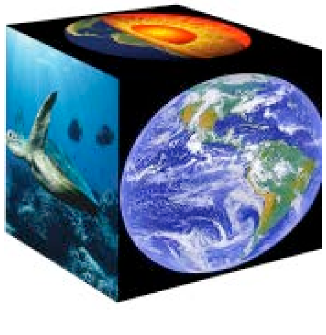
Embracing uncertainty: Creating a real-time University-based National Ensemble – a use case for EarthCube?

Michael C. Morgan

Division of Atmospheric and Geospace Sciences

National Science Foundation

Shaping the Development of EarthCube to Enable Advances in Data Assimilation and Ensemble Prediction



17 December 2012

Goals of presentation

- Motivate the idea for a national ensemble
- Encourage a discussion of
  - what’s possible?
  - community interest in participation
  - how such a system might work?
  - the necessary infrastructure – how might EarthCube facilitate the emergence of this ensemble?

Weather prediction as a deterministic initial value problem

Wilhelm Bjerknes

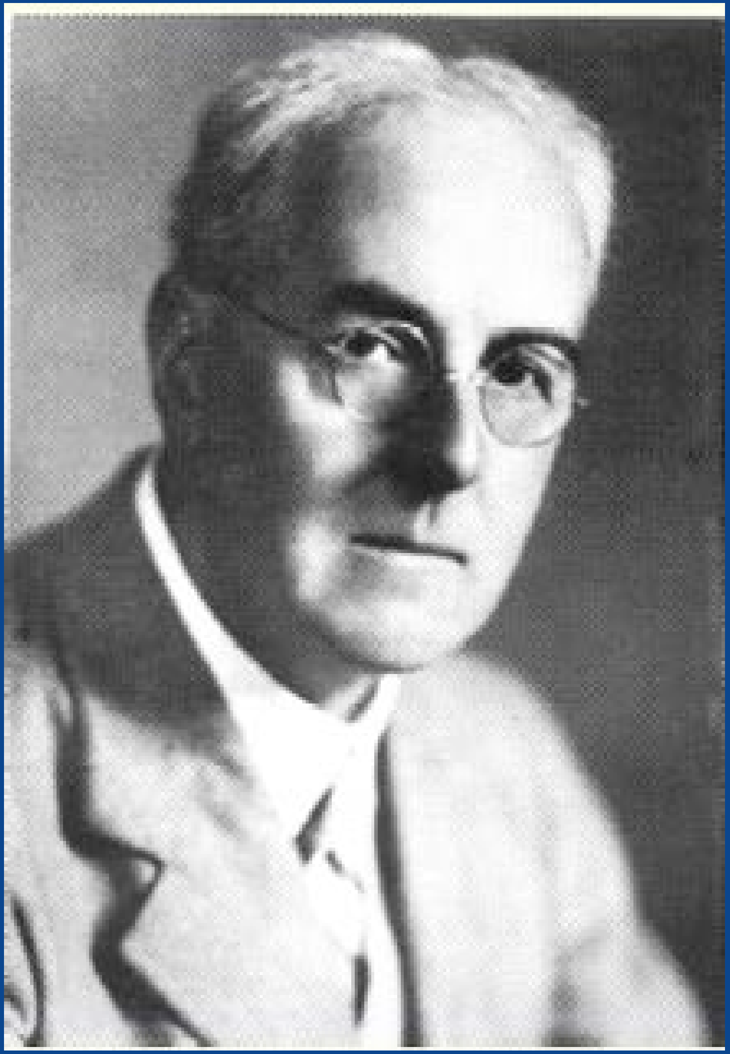
The prediction problem:

1. *One must know with sufficient accuracy the atmospheric condition at a particular time.*

