Reproductibilité et HPC

Bilan du Forum ORAP du 15 mars

Michel Kern

Inria & ORAP

Journée « Reproductibilité en Sciences »

Université Lyon I, 4 avril 2024

Outline

1. Presentation of Orap
2. Introduction
3. A social issue: improve trust
4. Technical tools: software
5. Scientific consequences
6. Conclusions

ORAP : ORGANISATION ASSOCIATIVE DU PARALLÉLISME

Promouvoir le Calcul Haute Performance

Structure de collaboration pour le développement et la diffusion du calcul parallèle, crée en 1994 sous l'impulsion de J.-L. Lions.

Organisation de deux Forums par an, sur des sujets d'actualité

- Forum 52: **Reproductibilité**
- Forum 51: **Vers l'exascale et au-delà : architectures et technologies**
- Forum 50: **HPC et Applications : quelques perspectives**
- Forum 47: **Vers un HPC frugal**
- Forum 46: **Calcul haute performance : vers une hybridation avec l'IA ?**

Conseil scientifique

- Sylvie Joussaume, CNRS/IPSL, présidente du conseil scientifique
- Edouard Audit, Maison de la simulation
- Marc Baaden, CNRS
- Christophe Biernacki, Inria
- François Bodin, Université Rennes 1
- Patrick Carribault, CEA
- Michel Kern, Inria
- Catherine Lambert, CERFACS
- Laurent Lellouch, CNRS
- Alain Lichnewsky
- Violaine Louvet, CNRS
- Jean-Philippe Nominé, CEA/DIF/DSSI
- Anne-Cécile Orgerie, CNRS
- Isabelle Perseil, INSERM
- Melanie Plainchault, Totalenergies
- Alain Refloch, ONERA
- Stéphane Requena, GENCI
- Guillaume Sylvand, AIRBUS Central R&T
- Denis Veynante, CNRS

Les présentations du 52e Forum

- Better Reproducibility: We do not want it, cannot afford it, but still need it and can have it (M. Héroux, Sandia)
- Présentation du réseau reproductibilité (A. Legrand, CNRS / LIG)

- **Session Environnements logiciels**
  - Reproducibility and performance: why choose (L. Courtes, Inria)
  - Reproductibilité et modèles de programmation; MPI et OpenMP (H. Taboada, CEA)
  - Floating-point Determinism, Reproducibility & Accuracy in HPC (D. Defour, Univ. Perpignan)

- **Session Enjeux pour les publications**
  - Problèmes de reproductibilité et incitations au changement (A. Legrand, CNRS / LIG)
  - Practical approaches to reproducibility in an era of rapid scientific evolution: An example with environmental Data Science (A. Fouilloux, Simula Research Laboratory)

- **Session Pratique des communautés**
  - Effet papillon et reproductibilité : climat et météo sont dans un ordinateur (O. Marti, J. Servonnat et A. Caubel, IPSL /LSCE)
  - Mini-apps: an effective tool to answer the question of reproducibility and performance (H. Calandra , J.Meng, TotalEnergies)

Présentations disponibles à: http://orap.irisa.fr/52ieme-forum-reproductibilite/

Outline

1. Presentation of Orap
2. Introduction
3. A social issue
4. Technical tools
5. Scientific consequences
6. Conclusions

An important topic

https://www.thetransmitter.org/brain-imaging/two-studies-fail-to-replicate-holy-grail-diana-fmri-method-for-detecting-neural-activity/

“The failed replications are not a dead end for DIANA, Yu says, because it’s a natural part of the scientific process to take steps forward and then backward.”

NEWS / **BRAIN IMAGING**

Two studies fail to replicate ‘holy grail’ DIANA fMRI method for detecting neural activity

The signal it flags is more likely the result of cherry-picking data, according to the researchers who conducted one of the new studies, but the lead investigator on the original work disputes that conclusion.

