - Détections de points d'intérêt
  - Visualisation rapide



48To

MELISSA



TensorFlow

Keras

CONCLUSION DEEP-MELISSA

Utilité pour un ingénieur de simulation

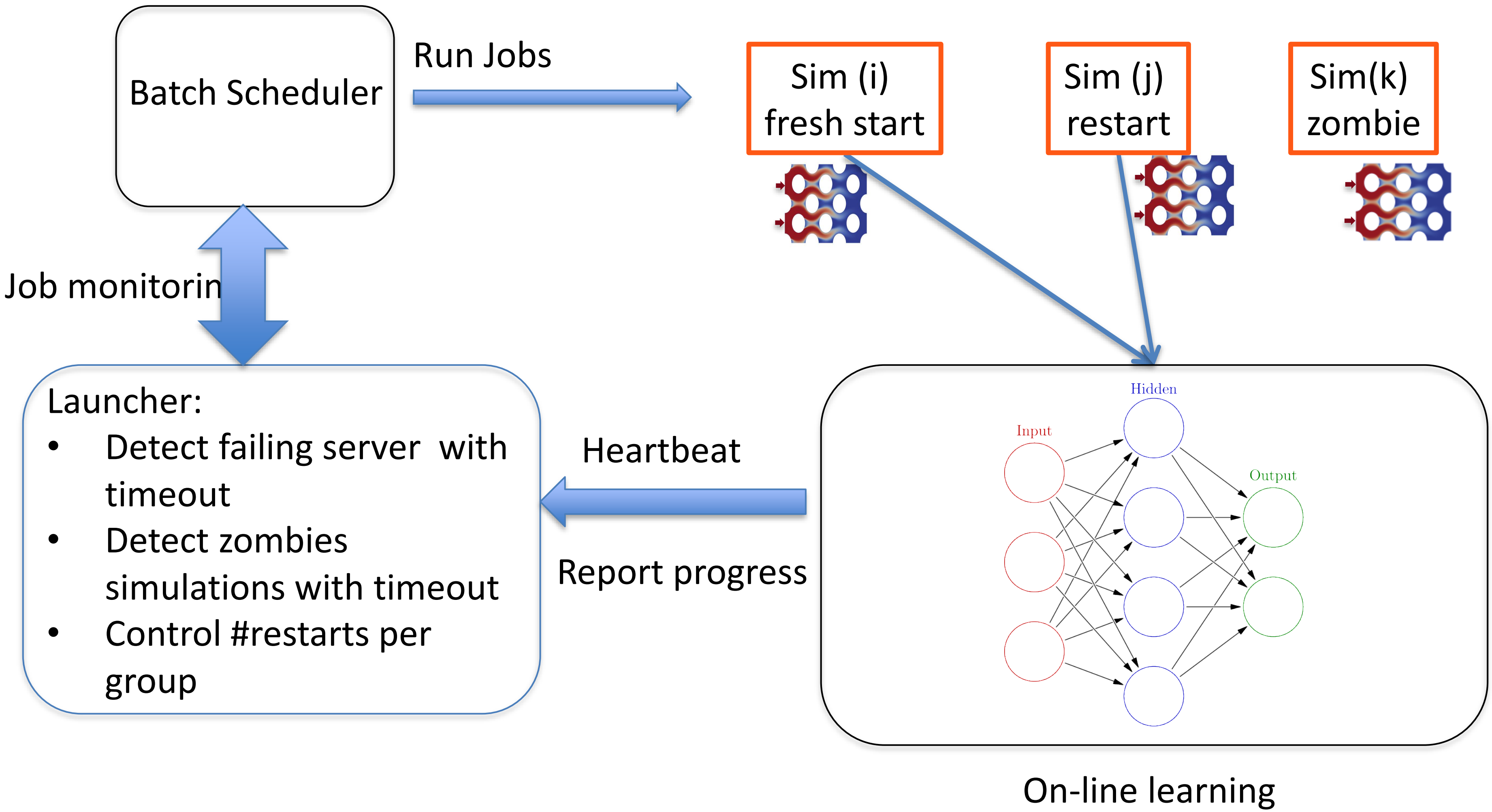
- **Génération de méta-modèles à base d'apprentissage**
  - Méta-modelés "ubiquitaires" (dans toutes les mailles et pour tous les pas de temps), ce type de méta-model n’existe pas encore
  - Visualisation et analyse des études paramétriques via des méta-modèles qui s'exécutent en temps réel sur un PC (avec une interface ParaView)
- **Réalisation de jumeaux numériques**
  - Méta-modeles deep-learning couplés à un IHM de visualisation remplissent cette fonction.
- **Accélération du cycle de l’étude par simulation via des méthodes ultra-rapides de post-traitement**
  - visualisation et analyse de résultats de simulations pour l'ingénieur d'études
  - deep-learning sur GPU : extraction de features, tracking de tourbillons ou réalisation de post-traitements lourds en temps réel.
- **Applications in-situ et in-transit**
  - permettre à l'ingénieur de gérer facilement les grands volumes de données issues des études numériques.

