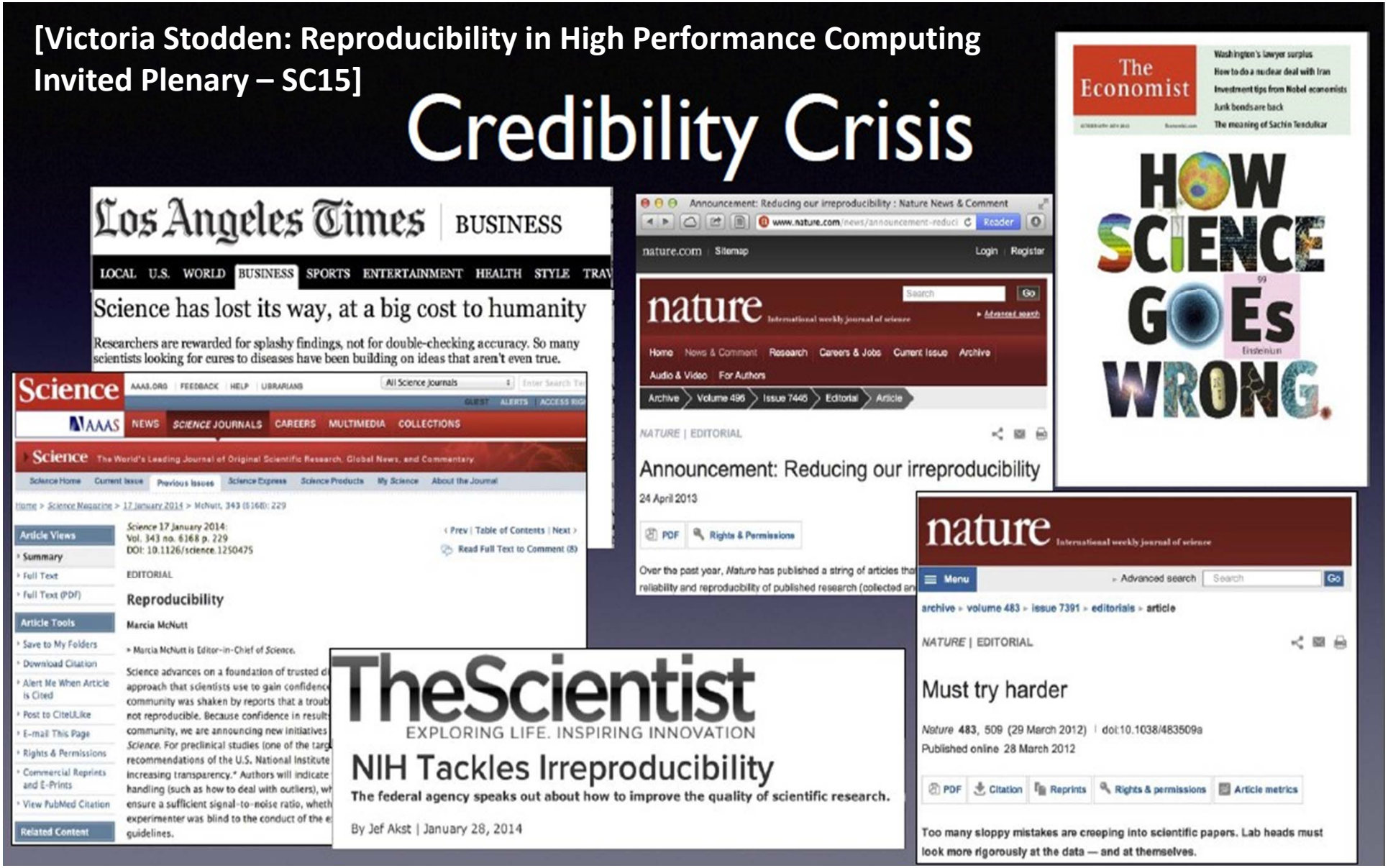
Bitwise Reproducibility (with Exascale Machines)