BY **CALLI MCMURRAY**
27 MARCH 2024 | 11 MIN READ

| HTTPS://DOI.ORG/10.53053/KINZ6963 | CITE THIS ARTICLE

But also for day to day's work (after A. Legrand)

```
my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking parameters and data

- *.dat? Ooh, you ran this in data/2091293-AJXQ37?
- Wasn't mymap.dat updated since then?
- That was for foo.csv. What about bar.csv? Is it reproducible?

Tracking the process (on short/long term)

- Why did I run this? What did I learn from it? When did I do this?

Tracking code version

- my_code is revision 21b95ecfa0911d6ca87668482b11ab9498edd8f3

Tracking software environment

- my_code depends on a dozen of libraries, which depend on dozens of libraries
- my_code was compiled with clang 1:9.0-49.1 and
  `-O3 -funroll-loops -fno-strict-aliasing -finline-functions` ...

In the beginning

*An article about computational science in a **scientific publication is not the scholarship itself, it is merely advertising** of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

J. B. Buckheit and D. L. Donoho (1995)

Electronic Documents Give Reproducible Research a New Meaning **RE1.3**

Jon F. Claerbout and Martin Karrenbach, Stanford Univ.

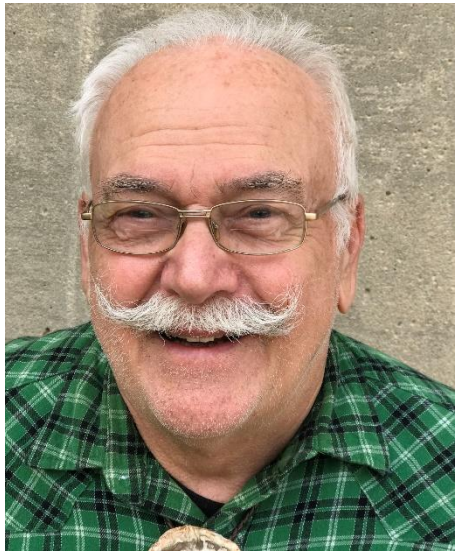
SUMMARY

A revolution in education and technology transfer follows from the marriage of word processing and software command scripts. In this marriage an author attaches to every figure caption a pushbutton or a name tag usable to recalculate the figure from all its data, parameters, and programs. This provides a concrete definition of reproducibility in computationally oriented research. Experience at the Stanford Exploration Project shows that preparing such electronic documents is little effort beyond our customary report writing; mainly, we need to file everything in a systematic way.