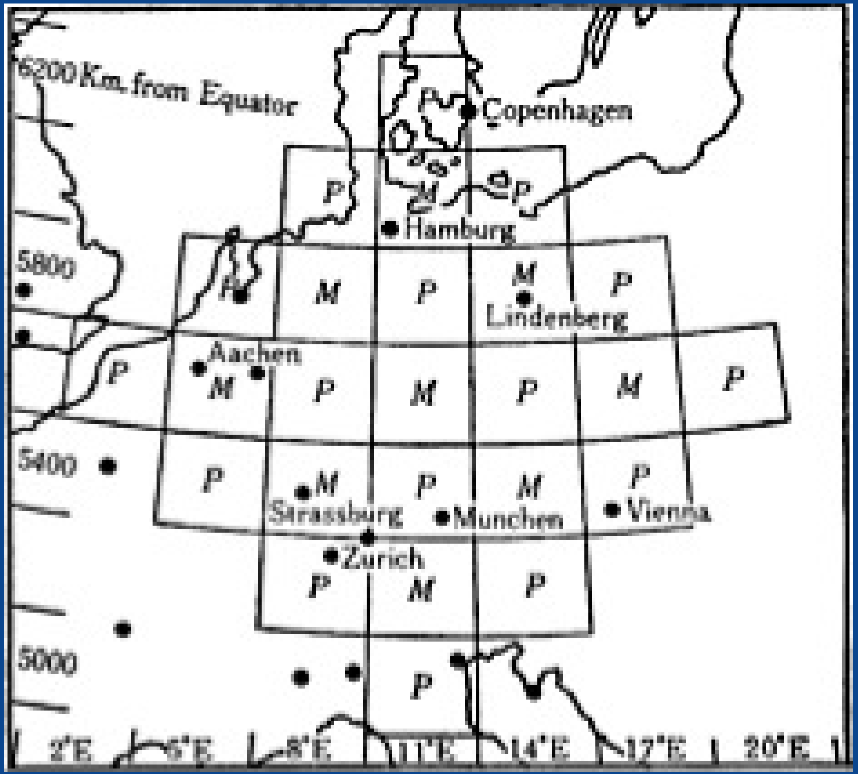
2. *One must know with sufficient accuracy the laws that govern the evolution of the atmosphere from one state to the next.*

Meteorologische Zeitschrift (1904)

Weather prediction as a deterministic initial value problem



Lewis Fry Richardson



"After so much hard reasoning, may one play with a fantasy? Imagine a large hall like a theatre . . . "

Richardson’s “forecast factory”

Integration of the Barotropic Vorticity Equation

VOLUME 2, NUMBER 4 **Tellus** NOVEMBER 1950

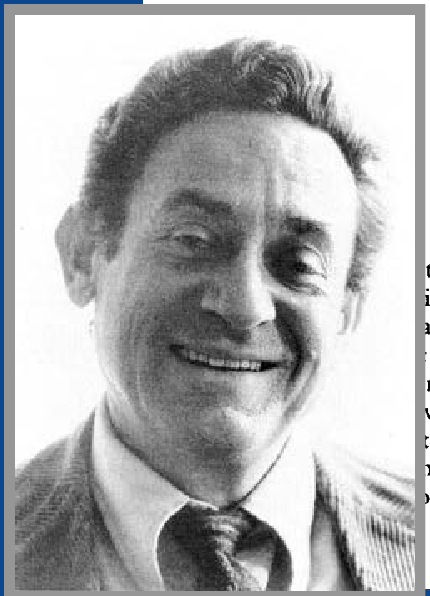
A QUARTERLY JOURNAL OF GEOPHYSICS

Numerical Integration of the Barotropic Vorticity Equation

By J. G. CHARNEY, R. FJÖRTOFT[1], J. von NEUMANN

The Institute for Advanced Study, Princeton, New Jersey[2]

(Manuscript received 1 November 1950)



Abstract

thod is given for the numerical solution of the barotropic vorticity equation imited area of the earth's surface. The lack of a natural boundary calls for an ation of the appropriate boundary conditions. These are determined by a argument and are shown to be sufficient in a special case. Approximate ns necessary to insure the mathematical stability of the difference equation ved. The results of a series of four 24-hour forecasts computed from actual he 500 mb level are presented, together with an interpretation and analysis. npt is made to determine the causes of the forecast errors. These are ascribed the use of too large a space increment and partly to the effects of baroclinicity. of the latter is investigated in some detail by means of a simple baroclinic model.