Prof. Dr. Thomas Ludwig
German Climate Computing Center (DKRZ)
University of Hamburg, Department for Computer Science (UHH/FBI)

[Victoria Stodden: Reproducibility in High Performance Computing Invited Plenary – SC15]

Credibility Crisis



[Victoria Stodden: Reproducibility in High Performance Computing]

Unpacking "Reproducibility"

"Empirical Reproducibility"

"Computational Reproducibility"

"Statistical Reproducibility"


nature

International weekly journal of science

Menu · Advanced search · Search · Go

archive » volume 483 » issue 7391 » editorials » article

NATURE | EDITORIAL

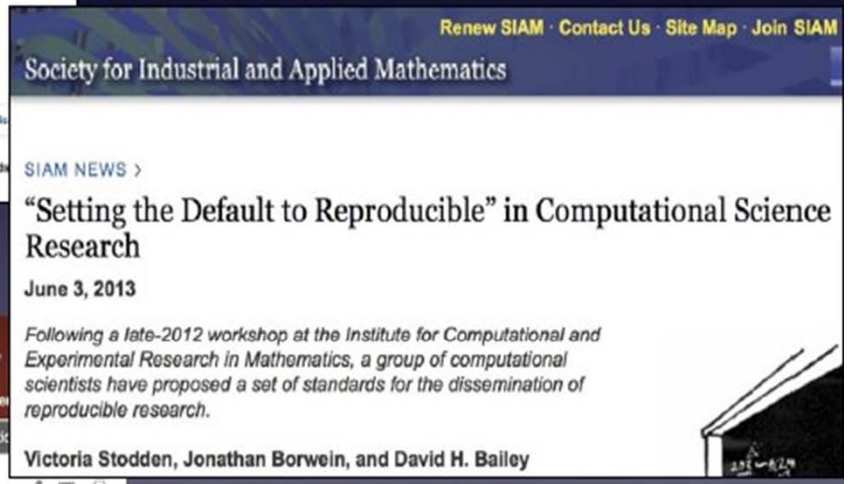
Must try harder

Nature 483, 509 (29 March 2012) | doi:10.1038/483509a

Published online 28 March 2012

PDF · Citation · Reprints · Rights & permissions · Article metrics

Too many sloppy mistakes are creeping into scientific papers. Lab heads look more rigorously at the data — and at themselves.


Renew SIAM · Contact Us · Site Map · Join SIAM

Society for Industrial and Applied Mathematics

SIAM NEWS ›

"Setting the Default to Reproducible" in Computational Science Research

June 3, 2013

Following a late-2012 workshop at the Institute for Computational and Experimental Research in Mathematics, a group of computational scientists have proposed a set of standards for the dissemination of reproducible research.

Victoria Stodden, Jonathan Borwein, and David H. Bailey


nature

International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Curre

Archive › Volume 506 › Issue 7487 › News Feature › Artic

NATURE | NEWS FEATURE

Scientific method: Statistical errors

***P* values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.**

Regina Nuzzo

12 February 2014

V. Stodden, IMS Bulletin (2013)

Non-Reproducibility with Computational Sciences

Many levels of abstractions

- Reproduce a scientific result
  - Find same insight
- Reproduce a computer based experiment
  - Run code again and check result
  - Good scientific practice
- Bitwise reproduce a single experiment
  - Run program with identical input data

All levels are present at the same time!

Reproducibility Contest at SC‘16

SCC Reproducibility Initiative Winner

Replication and reproducibility of experimental computer science results in peer-reviewed paper is gaining relevance in the HPC community. SC, the leading conference in the field, wants to promote and support replication and reproducibility through a new initiative that aims to integrate aspects of past technical papers into the Student Cluster Competition (SCC). The SCC is excited to announce “A parallel connectivity algorithm for de Bruijn graphs in metagenomic applications” as the winning paper for the inaugural reproducibility initiative. This paper and accompanying application will be reproduced in the SCC at SC16. This is the first time that students have been challenged to reproduce a paper rather than run prescribed data sets. Although they are doing similar tasks from previous competitions, they are seeing it from an entirely new perspective, as a component to the scientific process. “We want students to understand, early in their careers, the important role reproducibility plays in research.” explains the SCC Chair Stephen Harrell

(Follow-up activity to Stodden´s talk at SC‘15)

Computational Climate Science

What are the Problems Here?