In 1990 we began experimenting with electronic documents that merge our scientific software with our word-processing software. A year later we manufactured a CD-ROM containing a new textbook, Joe Dellinger's doctoral dissertation, and two progress reports of the Stanford Exploration Project. We distributed these CD-ROMs[1] to sponsors and

- make incremental improvements in electronic-document software
- seek partners for broadening standards (and making incremental improvements).

Our basic goal is reproducible research. The electronic document is our means to this end. In principle, reproducibility in research can be achieved without electronic documents and that is how we started. Our first nonelectronic reproducible document was a textbook in which the paper document contained the name of a program script in every figure caption. The program scripts were organized by book chapter and section so they could be correlated to an accompanying magnetic tape dump of the file system. The magnetic tape also contained all the necessary data to feed the program script.



Definitions

From M. Heroux, also
https://www.acm.org/publications/badging-terms

Reproducible

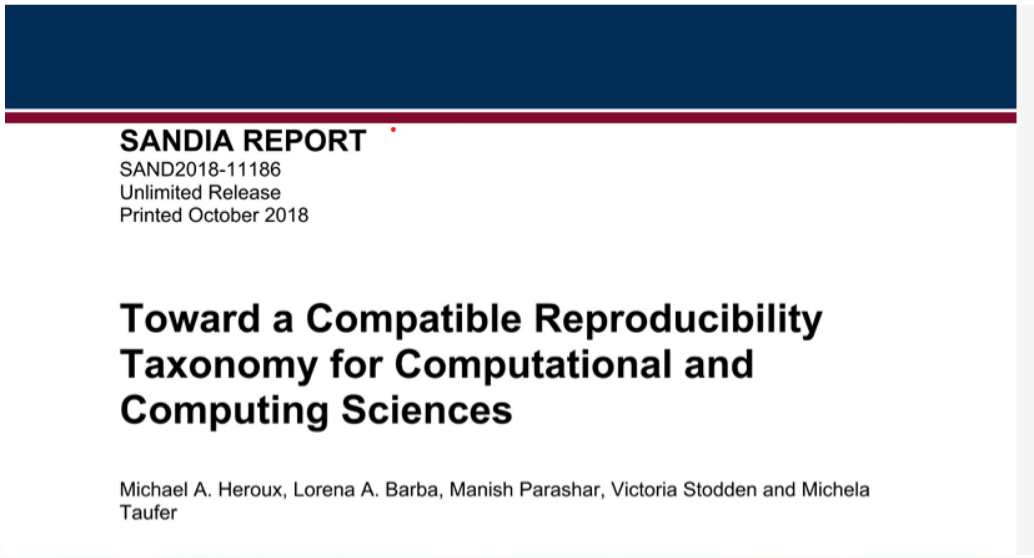
Authors provide all the necessary data and the computer codes to run the analysis again, recreating the results.
Different team, same computational setup

Replicable

A new study arrives at the same scientific findings as a previous study, collecting new data (with the same or different methods) and completes new analyses
Different team, different experimental setup

		Data: Same	Data: Different
Analysis	Same	Reproducible	Replicable
Analysis	Different	Robust	Generalisable

The Turing way,
- https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions


SANDIA REPORT
SAND2018-11186
Unlimited Release
Printed October 2018

Toward a Compatible Reproducibility Taxonomy for Computational and Computing Sciences

