

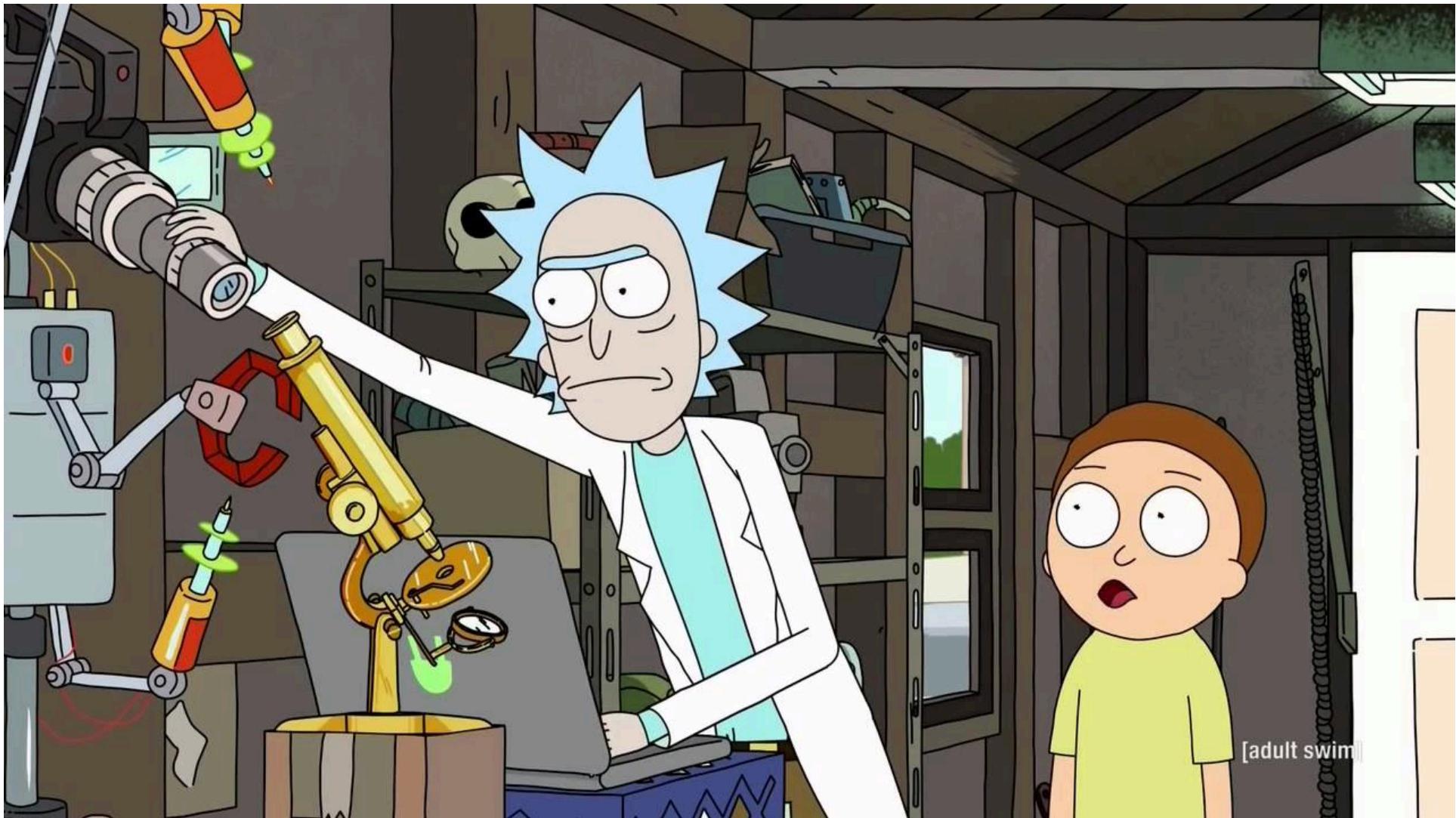
Utilisation de GRIF pour la modélisation des systèmes complexes

30 juin 2025

GRIF a 20 ans, LPNHE, Paris

Romain Reuillon, Mathieu Leclaire

L'expérimentation sur les modèles



Modélisation des systèmes complexes

Quoi ?

- multi-niveau : micro / macro
- dynamiques collectives
- intègrent souvent une composante sociale

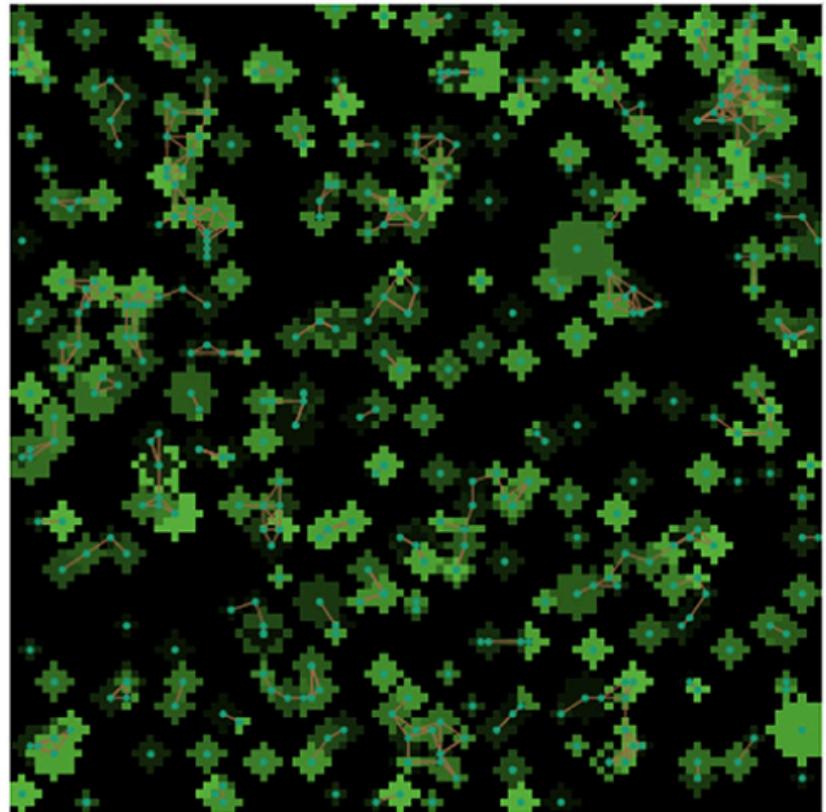
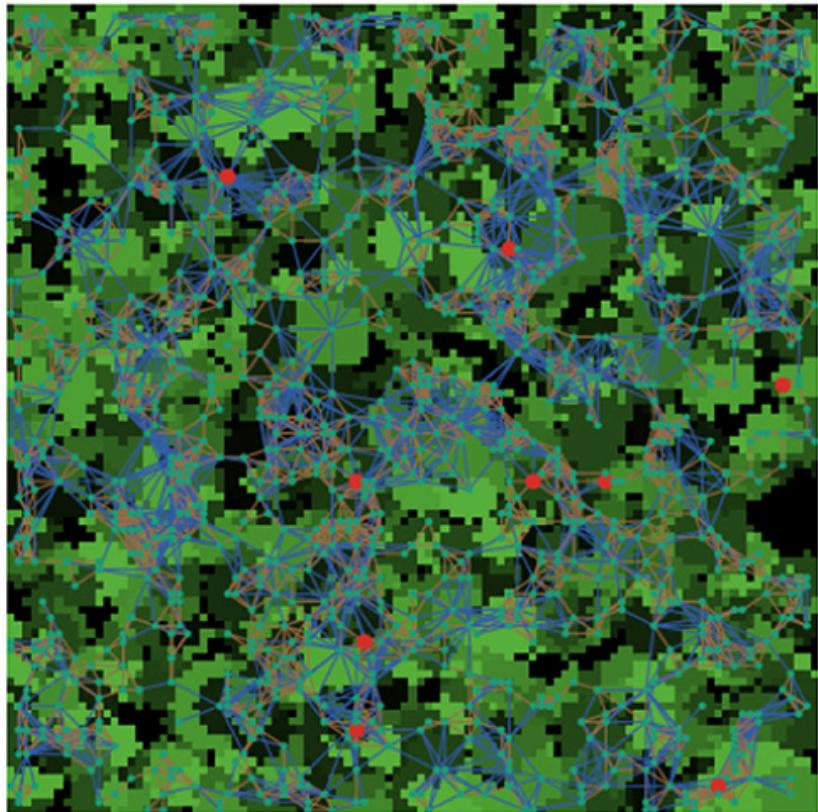
Pourquoi ?

