

# REPRODUCIBLE RESEARCH AND COMPUTER SCIENCE

---

Arnaud Legrand



20th Anniversary, May 2023



# NO TRANSPARENCY NO CONSENSUS



## Plan National pour la Science Ouverte (CoSO)

- France (CNRS, Inria, INRAE, ...) but also Europe and US
- Many flavors: *Citizen Science vs. Ethics and Societal Responsibility*

# TOWARD OPEN SCIENCE

## Plan National pour la Science Ouverte (CoSO)

- France (CNRS, Inria, INRAE, ...) but also Europe and US
- Many flavors: *Citizen Science vs. Ethics and Societal Responsibility*

### Main pillars:

1. Open access
2. Open data
3. Open source
  - Open hardware
4. **Open methodology (Reproducible Research)**
  - Open-notebook science
  - Open science infrastructures
5. Open peer review
6. Open educational resources

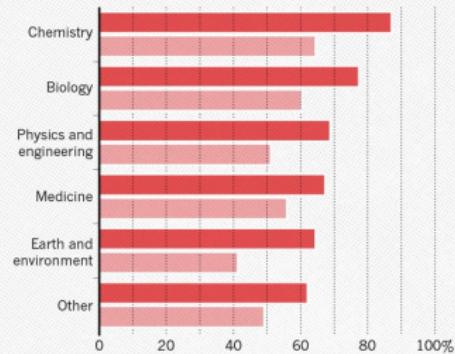


# SOCIO-TECHNICAL CHALLENGES

## 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



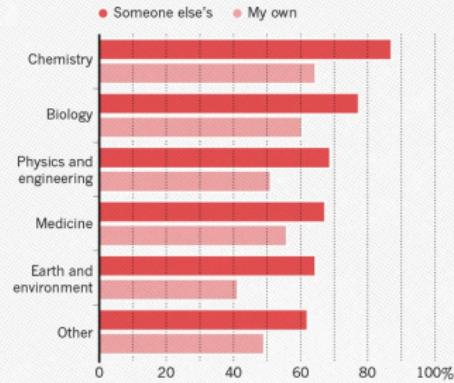
1,500 scientists lift the lid on reproducibility,

Nature, May 2016

# SOCIO-TECHNICAL CHALLENGES

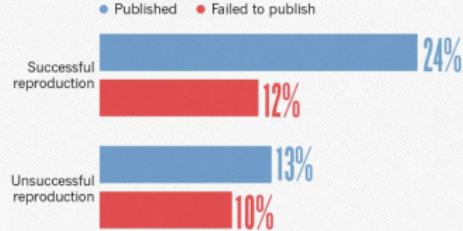
## HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

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.



1,500 scientists lift the lid on reproducibility,

Nature, May 2016

## Social causes

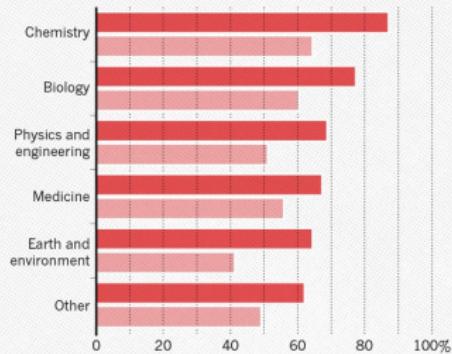
- Fraud, conflict of interest (pharmaceutic, ...)
- No incentive to reproduce/check our own work (afap), nor the work of others (big results!), nor to allow others to check (competition)
- Peer review does not scale: 1M+ articles per year!