Michael A. Heroux, Lorena A. Barba, Manish Parashar, Victoria Stodden and Michela Taufer

Outline

1. Presentation of Orap
2. Introduction
3. A social issue: improve trust
4. Technical tools
5. Scientific consequences
6. Conclusions

It is a matter of trust (M. Heroux)

Strategies for improving trust

- Build communities: share risk and reward
- Collaborate with stakeholders (publishers and funders): raise expectations
- Invest in software ecosystems: improve quality, share investments

Community in France (A. Legrand)

Réseau Français de Recherche Reproductible http://www.recherche-reproductible.fr/

Céline Acary-Robert (Grenoble), Frédéric Lemoine (Inst. Pasteur), …

Inter-disciplinary, support by Ministère de la Recherche

Stakeholders: Can we get the funders on-board ? ANR ?

Software development and maintenance must be recognized as a valuable activity

Barriers to reproducibility (M. Heroux, A. Fouilloux)

The natural selection of bad science

- Usual metrics (publi count, impact factor) do not promote reproducibility:
  Submit paper, complete project
- Change incentives to include value of better software, better science
- Must be done at community (institutional ?) level

For an individual researcher, reproducibility has a cost:
It takes longer, requires more work, and leads to fewer papers.

Must be a long term goal

Alternative point of view (AF)

Reproducibility is an enabler for better science when a group has limited resources: will let new students start faster

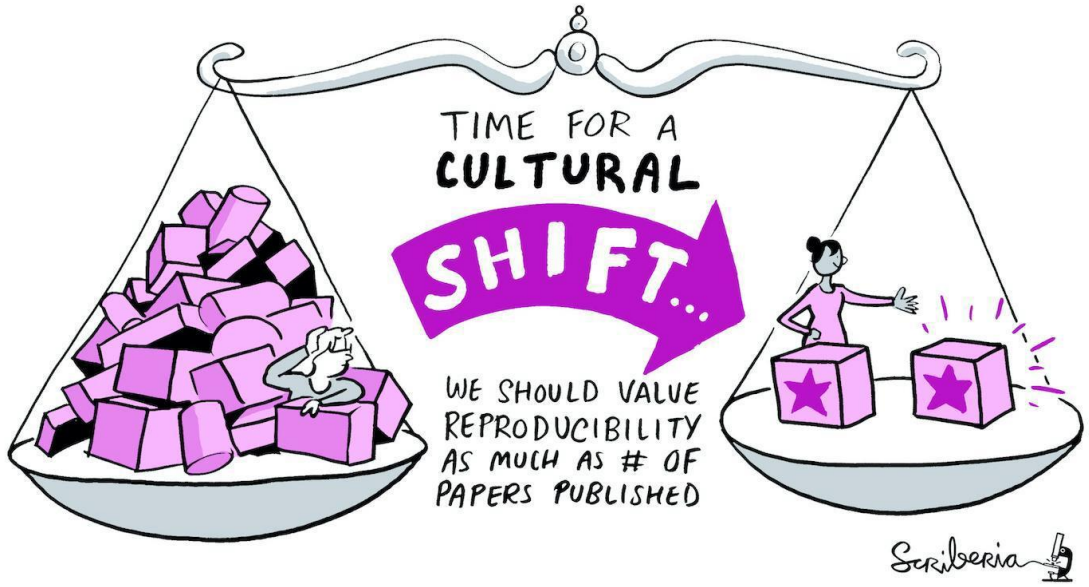


Image from A. Fouilloux's talk

Publications and badges (A. Legrand, M. Heroux)

Software artifacts reviewed in parallel with paper (independently)

Goals

- validate experimental results from published articles
- restore trust
- promote artifact sharing (benchmarks, data sets, tools, models)