Préparation d’une thèse en collaboration EDF-INRIA (Bruno RAFFIN)

Learning Meta Models

Control

Computations



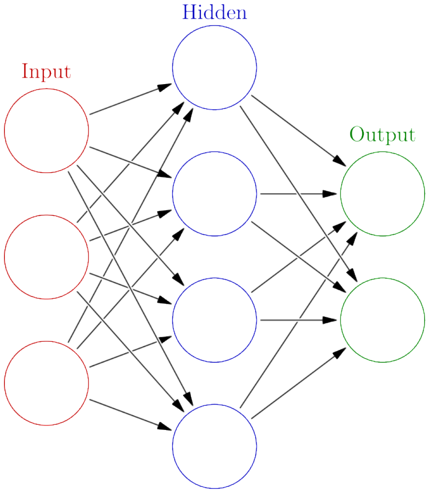
Learning Meta Models

Simulation:

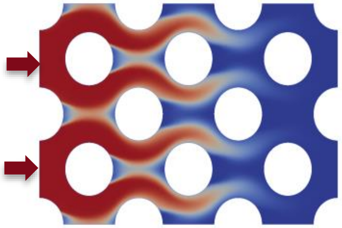
- Input parameters: temperatures, injector size, mesh,…
  - Ensemble runs: sample the parameter space
- Output: mesh + values associated to mesh cells at each time step

Neural Network:

Input parameters
+
time step t

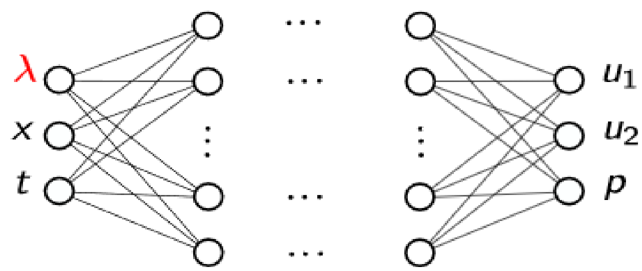


Values of each mesh cell
at time step t



Neural Architecture

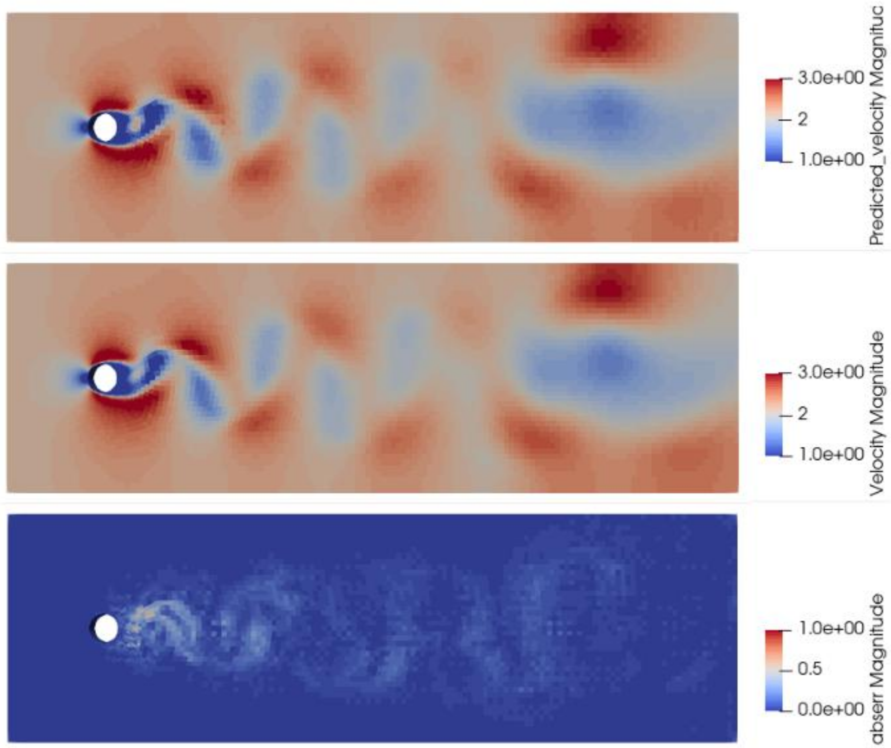
- Data-free PINN (Physics Informed NN)