The predictability of a flow which possesses many scales of motion

The predictability of a flow which possesses many scales of motion

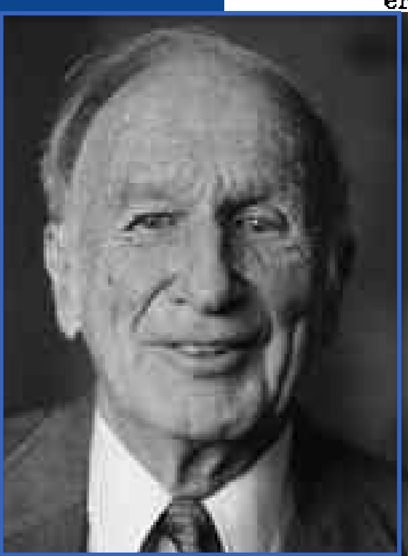
By EDWARD N. LORENZ, *Massachusetts Institute of Technology*[1]

(Manuscript received October 31, 1968, revised version December 13, 1968)

ABSTRACT

It is proposed that certain formally deterministic fluid systems which possess many scales of motion are observationally indistinguishable from indeterministic systems; specifically, that two states of the system differing initially by a small "observational error" will evolve into two states differing as greatly as randomly chosen states of the stem within a finite time interval, which cannot be lengthened by reducing the nplitude of the initial error. The hypothesis is investigated with a simple mathe- atical model. An equation whose dependent variables are ensemble averages of the rror energy" in separate scales of motion is derived from the vorticity equation which verns two-dimensional incompressible flow. Solutions of the equation are determined numerical integration, for cases where the horizontal extent and total energy of the stem are comparable to those of the earth's atomsphere.

It is found that each scale of motion possesses an intrinsic finite range of predictabi- y, provided that the total energy of the system does not fall off too rapidly with creasing wave length. With the chosen values of the constants, "cumulus-scale" otions can be predicted about one hour, "synoptic-scale" motions a few days, and the rgest scales a few weeks in advance. The applicability of the model to real physical stems, including the earth's atmosphere, is considered.



Prediction and data assimilation

- The atmospheric sciences occupies a unique place in the geosciences because of our maturing capacity to predict the complex interdependent physical system of the atmosphere with increasing fidelity and our increasing understanding of predictability.
- AS has exploited advances in optimal control/estimation theory to develop increasingly sophisticated schemes to assimilate a wide variety of state estimates of our environment to produce estimates of the global environmental state on a regular basis – retrospectively and in real-time.

Prediction and data assimilation

- As we move forward in an "age of observation," with increasing diversity of observing systems and their concomitant data volumes, DA research across the geosciences will be seen to have increasing intrinsic merit because DA is a fundamental tool for optimally determining the state of the *Earth system*.
- The goal of EarthCube is to transform the conduct of geoscience research by supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the geosciences.

Themes

1) Numerical weather prediction models, arguably one of the 20th century's greatest scientific achievements, are best used in an ensemble prediction mode rather than being run to produce a single deterministic forecast.

2) "Uncertainty" should be viewed not as an obstacle in our ability to predict, but rather viewed as an important source of information that may be used to enhance our ability to predict – known by many here, not so much by the public.

3) Just as observations and analyses have uncertainties, numerical forecast states should be considered *incomplete* without their concomitant uncertainties.

Themes

4) In order to make progress in this enterprise, tools must be developed to ease the use, exploration, and interrogation of observations, analyses, and forecasts from an ensemble prediction system.

5) Despite the emerging view that ensemble prediction and data assimilation have led to improvements in our Nation’s predictive skill, we are woefully undertraining and exciting interest in these endeavors and risk falling further behind internationally with respect to our Nation’s modeling infrastructure.

Themes

7) Considerable infrastructure exists for supporting DA and ensembling efforts at universities: models (community, university, lab), tools for assimilation (e.g., DART), tools for data transfer and data exploration (e.g., Unidata's LDM)

8) There exists no central repositories of geophysical data (relevant to data assimilation) that are easily accessible

Why probabilistic forecasts?