- enable fair comparison of results and techniques
- build upon others' research

Mentalities are evolving people care, make stuff available, errors are found and fixed

First experience with reviewing a paper that includes a link to the code

- Paper still reviewed on its own merit (applied maths / subsurface flow, so results are illustrative)
- Additional work for the reviewer (code does not work out of the box)

Reproducible results in journals and conferences

M. Heroux, A. Legrand

ACM TOMS Reproducible Computational Results

Submission: Optional RCR option

Standard reviewer assignment: Nothing changes

RCR reviewer assignment:

- Concurrent with standard reviews
- As early as possible in review process
- Known to and works with authors during the RCR process

RCR process:

- Multi-faceted approach, Bottom line: Trust the reviewer

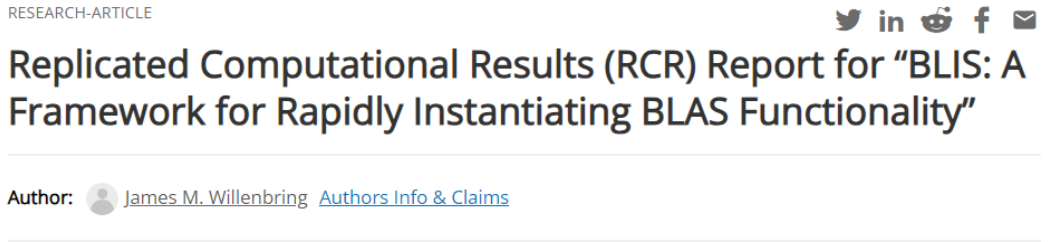
Publication:

- Reproducible Computational Results Designation
- The RCR referee acknowledged
- Review report appears with published manuscript


RESEARCH-ARTICLE

BLIS: A Framework for Rapidly Instantiating BLAS Functionality

Authors: Field G. Van Zee, Robert A. van de Geijn Authors Info & Claims


RESEARCH-ARTICLE

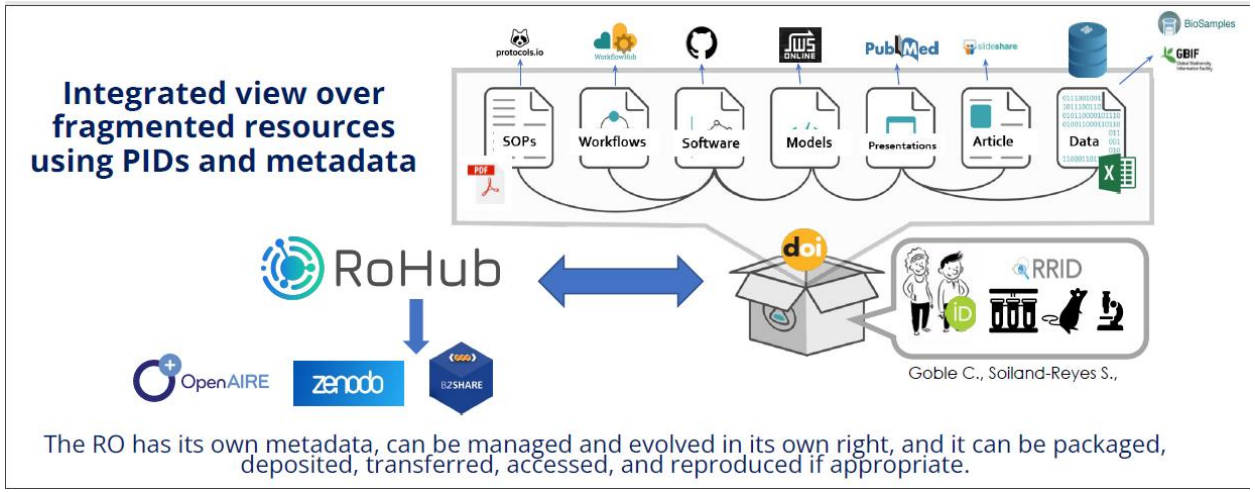
Replicated Computational Results (RCR) Report for "BLIS: A Framework for Rapidly Instantiating BLAS Functionality"

Author: James M. Willenbring Authors Info & Claims

Outline

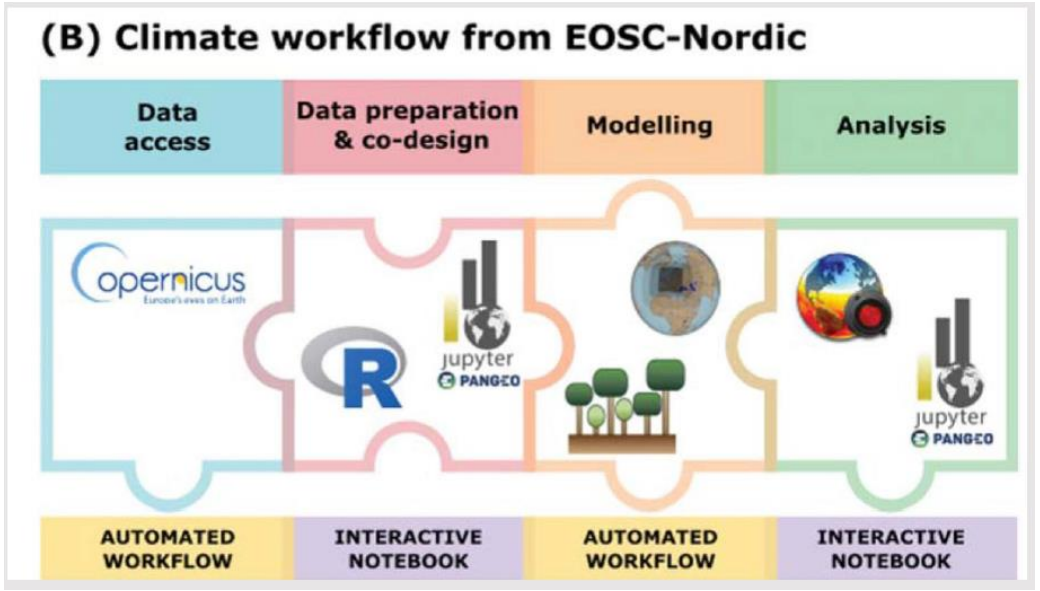
1. Presentation of Orap
2. Introduction
3. A social issue
4. Technical tools: software
5. Scientific consequences
6. Conclusions

Research objects, reproducible workflows (A. Fouilloux)



http://www.researchobject.org

- Research objects: link code, executable and data
- Galaxy for reproducible workflows
- ROHub to aggregate all resources



https://doi.org/10.1162/dint_a_00136

Guix: Reproducible package management / deployment (L. Courtes)

