

REPRODUCIBLE RESEARCH AND COMPUTER SCIENCE

Arnaud Legrand



L3 ENS Lyon
January 2024



A BIT ABOUT MYSELF

1997 ENS Lyon

2000-2003: Ph.D. @ LIP, *Heterogeneous parallel algorithms and scheduling: static and dynamic approaches* (O. BEAUMONT and Y. ROBERT)

2004-2005 "Post-doc" LIP and UC San Diego (H. CASANOVA)

2005 CR CNRS @ LIG (MESCAL team), Grenoble

2015 HDR *Scheduling for Large Scale Distributed Computing Systems: Approaches and Performance Evaluation Issues.*

2016 POLARIS research team

2018 DR CNRS

2021-2026 Section 6 CoNRS

Research Themes

Scientific computing cluster, grid, volunteer computing, clouds, supercomputers

Performance optimization scheduling, combinatorial/continuous optimization, game theory, reinforcement learning

Performance evaluation predictive simulation, trace analysis and visualization, statistical modeling

Research reproducibility and Open science

PUBLIC EVIDENCE FOR A LACK OF REPRODUCIBILITY

- J.P. Ioannidis. *Why Most Published Research Findings Are False* PLoS Med. 2005.
- *Lies, Damned Lies, and Medical Science*, The Atlantic. Nov, 2010
- *Reproducibility: A tragedy of errors*, Nature, Feb 2016.
- Steen RG, *Retractions in the scientific literature: is the incidence of research fraud increasing?*, J. Med. Ethics 37, 2011



Science has lost its way, at a big cost to humanity

Researchers are rewarded for splashy findings, not for double-checking accuracy. So many scientists looking for cures to diseases have been building on ideas that aren't even true.

The Science journal website is shown. The top navigation bar includes links for AAAS-ORG, FEEDBACK, HELP, LIBRARIANS, All Science Journals, and a search bar. The main content area features an article titled "Reproducibility" by Marcia McNutt. The article discusses the lack of reproducibility in science and the need for transparency. It includes a summary, full text, and a PDF link. The sidebar on the left provides links for Article Views, Article Tools, and Related Content.

The Nature journal website is shown. The top navigation bar includes links for Home, News & Comment, Research, Careers & Jobs, Current Issue, and Archive. The main content area features an announcement titled "Announcement: Reducing our irreproducibility" dated April 24, 2013. It discusses the journal's commitment to improving reproducibility. The sidebar on the left provides links for Article Views, Article Tools, and Related Content.

A large graphic titled "HOW SCIENCE GOES WRONG" in a stylized, colorful font. The background features various scientific illustrations like a globe, DNA helixes, and microscopes. The text "Economist" is visible in the top right corner of the graphic area.

The Nature journal website is shown again, this time with a different announcement. The main content area features an announcement titled "Announcement: Reducing our irreproducibility" dated April 24, 2013. The sidebar on the left provides links for Article Views, Article Tools, and Related Content.

The Scientist magazine website is shown. The main content area features an article titled "NIH Tackles Irreproducibility" by Jef Akst. The article discusses the National Institutes of Health's efforts to improve reproducibility in scientific research. The sidebar on the left provides links for Article Views, Article Tools, and Related Content.

Courtesy V. Stodden, SC, 2015

SCIENTIFIC MISCONDUCT

NEWSWORTHY STORIES ABOUT SCIENTIFIC MISCONDUCT

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received \approx \$19 million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. \$7.2 million!

NEWSWORTHY STORIES ABOUT SCIENTIFIC MISCONDUCT

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received \approx \$19 million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. \$7.2 million!

Diederik Stapel Professor, Social Psychology, Univ. Tilburg, 2011

I failed as a scientist. I adapted research data and fabricated research. Not once, but several times, not for a short period, but over a longer period of time. [...] I am aware of the suffering and sorrow that I caused to my colleagues... I did not withstand the pressure to score, to publish, the pressure to get better in time. I wanted too much, too fast. In a system where there are few checks and balances, where people work alone, I took the wrong turn.

58 retracted publications

NEWSWORTHY STORIES ABOUT SCIENTIFIC MISCONDUCT

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received \approx \$19 million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. \$7.2 million!

Diederik Stapel Professor, Social Psychology, Univ. Tilburg, 2011

I failed as a scientist. I adapted research data and fabricated research. Not once, but several times, not for a short period, but over a longer period of time. [...] I am aware of the suffering and sorrow that I caused to my colleagues... I did not withstand the pressure to score, to publish, the pressure to get better in time. I wanted too much, too fast. In a system where there are few checks and balances, where people work alone, I took the wrong turn.

58 retracted publications

Brian Wansink Professor, Psychological Nutrition, Cornell, 2016

I gave her a data set of a self-funded, failed study which had null results. I said "This cost us a lot of time and our own money to collect. There's got to be something here we can salvage because it's a cool (rich & unique) data set." I told her what the analyses should be. [...] Every day she came back with puzzling new results, and every day we would scratch our heads, ask "Why," and come up with another way to reanalyze the data with yet another set of plausible hypotheses

17 retracted publications

SCIENTIFIC MISCONDUCT? WHAT ARE THE CONSEQUENCES ?

Reinhart and Rogoff Professors of Economics at Harvard

gross debt [...] exceeding 90 percent of the economy has a significant negative effect on economic growth – Growth in a Time of Debt (2010)

While using RR's working spreadsheet, we identified coding errors, selective exclusion of available data, and unconventional weighting of summary statistics. – 2013: Herndon, Ash and Pollin

For 3 years, austerity was not presented as an option but as a necessity.

– 2013: Paul Krugman

At least, a scientific debate has been possible.

SCIENTIFIC MISCONDUCT? WHAT ARE THE CONSEQUENCES ?

Reinhart and Rogoff Professors of Economics at Harvard

gross debt [...] exceeding 90 percent of the economy has a significant negative effect on economic growth – Growth in a Time of Debt (2010)

While using RR's working spreadsheet, we identified coding errors, selective exclusion of available data, and unconventional weighting of summary statistics. – 2013: Herndon, Ash and Pollin

For 3 years, austerity was not presented as an option but as a necessity.

– 2013: Paul Krugman

At least, a scientific debate has been possible.

Bad science is deleterious

- It is used to backup stupid politics, it affects people's life, ...
- It blurs the frontier between scientists and crooks

Media attention inflates conspiracy opinions 😞

- *Scientific result are worthless.*
- *Scientists can't even agree with each others on economy/climate/vaccine/5G/...*
- *Stop the scientific dictatorship/lobby!*

COMMON REPRODUCIBILITY PITFALLS

GO READ THE PAPER BY SMITH ET. AL. 2009

Access through your institution Purchase PDF

Article preview
Abstract
Introduction
Section snippets
References (30)
Cited by (31)

 Future Generation Computer Systems
Volume 25, Issue 3, March 2009, Pages 315-325


Secure on-demand grid computing

M. Smith , M. Schmidt , N. Follenbeck , T. Dörnemann , C. Schridde ,
B. Freisleben 

Show more 

+ Add to Mendeley  Share 

<https://doi.org/10.1016/j.future.2008.03.002> [Get rights and content](#)

Abstract

In this paper, a novel approach for enabling Grid users to autonomously install and use custom software on demand using an image creation station is presented, while at the same time offering new security mechanisms to protect both software and data from other Grid users and external attackers. An automated dynamic firewalling mechanism

GO READ THE PAPER BY SMITH ET. AL. 2009

Purchase options

▼ Corporate

For R&D professionals working in corporate organizations.

▲ Academic and personal

For academic or personal use only.

US\$27.95

Local taxes may apply

Online access for 48 hours with the option to save or download the article in PDF format. [Learn more ↗](#)

Add to Cart

Looking for a customized option?

Contact sales for special pricing for your organization

Contact sales ↗

- Use your institution subscription... or Sci-Hub 😊,

GO READ THE PAPER BY SMITH ET. AL. 2009

Access through your institution | View Open Manuscript | Purchase PDF

Article preview
Abstract
Introduction
Section snippets
References (32)
Cited by (1)

 Journal of Parallel and Distributed Computing
Volume 166, August 2022, Pages 111-125


Simulation-based optimization and sensibility analysis of MPI applications: Variability matters

Tom Cornebize  , Arnaud Legrand  

Show more 

+ Add to Mendeley  Cite 

<https://doi.org/10.1016/j.jpdc.2022.04.002> 

Abstract

Finely tuning MPI applications and understanding the influence of key parameters (number of processes, granularity, collective operation algorithms, virtual topology, and

- Use your institution subscription... or Sci-Hub 😊, or HAL/Arxiv

GO READ THE PAPER BY SMITH ET. AL. 2009

The screenshot shows the HAL digital library interface. At the top, there is a search bar with the placeholder "Chercher un document, un auteur, un mot clé...". To the right of the search bar are a help icon (info) and a language dropdown set to "FR". Below the search bar, there is a navigation menu with "science ouverte" and a download button labeled "Télécharger pour visualiser".

The main content area displays a research paper titled "Simulation-based Optimization and Sensibility Analysis of MPI Applications: Variability Matters". The paper is from the journal "Journal of Parallel and Distributed Computing" for the year 2022. It is authored by Tom Cornebize (1, 2), Arnaud Legrand (3, 1). There is a "Afficher plus de détails" button.

On the left side, there is a sidebar with sections for "Dates et versions" and "Identifiants". The "Dates et versions" section lists two versions: "hal-03141988, version 1 (15-02-2022)" and "hal-03141988, version 2 (06-01-2022)". The "Identifiants" section provides the HAL Id ("hal-03141988, version 2"), ARXIV ID ("2102.07674"), and DOI ("10.1016/j.jpdc.2022.04.002").

The main content area also includes a "Résumé" section with the text: "Finely tuning MPI applications and understanding the influence of key parameters (number of processes, granularity, collective operation algorithms, virtual topology, and process placement) is critical to obtain". There is a "Mots clés" section with terms: "Simulation", "validation", "sensibility analysis", and "SimGrid".

- Use your institution subscription... or Sci-Hub 😊, or HAL/Arxiv

Rodriguez et al., CONCUR'15

Unfolding-based Partial Order Reduction*

César Rodríguez¹, Marcelo Sousa², Subodh Sharma³, and
Daniel Kroening⁴

¹ Université Paris 13, Sorbonne Paris Cité, LIPN, CNRS, France

^{2,4} Department of Computer Science, University of Oxford, UK

³ Indian Institute of Technology Delhi, India

Abstract

Partial order reduction (POR) and net unfoldings are two alternative methods to tackle state-space explosion caused by concurrency. In this paper, we propose the combination of both approaches in an effort to combine their strengths. We first define, for an abstract execution model, unfolding semantics parameterized over an arbitrary independence relation. Based on it, our main contribution is a novel stateless POR algorithm that explores at most one execution per Mazurkiewicz trace, and in general, can explore exponentially fewer, thus achieving a form of *super-optimality*. Furthermore, our unfolding-based POR copes with non-terminating executions and incorporates state-caching. Over benchmarks with busy-waits, among others, our experiments show a dramatic reduction in the number of executions when compared to a

JUST COMPARE TO THE ALGORITHM THEY PROPOSED

Rodriguez et al., CONCUR'15

Algorithm 1: An unfolding-based POR exploration algorithm.

```
1 Initially, set  $U := \{\perp\}$ , set  $G := \emptyset$ , and call Explore( $\{\perp\}, \emptyset, \emptyset$ ).  
2 Procedure Explore( $C, D, A$ )  
3   Extend( $C$ )  
4   if  $\text{en}(C) = \emptyset$  return  
5   if  $A = \emptyset$   
6     Choose  $e$  from  $\text{en}(C)$   
7   else  
8     Choose  $e$  from  $A \cap \text{en}(C)$   
9   Explore( $C \cup \{e\}, D, A \setminus \{e\}$ )  
10  if  $\exists J \in \text{Alt}(C, D \cup \{e\})$   
11    Explore( $C, D \cup \{e\}, J \setminus C$ )  
12  Remove( $e, C, D$ )  
13 Procedure Extend( $C$ )  
14   Add  $ex(C)$  to  $U$   
15 Procedure Remove( $e, C, D$ )  
16   Move  $\{e\} \setminus Q_{C,D,U}$  from  $U$  to  $G$   
17   foreach  $\hat{e} \in \#_U^i(e)$   
18     Move  $[\hat{e}] \setminus Q_{C,D,U}$  from  $U$  to  $G$ 
```

- Looks good!

JUST COMPARE TO THE ALGORITHM THEY PROPOSED

Rodriguez et al., CONCUR'15

We give some new definitions. Let C be a configuration of \mathcal{U} . The *extensions* of C , written $ex(C)$, are all those events outside C whose causes are included in C . Formally, $ex(C) := \{e \in E : e \notin C \wedge [e] \subseteq C\}$. We let $en(C)$ denote the set of events *enabled* by C , i.e., those corresponding to the transitions enabled at $state(C)$, formally defined as $en(C) := \{e \in ex(C) : C \cup \{e\} \in conf(\mathcal{U})\}$. All those events in $ex(C)$ which are not in $en(C)$ are the *conflicting extensions*, $cex(C) := \{e \in ex(C) : \exists e' \in C, e \#^i e'\}$. Clearly, sets $en(C)$ and $cex(C)$ partition the set $ex(C)$. Lastly, we define $\#^i(e) := \{e' \in E : e \#^i e'\}$, and $\#_U^i(e) := \#^i(e) \cap U$. The difference between both is that $\#^i(e)$ contains events from *anywhere* in the unfolding structure, while $\#_U^i(e)$ can only see events in U .

The algorithm is given in [Alg. 1](#). `Explore`(C, D, A), the main procedure, is given the configuration that is to be explored as the parameter C . The parameter D (for *disabled*) is the set of set of events that have already been explored and prevents that `Explore()` repeats work. It can be seen as a *sleep set* [7]. Set A (for *add*) is occasionally used to guide the direction of the exploration.

Additionally, a global set U stores all events presently known to the algorithm. Whenever some event can safely be discarded from memory, `Remove` will move it from U to G (for *garbage*). Once in G , it can be discarded at any time, or be preserved in G in order to save work when it is re-inserted in U . Set G is thus our *cache memory* of events.

The key intuition in [Alg. 1](#) is as follows. A call to `Explore`(C, D, A) visits all maximal configurations of \mathcal{U} which contain C and do not contain D ; and the first one explored will

- Looks good! Err... **not so simple**. Depending on how you do this, you quickly move from polynomial to exponential.

JUST COMPARE TO THE ALGORITHM THEY PROPOSED

Rodriguez et al., CONCUR'15

POSIX threads. The analyzer accepts deterministic programs, implements a variant of [Alg. 1](#) where the computation of the alternatives is memoized, and supports cutoffs events with the criteria defined in [§ 5](#).

We ran POET on a number of multi-threaded C programs. Most of them are adapted from benchmarks of the Software Verification Competition [17]; others are used in related works [8, 19, 2]. We investigate the characteristics of average program unfoldings (depth, width, etc.) as well as the frequency and impact of cutoffs on the exploration. We also compare POET with NIDHUGG [1], a state-of-the-art stateless model checking for multi-threaded C programs that implements Source-DPOR [2], an efficient but non-optimal DPOR. All experiments were run on an Intel Xeon CPU with 2.4 GHz and 4 GB memory. [Tables 1](#) and [2](#) give our experimental data for programs with acyclic and non-acyclic state spaces, respectively.

For programs with acyclic state spaces ([Table 1](#)), POET with and without cutoffs seems to perform the same exploration when the unfolding has no cutoffs, as expected. Furthermore, the number of explored executions also coincides with NIDHUGG when the latter reports 0 sleep-set blocked executions (cf., [§ 4](#)), providing experimental evidence of POET's optimality.

The unfoldings of most programs in [Table 1](#) do not contain cutoffs. All these programs are deterministic, and many of them highly sequential (STF, SPIN08, FIB), features known to make cutoffs unlikely. CCNF(n) are concurrent programs composed of $n - 1$ threads where thread i and $i + 1$ race on writing one variable, and are independent of all remaining

³ Source code and benchmarks available from: <http://www.cs.ox.ac.uk/people/marcelo.sousa/poet/>.

- Looks good! Err... **not so simple**. Depending on how you do this, you quickly move from polynomial to exponential.
- **Experiments!**

JUST COMPARE TO THE ALGORITHM THEY PROPOSED

Rodriguez et al., CONCUR'15

■ **Table 1** Programs with acyclic state space. Columns are: $|P|$: nr. of threads; $|I|$: nr. of explored traces; $|B|$: nr. of sleep-set blocked executions; $t(s)$: running time; $|E|$: nr. of events in \mathcal{U} ; $|E_{\text{cut}}|$: nr. of cutoff events; $|\Omega|$: nr. of maximal configurations; $\langle |U_\Omega| \rangle$: avg. nr. of events in U when exploring a maximal configuration. A * marks programs containing bugs. <7K reads as “fewer than 7000”.