$$l(\theta) = \underbrace{\frac{1}{N_c}\sum_{i=1} \|\Phi_\theta(x_c^i, t_c^i) - \Phi_{sim}(x_c^i, t_c^i)\|^2}_{donnees} + \underbrace{\frac{1}{N_f}\sum_{i=1} \|\mathcal{D}(\Phi_\theta(x_f^i, t_f^i))\|^2}_{EDP}$$

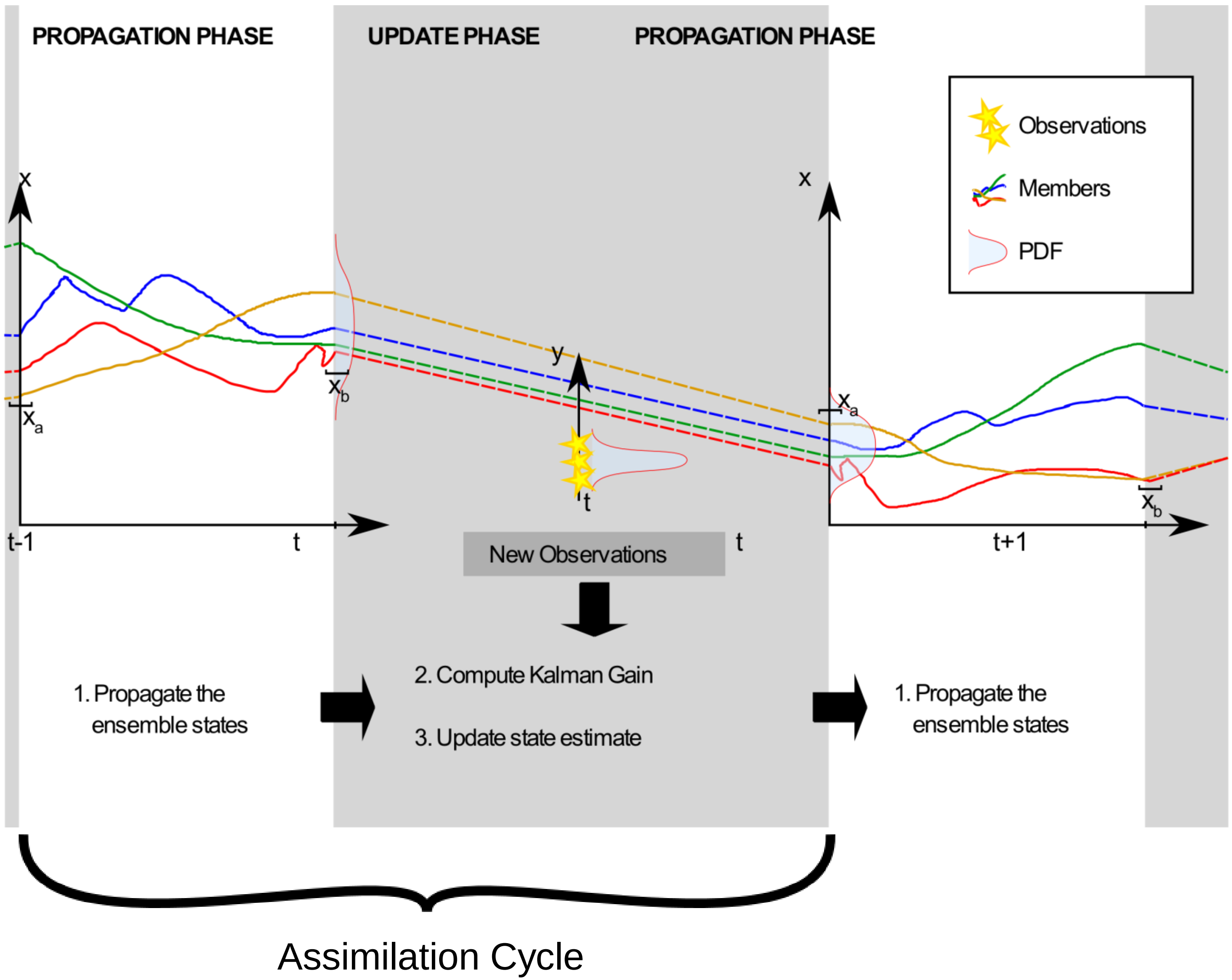
- Graph Neural Network (GNN):
  - Predict acceleration at t+1 from velocity at t on every cell of the mesh
  - Support irregular meshes