- formaliser des théories
- évaluer des théories
- aider la décision

Quels exemple de disciplines scientifiques

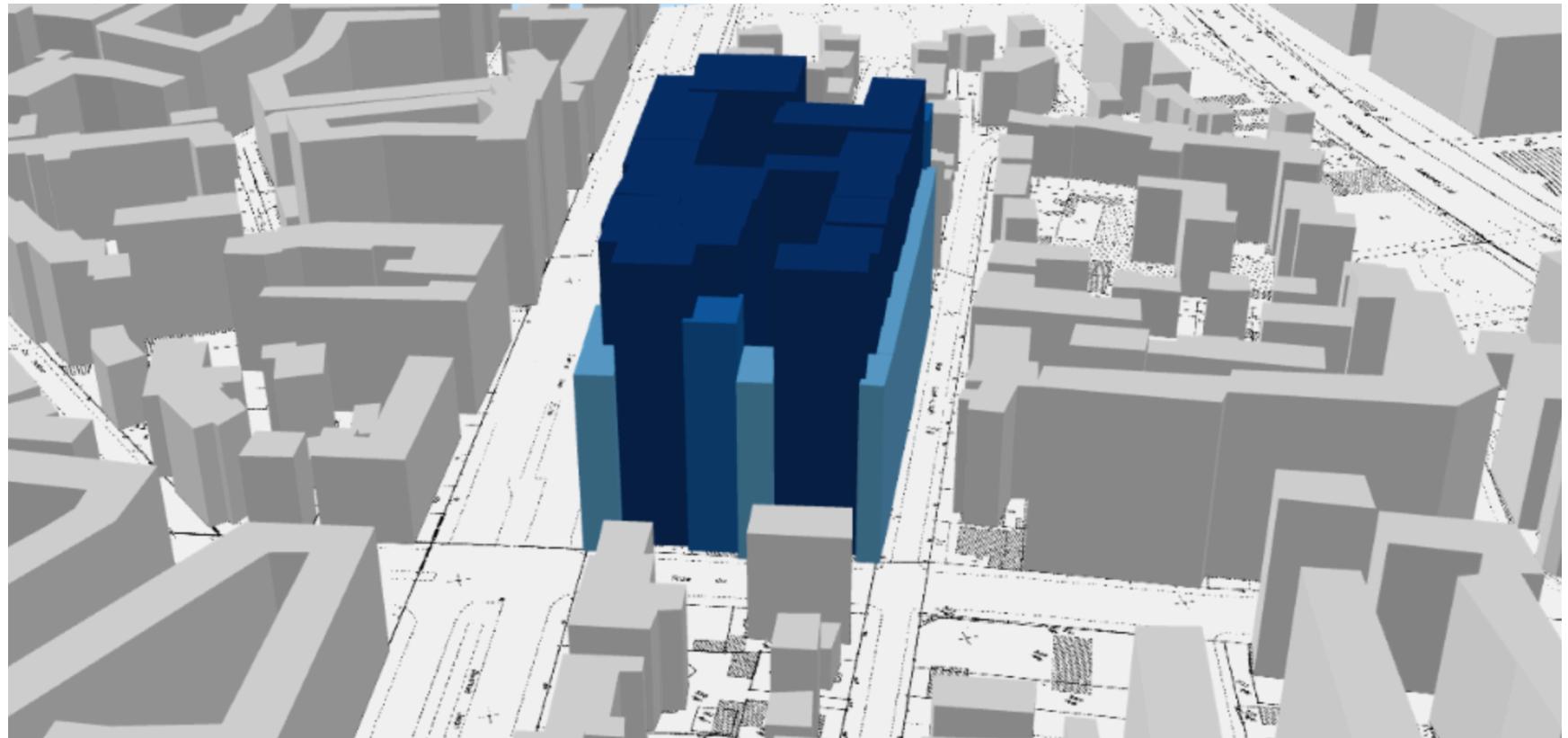
- Géographie
- Urbanisme
- Écologie
- Archéologie
- Agronomie
- Épidémiologie
- ...

BORNEO



Quel est l'impact de la déforestation sur le déplacement des grands singes ?

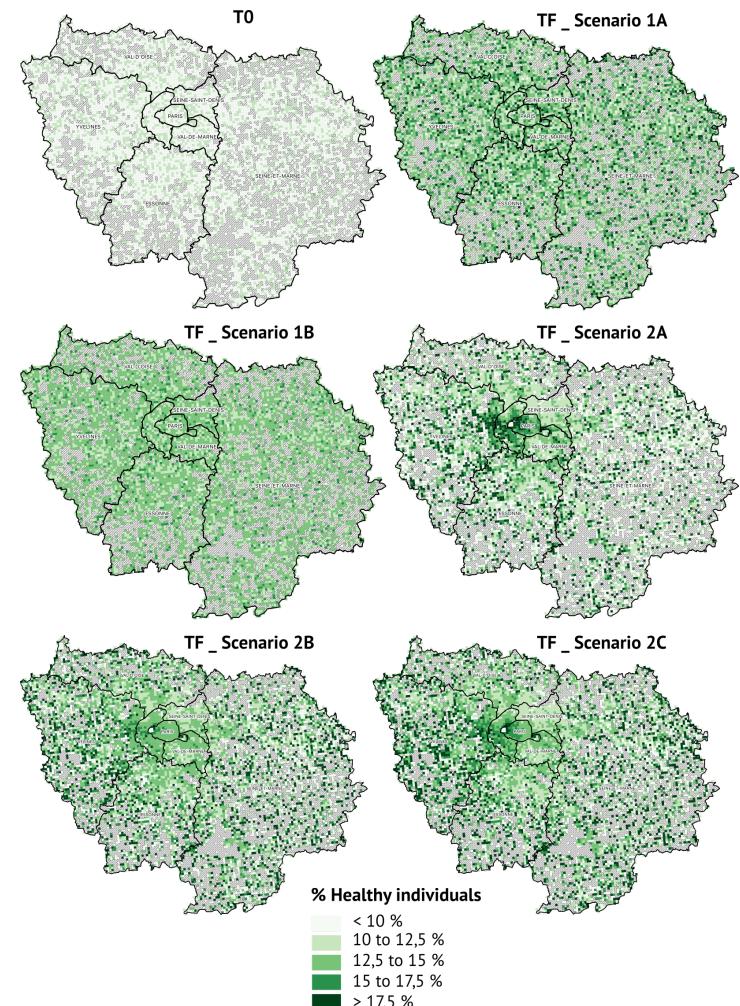
SimPLU3D



Quelles enveloppes bâties sous contrainte de la réglementation?