Benchmark	NIDHUGG				POET (without cutoffs)				POET (with cutoffs)				
	Name	$ P $	$ I $	$ B $	$t(s)$	$ E $	$ \Omega $	$\langle U_\Omega \rangle$	$t(s)$	$ E $	$ E_{\text{cut}} $	$ \Omega $	$\langle U_\Omega \rangle$
STF	3	6	0	0.06	121	6	79	0.04	121	0	6	79	0.06
STF*	3	-	-	0.05	-	-	-	0.02	-	-	-	-	0.03
SPIN08	3	84	0	0.08	2974	84	1506	2.04	2974	0	84	1506	2.93
FIB	3	8953	0	3.36	<185K	8953	92878	305	<185K	0	8953	92878	704
FIB*	3	-	-	0.74	-	-	-	81.0	-	-	-	-	133
CCNF(9)	9	16	0	0.05	49	16	46	0.07	49	0	16	46	0.06
CCNF(17)	17	256	0	0.15	97	256	94	5.76	97	0	256	94	6.09
CCNF(19)	19	512	0	0.28	109	512	106	22.5	109	0	512	106	22.0
SSB	5	4	2	0.05	48	4	38	0.03	46	1	4	37	0.03
Ssb(1)	5	22	14	0.06	245	23	143	0.11	237	4	23	140	0.11
SSB(3)	5	169	67	0.12	2798	172	1410	3.51	1179	48	90	618	0.90
SSB(4)	5	336	103	0.15	<7K	340	3333	20.3	2179	74	142	1139	2.07
SSB(8)	5	2014	327	0.85	<67K	2022	32782	4118	<12K	240	470	6267	32.1

Lastly, we note that this cutoff approach imposes no liability on what events shall be kept in the prefix, set G can be cleaned at discretion. Also, redefining (7) to use adequate orders [5] is straightforward, cf. App. C.1 (in our proofs we actually assume adequate orders).

- Looks good! Err... **not so simple**. Depending on how you do this, you quickly move from polynomial to exponential.
- **Experiments!** \triangle Possible 404, code not found! ahead!!!

JUST COMPARE TO THE ALGORITHM THEY PROPOSED

Rodriguez et al., CONCUR'15

- Looks good! Err... **not so simple**. Depending on how you do this, you quickly move from polynomial to exponential.
- **Experiments!** \triangle Possible 404, code not found! ahead!!!
- Wait, what's this language? Did this ever run one day?

JUST COMPARE TO THE ALGORITHM THEY PROPOSED

Rodriguez et al., CONCUR'15

- Looks good! Err... **not so simple**. Depending on how you do this, you quickly move from polynomial to exponential.
- **Experiments!** \triangle Possible 404, code not found! ahead!!!
- Wait, what's this language? Did this ever run one day?
- Wow, I'll need **CPLEX** and **Gurobi** but all I have is **lpsolve** or **glpk**

JUST COMPARE TO THE ALGORITHM THEY PROPOSED

Rodriguez et al., CONCUR'15

- Looks good! Err... **not so simple**. Depending on how you do this, you quickly move from polynomial to exponential.
- **Experiments!** \triangle Possible 404, code not found! ahead!!!
- Wait, what's this language? Did this ever run one day?
- Wow, I'll need **CPLEX** and **Gurobi** but all I have is **lpsolve** or **glpk**
- Sweet, they provided a **binary**! Oh, wait, MacOSX in 2015 ?!?

JUST COMPARE TO THE ALGORITHM THEY PROPOSED

Rodriguez et al., CONCUR'15

- Looks good! Err... **not so simple**. Depending on how you do this, you quickly move from polynomial to exponential.
- **Experiments!** \triangle Possible 404, code not found! ahead!!!
- Wait, what's this language? Did this ever run one day?
- Wow, I'll need **CPLEX** and **Gurobi** but all I have is **lpsolve** or **glpk**
- Sweet, they provided a **binary**! Oh, wait, MacOSX in 2015 ?!?
- The GitHub webpage says it requires Foo, Bar, and Baz, but none of the **versions** I find appear to work.

JUST COMPARE TO THE ALGORITHM THEY PROPOSED

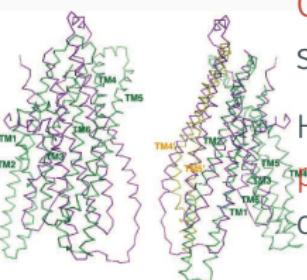
Rodriguez et al., CONCUR'15

- Looks good! Err... **not so simple**. Depending on how you do this, you quickly move from polynomial to exponential.
- **Experiments!** \triangle Possible 404, code not found! ahead!!!
- Wait, what's this language? Did this ever run one day?
- Wow, I'll need **CPLEX** and **Gurobi** but all I have is **lpsolve** or **glpk**
- Sweet, they provided a **binary**! Oh, wait, MacOSX in 2015 ?!?
- The GitHub webpage says it requires Foo, Bar, and Baz, but none of the **versions** I find appear to work.
- With which **parameters** and data set do you run this code? And Why?

In the end, **one new thesis** to understand this paper and contribute.

BLAMING "COMPUTER SCIENCE"

How COMPUTERS BROKE SCIENCE



Geoffrey Chang (Scripps, UCSD) works on crystallography and studies the structure of cell membrane proteins.

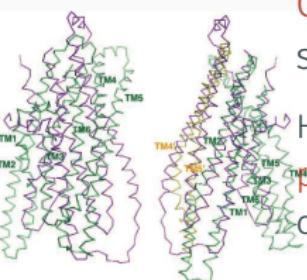
He specialized in structures of **multidrug resistant transporter proteins in bacteria**: MsbA de Escherichia Choli (Science, 2001), Vibrio cholera (Mol. Biology, 2003), Salmonella typhimurium (Science, 2005)

2006: Inconsistencies reveal a programming mistake

A homemade data-analysis program had flipped two columns of data, inverting the electron-density map from which his team had derived the protein structure.

5 retractions that motivate improved software engineering practices in comp. biology

How COMPUTERS BROKE SCIENCE



Geoffrey Chang (Scripps, UCSD) works on crystallography and studies the structure of cell membrane proteins.

He specialized in structures of **multidrug resistant transporter proteins in bacteria**: MsbA de Escherichia Choli (Science, 2001), Vibrio cholera (Mol. Biology, 2003), Salmonella typhimurium (Science, 2005)

2006: Inconsistencies reveal a programming mistake

A homemade data-analysis program had flipped two columns of data, inverting the electron-density map from which his team had derived the protein structure.

5 retractions that motivate improved software engineering practices in comp. biology

COMPUTERS...

How computers broke science – and what we can do about it

Most modern science is so complicated, and most journal articles so brief, it's impossible for the article to include details of many important methods and decisions made by the researcher as he analyzed his data on his computer. How, then, can another researcher judge the reliability of the results, or reproduce the analysis?



– Ben Marwick,
The conversation, 2015

Point-and-click procedures are rampant but they hinder reproducibility.

Spreadsheets are generalized and intensively used in biology:

- **Membrane-Associated Ring Finger (C3HC4) 1,**
E3 Ubiquitin Protein Ligase → **MARCH1** → 2016-03-01 →
1456786800
- **2310009E13** → 2.31E+19

And more recently, we had the **COVID tracing failure**.



Machine Learning: Trouble at the lab, The Economist 2013

According to some estimates, three-quarters of published scientific papers in the field of machine learning are bunk because of this "overfitting".

– Alex "Sandy" Pentland

The Reproducibility Crisis in ML-based science (Princeton workshop 2022)

Reproducibility failures in ML-based science are systemic. We found 20 reviews across 17 scientific fields (medicine, neuroimaging, autism diagnosis, genomics, computer security, ...) that find errors in a total of 329 papers that use ML-based science and in some cases leading to wildly overoptimistic conclusion. [...] complex ML models don't perform substantively better than decades-old LR models.

Data leakage: spurious relationship between the independent variables and the target variable that arises as an artifact of the data collection, sampling, or pre-processing strategy.

– S. Kapoor and A. Narayanan

THIS IS ABOUT COMPUTATIONAL SCIENCE. SHOULD MATHEMATICIANS CARE?

Computer Science is young and inherits from Mathematics, Engineering,
Linguistic, Nat. Sciences, ...

Purely theoretical scientists whose practice is close to mathematics may not be concerned (can't publish a math article without releasing the proofs).

THIS IS ABOUT COMPUTATIONAL SCIENCE. SHOULD MATHEMATICIANS CARE?

Computer Science is young and inherits from Mathematics, Engineering,
Linguistic, Nat. Sciences, ...

Purely theoretical scientists whose practice is close to mathematics may not be concerned (can't publish a math article without releasing the proofs).

Yet, incoherencies are common, especially in a fast moving field:

- E.g., definitions/concepts in book/article A and B are *slightly different* and the resulting theorems cannot be mixed
- Have a look at Vladimir Voevodsky's talk in 2014 at Princeton 😊
- ERC Nano bubbles: how, when and why does science fail to correct itself?

Flagging incorrect nucleotide sequence reagents in biomedical papers:

To what extent does the leading publication format impede automatic
error detection?
(Labbe et al., 2020)

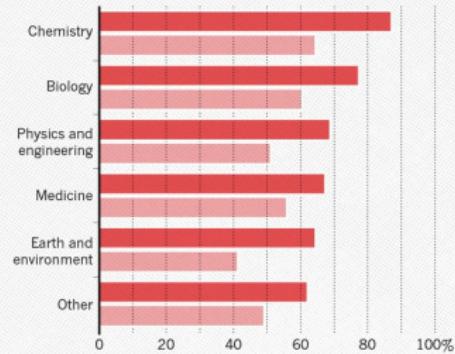
DIFFERENT KINDS OF REPRODUCIBILITY

SOCIO-TECHNICAL CHALLENGES

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.

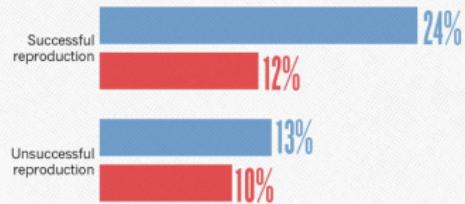
● Someone else's ● My own



HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.

● Published ● Failed to publish



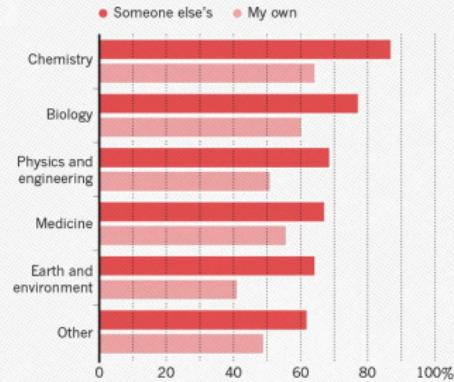
1,500 scientists lift the lid on reproducibility,

Nature, May 2016

SOCIO-TECHNICAL CHALLENGES

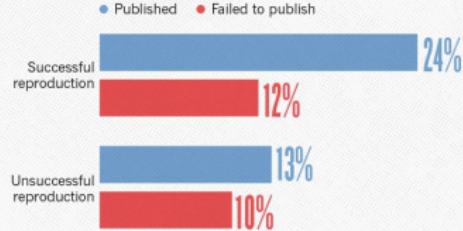
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.



Number of respondents from each discipline:
Biology 703, Chemistry 106, Earth and environmental 95,
Medicine 203, Physics and engineering 236, Other 233.

1,500 scientists lift the lid on reproducibility,

Nature, May 2016

Social causes

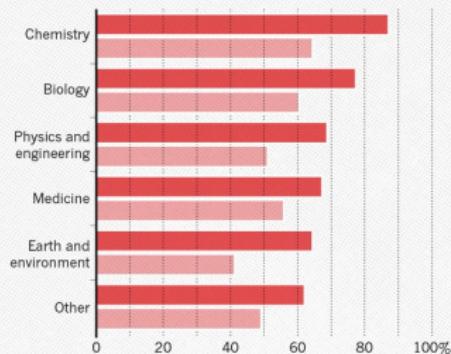
- Fraud, conflict of interest (pharmaceutic, ...)
- No incentive to reproduce/check our own work (afap), nor the work of others (big results!), nor to allow others to check (competition)
- Peer review does not scale: 1M+ articles per year!

SOCIO-TECHNICAL CHALLENGES

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.

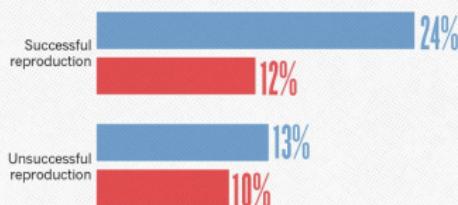
● Someone else's ● My own



HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.

● Published ● Failed to publish



Number of respondents from each discipline:

Chemistry 106, Biology 703, Physics and engineering 236, Medicine 203, Earth and environmental 95, Other 233.

1,500 scientists lift the lid on reproducibility,

Nature, May 2016

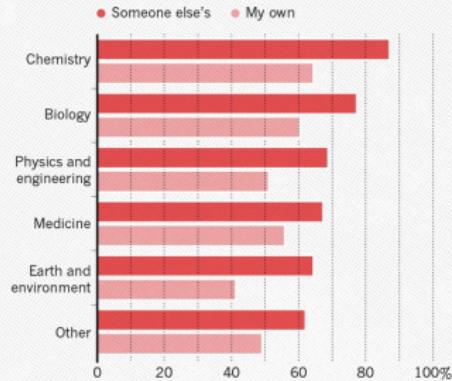
Social causes

- Fraud, conflict of interest (pharmaceutic, ...)
- No incentive to reproduce/check our own work (afap), nor the work of others (big results!), nor to allow others to check (competition)
- Peer review does not scale: 1M+ articles per year!
- Emerging practices: DORA/Plan S/COARA, DMP and FAIR data, artefact evaluation, reproducibility badges, reproducibility challenges, open reviews, ...

SOCIO-TECHNICAL CHALLENGES

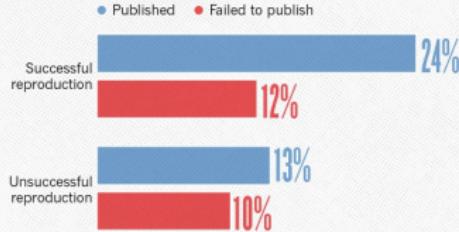
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.



Number of respondents from each discipline:
Biology 703, Chemistry 106, Earth and environmental 95,
Medicine 203, Physics and engineering 236, Other 233.

1,500 scientists lift the lid on reproducibility,

Nature, May 2016

Social causes

- Fraud, conflict of interest (pharmaceutic, ...)
- No incentive to reproduce/check our own work (afap), nor the work of others (big results!), nor to allow others to check (competition)
- Peer review does not scale: 1M+ articles per year!
- Emerging practices: DORA/Plan S/COARA, DMP and FAIR data, artefact evaluation, reproducibility badges, reproducibility challenges, open reviews, ...

Methodological/technical causes

- The many biases (apophenia, confirmation, hindsight, experimenter, ...): bad designs
- Selective reporting, weak analysis (statistics, data manipulation mistakes, computational errors)
- Lack of information, code/raw data unavailable

NO TRANSPARENCY NO CONSENSUS



DIFFERENT REPRODUCIBILITY CONCERNS IN MODERN SCIENCE

Biology, Oncology sample provenance, clinical trials \rightsquigarrow standardized protocols

Psychology, Nutrition HARKING, p-hacking \rightsquigarrow pre-registration

DIFFERENT REPRODUCIBILITY CONCERNS IN MODERN SCIENCE

Biology, Oncology sample provenance, clinical trials \rightsquigarrow standardized protocols

Psychology, Nutrition HARKING, p-hacking \rightsquigarrow pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical chaos, parallel architectures

DIFFERENT REPRODUCIBILITY CONCERN IN MODERN SCIENCE

Biology, Oncology sample provenance, clinical trials \rightsquigarrow standardized protocols

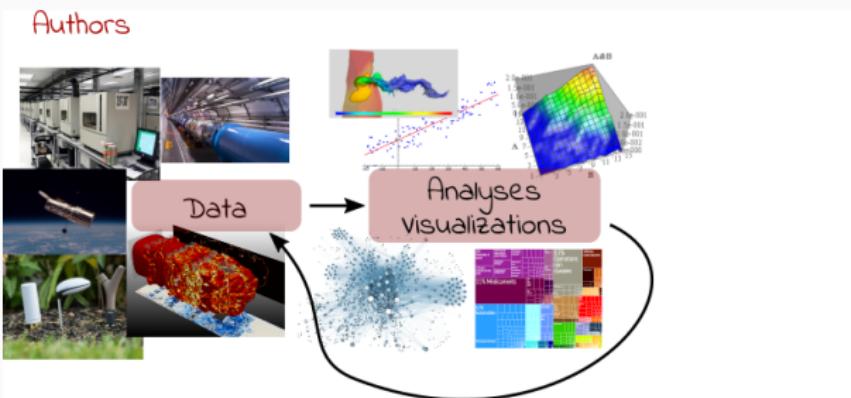
Psychology, Nutrition HARKING, p-hacking \rightsquigarrow pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical chaos, parallel architectures

Artificial Intelligence most of the above 😊

The processing steps between raw observations and findings have gotten increasingly numerous and complex



DIFFERENT REPRODUCIBILITY CONCERN IN MODERN SCIENCE

Biology, Oncology sample provenance, clinical trials \rightsquigarrow standardized protocols