# SOCIO-TECHNICAL CHALLENGES

## HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.

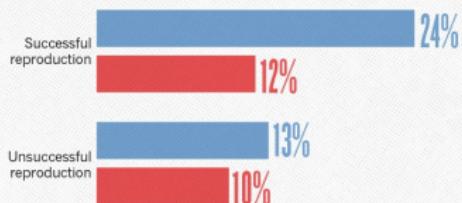
● Someone else's    ● My own



## 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

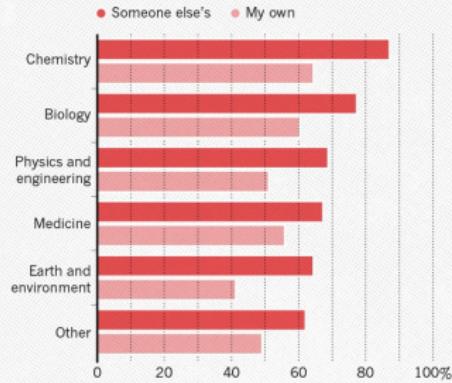
## Social causes

- 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

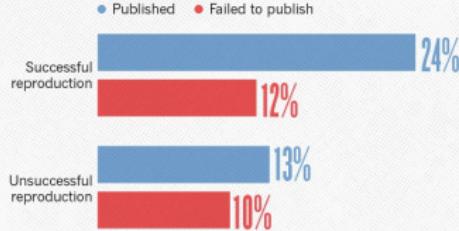
## HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

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

# 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

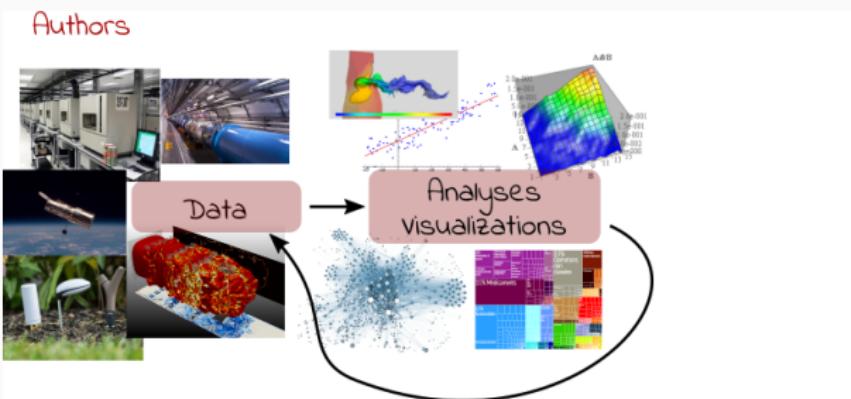
**Psychology, Nutrition** HARKING, p-hacking  $\rightsquigarrow$  pre-registration

**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

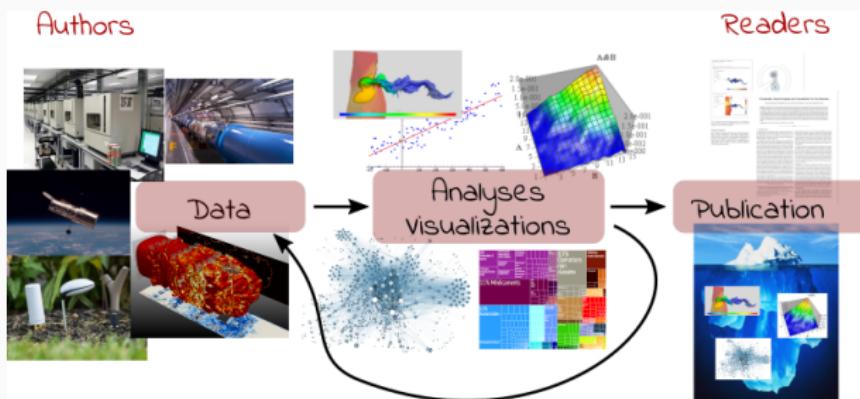
**Psychology, Nutrition** HARKING, p-hacking  $\rightsquigarrow$  pre-registration

**Genomics** software engineering, computational reproducibility, provenance

**Computational fluid dynamics** numerical chaos, parallel architectures

**Artificial Intelligence** most of the above 😊

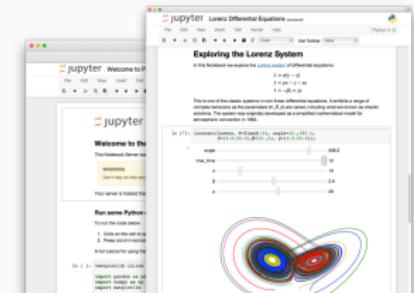
*The processing steps between raw observations and findings have gotten increasingly numerous and complex*