- Small uncertainties in the IC’s or the NWP model may evolve over time to meso- and synoptic-scale errors (Lorenz 1963).
- The predictability of the detailed weather evolution is limited (Lorenz 1969).
- “An objective way to estimate not only the most likely weather evolution, but also the uncertainty of the forecast, is to run an ensemble prediction system, which provides a probabilistic forecast of the atmospheric evolution.” Berner et al. (2011)

Design of an ensemble system

Effective perturbation selection

Selection of initial perturbations for ensembles requires that the perturbations grow (to ensure sufficient ensemble spread at future times) and that the initial perturbations share characteristics of analysis errors. Methods include: bred modes (NCEP), SVs (ECMWF)

Confronting model deficiencies

Ensemble spread may be diminished due to model errors. Ensembles generated from differing model physics (physics-based, stochastic perturbation parameterization schemes), BCs, and models have been implemented.

Ensemble limitations/issues?

- The *small size* of the ensemble can cause an "under-representation" problem when generating the background covariances for ensemble-based data assimilation. How many members? Is the ensemble under-dispersive?
- With both perturbed initial conditions and various physical parameterizations, ensemble forecasts can take into account both initial and model errors. Which approach is more effective? Would a hybrid approach be more effective?
- Can resources be harnessed nationally in creating the infrastructure to broaden community engagement in understanding and ultimately addressing these issues?

After so much [hard] reasoning, may one play with a fantasy? . . .

GOAL: A university-based national ensemble

To advance the national ensemble prediction capability by:

1. providing the tools necessary to accelerate research: a large, real-time ensemble database, tools to verify and interrogate ensemble output;
2. increasing accessibility and participation of the university community to this enterprise;
3. providing the non-atmospheric science community a sense of how uncertainty can be tapped for economic and societal benefit

CONCEPT: A university-based national ensemble

1. Ensemble would be created using the computing resources resident on desktops in university labs across the country (world?)
2. Initial conditions could be transferred to each computing system and model output (a single forecast) transferred to a central, cloud accessible, repository.
3. Each forecast challenged with an analysis of record
4. Analyses and output available to public/private sector for research (and perhaps operational) uses.

Potential "Deliverables"

- Once-a-day ensemble of O(100) members
- Analyses of record to initialize model runs and verify forecasts
- Characteristic errors associated with particular model configurations
- Background errors covariances for data assimilation

What could be accomplished?

- Foster education and training in:
  - NWP and data assimilation
  - Ensemble modeling
- Support and enhance research in:
  - Ensemble modeling
  - Conveying uncertainty
  - Model improvements
- Community engagement
- Support for
  - Operational requirements
  - Virtual field campaigns

How might this be accomplished?

- Consider using the WRF-ARW
  - Already thousands of users in the U.S.
  - Many are engaged in real-time forecasting and assimilation efforts for research and other purposes
- Many use the LDM to acquire data and /or analyses for initializing their model runs

WRF-ARW possible ‘physics’ configurations

14 - microphysics
3 - zero moisture options
2 - hail
2 - ice
2 - latent heating
7 - lw
7 - sw
8 - sfclay
5 - surface physics
5 - urban physics
12 - boundary layer
10 - cumulus
3 - shallow cumulus
2 - subsidence spread
2,2,2 - heat, snow, ice fluxes from surface