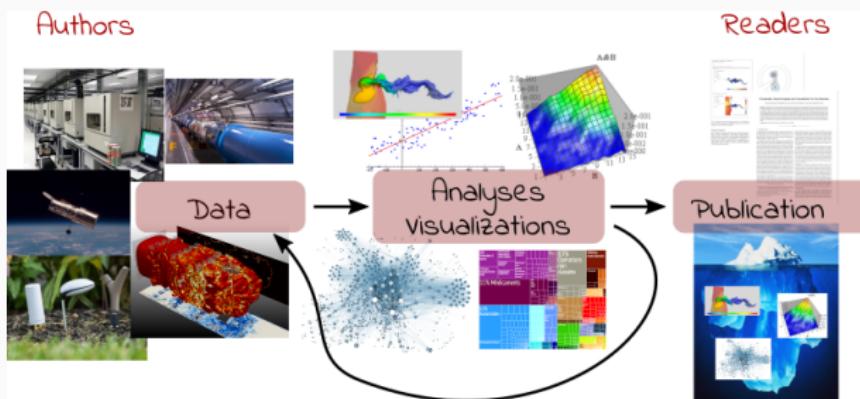
Psychology, Nutrition HARKING, p-hacking \rightsquigarrow pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical chaos, parallel architectures

Artificial Intelligence most of the above 😊

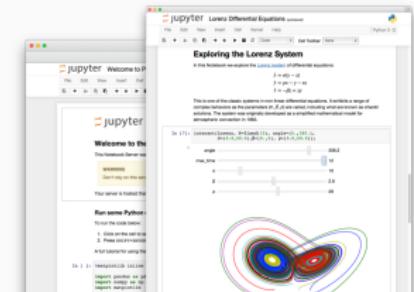
The processing steps between raw observations and findings have gotten increasingly numerous and complex



Reproducible Research = Bridging the Gap by working Transparently

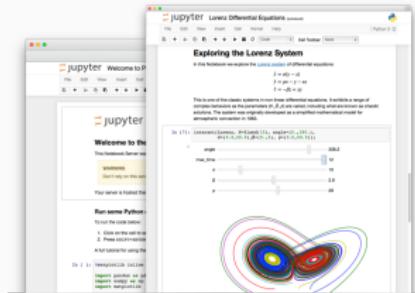
REPRODUCIBILITY ISSUES RELATED TO THE USE OF COMPUTERS

Computation provenance: notebooks and workflows

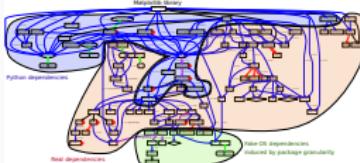


REPRODUCIBILITY ISSUES RELATED TO THE USE OF COMPUTERS

Computation provenance: notebooks and workflows



Software environments

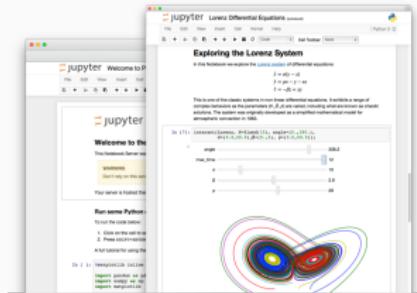


ReproZip



REPRODUCIBILITY ISSUES RELATED TO THE USE OF COMPUTERS

Computation provenance: notebooks and workflows



Software environments



Sharing and Archiving



GOOD PRACTICE #1

TAKING NOTES AND DOCUMENTING

FRUSTRATION AS AN AUTHOR/REVIEWER



Author

- I thought I used the same parameters but **I'm getting different results!**
- The new student wants to compare with **the method I proposed last year**
- My advisor asked me whether I took care of setting this or this but **I can't remember**
- The damned fourth reviewer asked for a major revision and wants me to change **Figure 3**. Which code and which data set did I use?
- **It worked yesterday!** 6 months later: Why did I do that?

Reviewer

- As usual, there is **no confidence interval**, I wonder about the variability and whether the difference is **significant** or not
- That can't be true, I'm sure **they removed some points**
- Why is this graph in logscale? **How would it look like otherwise?** I'm not even **sure** of what this value means. If only I could access the generation script

TOOL 1: COMPUTATIONAL NOTEBOOKS/LITTERATE PROGRAMMING

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

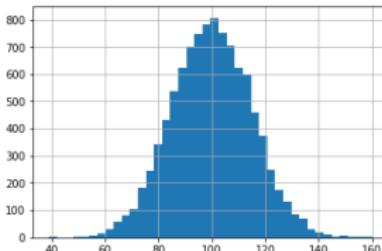
3.141592653589793

Mais calculé avec la méthode des [aiguilles de Buffon](#), on obtiendrait comme approximation :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/(sum((x+np.sin(theta))>1)/N)
```

3.1437198694098765

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1: COMPUTATIONAL NOTEBOOKS/LITTERATE PROGRAMMING

Document initial dans son environnement

The screenshot shows a Jupyter Notebook interface with the following details:

- Title:** # Un document computationnel
- In [1]:** A code cell containing:

```
from math import *
print(pi)
3.141592653589793
```

A note below it says: "Mais calculé avec la [méthode des aiguilles de Buffon](#) (https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtient l'air comme approximation :".
- In [2]:** A code cell containing:

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/sum((x+np.sin(theta))>1)/N
```

A note below it says: "On peut inclure des formules mathématiques comme $\sqrt{2/\pi} \exp(-x^2/2)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation...)."
- In [3]:** A code cell containing:

```
%matplotlib inline
import matplotlib.pyplot as plt
mu, sigma = 100, 15
x = mu + sigma*np.random.randn(10000)
plt.hist(x,40)
plt.grid(True)
plt.show()
```

A histogram plot showing a bell-shaped curve centered at 100.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

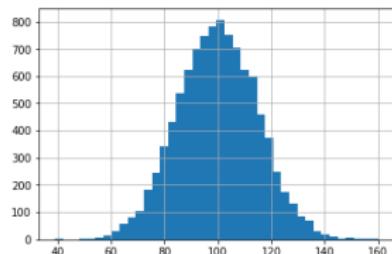
3.141592653589793

Mais calculé avec la [méthode des aiguilles de Buffon](#), on obtiendrait comme approximation :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/sum((x+np.sin(theta))>1)/N
```

3.1437198694998765

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation...).



TOOL 1: COMPUTATIONAL NOTEBOOKS/LITTERATE PROGRAMMING

Document initial dans son environnement

A screenshot of a Jupyter Notebook interface. The title cell contains the text '# Un document computationnel'. Below it, cell In [1] shows the Python code `from math import *` and `print(pi)`, resulting in the output 3.141592653589793. Cell In [2] contains code to calculate pi using theBuffon's needle method, with the result 3.14371986944998765. Cell In [3] shows a histogram of a normal distribution with mu=100 and sigma=15, centered at 100.

MarkDown

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

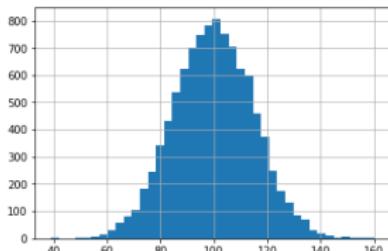
3.141592653589793

Mais calculé avec la [méthode des aiguilles de Buffon](#) (https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtiendrait comme approximation :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/(sum((x+np.sin(theta))>1))/N
```

3.14371986944998765

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1: COMPUTATIONAL NOTEBOOKS/LITTERATE PROGRAMMING

Document initial dans son environnement

The screenshot shows a Jupyter Notebook interface with three code cells:

- In [1]:** A cell containing a single line of Python code: `from math import * print(pi)`. The output is the value `3,141592653589793`.
- In [2]:** A cell containing a more complex script. It imports numpy, sets N=1000000, generates random points (x, y), calculates theta, and performs a summation. The output is the value `3,1437198694098765`. A red arrow points from this cell to the explanatory text below.
- In [3]:** A cell containing a histogram command. It imports matplotlib, sets mu=100, sigma=15, generates data, and plots it. The resulting histogram is shown as a blue bell-shaped curve centered at 100.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

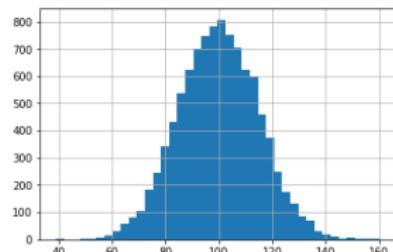
3.141592653589793

Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2*(sum((x+np.sin(theta))>1))/N
```

3.1437198694098765

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1: COMPUTATIONAL NOTEBOOKS/LITTERATE PROGRAMMING

Document initial dans son environnement

The screenshot shows a Jupyter Notebook interface with three code cells:

- In [1]:** Prints the value of pi as 3.141592653589793.
- In [2]:** Simulates Buffon's needle problem with N=1000000 needles. The output shows the calculated approximation of pi as 3.1437198694098765.
- In [3]:** Plots a histogram of 100,000 random numbers between 0 and 1. The plot is a bell-shaped curve centered around 0.5.

Annotations with red arrows point from the text "Résultats" to the output of each cell, indicating the final results of the computations.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

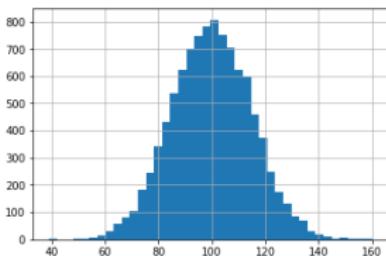
Mais calculé avec la méthode des [aiguilles de Buffon](#), on obtiendrait comme approximation :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2*(sum((x+np.sin(theta))>1))/N
```

3.1437198694098765

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et

des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1: COMPUTATIONAL NOTEBOOKS/LITTERATE PROGRAMMING

Document initial dans son environnement

Un document computationnel

```
In [1]:  
from math import *  
print(pi)  
3,141592653589793
```

Mais calculé avec la `_methodes_ des éimpulles de Buffon` (https://fr.wikipedia.org/wiki/Algille_de_Buffon), on obtiendrait comme `approximation` :

```
In [2]:  
import numpy as np  
N = 1000000  
x = np.random.uniform(size=N, low=0, high=1)  
theta = np.random.uniform(size=N, low=0, high=pi/2)  
2/(sum((x+np.sin(theta))>1))/N  
Out[2]: 3,1437198694098765
```

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).

```
In [3]:  
%matplotlib inline  
import matplotlib.pyplot as plt  
  
mu, sigma = 100, 15  
x = mu + sigma*np.random.randn(10000)  
  
plt.hist(x, 99)  
plt.grid(True)  
plt.show()
```

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

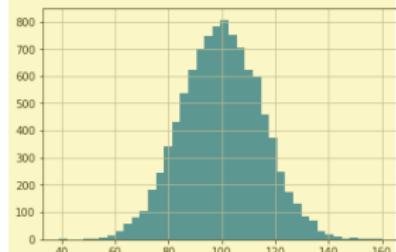
Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
import numpy as np  
N = 1000000  
x = np.random.uniform(size=N, low=0, high=1)  
theta = np.random.uniform(size=N, low=0, high=pi/2)  
2/(sum((x+np.sin(theta))>1))/N
```

3.1437198694098765

Export →

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1: COMPUTATIONAL NOTEBOOKS/LITTERATE PROGRAMMING

Document initial dans son environnement

The screenshot shows a Jupyter Notebook interface with three code cells:

- In [1]:** Prints the value of pi.
- In [2]:** Generates random points (x, y) and calculates the ratio of points inside the unit circle to the total number of points, which approximates pi.
- In [3]:** Plots a histogram of 100 random numbers between 0 and 150, showing a bell-shaped distribution centered around 100.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

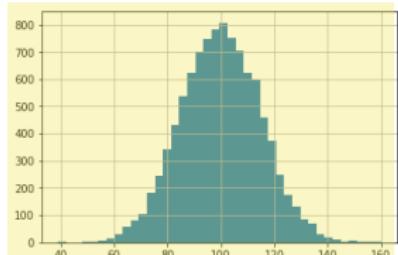
Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/(sum((x+np.sin(theta))>1)/N)
```

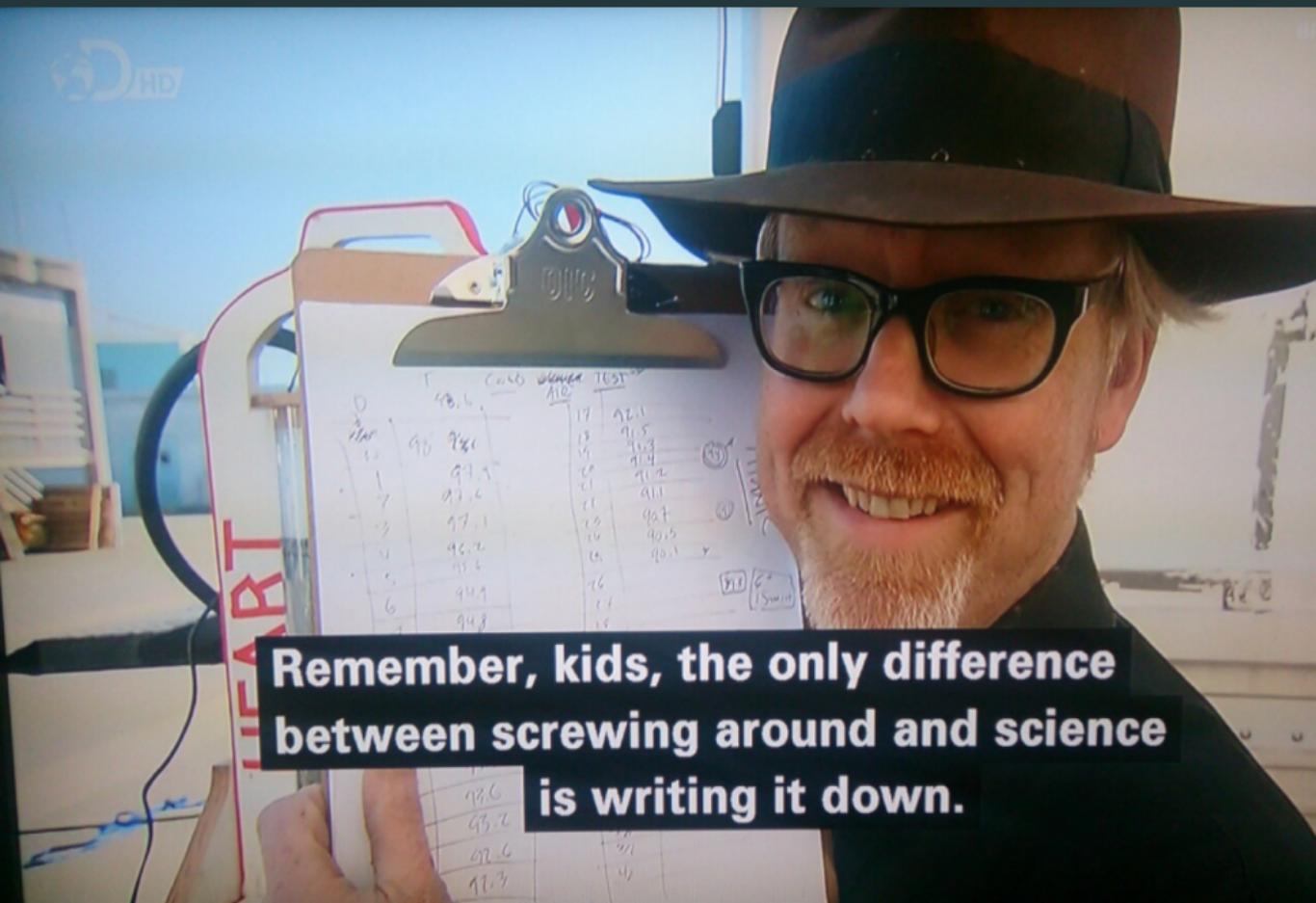
3.1437198694098765

Export

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1 BIS: LABORATORY NOTEBOOKS, COMPUTATIONAL DOCUMENTS



**Remember, kids, the only difference
between screwing around and science
is writing it down.**

TOOL 1 TER: WORKFLOWS

Notebooks are no panacea and do not help developing clean code

The screenshot shows a Jupyter Notebook interface with the title bar "jupyter example_pi.ipynb". The notebook contains several cells:

- In [1]:** A code cell containing:

```
# Un document computationnel

# Mon ordinateur n'indique que j'suis vers *approximativement*.

In [1]:
```

```
from math import *
print(pi)
3.141592653589793
```

Annotations above this cell say "Hide Prompt" and "Hide Code". Below the output, a note says "Mais calculé avec la `__method__` des (ajoutées de Buffet) `math.cmath` (http://hg.python.org/wk2/ajoute_le_Buffer), on obtiendrait comme approximation...".
- In [2]:** A code cell containing:

```
In [2]:
```

```
import numpy as np
n = 1000000
x = np.random.uniform(-np.pi, np.pi)
theta = np.random.uniform(0, np.pi/2)
if (x**2 + np.sin(theta)**2) > 1/n
```

Annotations above this cell say "Hide Prompt" and "Hide Code". Below the output, a note says "On peut inclure des formules mathématiques comme $\sqrt{frac(1)(pi)}$ ou $sqrt(2)$ dans les cellules de code. Le résultat sera alors normalisé, dessiné ou n'est rien à voir avec tout (si ce n'est une constante de normalisation...)."
- In [3]:** A code cell containing:

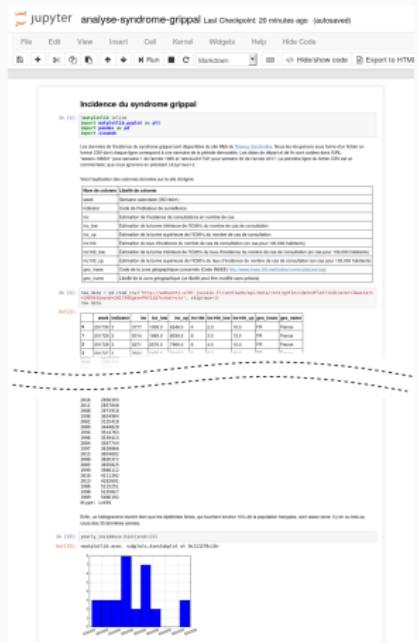
```
In [3]:
```

```
%matplotlib inline
import matplotlib.pyplot as plt
Nn, sigma = 100, 33
x = np.random.normal(0, sigma, Nn)
plt.hist(x, 40)
plt.xlabel('x')
plt.ylabel('y')
plt.show()
```

Annotations above this cell say "Hide Prompt" and "Hide Code". Below the output is a histogram plot showing a bell-shaped curve centered at zero, with x-axis values from -40 to 40 and y-axis values from 0 to 800.