GuixHPC aims at providing reproducibility without sacrificing performance

Specific Guix-HPC project: https://hpc.guix.info/
Possible alternative to Spack

Link to Software Heritage

Important example: portable and high-performance MPI
https://hpc.guix.info/blog/2019/12/optimized-and-portable-open-mpi-packaging/

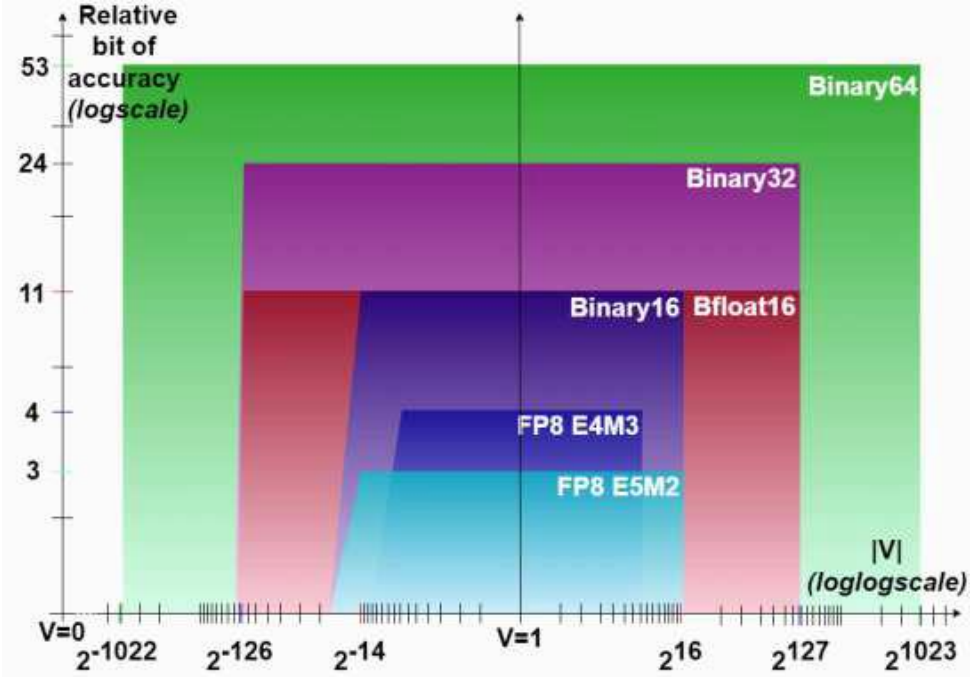
Floating point computations (D. Defour)

Since 2012, thanks to (because of ?) machine learning, new floating point formats
Both range and precision reduced

As FP is not associative, order of computations matters

Repeatable computations

- Make addition associative again
  - Ensure bitwise repeatability
  - Use long accumulator
  - Parallel algorithm suitable for today'sarchitectures
- Enforce execution order



Language, compiler version and processor generation can all lead to different results

Parallel computing and its surprises (H. Taboada)

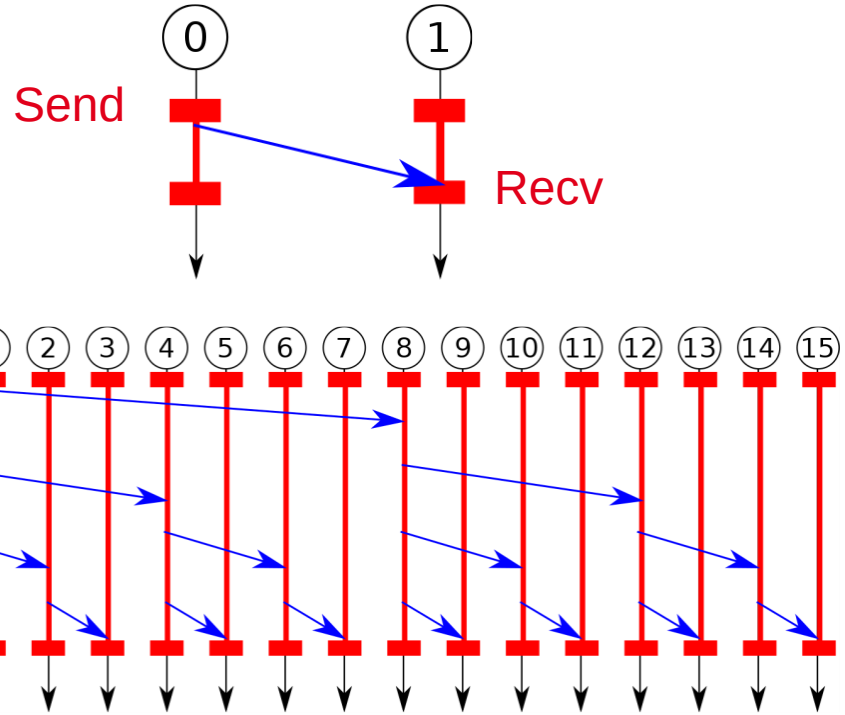
MPI

- Standard (since 1993, current: 4.1) for parallel computing with distributed memory.
- Programming model by explicit message exchanges

Point to point communications: `send, receive`

Messages between the same source and the same destination must be received in the order they were sent
No longer true if threads are used

Collective communications: `broadcast, reduce`, …
Multiple algorithms, binomial, linear, …
Each algorithm may realize the operations
in its own order



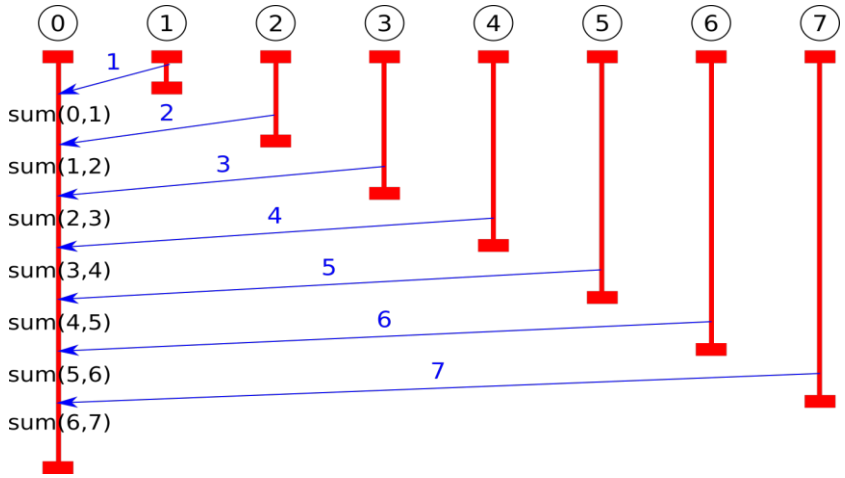
MPI: the case of reductions

The same kind of issues occur with OpenMP / threads (shared memory)

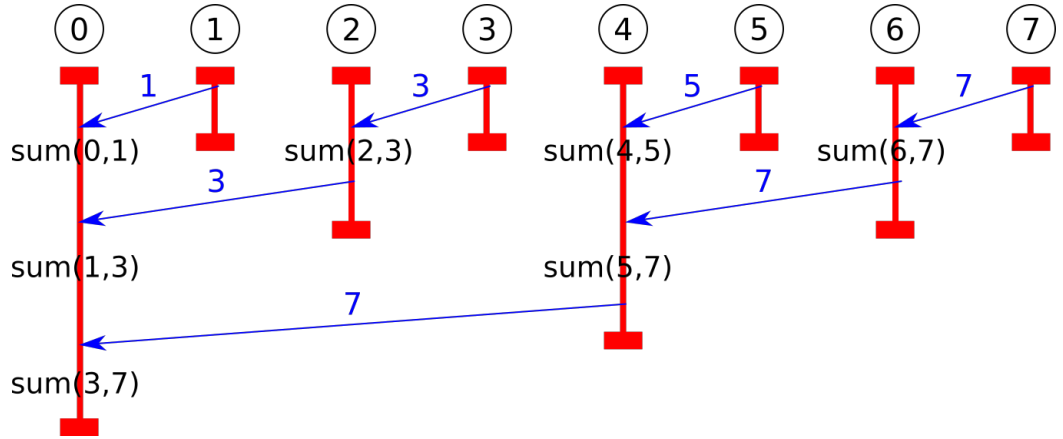
Reduce: compute the sum (say) of values distributed over all procs

Will successive call to reduce deliver the same result ? Not always !

Linear



Binomial



Each of these algorithms could be executed at each call of `MPI_Reduce`