Early Results

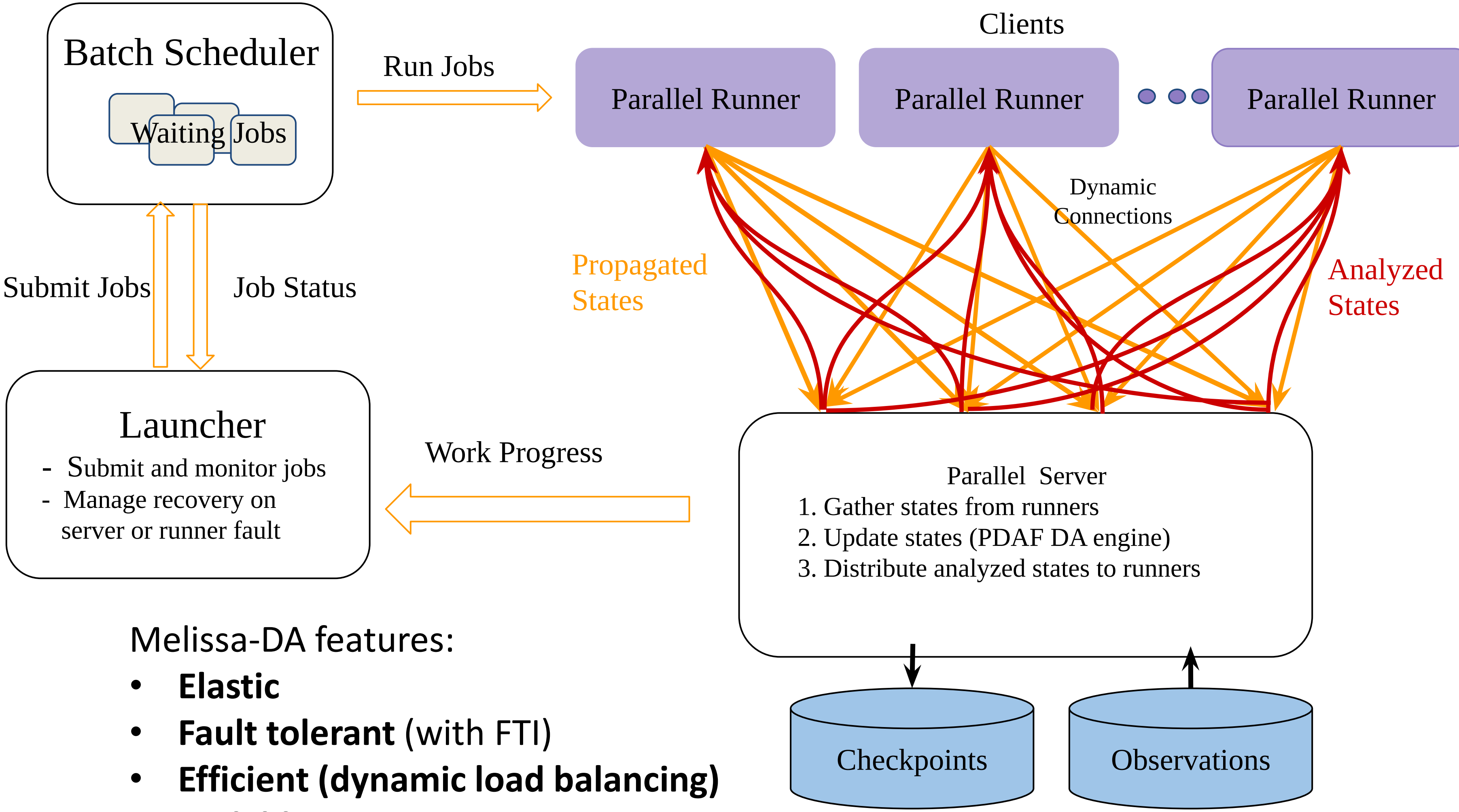


https://www.youtube.com/watch?v=gPH8ioG88R8

Data Assimilation: the Ensemble Kalman Filter (EnKF)



Melissa for Data Assimilation

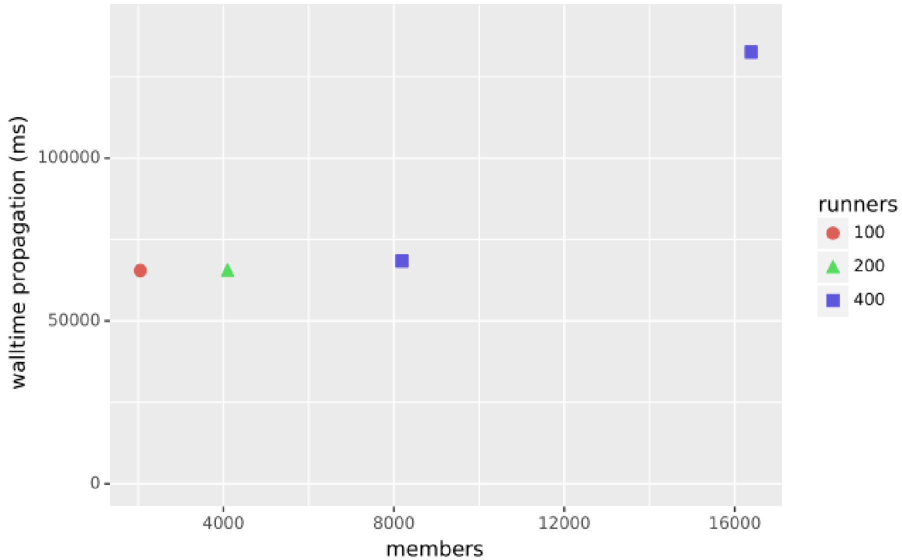


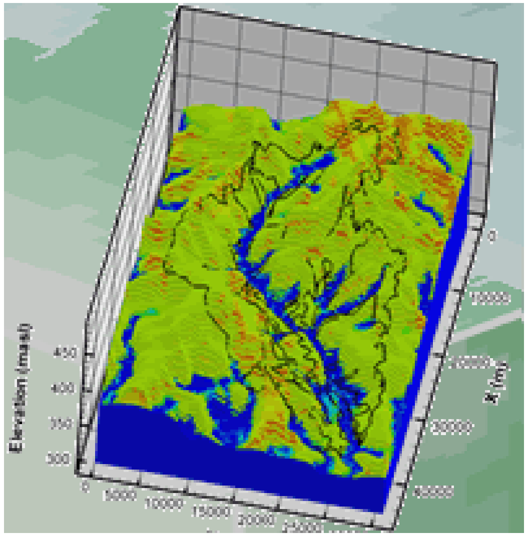
Melissa-DA features:

- **Elastic**
- **Fault tolerant** (with FTI)
- **Efficient (dynamic load balancing)**
- **Scalable**
- **Modular**

Early Results

90% efficiency (update phase)