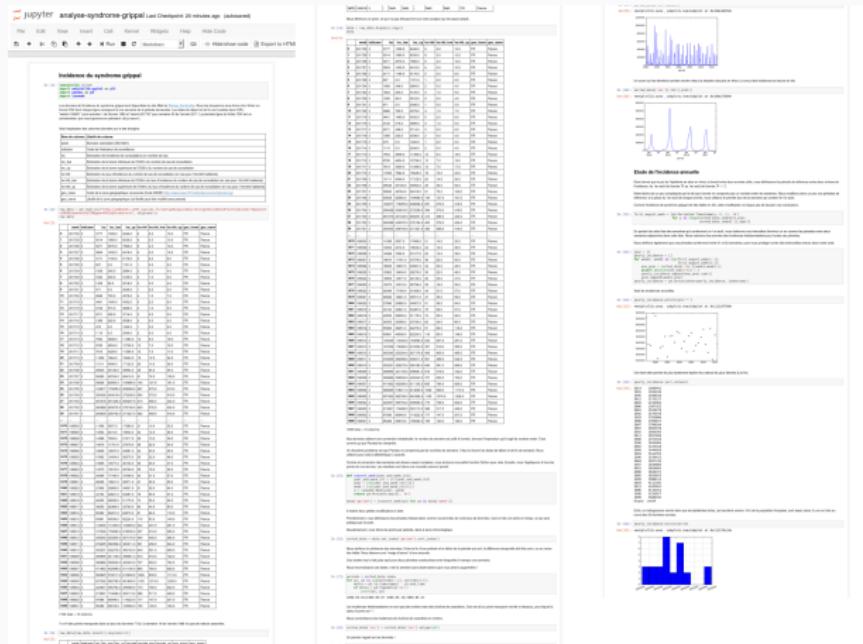
TOOL 1 TER: WORKFLOWS

Notebooks are no panacea and do not help developing clean code



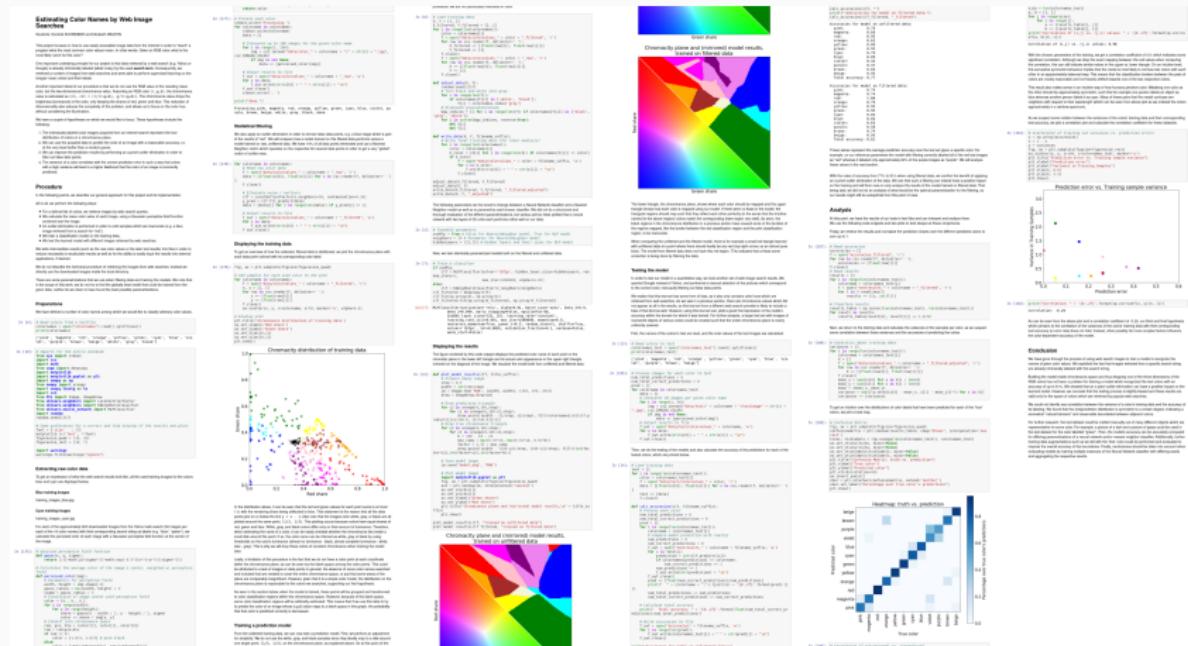
TOOL 1 TER: WORKFLOWS

Notebooks are no panacea and do not help developing clean code

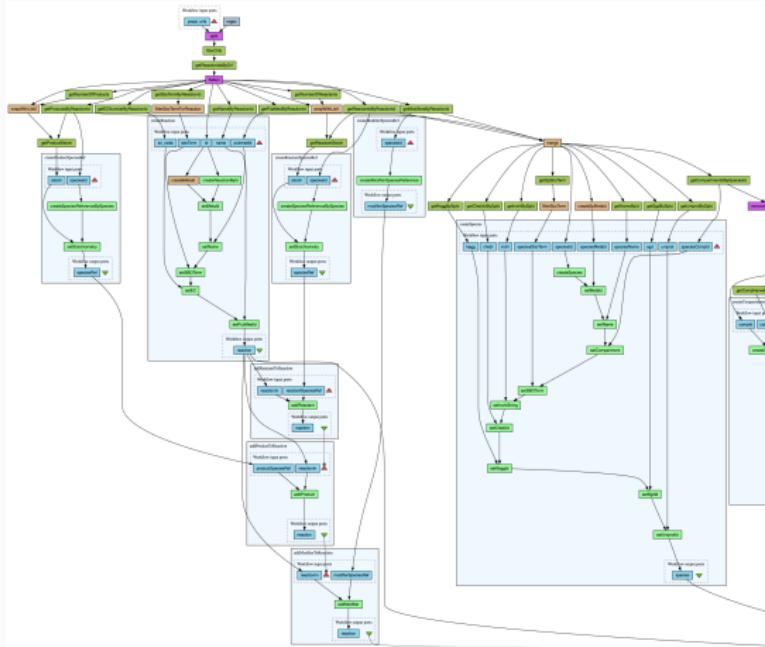
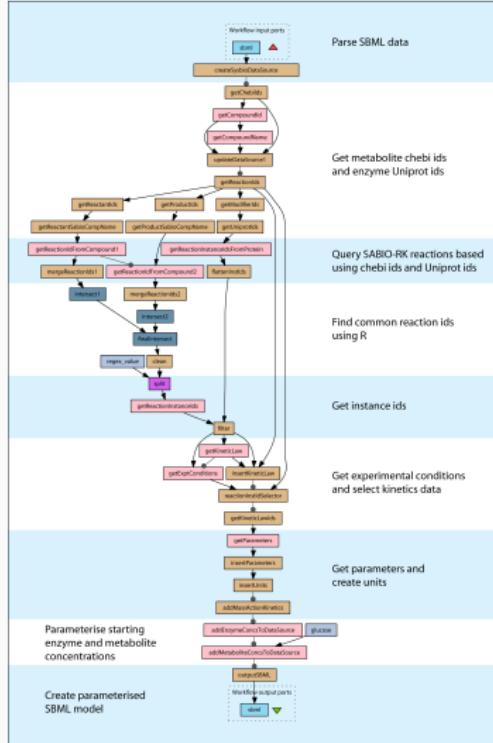


TOOL 1 TER: WORKFLOWS

Notebooks are no panacea and do not help developing clean code



TOOL 1 TER: WORKFLOWS



TOOL 1 TER: WORKFLOWS

Workflows:

- Clearer high-level view
- **Explicit** composition of codes and data movement
- Safer sharing, reusing, and execution
- Notebooks are a variant that is both impoverished and richer
 - No simple/mature path from a notebook to a workflow

Examples:

- Galaxy, Kepler, Taverna, Pegasus, Collective Knowledge, VisTrails
- Light-weight: `make`, dask, drake, swift, `snakemake`, ...
- Hybrids: SOS-notebook, ...

GOOD PRACTICE #2

CONTROLLING SOFTWARE ENVIRONMENT

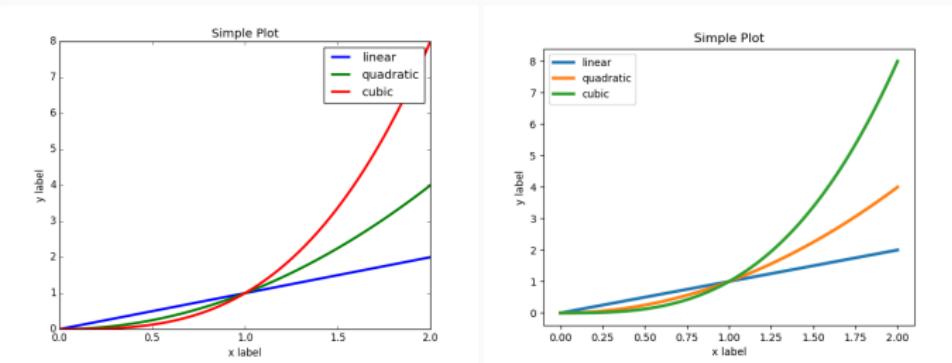
ARGH... DAMNED COMPUTERS

- Alice: I got 3.123123 Bob: I got segfault
- Damned! It used to work!!! Whenever I upgrade my computer, things break so I try to stay away from this 😞
- Whenever trying the code of my colleague, I had to install `libFoo-1.5c` and `pip install blah` but I broke everything and now neither his code nor mine works! 😞
- But hey! Here is my code. It's on GitHub so feel free to play with it! I'm doing open science 😊
 1. No one will ever run/use your code if it isn't **easy to install**
 2. No one will ever manage to run your code if you don't **document how to run it**
 3. Others (even you) are unlikely to get the same results unless you **control and share your software environment**

SOFTWARE DEPENDENCIES: HORROR STORIES

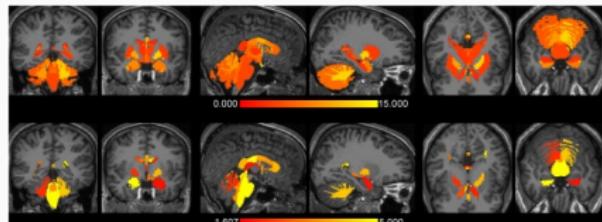
SOFTWARE DEPENDENCIES: HORROR STORIES

- Software environment evolution



SOFTWARE DEPENDENCIES: HORROR STORIES

- Software environment evolution
- OS heterogeneity



The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements (PLOS ONE, 2012)

Significant differences in volume and cortical thickness were revealed across FreeSurfer versions:

- volume: $8.8 \pm 6.6\%$ (range 1.3-**64.0%**)
- cortical thickness: $2.8 \pm 1.3\%$ (range 1.1-7.7%)

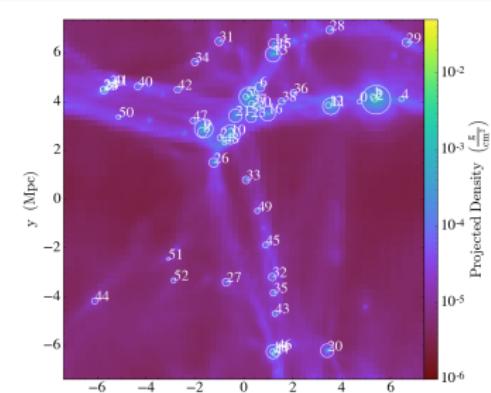
About a factor two smaller differences were found between the Mac and HP workstations and between Mac OSX 10.5 and OSX 10.6.

In the context of an ongoing study, users are discouraged to update to a new major release of either FreeSurfer or operating system.

Formal assessment of the accuracy of FreeSurfer is desirable.

SOFTWARE DEPENDENCIES: HORROR STORIES

- Software environment evolution
- OS heterogeneity
- Impact of the compiler

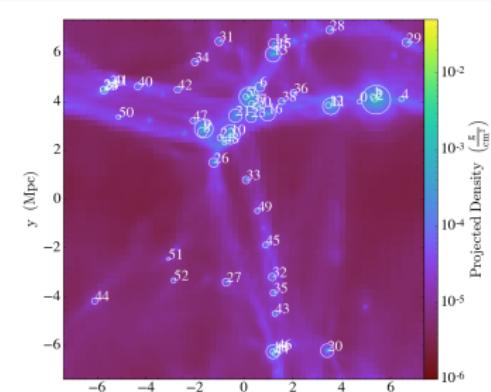


Assessing Reproducibility: An Astrophysical Example of Computational Uncertainty in the HPC Context (ResCuE-HPC, 2018)

Compiler	Optim.	Largest Halo Avg Mass.	Std. Err	Walltime
gcc@6.2.0	None	2.273E 46	1.069E 44	22h

SOFTWARE DEPENDENCIES: HORROR STORIES

- Software environment evolution
- OS heterogeneity
- Impact of the compiler

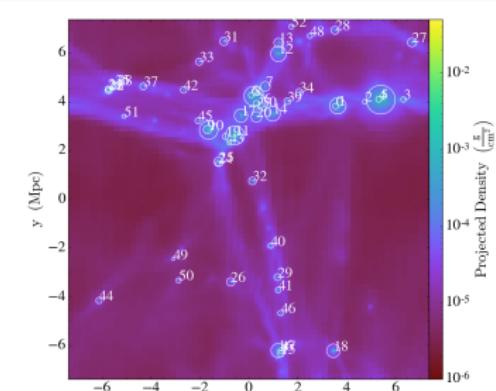


Assessing Reproducibility: An Astrophysical Example of Computational Uncertainty in the HPC Context (ResCuE-HPC, 2018)

Compiler	Optim.	Largest Halo Avg Mass.	Std. Err	Walltime
gcc@6.2.0	None	2.273E 46	1.069E 44	22h

SOFTWARE DEPENDENCIES: HORROR STORIES

- Software environment evolution
- OS heterogeneity
- Impact of the compiler

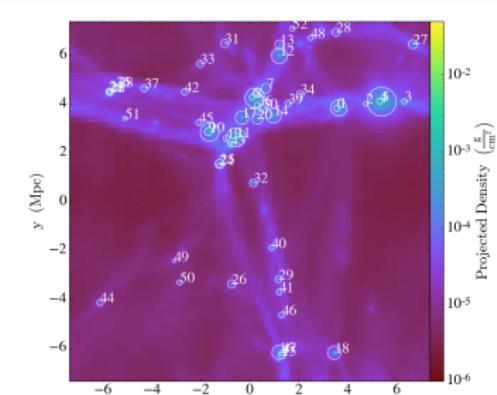


Assessing Reproducibility: An Astrophysical Example of Computational Uncertainty in the HPC Context (ResCuE-HPC, 2018)

Compiler	Optim.	Largest Halo		Walltime
		Avg Mass.	Std. Err	
gcc@6.2.0	None	2.273E 46	1.069E 44	22h
gcc@6.2.0	Normal	2.266E 46	1.218E 44	10h
gcc@6.2.0	High	2.275E 46	1.199E 44	9h

SOFTWARE DEPENDENCIES: HORROR STORIES

- Software environment evolution
- OS heterogeneity
- Impact of the compiler



Assessing Reproducibility: An Astrophysical Example of Computational Uncertainty in the HPC Context (ResCuE-HPC, 2018)

Compiler	Optim.	Largest Halo		Walltime
		Avg Mass.	Std. Err	
gcc@6.2.0	None	2.273E 46	1.069E 44	22h
gcc@6.2.0	Normal	2.266E 46	1.218E 44	10h
gcc@6.2.0	High	2.275E 46	1.199E 44	9h
intel@16.0.3	None	22.71 E 46	1.587E 44	39h
intel@16.0.3	Normal	43.30 E 46	1.248E 44	7h
intel@16.0.3	High	2.268E 46	1.414E 44	6h
cce@8.5.5	Low	43.11 E 46	1.353E 44	16h
cce@8.5.5	Normal	2.271E 46	1.261E 44	6h
cce@8.5.5	High	2.272E 46	1.341E 44	5h

COMPLEX ECOSYSTEMS

```
1 import matplotlib  
2 print(matplotlib.__version__)
```

3.5.1

COMPLEX ECOSYSTEMS

```
1 import matplotlib  
2 print(matplotlib.__version__)
```