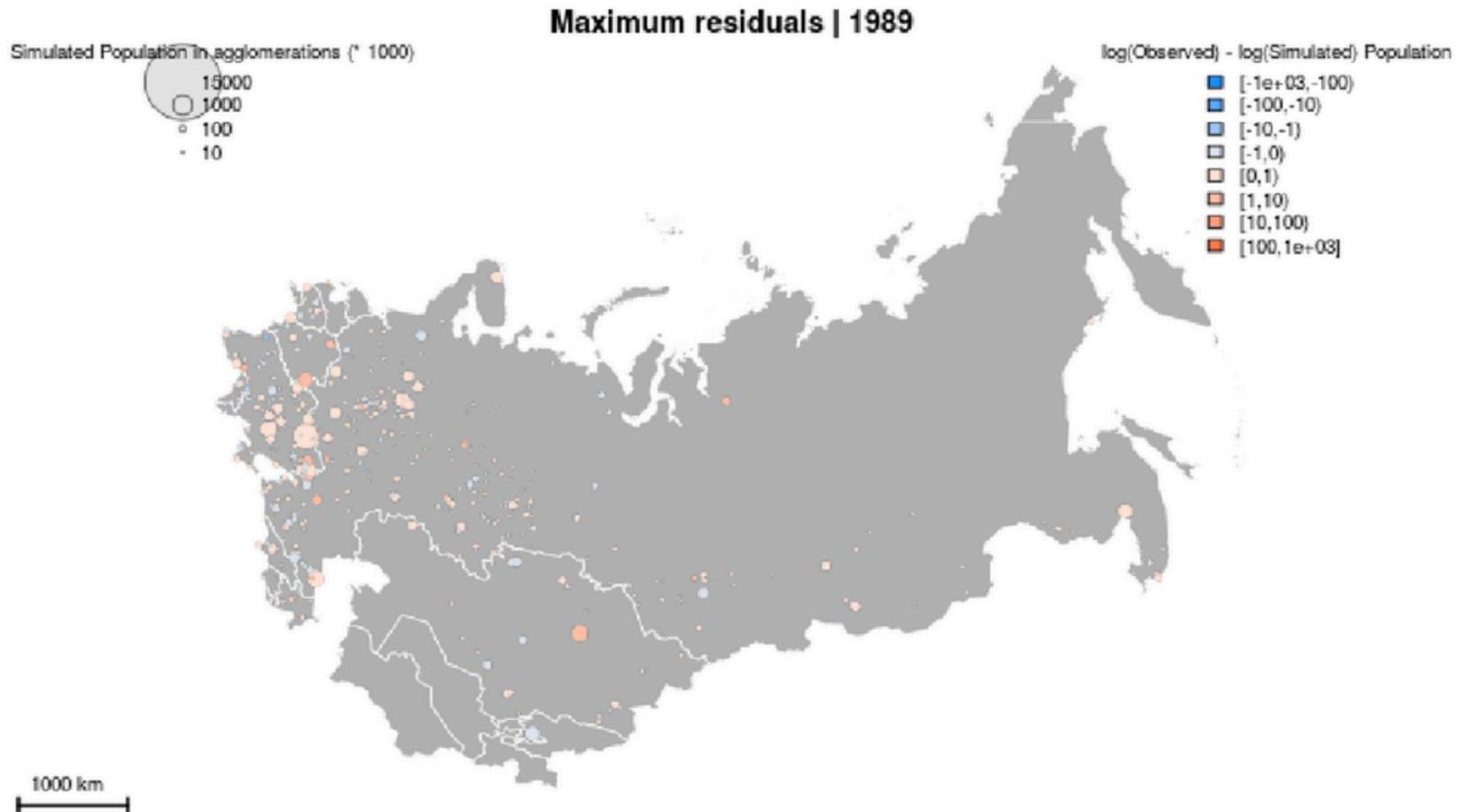
5aDay

Quel impact des déplacements quotidiens sur les habitudes de santé ?



Marius

Urban Transition model 1959-1989 ($|residuals| \geq 0.3$)



Comment la chute du mur de Berlin a-t-elle modifiée l'évolution du système de villes Russe ?

L'évaluation

Un défi:

- Peu de données
- Beaucoup d'incertitudes

Il faut tout de même évaluer leur *pertience/validité* !

=> Trouver des régularités malgré les incertitudes

Méthodes d'évaluation

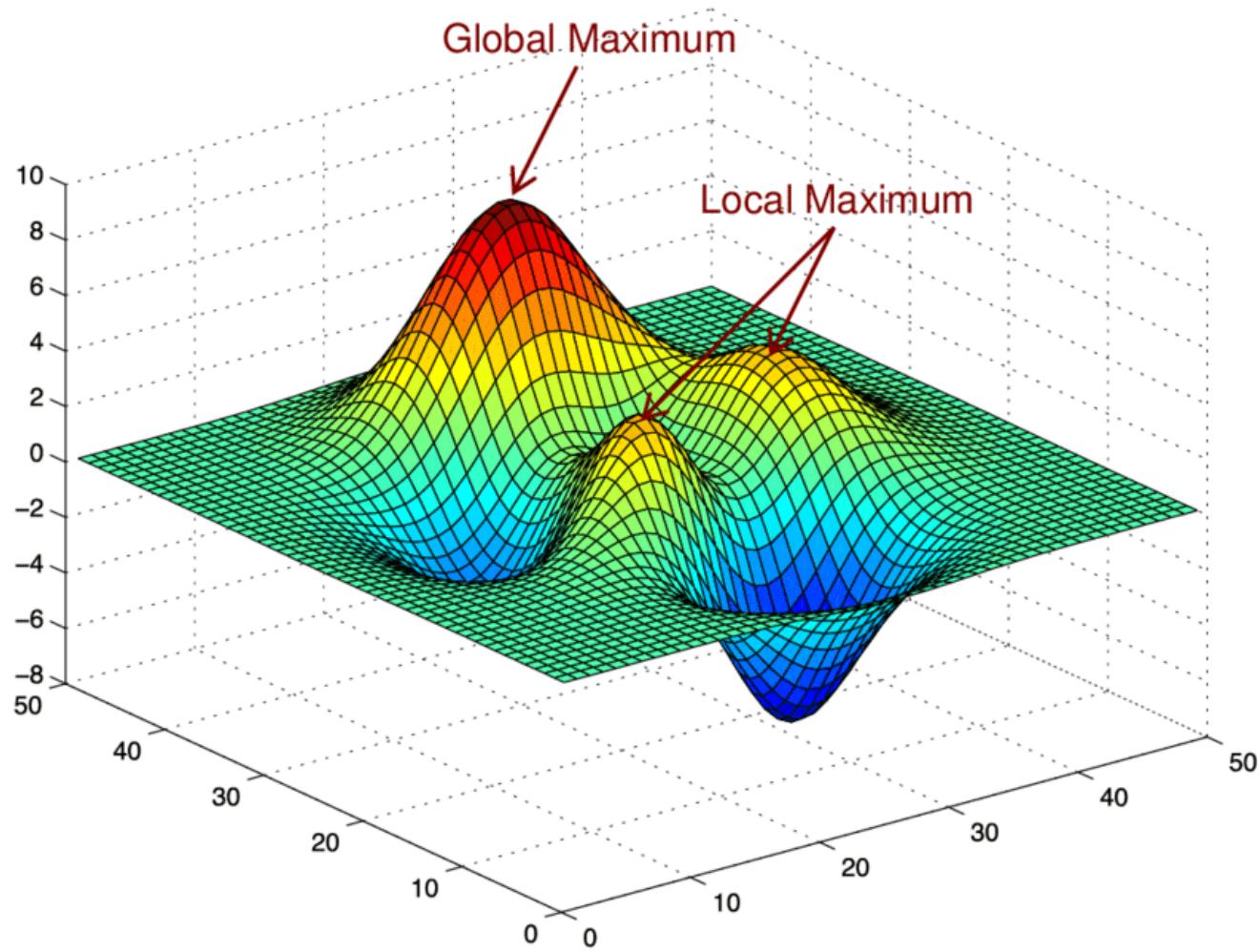
Utilisation d'algorithmes de recherche itératifs dans l'espace des inconnues du modèles:

- Recherche d'optimums
- Recherche de diversité
- Algorithmes hybrides

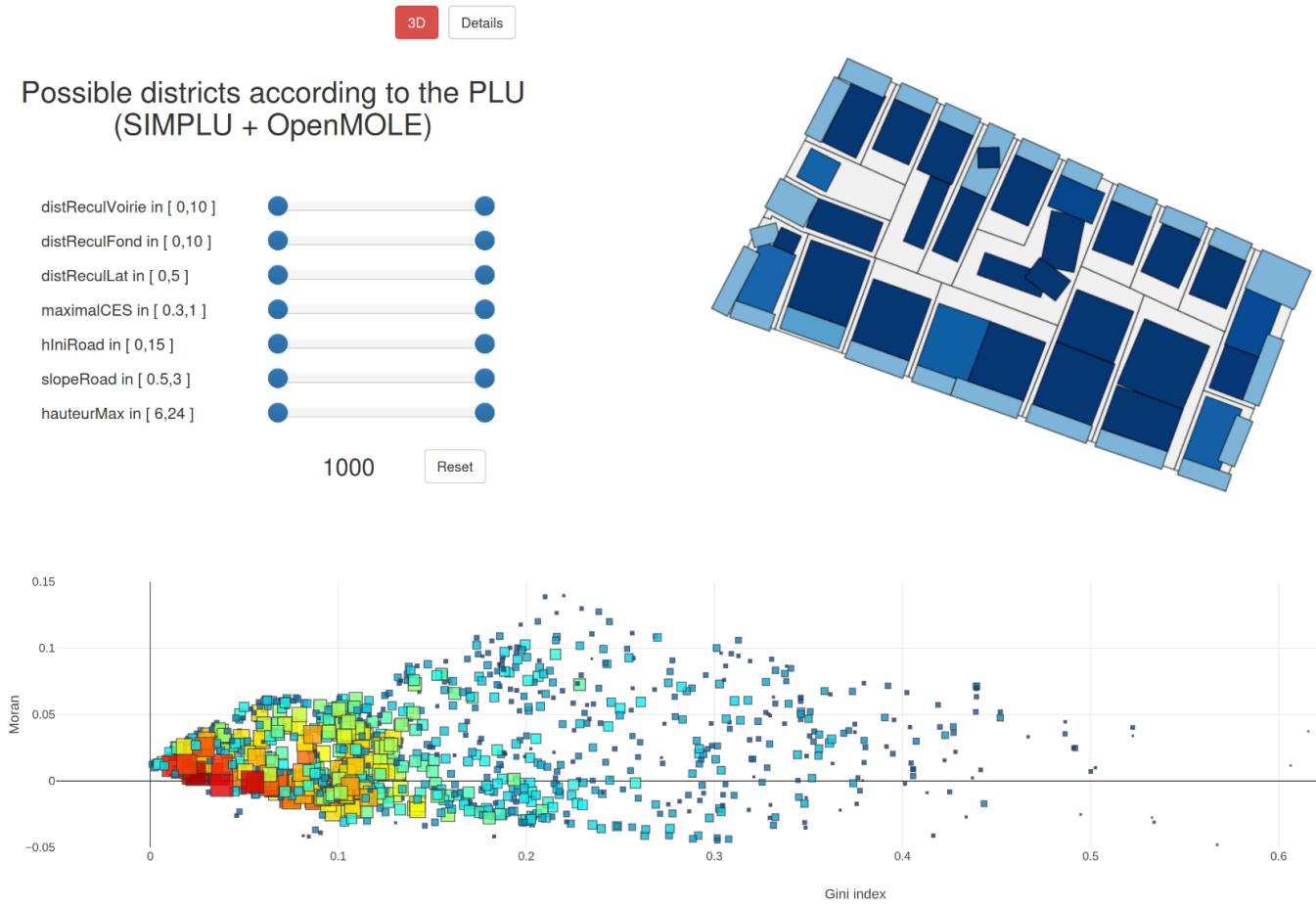
Algorithmes disponibles dans OpenMOLE:

NSGA2, ABC, Profile, PSE, OSE, HDOSE, PPSE...