- In Climate Science
  - Highly non-linear mathematics
  - Simulations over long time periods

- In Computer Science
  - Associative property does not hold
  - Machines are non-deterministic

Bitwise Reproducibility

What might harm bitwise reproducibility?

- Processors are exchanged
- Libraries are exchanged
- Compilers are exchanged
- Compiler options are changed
- Domain region is differently partitioned
- Number of compute nodes is changed
- Non-determinism in libraries with e.g. reduction operations
- Non-determinism in programs with e.g. load balancing

This is a non exhaustive list

- The order of items is non-deterministic

Procedure for Climate Modelling

Example: ICON (icosahedral non-hydrostatic general circulation model)

- Requirement

  Must deliver bitwise reproducible results for all possible domain compositions (strong requirement!)

- In fact: sequential and parallel must be identical

  They really compare small data sets
  (single core vs. multiple core)

Benchmarking Problem During Procurement

- Requirements
  - Vendor is allowed to use optimal compiler options
  - Vendor is not allowed to change the scientific result of the climate model programs
- Consequences
  - Bitwise reproducibility cannot be guaranteed
  - However, result must not be biased

Example of a Biased Result

Procedure

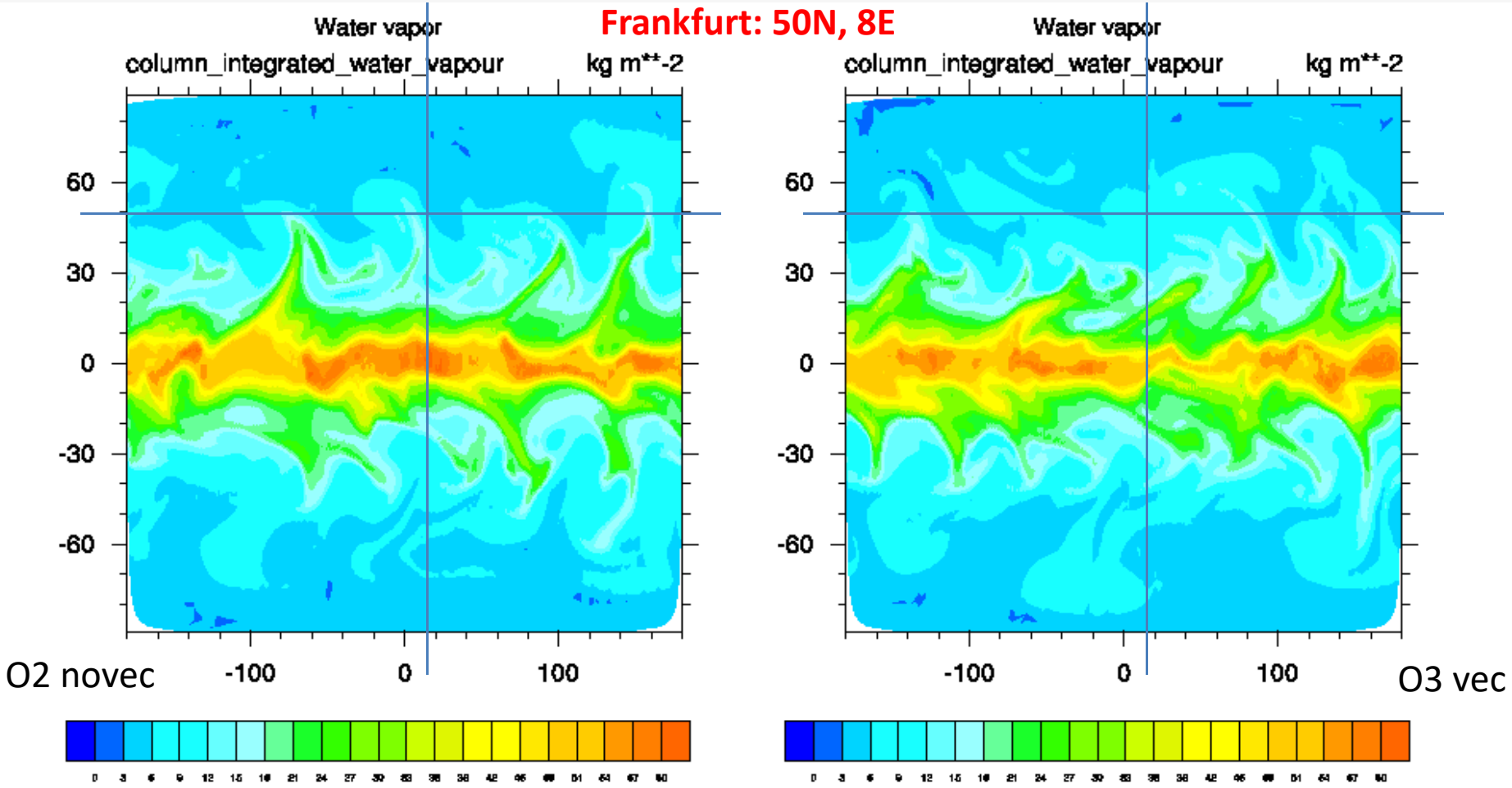
- Vendors use -O2 and -O3 with vectorization