3.5.1

```
1 apt show python3-matplotlib
```

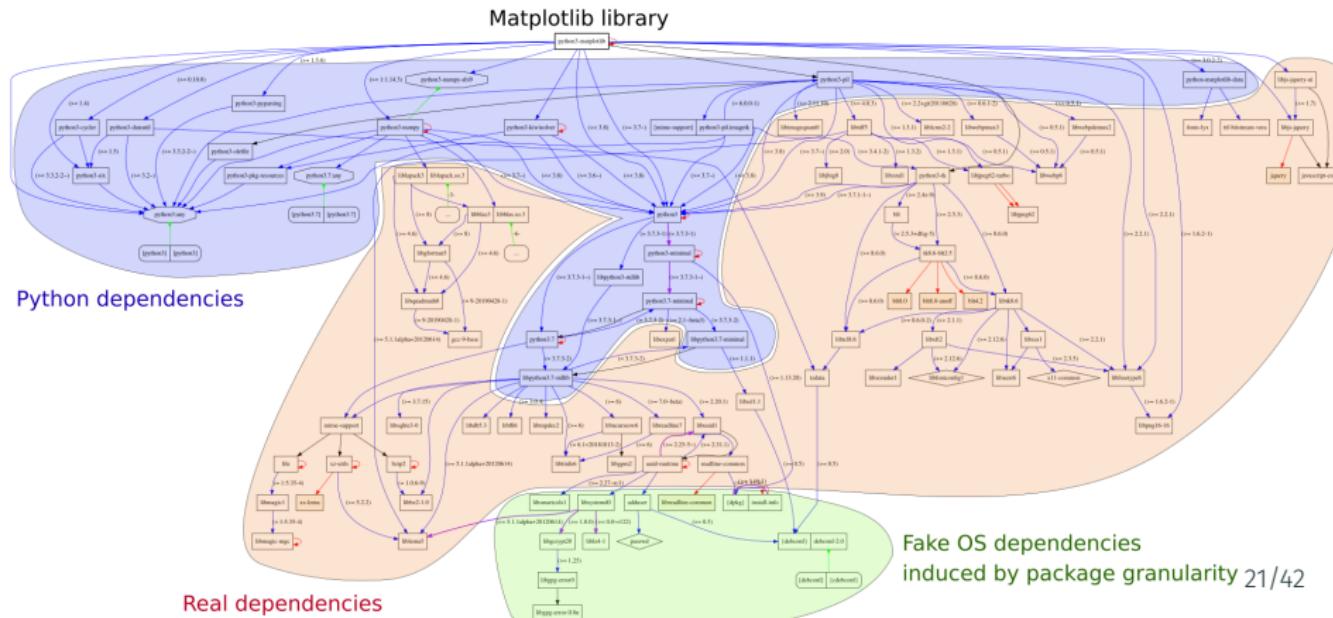
Package: python3-matplotlib
Version: 3.5.1-2+b1
Source: matplotlib (3.5.1-2)
Maintainer: Sandro Tosi <morph@debian.org>
Installed-Size: 27.6 MB
Depends: libjs-jquery, libjs-jquery-ui, python-matplotlib-data (>= 3.5.1),
 python3-dateutil, python3-pil.imagetk, python3-pyparsing (>= 1.5.6),
 python3-six (>= 1.4), python3-numpy (>= 1:1.20.0), python3-numpy-
 abi9,
 python3 (<< 3.11), python3 (>= 3.9~), python3-cycler (>= 0.10.0),
 python3-fonttools, python3-kiwisolver, python3-packaging, python3-
 pil,
 python3:any, libc6 (>= 2.29), libfreetype6 (>= 2.2.1),
 libgcc-s1 (>= 3.3.1), libqhull-r8.0 (>= 2020.1), libstdc++6 (>= 11)
Recommends: python3-tk
Suggests: dvipng, ffmpeg, fonts-staypuft, ghostscript, gir1.2-gtk-3.0, inkscape,
 ipython3, librsvg2-common, python-matplotlib-doc, python3-cairoffi,
 python3-gi, python3-gi-cairo, python3-gobject, python3-pyqt5,
 python3-scipy, python3-sip, python3-tornado, texlive-extra-utils

COMPLEX ECOSYSTEMS

```
1 import matplotlib  
2 print(matplotlib.__version__)
```

3.5.1

```
1 apt show python3-matplotlib
```



TOOL 2: CONTAINERS AND PACKAGE MANAGERS

The good



The bad



The ugly



Automatic tracking

TOOL 2: CONTAINERS AND PACKAGE MANAGERS

The good



The bad



The ugly



Automatic tracking

Containers

- Pros: Lightweight, Good isolation, Easy to use
 - Running as easy as `docker run <cmd>`
 - Building images: `docker build -f <Dockerfile>`
 - Sharing through the Docker Hub: `docker pull/push `

TOOL 2: CONTAINERS AND PACKAGE MANAGERS

The good



The bad



The ugly



Automatic tracking

Containers

- **Pros:** Lightweight, Good isolation, Easy to use
- **Cons:** Opaque, Container build is generally not reproducible
 - Recipes rarely follow *reproducible good practices*

```
1   FROM ubuntu:20.04
2   RUN apt-get update
3       && apt-get upgrade -y
4       && apt-get install -y ...
```

- Choose a stable image (and the smallest possible)
- Include only the necessary libraries (e.g. no graphics libs)
- Avoid system updates (instead freeze sources)

TOOL 2: CONTAINERS AND PACKAGE MANAGERS

The good



The bad



The ugly



Automatic tracking

Containers

- Pros: Lightweight, Good isolation, Easy to use
- Cons: Opaque, Container build is generally not reproducible

Package managers (the ugly and the good)

- Language specific: `pip/pipenv/virtualenv`, `conda`, `CRAN/Bioconductor`
 - Limits: version management, durability, permeable, language centric
- **GUIX/NiX** = Full-fledged functional package manager
 - Native support for environment (*à la git*)
 - Isolation through `--pure`
 - Recompile from source (cache recommended)

DEBIAN DEPENDENCIES

```
1 dpkg --status python3-matplotlib
```

```
Package: python3-matplotlib
Version: 3.6.3-1+b1
Source: matplotlib (3.6.3-1)
Depends: libjs-jquery, libjs-jquery-ui, python-matplotlib-data (>= 3.6.3),
          python3-dateutil, python3-pil.imagetk, python3-pyparsing (>= 1.5.6),
          python3-six (>= 1.4), python3-numpy (>= 1:1.22.0), python3-contourpy,
          python3 (<< 3.12), python3 (>= 3.11~), python3-numpy-abi9,
          python3-cycler (>= 0.10.0), python3-fonttools, python3-kiwisolver,
          python3-packaging, python3-pil, python3:any, libc6 (>= 2.34),
          libfreetype6 (>= 2.2.1), libgcc-s1 (>= 3.3.1),
          libqhull-r8.0 (>= 2020.1), libstdc++6 (>= 11)
```

DEBIAN DEPENDENCIES

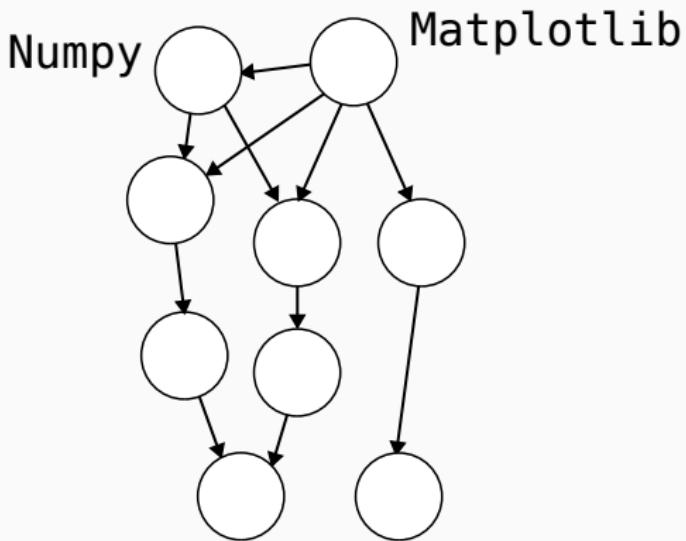
```
1 dpkg --status python3-matplotlib
```

```
Package: python3-matplotlib
Version: 3.6.3-1+b1
Source: matplotlib (3.6.3-1)
Depends: libjs-jquery, libjs-jquery-ui, python-matplotlib-data (>= 3.6.3),
          python3-dateutil, python3-pil.imagetk, python3-pyparsing (>= 1.5.6),
          python3-six (>= 1.4), python3-numpy (>= 1:1.22.0), python3-contourpy,
          python3 (<< 3.12), python3 (>= 3.11~), python3-numpy-abi9,
          python3-cycler (>= 0.10.0), python3-fonttools, python3-kiwisolver,
          python3-packaging, python3-pil, python3:any, libc6 (>= 2.34),
          libfreetype6 (>= 2.2.1), libgcc-s1 (>= 3.3.1),
          libqhull-r8.0 (>= 2020.1), libstdc++6 (>= 11)
```

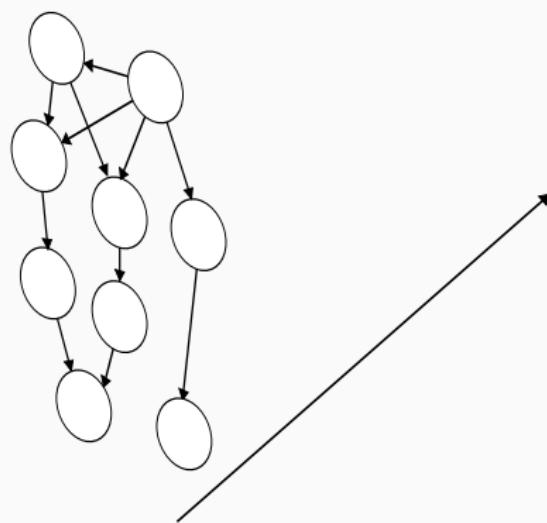
On a given day:

- Several versions of each package are available on the server
- Installing the latest version of a package may require upgrading some other packages

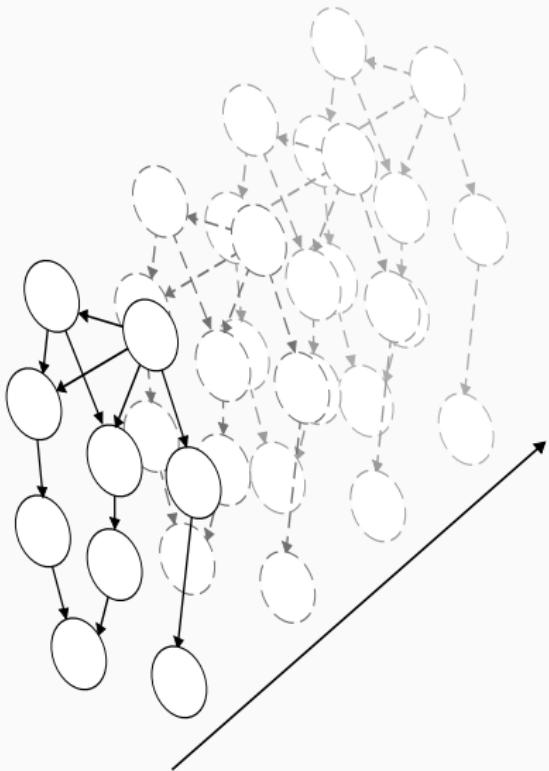
LOOSE VS. STRICT DEPENDENCIES IN PICTURE



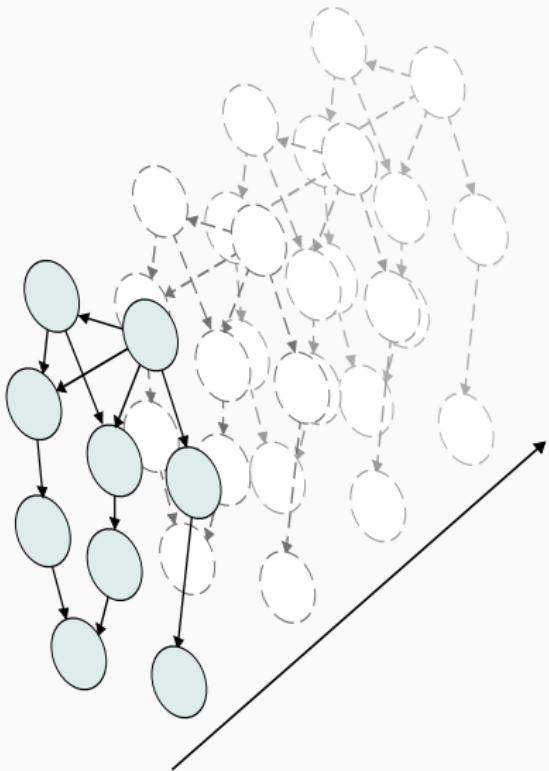
LOOSE VS. STRICT DEPENDENCIES IN PICTURE



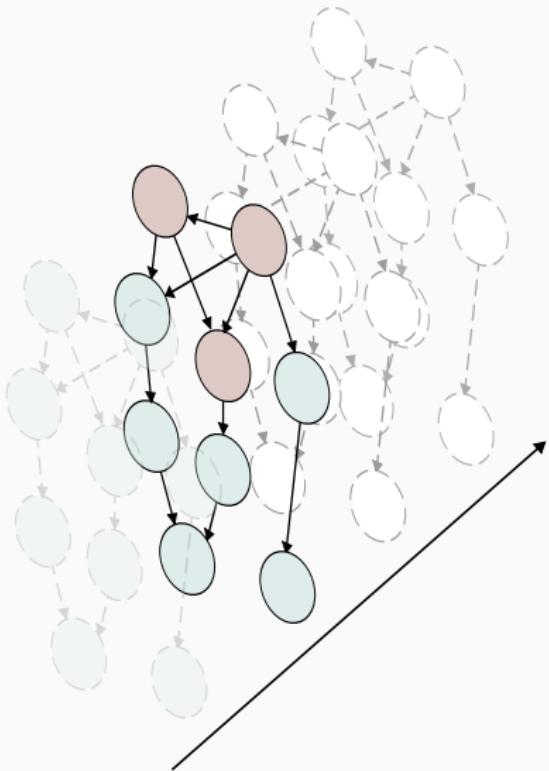
LOOSE VS. STRICT DEPENDENCIES IN PICTURE



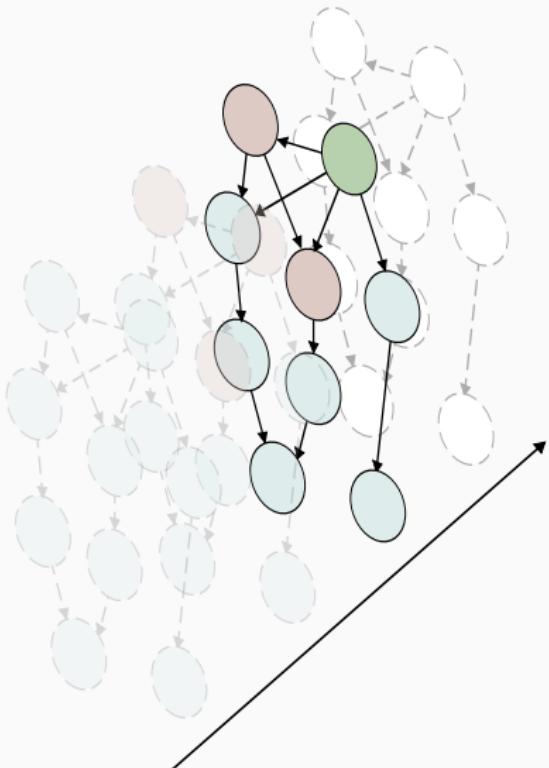
LOOSE VS. STRICT DEPENDENCIES IN PICTURE



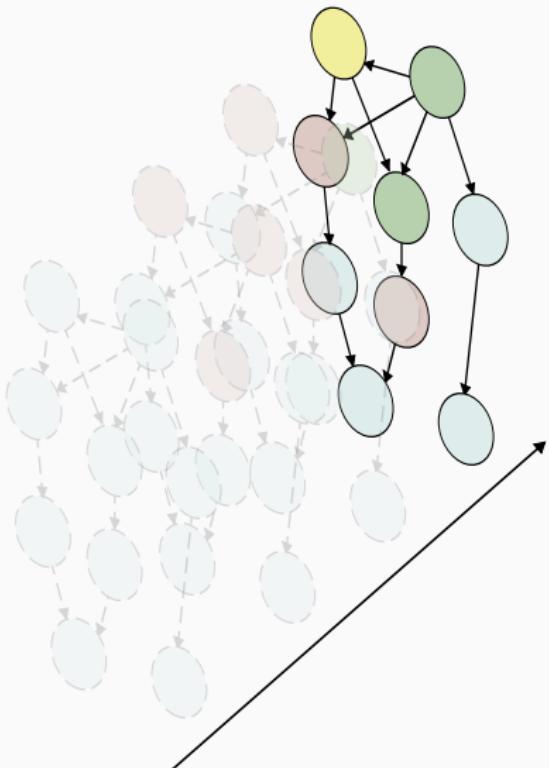
LOOSE VS. STRICT DEPENDENCIES IN PICTURE