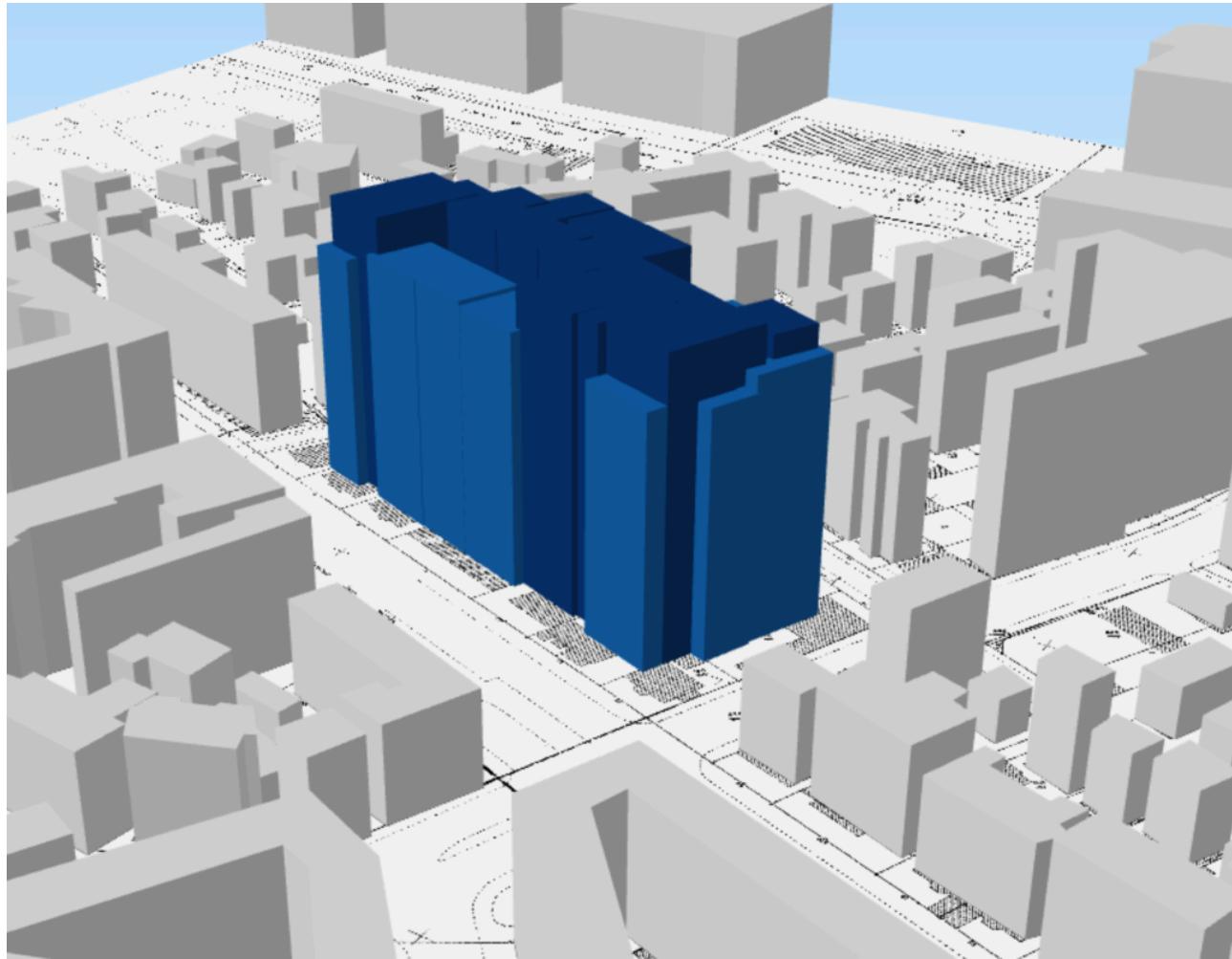
Exemple calibrer avec NSGA2



Exemple recherche de motifs avec PSE



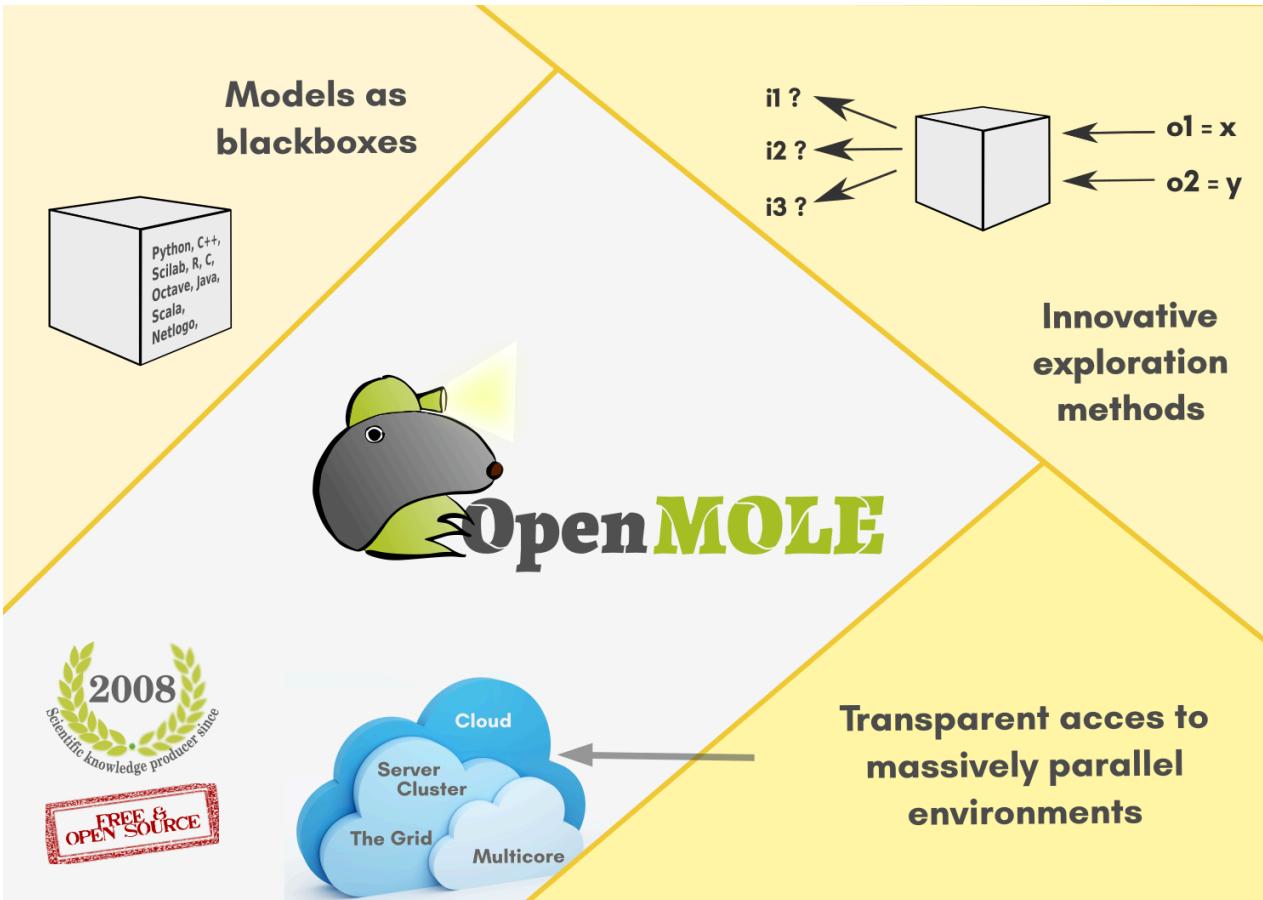
Exemple PSE



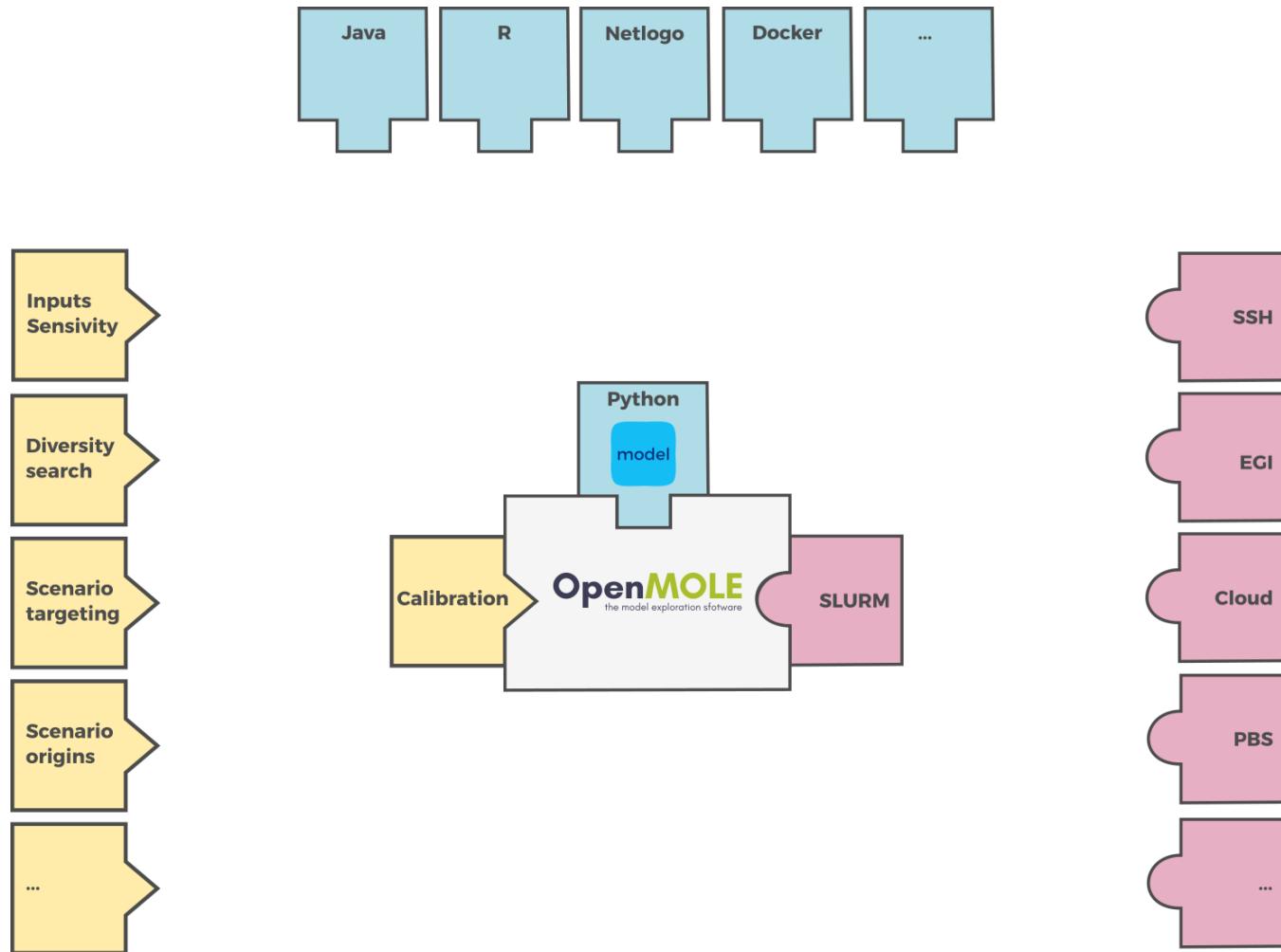
Qui ?

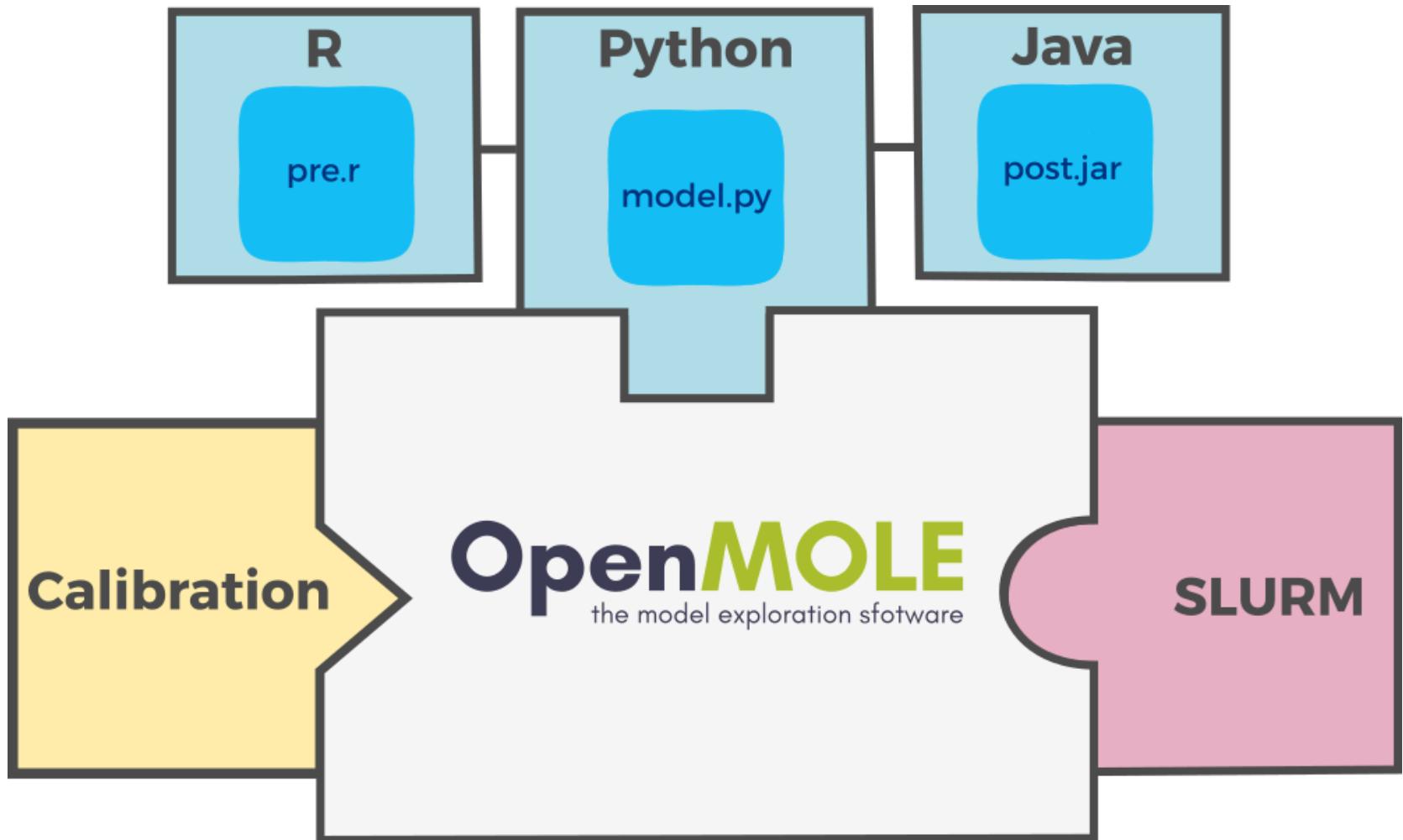
- Doctorant, post-doctorant, chercheurs
- Dans des laboratoires sans support technique
- Sachant programmer dans un environnement de modélisation (ex R, NetLogo,...)
- Pas ou peu de connaissances systèmes
- Pas de notion de calcul distribué
- Des pratiques et outils de modélisation très hétérogènes
- Peu de connaissance sur comment évaluer leurs modèles