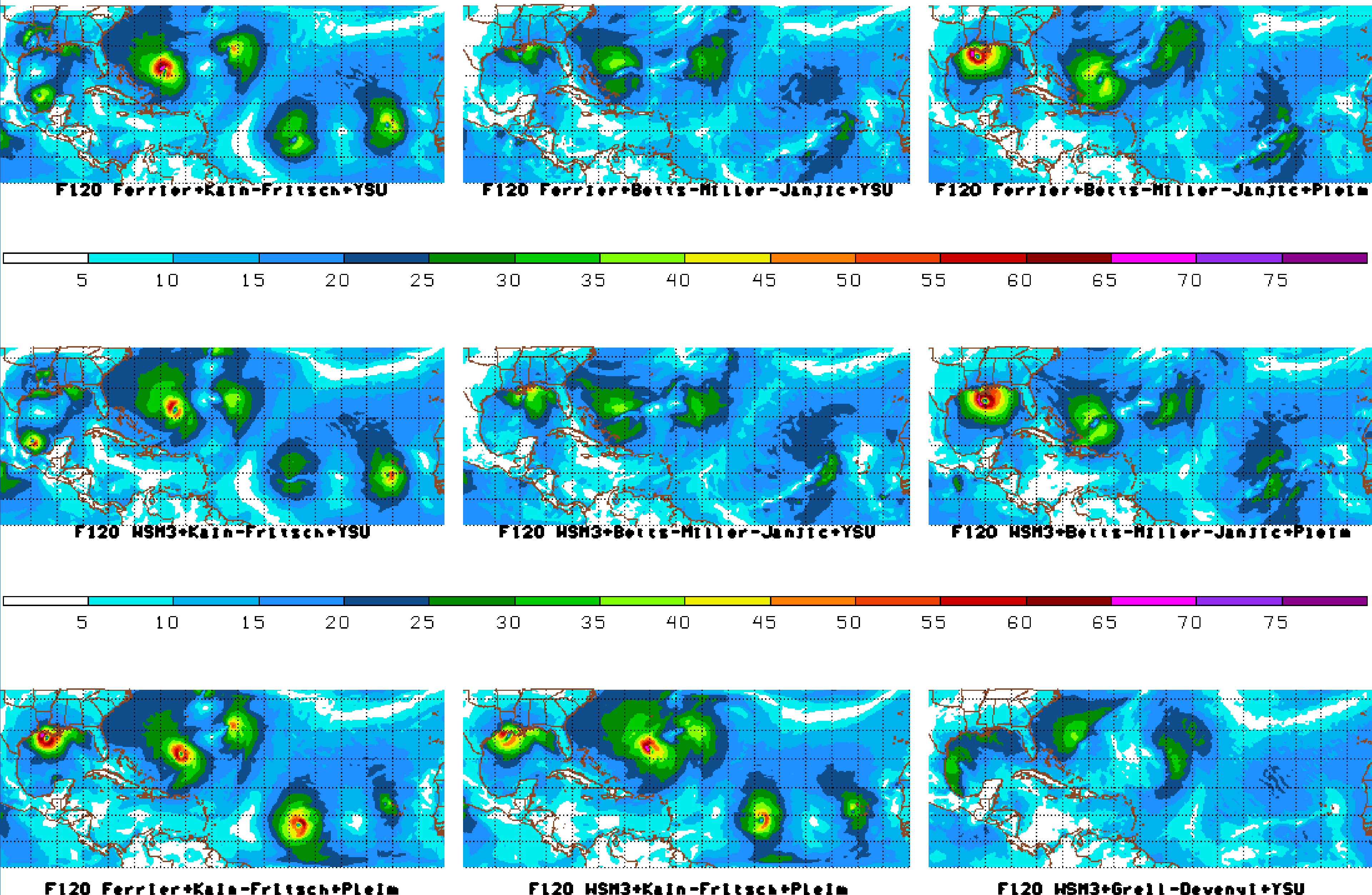
WRF-ARW possible ‘physics’ configurations

14 - microphysics
3 - zero moisture options
2 - hail
2 - ice
2 - latent heating
7 - lw
7 - sw
8 - sfclay
5 - surface physics
5 - urban physics
12 - boundary layer
10 - cumulus
3 - shallow cumulus
2 - subsidence spread
2,2,2 - heat, snow, ice fluxes from surface

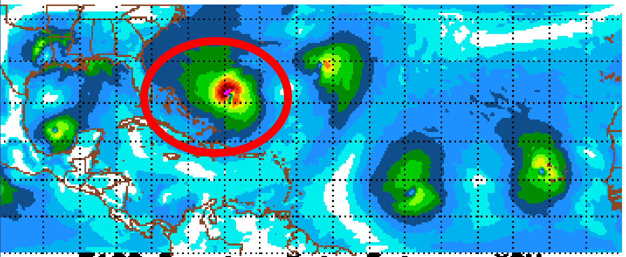
Just considering the microphysics, BL schemes and CP’s, you could conceivably get an ensemble of 1680 members (or so)