Reproducible Research = Bridging the Gap by working Transparently

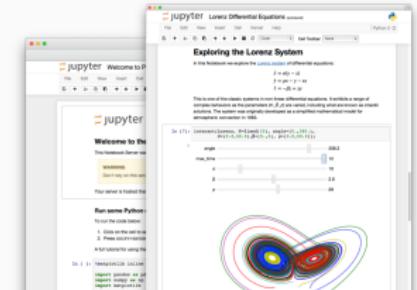
# REPRODUCIBILITY ISSUES RELATED TO THE USE OF COMPUTERS

## Computation provenance: notebooks and workflows

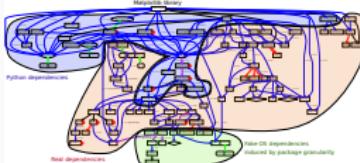


# REPRODUCIBILITY ISSUES RELATED TO THE USE OF COMPUTERS

## Computation provenance: notebooks and workflows

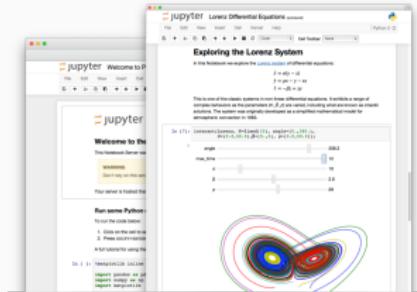


## Software environments



# REPRODUCIBILITY ISSUES RELATED TO THE USE OF COMPUTERS

## Computation provenance: notebooks and workflows



## Software environments



## Sharing and Archiving



# GRID'5000 AND SIMGRID: A LONG LOVE-STORY

---

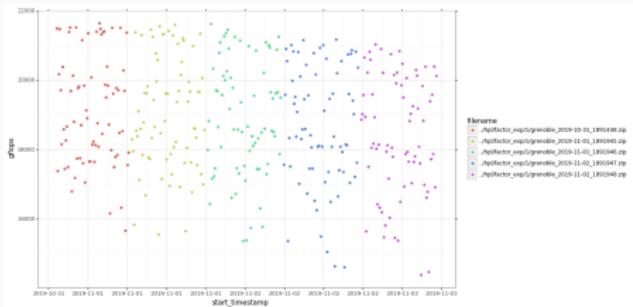
## GRID'5000 AND SIMGRID: A LONG LOVE-STORY

---

- HPL performance (32 nodes, 70 cfg., 5 repetitions) · Time scale = 3 days

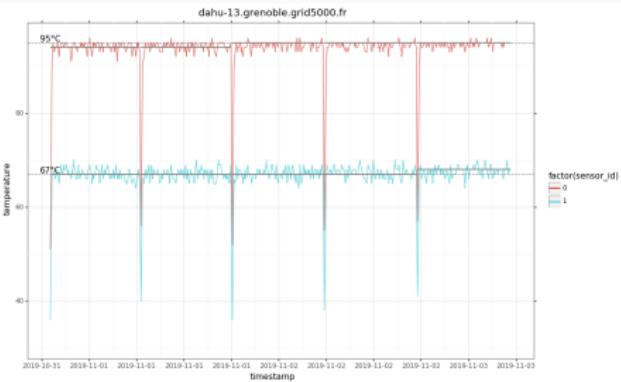
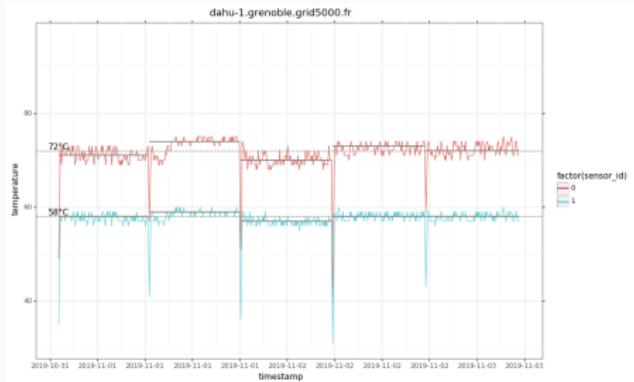
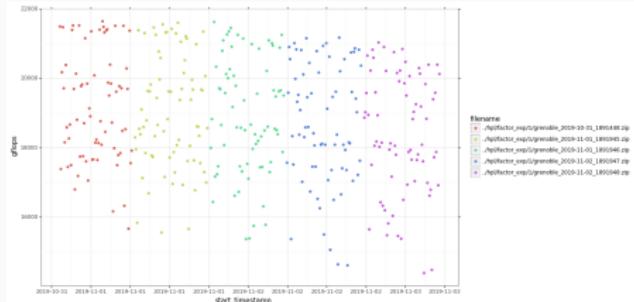
# GRID'5000 AND SIMGRID: A LONG LOVE-STORY