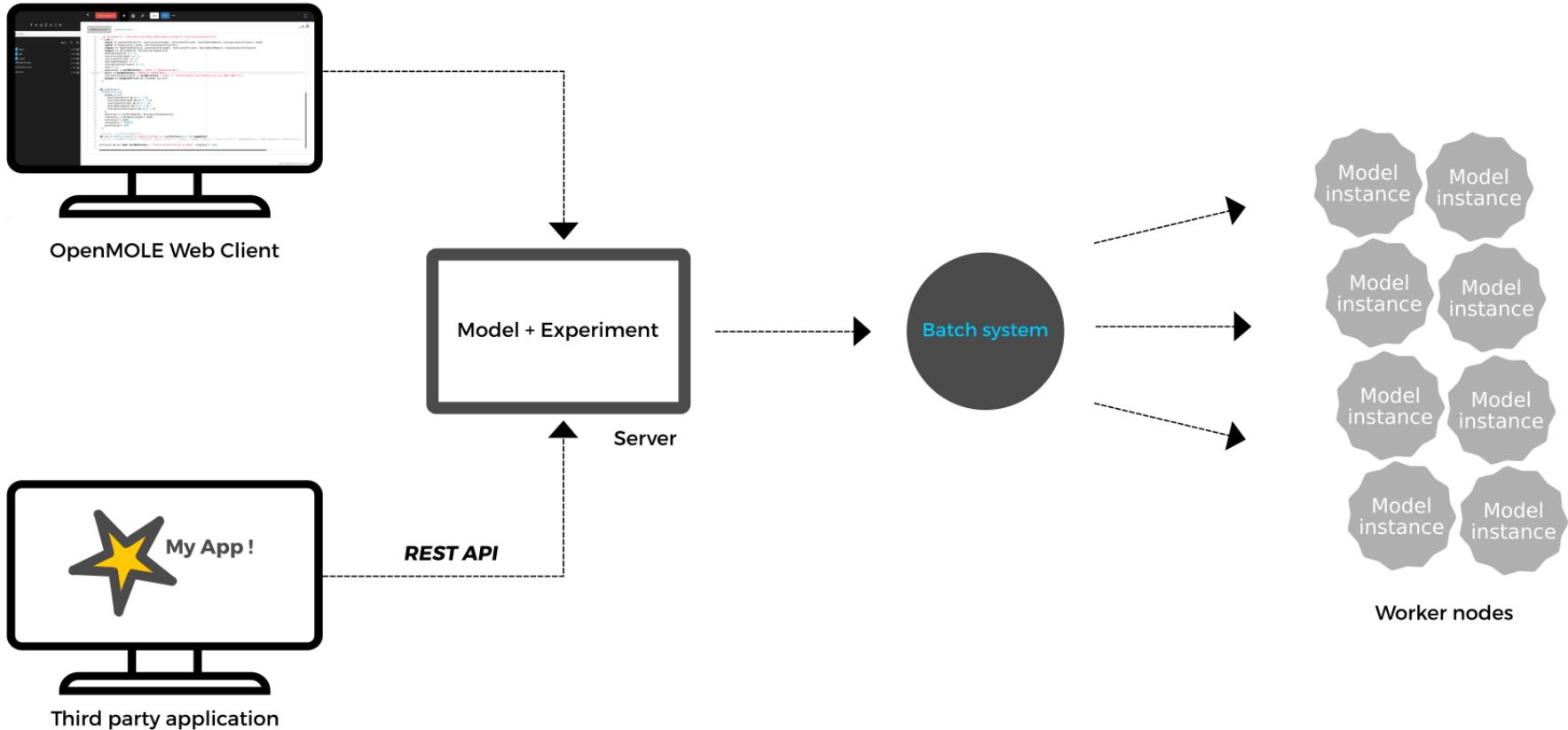
Notre proposition



OpenMOLE: une plateforme methodologique et technique







Délegation du calcul

Chaque utilisateur a son/ses propre(s):

- code / plateforme
- ses données
- infra(s) de calcul

=> Calculs dans des taches basées sur singularity, dont l'execution est relocalisables

Une partie du workflow est déléguée, à une ou plusieurs infra

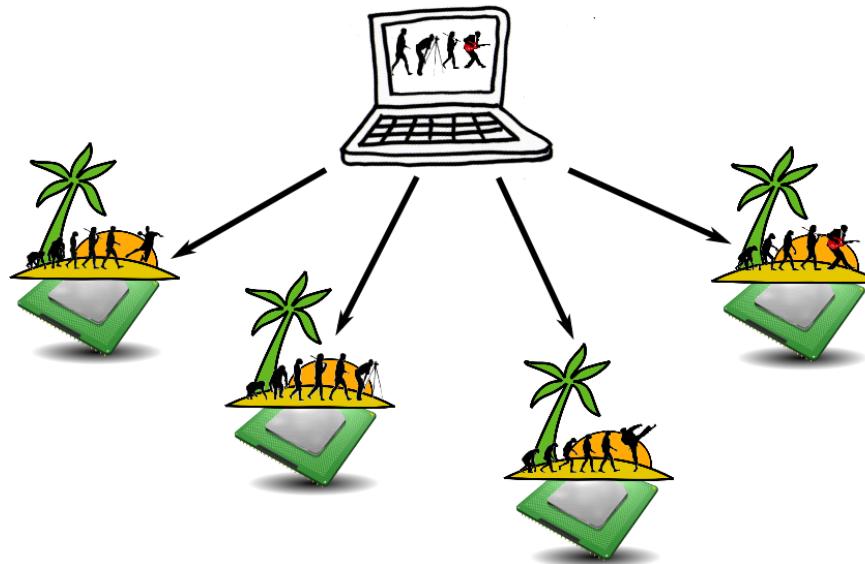
Soumission de Jobs

- replication du runtime OpenMOLE (avec cache)
- replication des données/programmes/containers (avec cache)
- soumission d'un job:
 - execute une partie du workflow
 - copie le résultat sur un stockage
- recréation du résultat
- réinjections du résultat dans le workflow

Parallélisme du point de vue des méthodes

Les algorithmes d'évaluation d'OpenMOLE comportent:

- une partie itérative
- une partie permettant de distribuer efficacement les simulations



Échelle d'une experience de validation

- execute 100 à 500 mille jobs
- un job dure entre 2 minutes et 1 heure
- 1000 à 2000 en parallèle à chaque instant
- déplace 10 à 100 de GO de données
- requiert environ 5 années/CPU

Les infrastructures



Les quelles?

