I GUESS I SHOULDN'T DO THAT

I WONDER IF THAT HAPPENS EVERY TIME.
Reproducibility: An Insight from the AGORA High-resolution Galaxy Simulations Comparison

Ji-hoon Kim (SLAC/Stanford/KIPAC)


Special thanks to: T. Abel (Stanford), O. Agertz (Surrey), N. Gnedin (Fermilab), R. Feldmann (Zurich), O. Hahn (Nice), B. Keller (McMaster), A. Lupi (IAP), P. Madau (UCSC), L. Mayer (Zurich), K. Nagamine (Osaka/UNLV), J. Primack (UCSC), B. Smith (Edinburgh), R. Teyssier (Zurich), M. Turk (NCSA), J. Wadsley (McMaster)
Fundamental Principle of Scientific Method

- Experiments must be **reproducible** to be established as knowledge.
“Reproducibility Crisis” - Nature Magazine

**How Much Published Work in Your Field is Reproducible?**

Physicists and chemists were most confident in the literature.

- **Chemistry**
- **Physics and Engineering**
- **Earth and Environment**

25% of respondents

**What Factors Contribute to Irreproducible Research?**

Many top-rated factors relate to intense competition and time pressure.

- Always/often contribute
- Sometimes contribute

- Selective reporting
- Pressure to publish
- Low statistical power or poor analysis
- Not replicated enough in original lab
- Insufficient oversight/mentoring
- Methods, code unavailable
- Poor experimental design
- Raw data not available from original lab
- Fraud
- Insufficient peer review
- Problems with reproduction efforts
- Technical expertise required for reproduction
- Variability of standard reagents
- Bad luck

Nature Survey of 1576 researchers (2016)
How Other Fields Are Dealing With It

- e.g. Reproducibility Project: Psychology / Cancer Biology

[Screen shot of Open Science Framework page]

Marshmallow experiment, pbs.org

http://osf.io/ezcuj

www.jihoonkim.org
“Reproducibility Crisis” in Psychology

- Only 36% of replicated studies show statistically significant results.

Nosek et al. for the Open Science Collaboration (2015, Science)
How About Galactic Astronomy?

• The success of our galaxy formation theory relies heavily on robust and reproducible numerical experiments.
How About Galactic Astronomy?

- The task of reproducing numerical experiments, or comparing simulations across platforms, has not received the highest priority.
Reproducible Simulation Raises Realism

- To increase the predictive power of numerical simulations, and the field itself, let us compare simulations across code platforms.
AGORA Comparison Infrastructure

- Includes key components necessary to run galaxy-scale simulations in a reproducible manner: code-independent and available to public
AGORA Dark Matter-Only Comparison

- Flagship paper with a proof-of-concept test (Kim et al. 2014)

- Fully established comparison pipeline

- Runtime parameters identified that make codes compatible with one another

- Publicly available ICs are being used to build a library of AGORA simulations making future comparisons trivial

~$10^{11}$ M$_\odot$ halo at $z=0$, projected DM density, Kim et al. (2014)

~$10^{12}$ M$_\odot$ halo at $z=0$, AGORA IC used in FIRE Collaboration
Gravito-Hydrodynamics Comparison

- **Second paper** with an isolated MW-mass disk test (Kim et al. 2016)

  - Subgrid physics models such as Jeans pressure floor, star formation, supernova feedback energy, and metal production **carefully constrained across code platforms**
  
  - High spatial resolution to minimize dependence on a phenomenological model

---

**Figure 2.** The 500 Myr composite of gas surface densities from Sim-noSF with radiative gas cooling but without star formation or supernova feedback. Each frame is centered on the galactic center – location of maximum gas density within 1 kpc from the center of gas mass. For visualizations of the particle-based codes hereafter (Figures 1-3, 14-15, 32, 34, 35) – but not in any other analyses except these figures – yt uses an in-memory octree on which gas particles are deposited using smoothing kernels. See Section 5 for descriptions of participating codes in this comparison, and Section 6.1 for a detailed explanation of this figure. Compare with Figure 14. Simulations performed by: Daniel Ceverino (ART-I), Robert Feldmann (ART-II), Mike Butler (ENZO), Romain Teyssier (RAMSES), Spencer Wallace (CHANGA), Ben Keller (GASOLINE), Jun-Hwan Choi (GADGET-3), Yves Revaz (GEAR), and Alessandro Lupi (GIZMO). The full color version of this figure is available in the electronic edition. The high-resolution versions of this figure and article are available at the Project website, http://www.AGORAsimulations.org/.
Convergence Among All 9 Codes