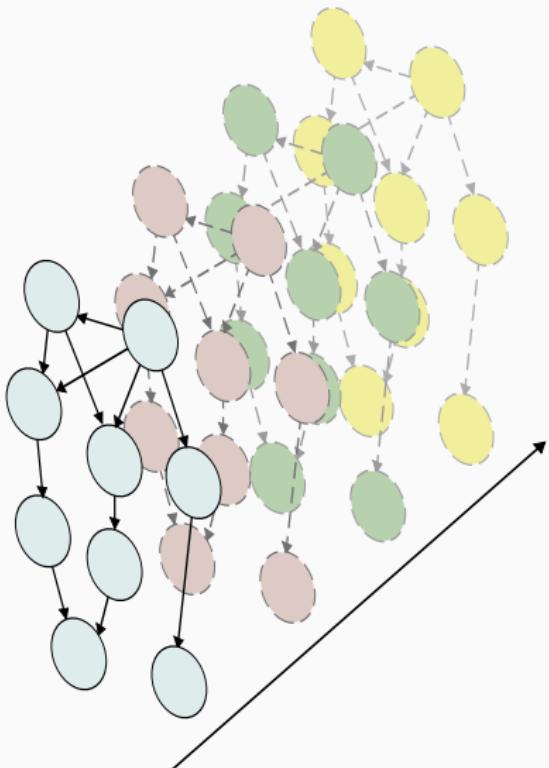
LOOSE VS. STRICT DEPENDENCIES IN PICTURE



LOOSE VS. STRICT DEPENDENCIES IN PICTURE



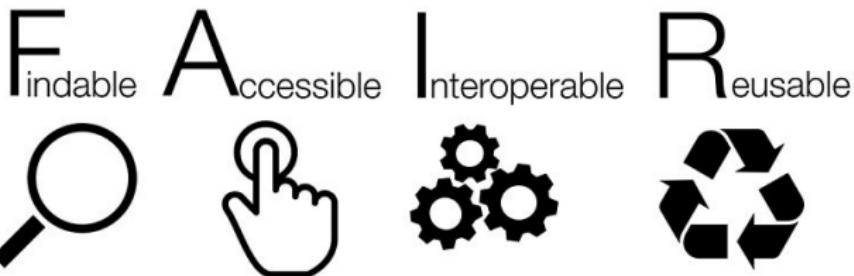
LOOSE VS. STRICT DEPENDENCIES IN PICTURE



GOOD PRACTICE #3

VERSION CONTROL AND ARCHIVING

FAIR PRINCIPLES



<https://www.go-fair.org/fair-principles/>

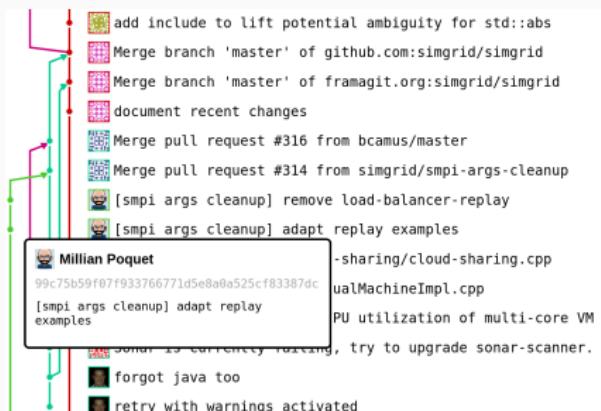
- "*Open as much as possible and close as much as necessary*"
- Management, publication, annotation (metadata), archiving
- Source code = specific data with specific consideration

Let's go beyond general principles!

TOOL 3: VERSION CONTROL AND FORGE

Git = version control

- Developed in 2005 by Linus Torvalds for the kernel development
- Local and efficient rollbacks
- Distributed: everyone has a full copy of the history



GitHub, GitLab, and Co

- Free hosting of public projects, social network



Limitation

- Managing large data: **Git LFS** **Git Annex** (or DataLad)

TOOL 3BIS: FIGHTING INFORMATION LOSS WITH ARCHIVES



or



= awesome collaborations (\neq archive)

- D. Spinellis. *The Decay and Failures of URL References*. CACM, 46(1), 2003
The half-life of a referenced URL is approximately 4 years from its publication date.
- P. Habibzadeh. *Decay of References to Web sites in Articles Published in General Medical Journals: Mainstream vs Small Journals*. Applied Clinical Informatics. 4 (4), 2013
half life ranged from 2.2 years in EMHJ to 5.3 years in BMJ
- Discontinued forges: Code Space, Gitorious, Google code, Inria Gforge

TOOL 3BIS: FIGHTING INFORMATION LOSS WITH ARCHIVES



or



= awesome collaborations (\neq archive)

- D. Spinellis. *The Decay and Failures of URL References*. CACM, 46(1), 2003
The half-life of a referenced URL is approximately 4 years from its publication date.
- P. Habibzadeh. *Decay of References to Web sites in Articles Published in General Medical Journals: Mainstream vs Small Journals*. Applied Clinical Informatics. 4 (4), 2013
half life ranged from 2.2 years in EMHJ to 5.3 years in BMJ
- Discontinued forges: Code Space, Gitorious, Google code, Inria Gforge

Article archives



Data archives



figshare



Software Archive

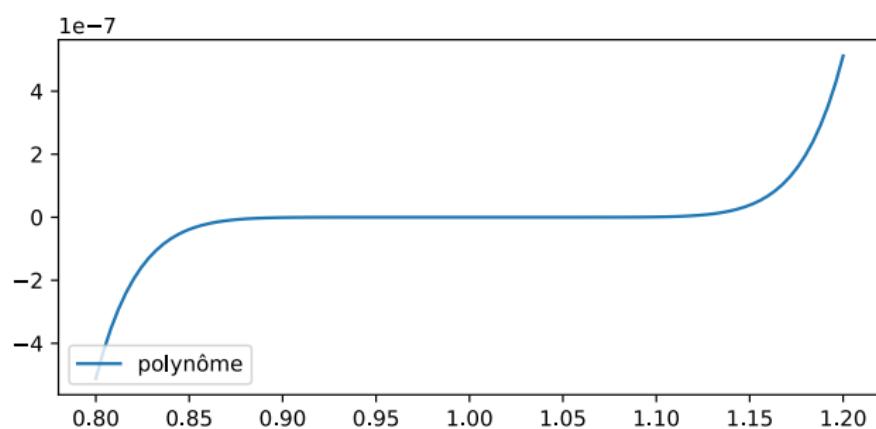


Software Heritage

Collect/Preserve/Share

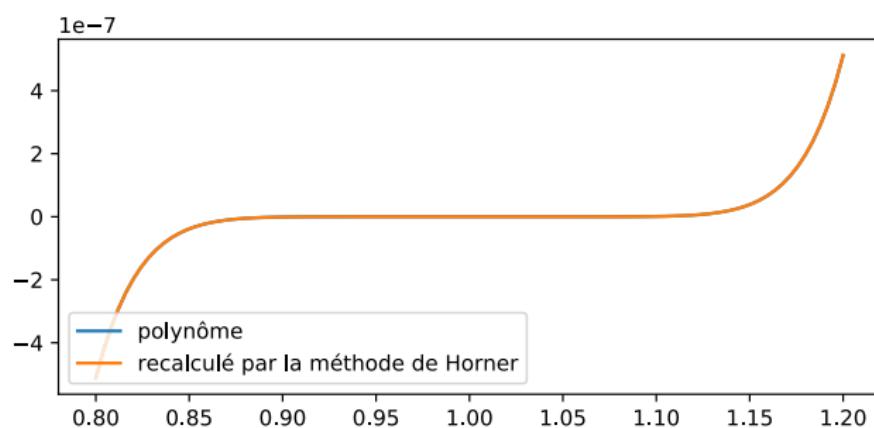
CONTROLLING THE WHOLE
SOFTWARE/COMPIILING STACK IS NOT
SUFFICIENT

ALL I CARE ABOUT IS THE ALGORITHM OUTPUT (FP)



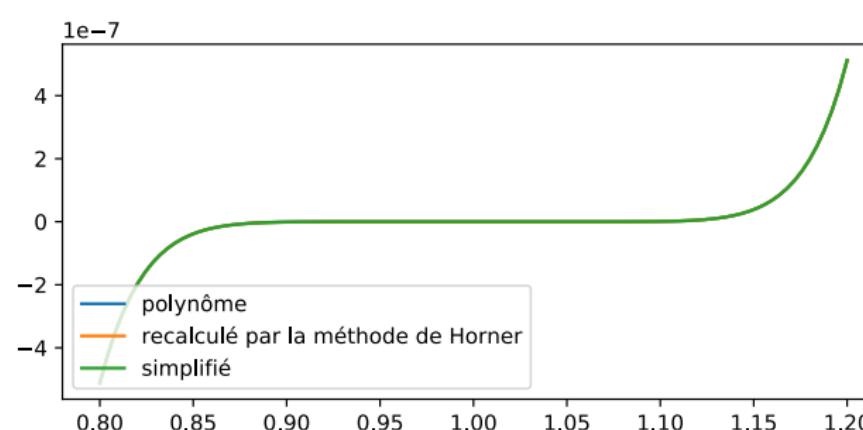
```
1 def polynome(x):  
2     return x**9 - 9.*x**8 + 36.*x**7 - 84.*x**6 + 126.*x**5 \  
3         - 126.*x**4 + 84.*x**3 - 36.*x**2 + 9.*x - 1.
```

FLOATING-POINT ARITHMETIC



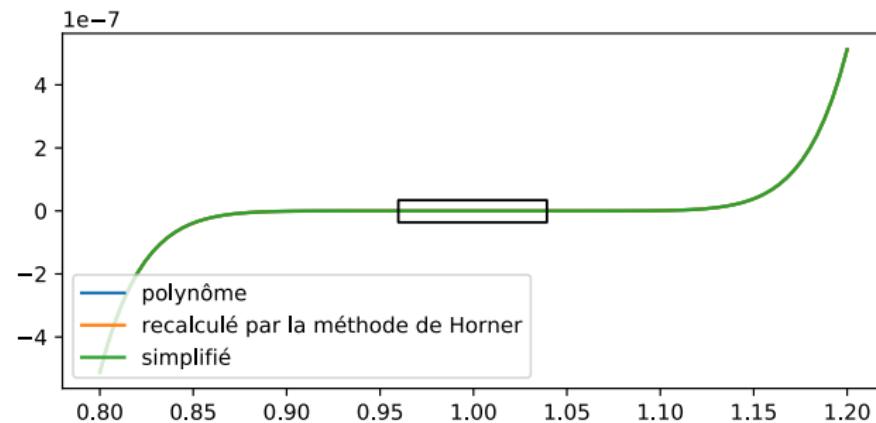
```
1 def horner(x):  
2     return x*(x*(x*(x*(x*(x*(x*(x - 9.) + 36.) - 84.) + 126.) \  
3             - 126.) + 84.) - 36.) + 9.) - 1.
```

FLOATING-POINT ARITHMETIC

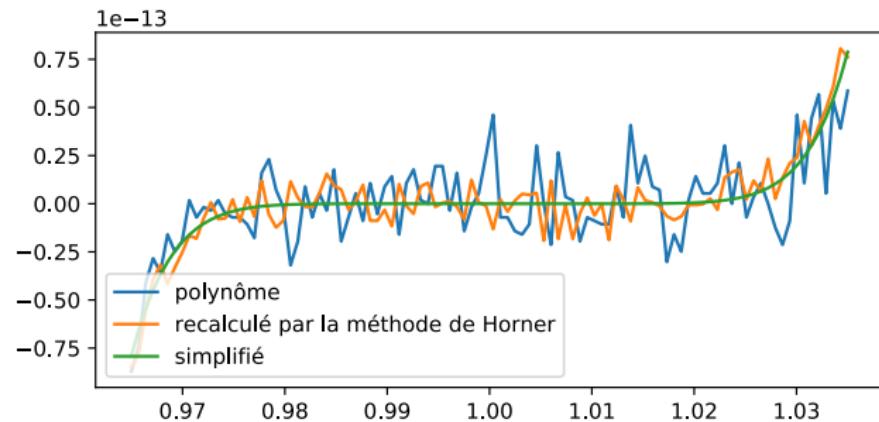


```
1 def simple(x):  
2     return (x-1.)**9  
3 # Easy! ;)
```

FLOATING-POINT ARITHMETIC



FLOATING-POINT ARITHMETIC



ROUNDING

- Every operation includes implicit rounding.
- $a+b$ is actually `round`($a+b$).
- Unfortunately:

$$\text{round}(\text{round}(a+b)+c) \neq \text{round}(a+\text{round}(b+c)).$$

- Operation order therefore matters.

For a reproducible computation, operation order must be preserved!!!

HOW TO EXPLAIN IT TO MY COMPILER?

To speed up computations, compilers may change operation order, and thus results.

Two options for computing reproducibly:

1. Insist on the preservation of operation order,
 - if the language permits it.
 - Example: Module 'ieee_arithmetic' in Fortran 2003
2. Make compilation reproducible:
 - Record the precise compiler version
 - Record all compilation options

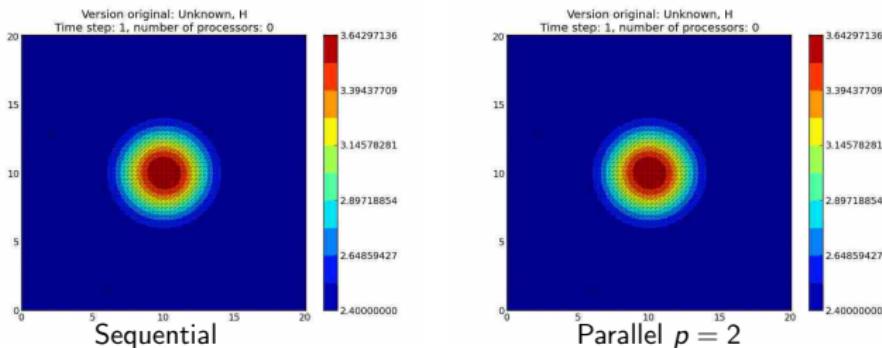
DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

Telemac2D: the simplest gouttedeo simulation

The gouttedeo test case

- 2D-simulation of a water drop fall in a square bassin
- Unknown: water depth for a 0.2 sec time step
- Triangular mesh: 8978 elements and 4624 nodes

Expected numerical reproducibility (time step = 1, 2, ...)

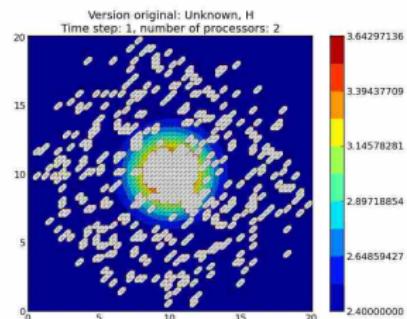
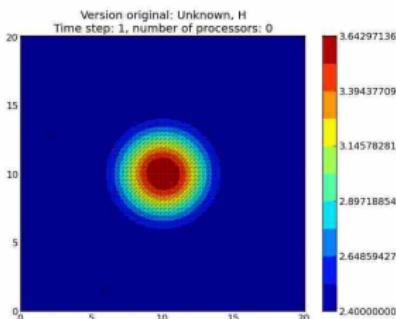


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 1

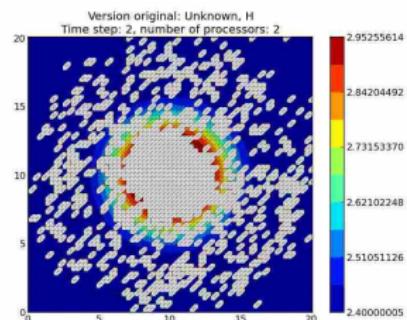
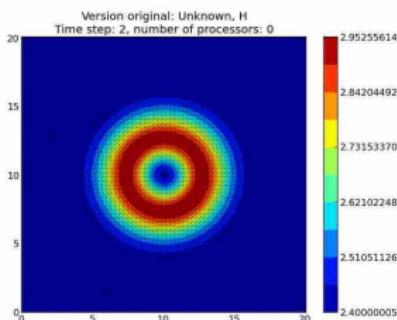


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 2

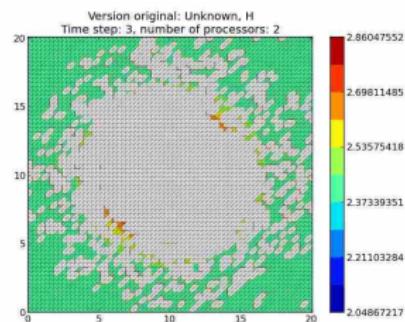
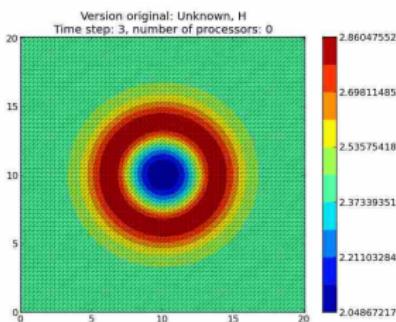