- GRIF, Grille et Cloud
- Mésocentres (notamment le CRIANN, MESO@LR)
- MESONET
- Clusters / Serveurs dans les labo

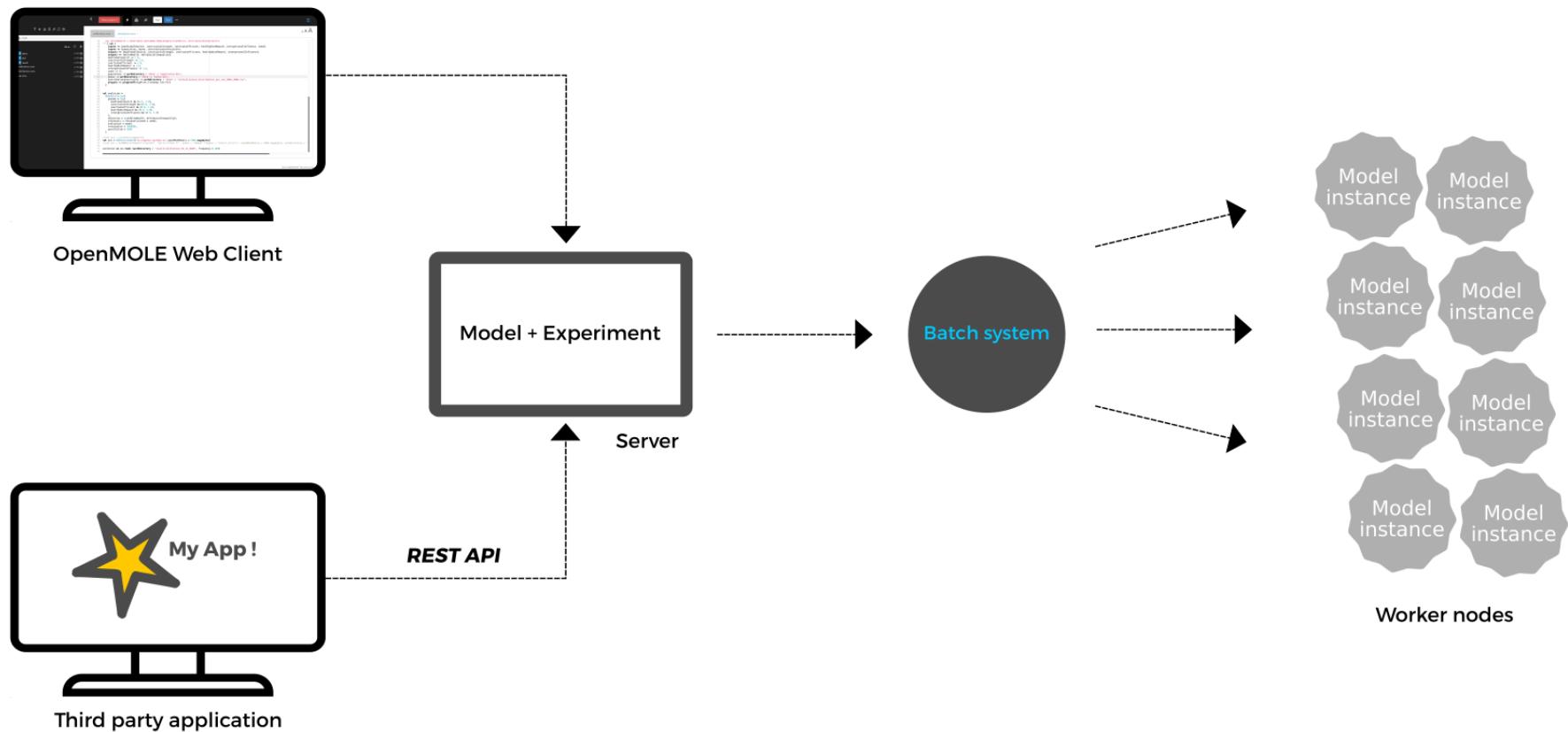
Cloud @ GRIF

Utilisation depuis 2011:

- StratusLAB
- FedCloud
- VirtualData

Depuis 2016: utilisation en production pour l'hébergement de services liés à OpenMOLE (site web, forge, chat, backup, machine de calcul, CI, ...)

Depuis 2024: OpenMOLE dans le Cloud



Cloud OpenMOLE

The screenshot shows the Cloud OpenMOLE dashboard. At the top, there are three buttons: Settings, Info, and Admin. The Admin button is highlighted. Below these are resource statistics: Role (Admin), OpenMOLE version (latest), OpenMOLE memory (2.00 GB), Max Memory (4.00 GB), Max CPU (4), and Storage (24.98/29 GB). A 'Stop OpenMOLE' toggle switch is shown as off. The status is 'Running' with a green dot. The date and time are listed as Tue, 23 Jun 2025 16:02:29. A 'Go to OpenMOLE' button is at the bottom.

The screenshot shows the OpenMOLE graphical interface. At the top, there are buttons for New project, settings, and help. The main area has tabs for Hello World in Python / +, RUN, and CHECK. On the left, there's a tree view with nodes: result, hello.py, and python.oms. The right side is a code editor with the following Scala code:

```
1 // Declare the variable
2 val arg = Val[Int]
3 val output = Val[File]
4
5 // python task
6 val pythonTask =
7   PythonTask(workDirectory / "hello.py") set (
8     inputs += arg.mapred,
9     outputs += arg,
10    outputs += output mapped "output.txt",
11  )
12
13 val copy = CopyFileHook(output, workDirectory / "result/hello$arg.txt")
14 val env = LocalEnvironment(2)
15
16 DirectSampling(
17   evaluation = pythonTask,
18   sampling = arg in (0 to 10)
19 ) hook copy on env
20
```

Construction des expériences, tests, orchestration des calcul massifs

Cloud OpenMOLE

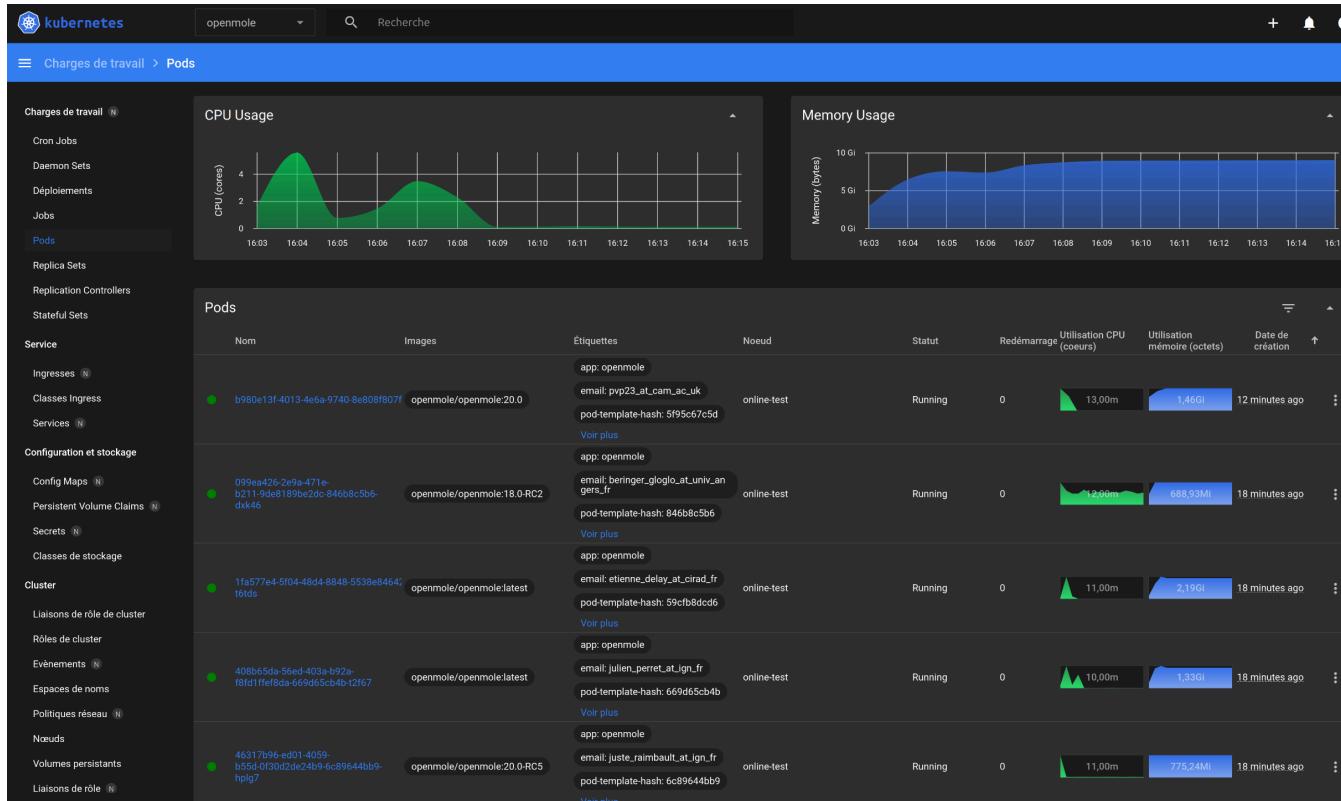
Construit avec:

- K3S
- Longhorn
- Minio (backup)
- Scala / Scala.js

Facilement déployable: <https://github.com/openmole/connect>

Architecture K3S

1 pod central, qui lance 1 pods OpenMOLE par utilisateur actifs



Création/extinction automatique des pods, 1 volume par utilisateur, utilisation d'ephemeral storage

Grille @ GRIF

ISC-PIF achat de machine de grille en 2007 et en 2017

Objectifs:

- Construire un environnement de calcul à grande échelle pour l'étude des systèmes complexes
- Mutualiser des ressources de calcul des laboratoires travaillant sur les systèmes complexes

Historique d'utilisation

Début d'utilisation en 2008, avec 2 projets: OpenMOLE, Bioemergence

Seul OpenMOLE a utilisé la grille à grande échelle:

- Production de résultats en 2010
- Explorations massives en 2013

Publication en 2015:

The screenshot shows a journal article page. At the top, it displays the Impact Factor: 3.1 / 5-Year Impact Factor: 3.8, a Journal Homepage button, and a Submit button. Below this, there's a lock icon indicating Restricted access, followed by Research article and First published online January 1, 2015. The title of the article is "Half a Billion Simulations: Evolutionary Algorithms and Distributed Computing for Calibrating the Simpoplocal Geographical Model". It lists authors: Clara Schmitt, Sébastien Rey-Coyrehourcq, and Denise Pumain, with a link to view all authors and affiliations. Below the title, it says Volume 42, Issue 2 | <https://doi.org/10.1068/b130064p>. A horizontal menu bar includes Contents, Get access, Cite article, Share options, Information, rights and permissions, and Metrics and citations. The Abstract section begins with: "Multiagent geographical models integrate very large numbers of spatial interactions. In order to validate these models a large amount of computing is necessary for their simulation and calibration. Here a new data-processing chain, including an automated calibration procedure, is tested on a computational grid using evolutionary algorithms. This is applied for the first time to a geographical model designed to simulate". To the right, there's a sidebar titled "Similar articles:" with a lock icon and the text "Restricted access" followed by a link to "Calibration of Micro Simulation Models".

Bilan sur l'utilisation de la Grille

La grille a permis la production de résultats impossible autrement, pour quelques dizaines d'utilisateurs

=> *malgré nos efforts, pas d'adoption massive*

Freins

Freins à l'utilisation:

- difficulté d'obtenir des certificats
- haute probabilité de pannes (DIRAC/VOMS/Storage/Bugs OpenMOLE)
- manque d'accompagnement/référent sur l'adoption/utilisation d'EGI

Freins à la mutualisation:

- investissement de départ technique et matériel important pour des sites hors GRIF/EGI
- investissement unitaire élevé

État actuel

Plus la possibilité d'obtenir des certificats pour le CNRS

API REST DIRAC via token n'est pas encore disponible et documentée

Machines de grille vieillissantes

=> Possibilité d'utiliser EGI dans OpenMOLE en suspens

Strategies alternatives

Orientation des utilisateurs vers:

- les mésocentres
- MESONET

+ federation des resources lattentes des laboratoires

Les centres de calculs

Mésocentres:

- l'accompagnement des utilisateurs est inégale
- des incompréhensions sur les besoins d'utilisation

MESONET:

- infra bien adaptée
- bon accompagnement
- mais pour des projets avec une composante formation

Les resources latentes des laboratoires

Développement d'un système de batch : MiniClust

Adapté aux calculs d'évaluation de modèles:

- mono-coeur / faiblement parallel
- court (quelques minutes)
- nombreux, plusieurs centaines de milliers par expérience

<https://github.com/openmole/miniclust>

MiniClust

Système de desktop grid:

- serveur central est un serveur minio
- les workers se lance en une commande (avec docker)
- les workers n'ont besoin que d'une connexion https sortante
- les workers peuvent rejoindre/partir à tout moment
- les workers utilisent la machine quand elle n'est pas trop chargée
- pas de système de fichier partagé
- système de cache de données sur les workers

Une instance MiniClust : Babar

Service lancé il y a 2 mois

- Un serveur Minio sur un vd.4, peut fédérer plusieurs centaines de machines
- Des workers, ISC-PIF, IGN, CIRAD:

The screenshot shows a GitHub README page for the 'miniclust-worker' repository. The title is 'MiniClust Worker'. It contains instructions to clone the repository and edit the config.yml file. It also shows a sample config.yml file and a command to run the service.

MiniClust Worker

To deploy a [miniclust](#) worker on a computer, just clone this repo.

```
git clone https://github.com/openmole/miniclust-worker.git
```

Edit the config.yml file, and change the service url, key and secret:

```
minio:  
  url: https://babar.openmole.org  
  key: key  
  secret: secret
```

Then run:

```
./start
```

Conclusion

- De nombreux services dans le cloud
- Nos services dans le cloud fonctionnent très bien
- Nous n'avons plus d'accès à la grille possible
- A la recherche d'une stratégie afin de permettre un accès à des coeurs de calcul pour les utilisateurs d'OpenMOLE

Limites / Questions

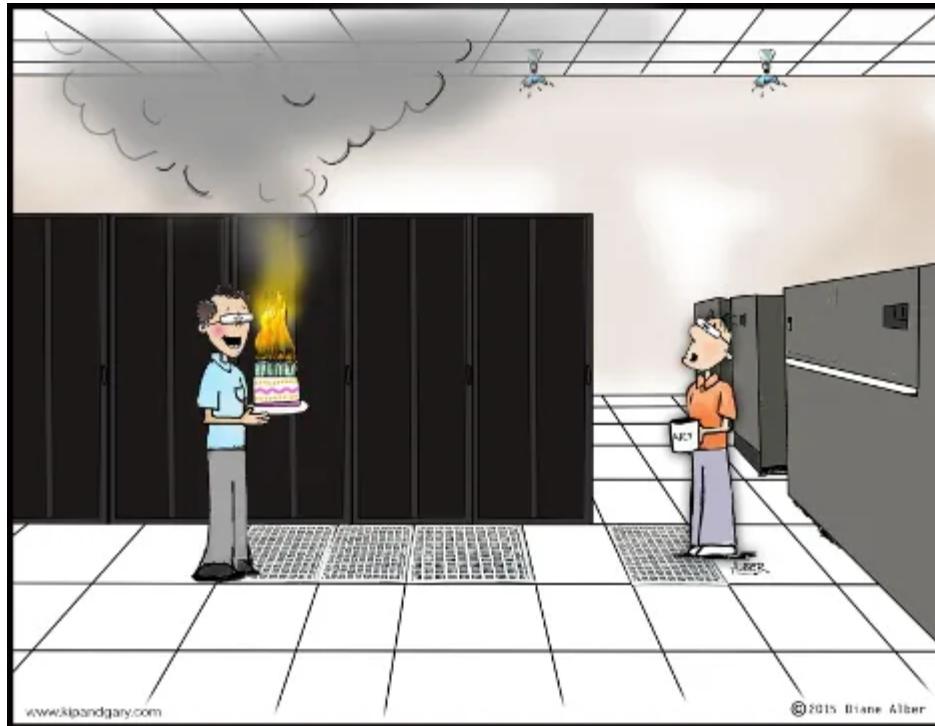
Limites:

- Malgré toutes nos tentatives pas d'adoption massive de la grille, trop de barières
- IOPS bas sur certaines machines de cloud, mais résolu avec l'ephemeral storage pour notre cas

Questions:

- Y a t'il un mésocentre/cluster facilement accessible pour les utilisateurs SHS France-Iliens ?
- Quelle infrastructure pour les utilisateurs rattachés à un mesocentre qui ne supporte pas OpenMOLE ?
- Est-il possible pour un projet de recherche d'investir 5-10k€ dans GRIF ?
- Comment expérimenter avec des protocoles type IPFS, sur le cloud ? (alertes Renater)

Joyeux anniversaire !!!



Merci pour ces 20 ans de travail et d'accompagnement !

ExModelo 2025



About

Models and simulations

The use of simulation models has widely spread in recent years, in various fields of academic research. Models are being developed to represent and try to better understand all kinds of systems: population dynamics, epidemics, transportation systems, macro-scale systems, micro-scale systems, etc. In some scientific areas, models and *in silico* simulations have become essential to help study *in vivo* situations.

<https://exmodelo.org>