- Modern galaxy simulation codes agree very well with one another in many dimensions (agreement as good as within <10% at all radii).

![Graphs showing convergence among 9 galaxy simulation codes for surface density, rotational velocity, velocity dispersion, and Kennicutt-Schmidt relation profiles.](image)
Convergence Among All 9 Codes

- Intrinsic code differences are small and generally dwarfed by variations in the implementation of the common subgrid physics.

→ Predictions made from a modern high-resolution galaxy formation simulation are likely robust and reproducible.
Great, but how did we get here?

- Inter-code convergence achieved only after a Herculean effort by passionate participants, aided by many workshops and telecons.

Density-temperature PDF (run with no star formation)
Great, but how did we get here?

- Inter-code convergence achieved **only after** a Herculean effort by passionate participants, aided by many workshops and telecons.

![Density-temperature PDF (code order scrambled)](July 2015)

![Density-temperature PDF (run with no star formation)](October 2016)
Great, but how did we get here?

- Inter-code convergence achieved **only after** a Herculean effort by passionate participants, aided by many workshops and telecons.

7. DISCUSSION AND CONCLUSION

Through workshops and teleconferences, and via common languages and infrastructure built together, Project participants were able to better understand other codes, and improve their own. Participants found an optimal set of simulation parameters that makes their code to be best compatible with others. We came to understand how seemingly identical parameters differ in their meanings in different codes, and how seemingly different parameters have in fact identical meanings. In some comparisons, numerical errors were discovered and fixed in participating codes. The AGORA framework, now tested with the common physics and subgrid models, are serving as a launchpad to initiate *astrophysically-motivated* comparisons aimed at raising the predictive power of galaxy simulations, especially as we run the zoom-in cosmological simulations outlined in our flagship paper (Kim et al. 2014). In the coming years, we expect AGORA to continue to provide a sustainable and fertile platform on which numerical experiments are readily validated and cross-calibrated, and ambitious multi-platform collaborations are forged.

from Kim et al. (2016)

- (1) Human errors fixed
  - (2) Runtime parameters found that make the codes compatible with one another
  - (3) Errors in (some) codes fixed
A Human Experiment In Itself

- We have founded an **one-of-a-kind, open forum** where numerical astrophysicists can talk to and learn from one another.

Kim et al. (2016)
We have founded an one-of-a-kind, open forum where numerical astrophysicists can talk to and learn from one another.
Conclusion

- We strive to promote **collaborative and reproducible research** in the numerical galaxy formation community.

**Goals and challenges**

AGORA aims to increase the predictive power of numerical simulations through a multi-platform approach.

**New possibilities**

AGORA offers a unique opportunity to validate answers to long-standing problems in galaxy formation.
Thank you!
Supplemental Slides
AGORA
A High-resolution Galaxy Simulations Comparison Initiative: www.AGORAsimulations.org

AGORA Project: Goal and Team

- **GOAL:** A collaborative, multi-code platform to **raise the realism and predictive power** of high-res galaxy simulations

- **TEAM** - 140+ participants from 60+ institutions, 10/2016
  - 10+ groups each with variations of 9+ codes
  - 5 conferences & 11 web conferences organized
  - Project Coordinator: Ji-hoon Kim (Stanford/SLAC)

- **DATA SHARING:** Initial conditions, astrophysics modules, analysis software, and simulation outputs all to be public


High-res Galaxy Simulation

Variation of the official AGORA intro slide (credit: Kim & Governato) / Project funded in part by:

[www.jihoonkim.org](http://www.jihoonkim.org)