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 3

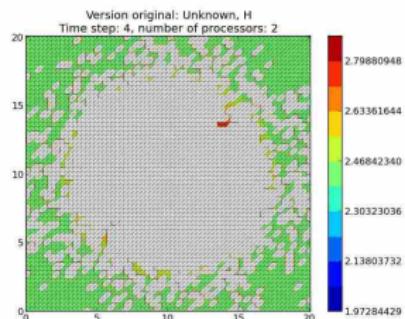
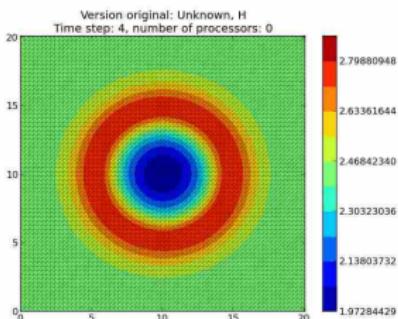


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 4

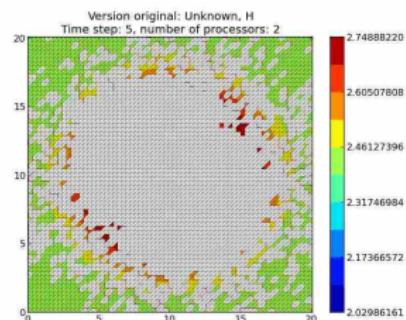
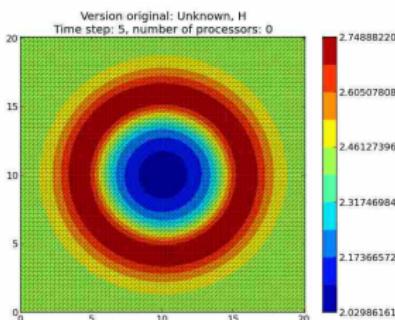


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 5

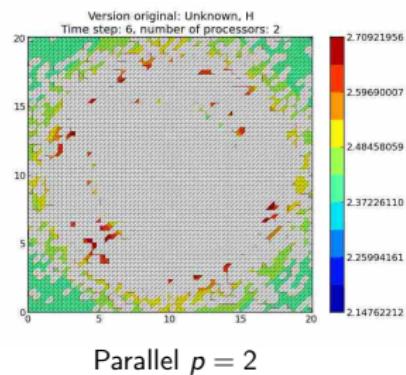
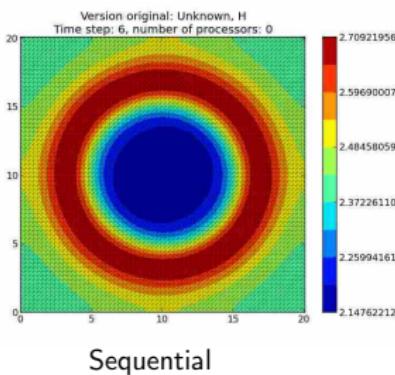


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 6

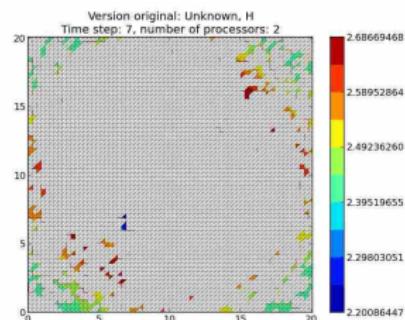
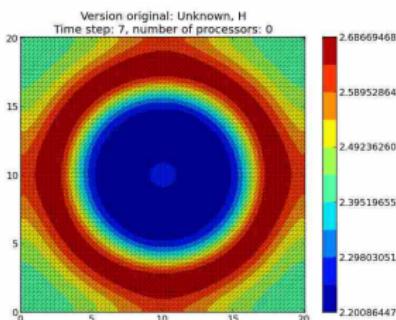


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 7

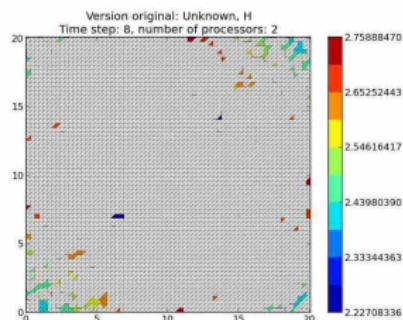
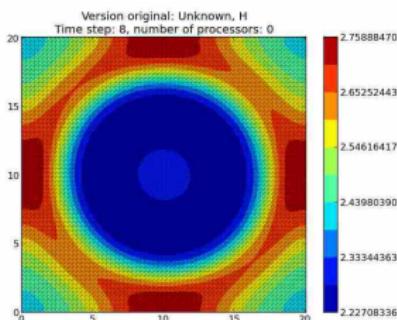


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 8

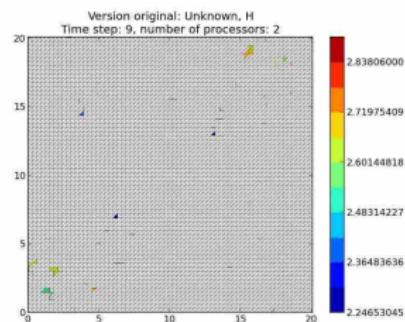
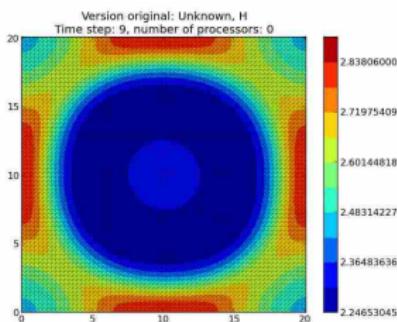


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 9

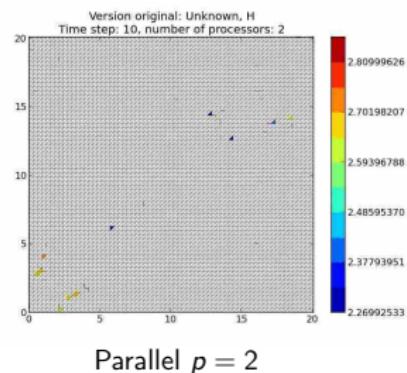
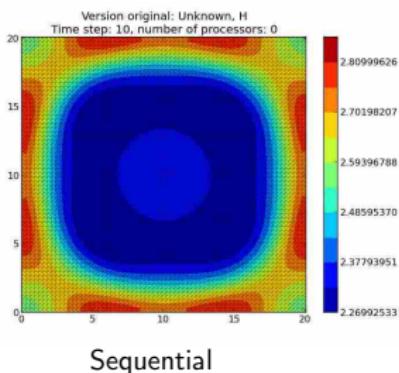


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 10

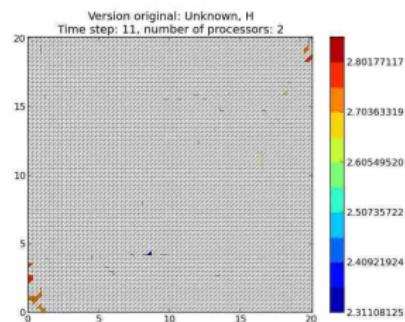
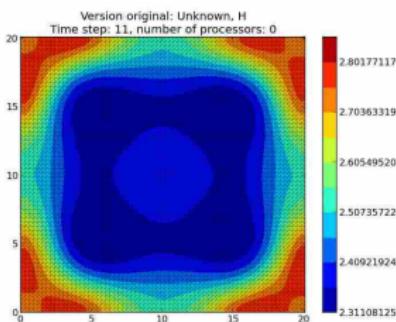


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 11

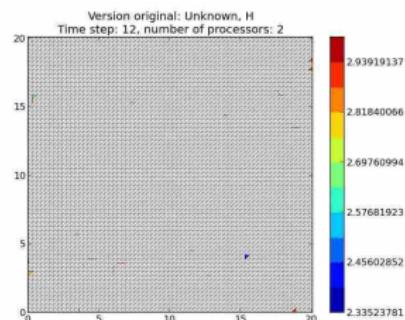
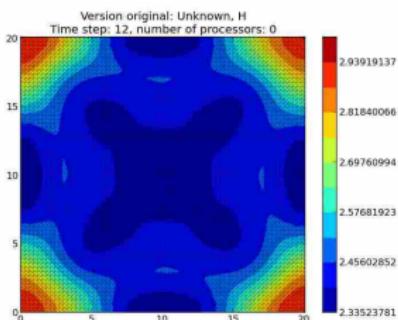


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 12

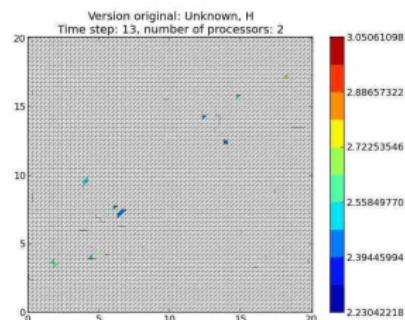
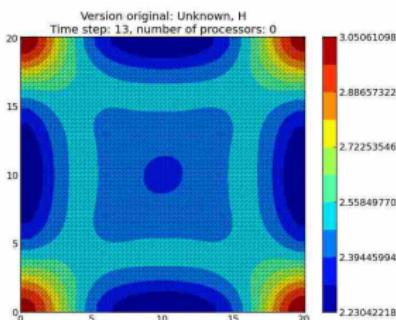


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 13

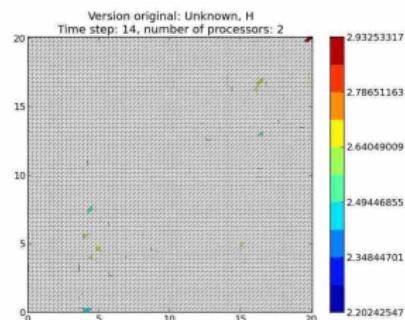
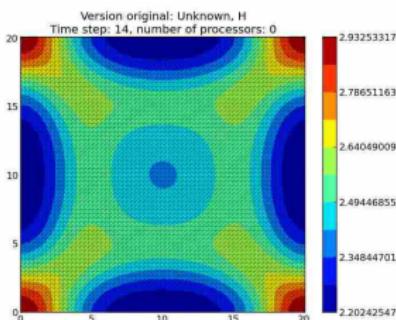


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 14

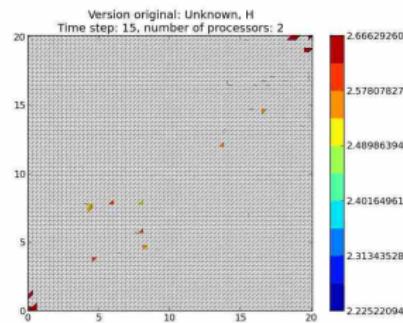
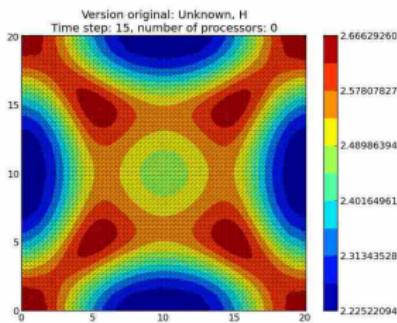


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

A white plot displays a non-reproducible value

NO numerical reproducibility!

time step = 15

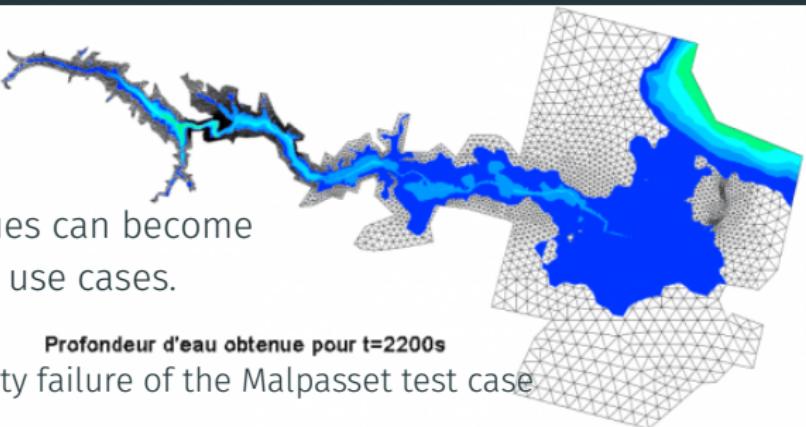


DID I MENTION WE HAVE PARALLEL MACHINES NOWADAYS?

These numerical issues can become quite harmful in real use cases.

Profondeur d'eau obtenue pour t=2200s

TABLE 1.1: Reproducibility failure of the Malpasset test case



	The sequential run	a 64 procs run	a 128 procs run
depth H	0.3500122E-01	0.2748817E-01	0.1327634E-01
velocity U	0.4029747E-02	0.4935279E-02	0.4512116E-02
velocity V	0.7570773E-02	0.3422730E-02	0.7545233E-02

Numerical reproducibility?: Approximations in the model, in the algorithm, in its implementation, in its execution.

The whole chain needs to be revisited.

Courtesy of P. Langlois and R. Nheili

SOFTWARE/HARDWARE DEPENDENCIES

Runtime dependencies interpreter, libraries, other programs

Build dependencies compilers, headers, `autotools/cmake`, etc.

SOFTWARE/HARDWARE DEPENDENCIES

Runtime dependencies interpreter, libraries, other programs

Build dependencies compilers, headers, `autotools/cmake`, etc.

Version a loose description

- `v_1.4.2`

SOFTWARE/HARDWARE DEPENDENCIES

Runtime dependencies interpreter, libraries, other programs

Build dependencies compilers, headers, `autotools/cmake`, etc.

Version a loose description

- `v_1.4.2` (commit `78c293967830979f988b39cea2587a7c26b`),

SOFTWARE/HARDWARE DEPENDENCIES

Runtime dependencies interpreter, libraries, other programs

Build dependencies compilers, headers, `autotools/cmake`, etc.

Version a loose description

- `v_1.4.2` (commit `78c293967830979f988b39cea2587a7c26b`),
(built/run in *this environment* with *such and such options*)

SOFTWARE/HARDWARE DEPENDENCIES

Runtime dependencies interpreter, libraries, other programs

Build dependencies compilers, headers, `autotools/cmake`, etc.

Version a loose description

- `v_1.4.2` (commit `78c293967830979f988b39cea2587a7c26b`,
(built/run in this environment with such and such options)
- `v_1.5` or `v_1.4.3` are often *better* than `v_1.4.2`

SOFTWARE/HARDWARE DEPENDENCIES

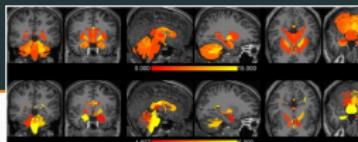
Runtime dependencies interpreter, libraries, other programs

Build dependencies compilers, headers, `autotools/cmake`, etc.

Version a loose description

- `v_1.4.2` (commit `78c293967830979f988b39cea2587a7c26b`),
(built/run in *this environment* with *such and such options*)
- `v_1.5` or `v_1.4.3` are often *better* than `v_1.4.2`
- Convenient but very problematic for reproducibility

SOFTWARE/HARDWARE DEPENDENCIES

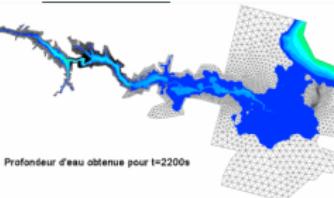
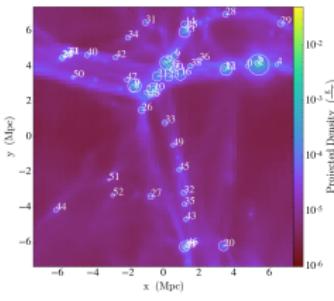


Runtime dependencies interpreter, libraries, other pro

Build dependencies compilers, headers, `autotools/cm`

Version a loose description

- `v_1.4.2` (commit [78c293967830979f988b39cea258](#) (built/run in this environment with such and su
- `v_1.5` or `v_1.4.3` are often *better* than `v_1.4.2`
- Convenient but very problematic for reproducit



It is not a good sign if your code is sensitive.

SOFTWARE/HARDWARE DEPENDENCIES

Runtime dependencies interpreter, libraries, other programs

Build dependencies compilers, headers, `autotools/cmake`, etc.

Version a loose description

- `v_1.4.2` (commit `78c293967830979f988b39cea2587a7c26b`,
(built/run in *this environment* with *such and such options*)
- `v_1.5` or `v_1.4.3` are often *better* than `v_1.4.2`
- Convenient but very problematic for reproducibility

It is not a good sign if your code is sensitive. You need:

1. Variation generation (test)
2. Perfect control (debug)

WHAT WILL IT TAKE ?

CHANGING RESEARCH PRACTICES

Soft. Engineering, Statistics, and Reproducible Research in the curricula

Manifesto: "*I solemnly pledge*" ([WSSSPE](#), [Lorena Barba](#), [FAIR](#))

1. I will teach my graduate students about reproducibility
2. All our research code (and writing) is under version control
3. We will always carry out verification and validation
4. We will share data, plotting script & figure under CC-BY
5. We will upload the preprint to arXiv at the time of submission of a paper
6. We will release code at the time of submission of a paper
7. We will add a "Reproducibility" declaration at the end of each paper
8. I will keep an up-to-date web presence



Learn and Teach using online resources like

- [Software Carpentry](#), [The Turing Way](#), ...