Results

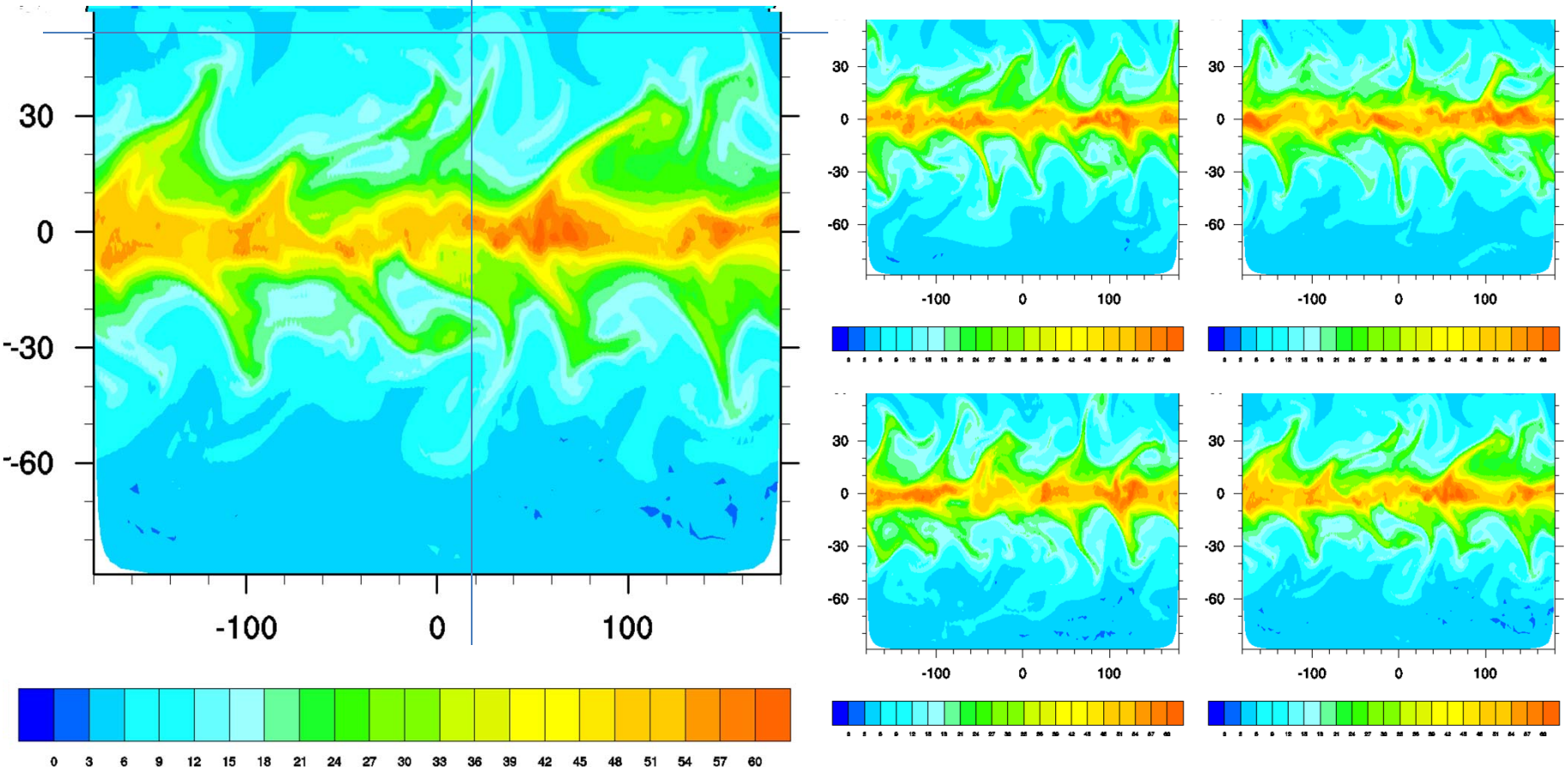
- Exhibits differences between
  - -O2 novec and -O3 vec
  - -O3 AVX from run to run