- HPL performance (32 nodes, 70 cfg., 5 repetitions) • Time scale = 3 days



# GRID'5000 AND SIMGRID: A LONG LOVE-STORY

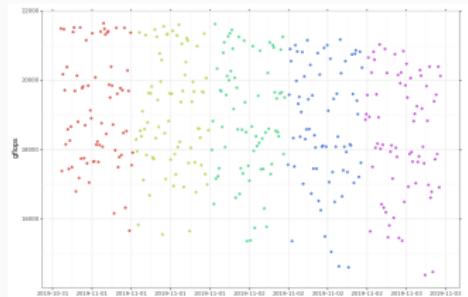
- HPL performance (32 nodes, 70 cfg., 5 repetitions) • Time scale = 3 days



Courtesy of T. Cornebize

# GRID'5000 AND SIMGRID: A LONG LOVE-STORY

- HPL performance (32 nodes, 70 cfg., 5 repetitions)
- Time scale = 3 days



factorName

...  
\_applyFactor\_exp1grid5000\_2019-10-31\_1801448.ip  
\_applyFactor\_exp1grid5000\_2019-11-01\_1801451.ip  
\_applyFactor\_exp1grid5000\_2019-11-01\_1801461.ip  
\_applyFactor\_exp1grid5000\_2019-11-01\_1801464.ip  
\_applyFactor\_exp1grid5000\_2019-11-02\_1801460.ip



Temperature

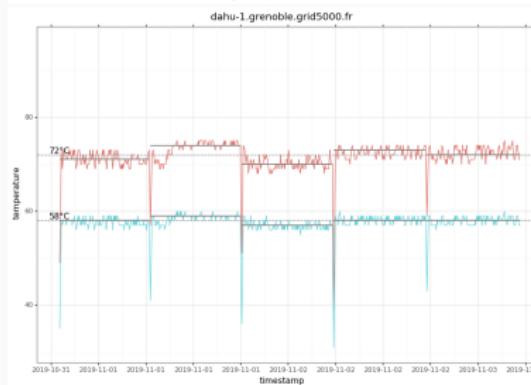
20 25 30 35

timestamp

CPU temperatures (average "per timestamp")

CPU 0

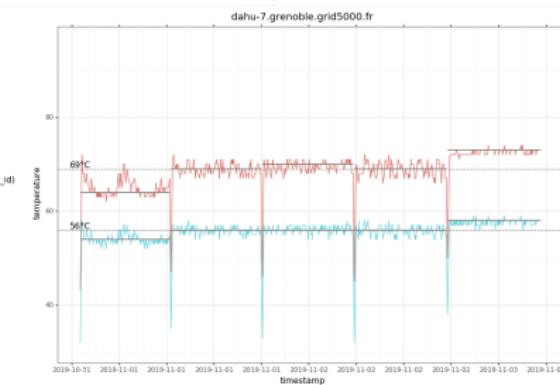
factorName  
\_applyFactor\_exp1grid5000\_2019-10-31\_1801448.ip  
\_applyFactor\_exp1grid5000\_2019-11-01\_1801451.ip  
\_applyFactor\_exp1grid5000\_2019-11-01\_1801461.ip  
\_applyFactor\_exp1grid5000\_2019-11-01\_1801464.ip  
\_applyFactor\_exp1grid5000\_2019-11-02\_1801460.ip



factor(sensor\_id)