Code not reproducible

Takeaways for floating-point and parallelism

- Most research today is done on 64/32 bits systems
- Tomorrow's HPC system will most likely be built from IA processors
- Can we obtain reproducibility with 16/8 bits ? Do we want to ?
- How to certify that a computed solution is numerically meaningful ?

Reproducibility for FP and parallel computing is possible,
but usually implies some loss in performance

Is bitwise reproducibility really needed ?

Yes, for debugging

What is the relationship between reproducibility and correctness ?

Outline

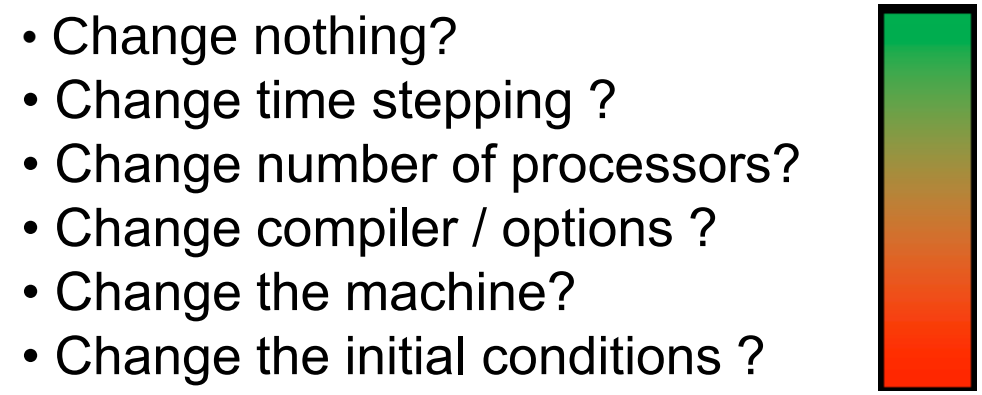
1. Presentation of Orap
2. Introduction
3. A social issue
4. Technical tools
5. Scientific consequences
6. Conclusions

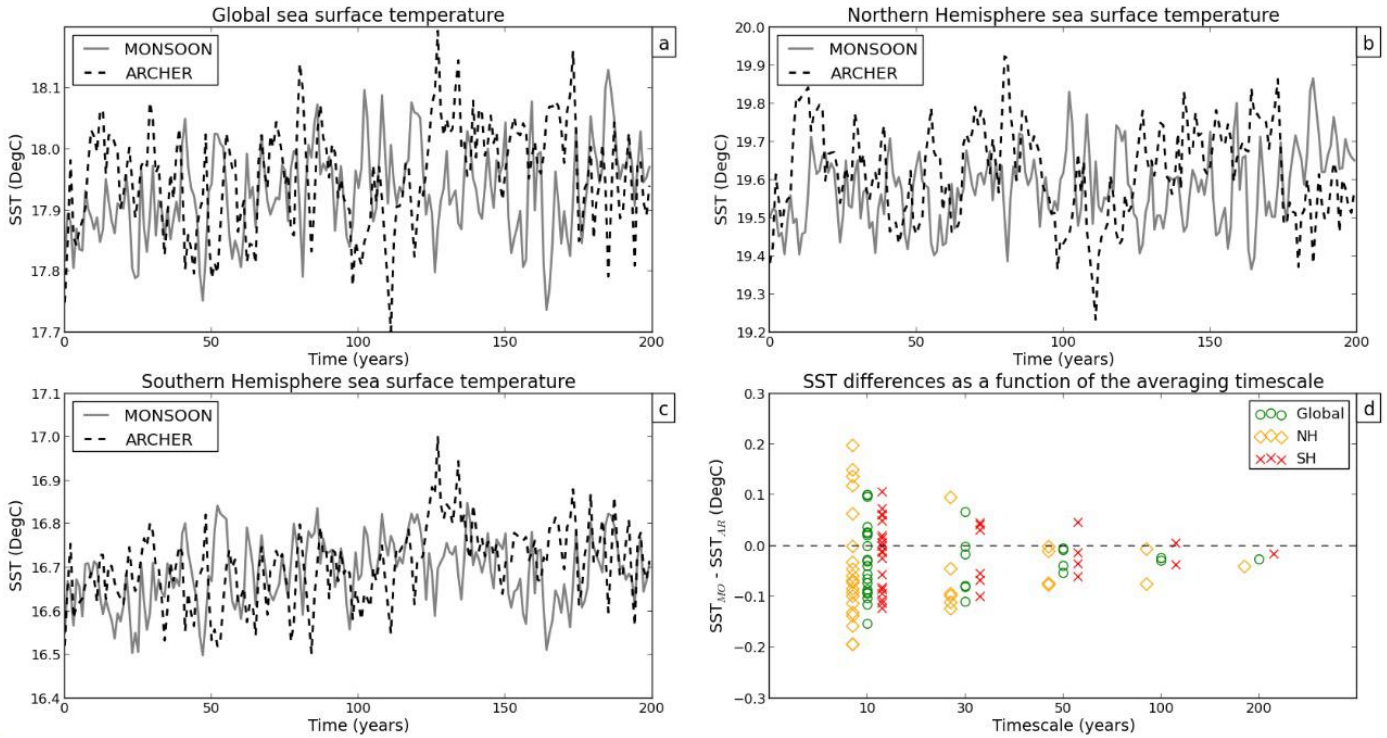
Reproducibility and climate (O. Marti)

- **Computers**
  - If nothing changes, nothing changes
  - Code version, input data, hardware, number of cores / partition geometry, compiler / libraries, compilation options, runtime options, Age du capitaine, …
- **Physical result**
  - Must be redefined for each domain
- **Who can reproduce my result?**
  - All other things being equal
  - Climatology : many robust results, replicated by several teams, several models, several super-computers, …

Twice the « same » simulation == 1 or 2 results ?

- Change nothing?
- Change time stepping ?
- Change number of processors?
- Change compiler / options ?
- Change the machine?
- Change the initial conditions ?





Reproducibility and supercomputing (M. Heroux)

Scenario:

- You compute a “hero” calculation using 5M node-hours on Jean Zay and submit your results for publication.
- During the review process, a referee questions the validity of your results.
- What options are feasible:
  - The reviewer re-runs your code on a laptop or cluster.
  - The reviewer re-runs your code on Jean Zay.