Clearly could create a large-member, multi-member physics-based ensemble using a single dynamical core and various parameterization packages for the unresolved physics

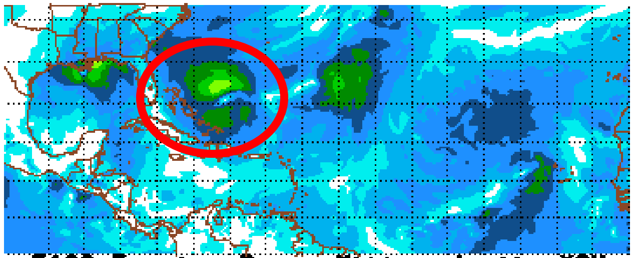
Inner Grid 120 Hour Forecast Initialized 1200 UTC 28 August



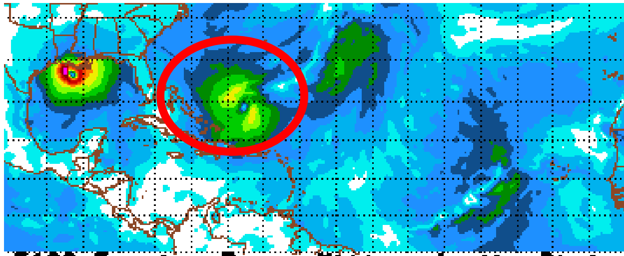
Future Hurricane Hanna



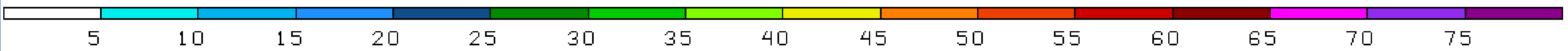
F120 Ferrier+Kain-Fritsch+YSU

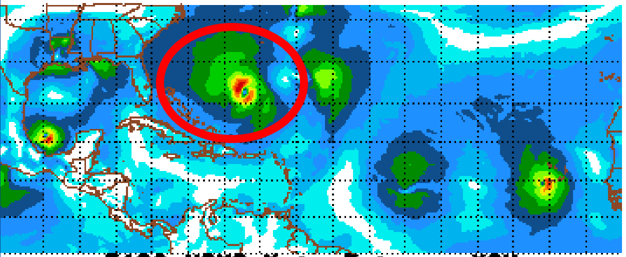


F120 Ferrier+Betts-Miller-Janjic+YSU

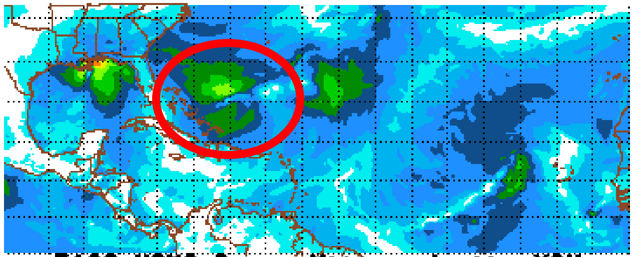


F120 Ferrier+Betts-Miller-Janjic+Pleim

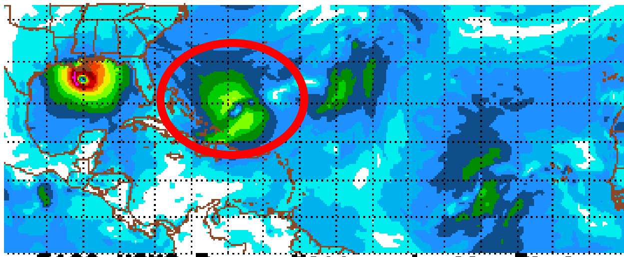




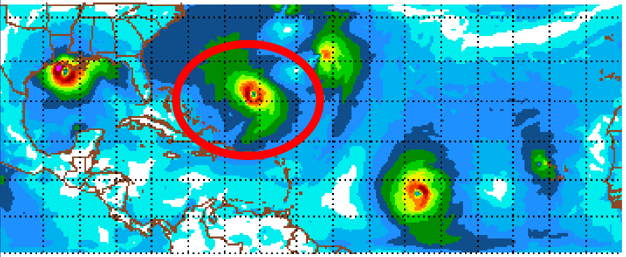
F120 WSM3+Kain-Fritsch+YSU



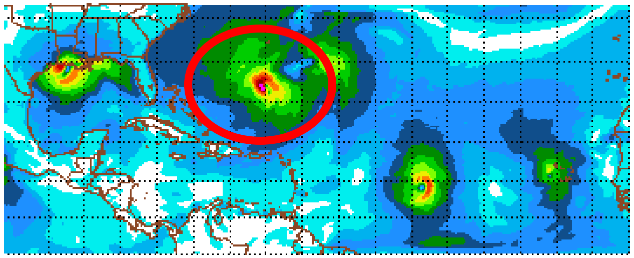
F120 WSM3+Betts-Miller-Janjic+YSU



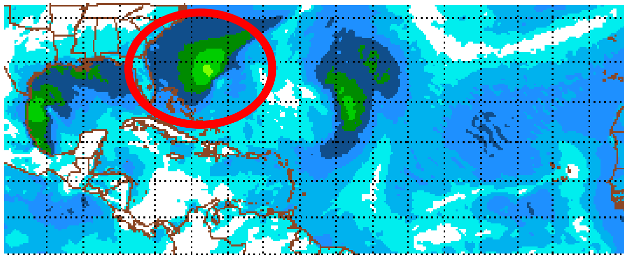
F120 WSM3+Betts-Miller-Janjic+Pleim



F120 Ferrier+Kain-Fritsch+Pleim



F120 WSM3+Kain-Fritsch+Pleim



F120 WSM3+Grell-Devenyi+YSU

Potential opportunities for advancing

- **Science**
  - Improvement in the tools used to do science
  - What ensemble size for error statistics to converge?
- **Operations**
  - New uncertainty products
  - Development of forecast scenarios
- **Education**
  - Allows participants to understand, NWP, assimilation, uncertainty evolution and communication

Issues/Challenges/Questions

- What sort of ensemble (IC, MP, MM)?
- What domain? Limited-area or global: dictates length of integration – trade-offs (grid-spacing, size of community participation, "research-worthiness")
- Who decides all of the above?
- Volume and quality of data?
- What tools are necessary to interrogate this massive data set?
- Would such a data set be of use to the research and operational communities?

What's necessary to begin?