0

1



factor(sensor\_id)

0

1

Courtesy of T. Cornebize

## SYSTEMATIC TESTS ON GRID 5000

---

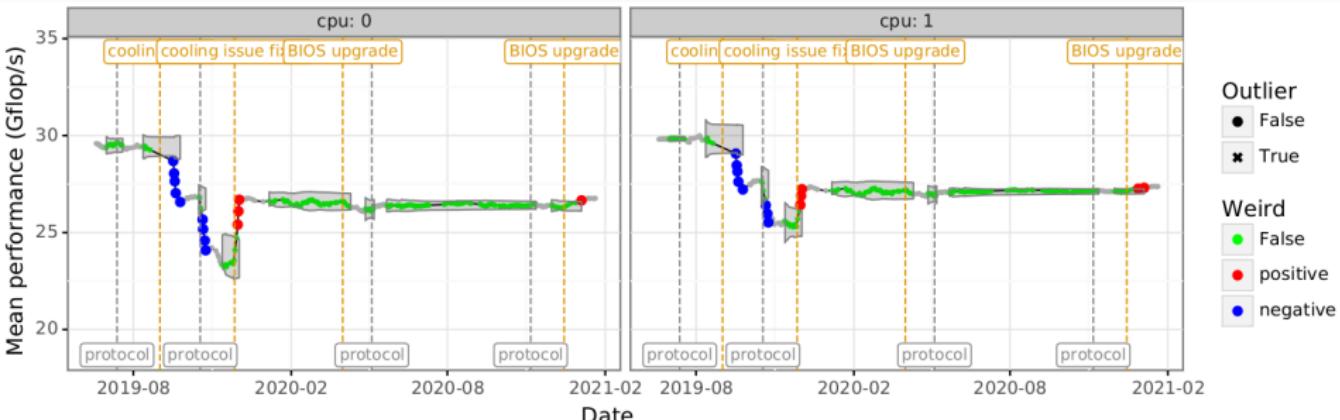
≈ Daily measurement of the whole cluster(s) over ≈ 2 years

[https://cornebize.net/g5k\\_test/](https://cornebize.net/g5k_test/)

# SYSTEMATIC TESTS ON GRID 5000

≈ Daily measurement of the whole cluster(s) over ≈ 2 years

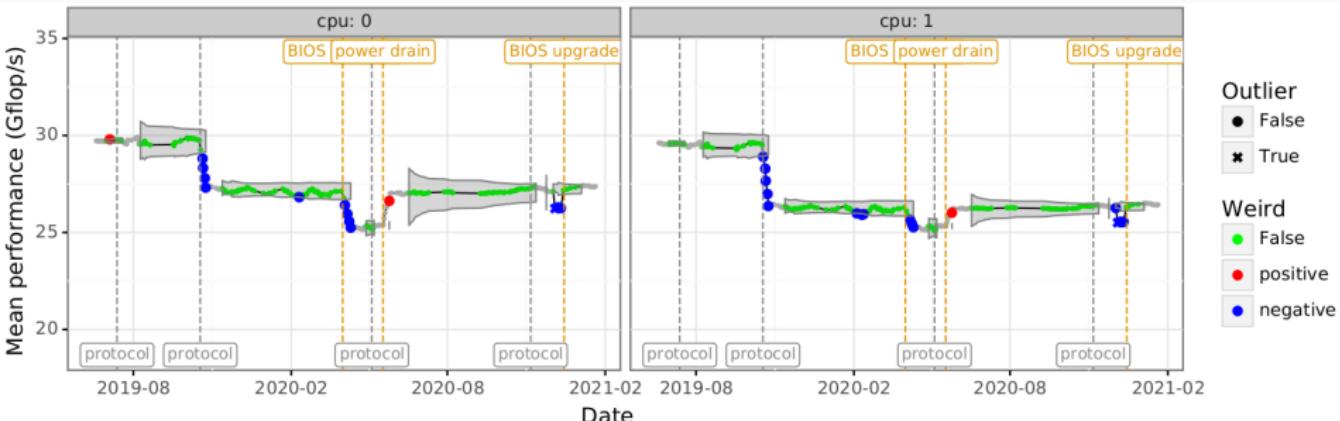
[https://cornebize.net/g5k\\_test/](https://cornebize.net/g5k_test/)



