What Do We Do Next?

Kenneth P. Bowman

Department of Atmospheric Sciences

Texas A&M University

TEXAS A&M
UNIVERSITY

ATMOSPHERIC
SCIENCES

- *Observations* - distinguish modern science from the classical understanding of the universe
- *Communication* - observations, theories, experimental results
- *Continuity* - preserving and building on existing data and ideas
- *Reproducibility* - repeated testing of both theory and observations

- We have a wonderful problem
  - Enormous new high-quality data sets
  - Amazing ability to simulate the physical world
- How can we take advantage of the opportunities that technology have brought?

TEXAS A&M
UNIVERSITY

Data

- Observations tell us something about the real world
- Because observations can never be recreated, it makes sense to save all observational data
- Challenges of size, access, discoverability, usability, preservation
- Difficult issues, but archiving data addresses the fundamental issues of communication, continuity, and reproducibility

TEXAS A&M
UNIVERSITY

Modeling and Simulation

- Simulations are just calculations (based on theory, often with empirical components where theory is inadequate)

- Allow controlled experiments (with approximations to the real world) and predictions of the future

- Output from numerical models faces the same challenges as observational data (size, access, discoverability, usability, preservation), but … in my opinion … few, if any, results from simulations need to be preserved *permanently*

- Models improve with time, so model output has a limited useful lifetime

- Because models can generate output in gigantic volumes, it is impossible to store everything and would be a waste of resources

Archiving Model Output

- How much model output should we preserve and for how long?
  - Suggestion: (*thanks to Gretchen Mullendore*) a workshop for modelers and data scientists to discuss priorities and develop guidelines for what should be saved from different types of models and different experiments and for how long

- Why not just archive the model code and re-run it if necessary?

What about Code?

- Analysis code and 'small' models
- Do these have to be made public? Do they have to be archived?
- Example: TRAJ3D
  - Developed and improved over many years
  - Highly tested
  - Fast, accurate, and flexible
  - Algorithms and results have been described in multiple publications
- Code is intangible and essentially free to duplicate
- Should instruments be required to be 'public'?

TEXAS A&M
U N I V E R S I T Y

Requests (as a modeler)

- We need guidelines on what to archive and for how long
  - Should be worked out in consultations among agencies, publishers, and scientists
- Consistency among the funding agencies and publishers would be *very* helpful
- Guidelines will evolve with time
- Start with the minimum requirements
- We probably need discipline-specific repositories. PI, departmental, or university resources are not likely to be sufficient
- This will cost money and time