CHANGING PUBLISHING PRACTICES

Artifact evaluation and ACM badges



Major conferences

- Supercomputing: Artifact Description (AD) mandatory, Artifact Evaluation (AE) still optional, Double blind vs. RR
- NeurIPS, ICLR: open reviews, reproducibility challenge



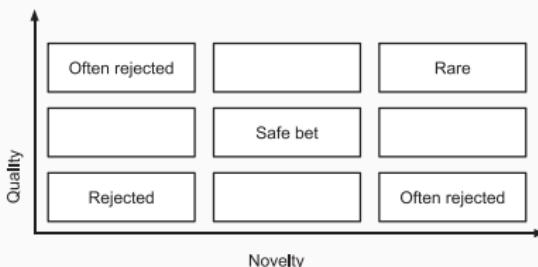
Joelle Pineau @ NeurIPS'18

- ACM SIGMOD 2015-2019, Most Reproducible Paper Award...

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

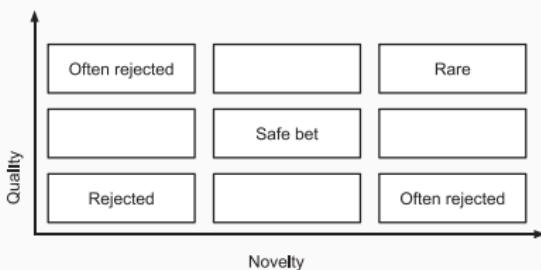
CHANGING ACADEMIC PRACTICES (PUBLISH OR PERISH)

- Goodhart's Law: Are Academic Metrics Being Gamed?, M. Fire 2019
 - AI: over 1,000 ranked journals ($\times 10$ in 15 years)
 - Shorter papers with increasing self references
 - More and more papers without any citation
 - Sharp increase in the number of new authors publishing at a much faster rate given their career age
- The Truth, The Whole Truth, and Nothing But the Truth: A Pragmatic, Guide to Assessing Empirical Evaluations, TOPLAS 2016



CHANGING ACADEMIC PRACTICES (PUBLISH OR PERISH)

- Goodhart's Law: Are Academic Metrics Being Gamed?, M. Fire 2019
 - AI: over 1,000 ranked journals ($\times 10$ in 15 years)
 - Shorter papers with increasing self references
 - More and more papers without any citation
 - Sharp increase in the number of new authors publishing at a much faster rate given their career age
- The Truth, The Whole Truth, and Nothing But the Truth: A Pragmatic, Guide to Assessing Empirical Evaluations, TOPLAS 2016



- Impact factor abandoned by Dutch university in hiring and promotion, decisions. Nature, June 2021. Faculty and staff members at Utrecht University will be evaluated by their commitment to open science

REPRODUCIBLE RESEARCH = RIGOR AND TRANSPARENCY

Good research requires time and resources

1. Train yourself and your students: RR, statistics, experiments

- Beware of checklists and norms Understand what's at stake

MOOC Reproducible Research: Methodological principles for a transparent science, Inria Learning Lab

- Konrad Hinsen, Christophe Pouzat
- 3rd Edition: March 2020 – ... (16,800+)



REPRODUCIBLE RESEARCH = RIGOR AND TRANSPARENCY

Good research requires time and resources

1. Train yourself and your students: RR, statistics, experiments

- Beware of checklists and norms Understand what's at stake

MOOC Reproducible Research: Methodological principles for a transparent science, Inria Learning Lab

- Konrad Hinsen, Christophe Pouzat
- 3rd Edition: March 2020 – ... (16,800+)



MOOC "Advanced RR" planned for May 2024

- Managing data (`FITS/HDF5, git annex`)
- Software environment control (`docker, singularity, guix`)
- Scientific workflow (`make, snakemake`)

REPRODUCIBLE RESEARCH = RIGOR AND TRANSPARENCY

Good research requires time and resources

1. Train yourself and your students: RR, statistics, experiments
 - Beware of checklists and norms Understand what's at stake

MOOC Reproducible Research: Methodological principles for a transparent science, Inria Learning Lab

- Konrad Hinsen, Christophe Pouzat
- 3rd Edition: March 2020 – ... (16,800+)



MOOC "Advanced RR" planned for May 2024

- Managing data (`FITS/HDF5, git annex`)
- Software environment control (`docker, singularity, guix`)
- Scientific workflow (`make, snakemake`)

2. Change the norm: make publication practices evolve
3. Incentive: consider RR/open science when hiring/promoting

REPRODUCIBLE RESEARCH = RIGOR AND TRANSPARENCY

Good research requires time and resources

1. Train yourself and your students: RR, statistics, experiments

- Beware of checklists and norms Understand what's at stake

MOOC Reproducible Research: Methodological principles for a transparent science, Inria Learning Lab

- Konrad Hinsen, Christophe Pouzat
- 3rd Edition: March 2020 – ... (16,800+)



MOOC "Advanced RR" planned for May 2024

- Managing data (`FITS/HDF5`, `git annex`)
- Software environment control (`docker`, `singularity`, `guix`)
- Scientific workflow (`make`, `snakemake`)

2. Change the norm: make publication practices evolve

3. Incentive: consider RR/open science when hiring/promoting

4. Prepare the Future: Toward literate experimentation?

- Reuse, reuse, reuse!
- Shared and controlled testbeds
- How to share Experiments ?



THE SCIENCE IS CLEAR

Why are we
ignoring it?

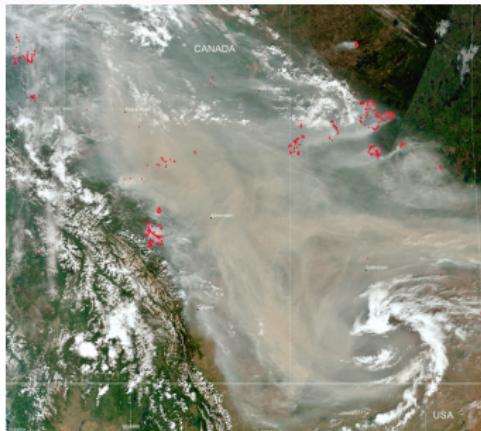
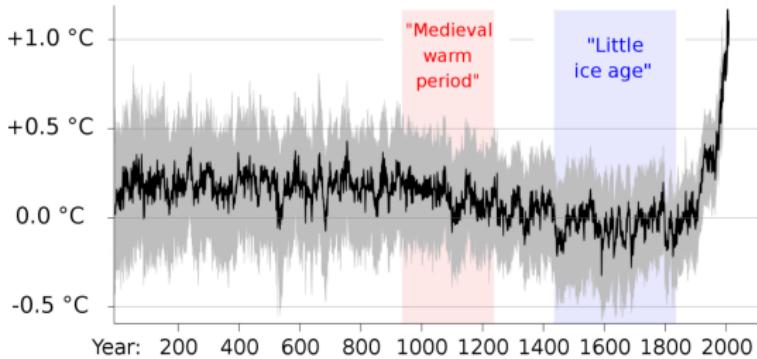
scientist rebellion



IPCC, IPBES, <https://climate.nasa.gov/>

1. Global climate change is not a future problem

Global Average Temperature Change



https://en.wikipedia.org/wiki/Global_temperature_record

2023 Alberta wildfires (> 1 Mha)

THE SCIENCE IS CLEAR

scientist rebellion

Why are we ignoring it?

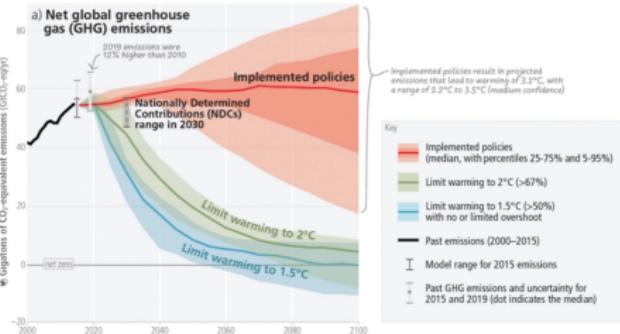


IPCC, IPBES, <https://climate.nasa.gov/>

1. Global climate change is **not** a future problem
2. It is **entirely** due to human activity

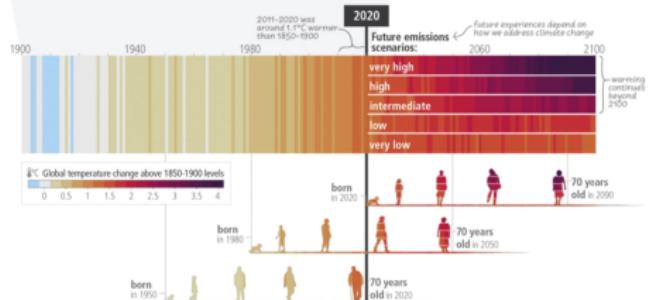
Limiting warming to **1.5°C** and **2°C** involves rapid, deep and in most cases immediate greenhouse gas emission reductions

Net zero: CO₂ and net zero GHG emissions can be achieved through strong reductions across all sectors



Paris Agreement'15 ~ Net Zero by 2050

c) The extent to which current and future generations will experience a hotter and different world depends on choices now and in the near-term



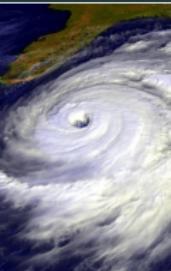
Latest IPCC report

41/42

THE SCIENCE IS CLEAR

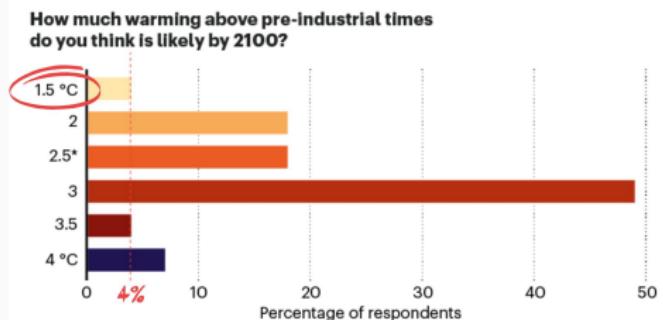
scientist rebellion

Why are we
ignoring it?



IPCC, IPBES, <https://climate.nasa.gov/>

1. Global climate change is **not** a future problem
2. It is **entirely** due to human activity
3. **9 out of 10 IPCC scientists believe overshoot is likely**

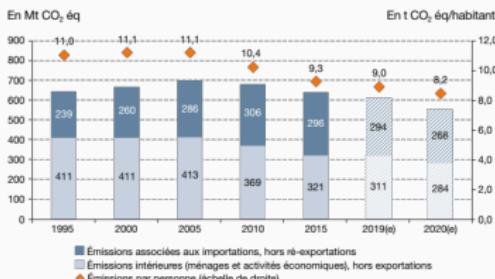


@natu Nature survey, Nov. 2021

THE ELEPHANT IN THE ROOM: CLIMATE CHANGE

Put aside biodiversity loss, pollution, freshwater, land system change...

ÉVOLUTION DE L'EMPREINTE CARBONE DE LA FRANCE



Empreinte carbone moyenne en France
10 tonnes de CO₂e/an/pers.



÷2
d'ici
2030

<2t CO₂e

Objectif d'ici 2050

- de 2 t de CO₂e/an/pers.

+ Faire plus d'activités bas carbone !

Danser, chanter, jardiner, rêver, écire, lire, courir, randonner, planter des arbres, discuter, marcher en forêt, méditer, passer du temps avec ceux qu'on aime, lire...

Bref, inventer nos vies bas carbone désirables !

Par exemple :

0,5 t CO₂e/Annee : à la maison, préférence à l'électrique et sans produits dérivés

0,5 t CO₂e/Annee : faire une partie route (30%) de votre voiture (30%) de fabrication ancienne sur 30 ans, importer moins de 10% de vos denrées et faire moins de transports en commun.

0,5 t CO₂e/Annee : faire moins de 40 km de trajet en voiture, faire diverses expérimenter des alternatives et interroger les acteurs, sobriété dans l'usage des denrées et services publics.

0,2 t CO₂e/Annee : échafaudage sur un tapis de puzzle, 100% recyclé et d'un emballage bas carbone, faire moins de 10 km de trajet en voiture, faire moins de 40 km de trajet en voiture, faire diverses expérimenter des alternatives et interroger les acteurs, sobriété dans l'usage des denrées et services publics.

0,2 t CO₂e/Annee : faire moins de 40 km de trajet en voiture, faire diverses expérimenter des alternatives et interroger les acteurs, sobriété dans l'usage des denrées et services publics.



<https://www.nosviesbascarbonne.org/>

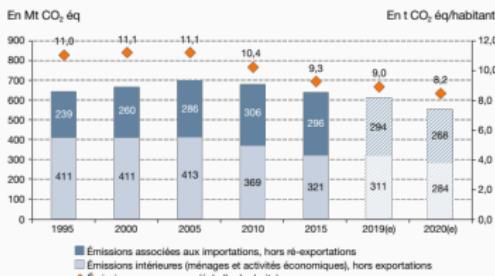
INVENTONS
NOUS VIES
BAS CARBONE

Sources : Kit Inventons nos vies bas carbone (Fév. 2021), Rapport sur l'état de l'environnement en France (Déc. 2020)

THE ELEPHANT IN THE ROOM: CLIMATE CHANGE

Put aside biodiversity loss, pollution, freshwater, land system change...

ÉVOLUTION DE L'EMPREINTE CARBONE DE LA FRANCE



(e) = estimations.

Note : l'empreinte carbone porte sur les trois principaux gaz à effet de serre (CO₂, CH₄, N₂O). En 2021, la méthodologie a été ajustée afin de mieux tenir compte de l'évolution des coûts du pétrole brut, du gaz et du charbon. L'ensemble de la série a ainsi été révisé, l'essentiel des ajustements portant sur les émissions importées de CH₄.

Champ : périmètre Kyoto (Île-de-France et outre-mer appartenant à l'UE).

Sources : Citepa ; AIE ; FAO ; Douanes ; Eurostat ; Insee. Traitement : SDES, 2021



Empreinte carbone moyenne en France
10 tonnes de CO₂e/an/pers.



÷2
d'ici
2030

<2t CO₂e

Objectif d'ici 2050

- de 2 t de CO₂e/an/pers.

+ Faire plus d'activités bas carbone !

Danser, chanter, jardiner, rêver, écire, lire, courir, randonner, planter des arbres, discuter, marcher en forêt, méditer, passer du temps avec ceux qu'on aime, lire...

Bref, inventer nos vies bas carbone désirables !

Par exemple :

0,5 t CO₂e : Activité alimentaire : à tendance régulière à diminuer le niveau produits végétaux

0,5 t CO₂e : Activité Transport : 2000km en voiture (équivalent annexe sur 30 ans importé électricité) ou 1000km en train ou 2000km transports en commun

0,5 t CO₂e : Activité Consommation : Basé sur les 4,4 t de consommation moyenne, faire diverses expérimentations et innovations, notamment dans les transports et les énergies renouvelables

0,2 t CO₂e : Activité Logement : Chaque fois sur un emppc en parallèle, "0% carbone" d'un logement peut isolé et équilibrer l'effet de son chauffage à la biomasse ou solaire thermique

0,2 t CO₂e : Activité Services publics : bonnes enseignances, éducation, etc.

INVENTONS
NOS VIES
BAS CARBONE

<https://www.nosviesbascarbone.org/>

Sources : Kit Inventons nos vies bas carbone (Fév. 2021), Rapport sur l'état de l'environnement en France (Déc. 2020)

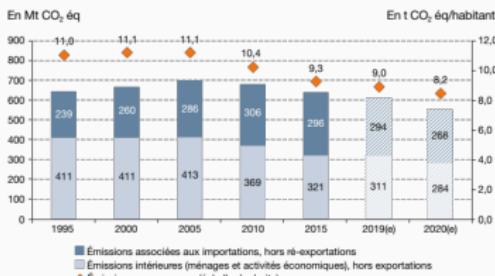
French government response

- Verdissement de l'industrie: « pause » sur les normes environnementales
- Loi de programmation militaire (+41%)
- Nous devons préparer la France à une élévation de la température de 4 °C
- Academia ? PEPR 5G, Cloud, NUMPEX, Quantique, IA, Agroécologie et numérique

THE ELEPHANT IN THE ROOM: CLIMATE CHANGE

Put aside biodiversity loss, pollution, freshwater, land system change...

ÉVOLUTION DE L'EMPREINTE CARBONE DE LA FRANCE



Empreinte carbone moyenne en France
10 tonnes de CO₂e/an/pers.

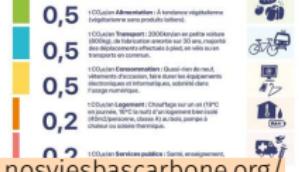


÷2
d'ici
2030

Objectif d'ici 2050
- de 2 t de CO₂e/an/pers.



Par exemple :



<https://www.nosviesbascarbonne.org/>

INVENTONS
NOUS VIES
BAS CARBONE

Sources : Kit Inventons nos vies bas carbone (Fév. 2021), Rapport sur l'état de l'environnement en France (Déc. 2020)



French government response

- Verdissement de l'industrie: « pause » sur les normes environnementales
- Loi de programmation militaire (+41%)
- Nous devons préparer la France à une élévation de la température de 4 °C
- Academia ? PEPR 5G, Cloud, NUMPEX, Quantique, IA, Agroécologie et numérique

Several scenarios on the table

- What will research/CS look like/be used for in such a world?
- Energy optimization/saving ≠ sobriety and frugality