  - You re-run your code on Jean Zay.
  - Your results are rejected.
  - Your results are accepted, but with risk.

Improve trust in your computations:

Linear algebra HPCG benchmark (solve sparse linear system with conjugate gradient)

- Exploit structural spectral properties of the matrix
- Can be checked during the computation

Fluid dynamics (subsurface flow)

- Check conservation of mass, number of particles

Application dependent

How different is it from using any research infrastructure ? (MK)

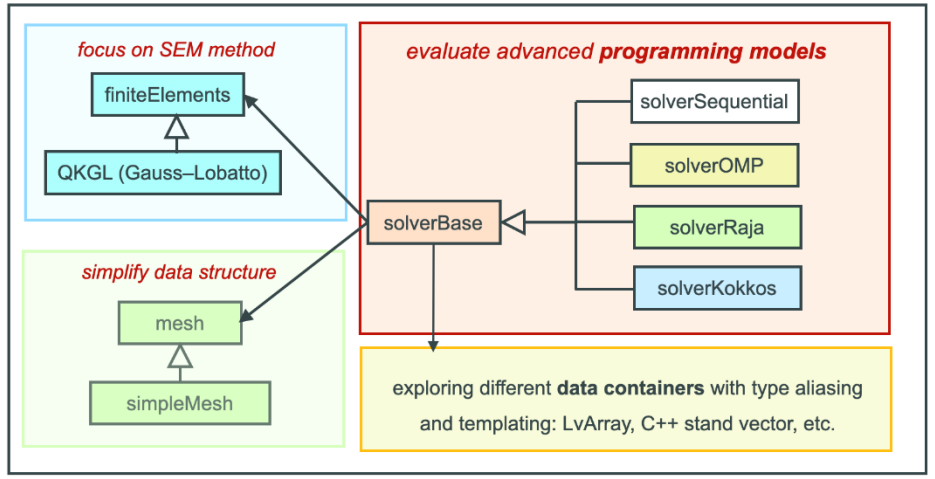
Dans l'industrie: exemple de TotalEnergies (H. Calandra)

Mini-apps: to answer the **reproducibility** and **performance** of HPC

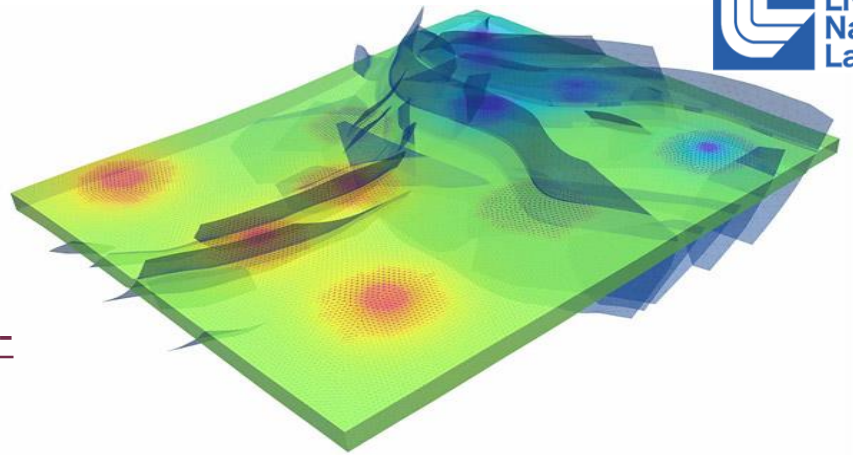
Portability: how you application can run on different HPC Hardware using different Programming Models?

Reproducibility: whether you can reproduce the same results?

Repeatability: whether the performance of your application changes under different conditions?



Lawrence Livermore National Laboratory



https://www.llnl.gov/news/llnl-partners-open-access-co2-storage-simulator

Outline

1. Presentation of Orap
2. Introduction
3. A social issue
4. Technical tools
5. Scientific consequences
6. Conclusions

Takeaways from the round-table and personal musings

Importance of community to:

- Define what it means to be reproducible
- Change incentives to reward reproducibility
- Train new / young researchers to invest in reproducibility

Importance of software and tools

- Invest in maintainable software
- Make everything automatic through scripts
- With the right tools, reproducibility leads to gains in productivity

Involving stakeholders (publishers, funders) must be gradual

- Reproducibility in 5 or 10 years ?
- Are computational experiement different from physical ones ?
- Reproducibility for whom ? You, your students, your team, reviewers, colleagues