# SYSTEMATIC TESTS ON GRID 5000

≈ Daily measurement of the whole cluster(s) over ≈ 2 years

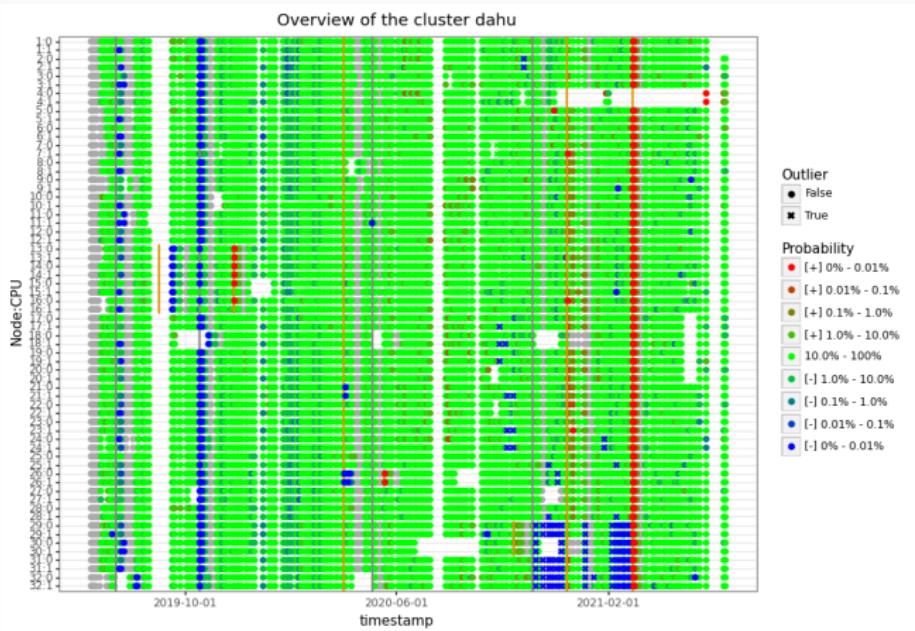
[https://cornebize.net/g5k\\_test/](https://cornebize.net/g5k_test/)