    AVX + FMA + dynamic memory management = non-deterministic

t_{sim}=100d O2: 10min30sec -> O3: 8m41sec => (+17,3%)



t_{sim}=100d O3 -x AVX – 4 runs with identical input



Some Observations

- Bitwise reproducibility difficult to implement
- Load balancing not possible
- Usage of optimized libraries prohibited
- High compiler optimization impossible
- Accelerator etc. presumably harmful

- And still not a reproducible scientific result
  (because you will never have the same computer with the same libraries, compilers etc.)
- And presumably still not a reproducible computer based experiment because of lack of description
  (a third party will not be able to reproduce it even on an identical environment)

Some Questions More Related to Science in General

Let‘s assume you really can reproduce an HPC based program run in a bitwise manner

- Why would you as the program author want to do that?

Answer from Max Planck Institute for Meteorology

- Find errors in parallelization
  - E.g. races in OpenMP-based implementations
- Understand influences of changes with
  - Compilers, libraries, global sums, ...

Some Questions More Related to Science in General

Let‘s assume you really can reproduce an HPC based program run in a bitwise manner

- Why would you as the program author want to do that?
- Why would someone else want to do that?
- Who would want to pay for doing this if it implies to copy e.g. the DKRZ infrastructure?
- Which progress in the history of mankind was based on being able to reproduce someone else’s scientific result in an identical way?

A Personal Pre-Final Observation

- More questions than answers
  - Different answers from different people
- Exascale will heavily intensify the problems
- Data intensiveness will intensify the problem
- The production of new scientific results increases exponentially

- Science needs to investigate the issue of how to evaluate the correctness and validity of its results
- Fortunately, we see more and more workshops on this issue

References

- Victoria Stodden
  https://web.stanford.edu/~vcs/talks/SC15-Nov182015-STODDEN.pdf
- Heise-Newsticker: Open Science
  https://www.heise.de/newsticker/meldung/Open-Science-Forscher-brechen-aus-dem-Elfenbeinturm-aus-3630380.html
- Dagstuhl-Seminar on Reproducibility
  http://drops.dagstuhl.de/opus/volltexte/2016/5817/
- Allison Baker et al.: Evaluating Lossy Data Compression
  http://www.geosci-model-dev.net/9/4381/2016/
- Miriam Leeser, Panel at SC16: Reproducibility and Repeatability
  http://sc16.supercomputing.org/presentation/?id=pan109&sess=sess177
- SC'16: SCC Reproducibility Initiative
  http://sc16.supercomputing.org/studentssc/scc-reproducibility-initiative-winner/