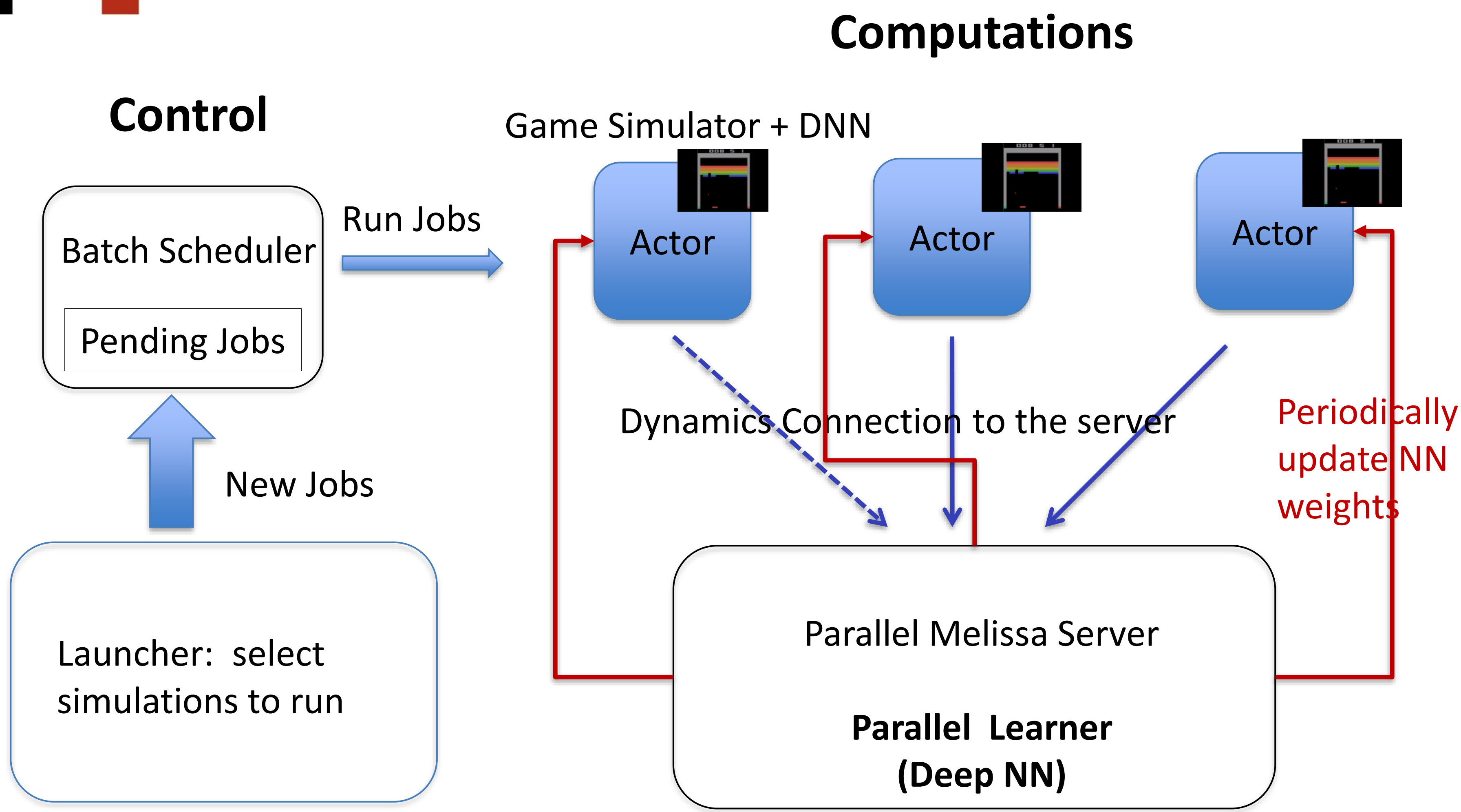




ParFlow Hydrologic Model:
4M grid cells, 25 obs
(800 m res.)

- Up to **16384 members distributed on up to 400 runners running with 16 960 cores**
- **2.9 TiB of Data transfer** between runners and server per assimilation cycle

Melissa for DRL at Scale



Join work between the DataMove
And Sequel teams

Goal: to go beyond AlphaGoZero