# SYSTEMATIC TESTS ON GRID 5000

≈ Daily measurement of the whole cluster(s) over ≈ 2 years

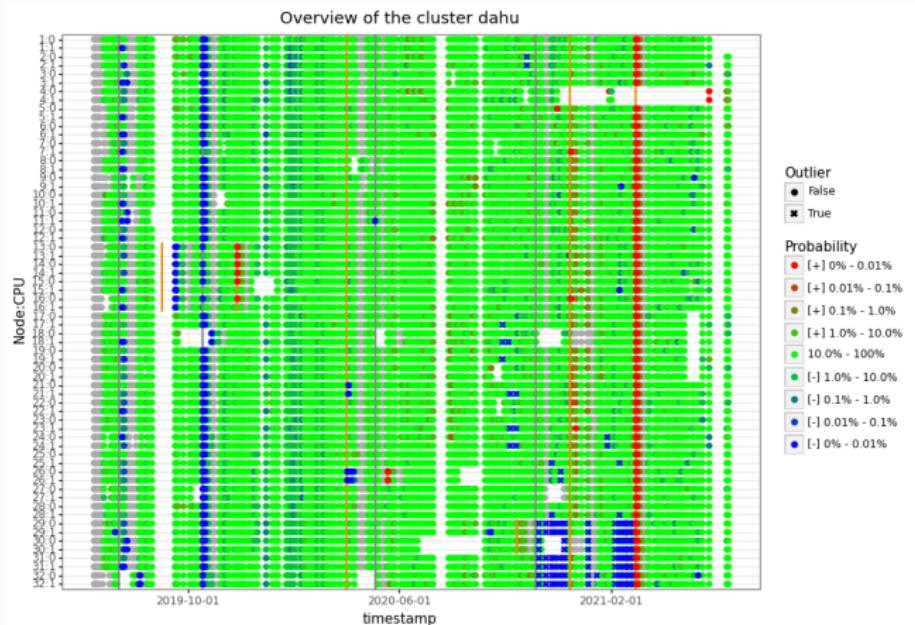
[https://cornebize.net/g5k\\_test/](https://cornebize.net/g5k_test/)



# SYSTEMATIC TESTS ON GRID 5000

≈ Daily measurement of the whole cluster(s) over ≈ 2 years

[https://cornebize.net/g5k\\_test/](https://cornebize.net/g5k_test/)



We will **not** address this with a better control/software/hardware

# REPRODUCIBILITY CHALLENGES IN PARALLEL/DISTRIBUTED SYSTEMS

**Challenge:** A physical system that evolves differently from other physical/biological systems

- Cultural change
  - Data/code/env/experiment sharing (like everyone else)
  - Publication/recognition practices (like everyone else)
  - **Metrology, experiment design, statistics**
- Beware of checklists and norms. Understand what's at stake

# REPRODUCIBILITY CHALLENGES IN PARALLEL/DISTRIBUTED SYSTEMS

**Challenge:** A physical system that evolves differently from other physical/biological systems

- Cultural change
  - Data/code/env/experiment sharing (like everyone else)
  - Publication/recognition practices (like everyone else)
  - Metrology, experiment design, statistics
- Beware of checklists and norms. Understand what's at stake

**Prepare the Future:** How to share/reuse Experiments?

- Reuse, reuse, reuse!
- Shared and controled **testbeds** (Grid'5000/SILECS/SLICES  $\approx$  LHC)
- Toward **literate experimentation?**



# THE ELEPHANT IN THE ROOM



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

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

# THE ELEPHANT IN THE ROOM

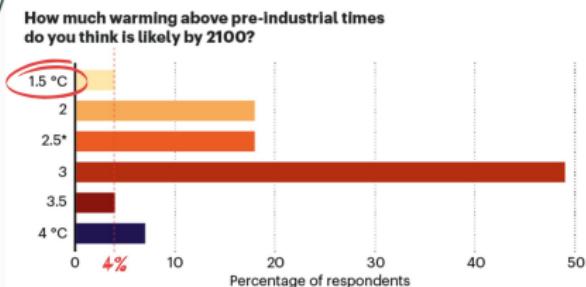


IPCC, IPBES, <https://climate.nasa.gov/>

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*

Nature survey, Nov. 2021

"Several" scenarios on the table



©nature

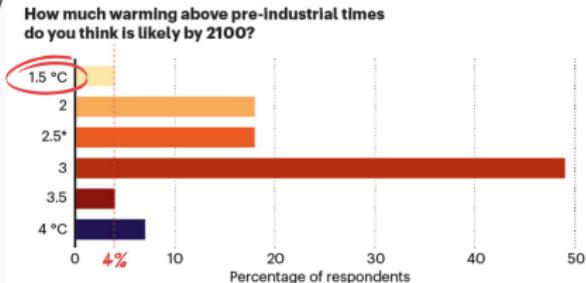
# THE ELEPHANT IN THE ROOM



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*

Nature survey, Nov. 2021



"Several" scenarios on the table

- What will CS look like/be used for in such a world?
- Let's not confuse energy optimization/saving with sobriety