- Community interest with a compelling science case. What science could be accomplished with this data set?
- Community organization and coordination
- Community plan for work- and data-flows
- Community plan for education and training

Possible implementation path

- Organizing workshop
- Pilot examples . . .
  – North American domain for a year?
  – N. American domain for winter/spring and Tropical domain for summer and fall?
  – Global domain for a year?
- Post-pilot
  – Regional virtual field campaigns triggered by actual field campaigns or events of special interest (e.g., “high-impact” events)
  – Full-scale spin-up with a larger community

What would NSF *possibly* support?

NSF, broadly

- The National Science Foundation Act of 1950 (Public Law 81-507) set forth NSF's mission and purpose:
  - *To promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense....*
- The Act authorized and directed NSF to initiate and support:
  - basic scientific research and research fundamental to the engineering process,
  - programs to strengthen scientific and engineering research potential,
  - science and engineering education programs at all levels and in all the various fields of science and engineering,
  - programs that provide a source of information for policy formulation,
  - and other activities to promote these ends.

Mission of AGS

To extend intellectual frontiers in atmospheric and geospace sciences by making responsible investments in fundamental research, technology development, and education that enable discoveries, nurture a vibrant, diverse scientific workforce, and help attain a prosperous and sustainable future.

What would NSF *possibly* support?

- Necessary infrastructure to support a pilot
  - DA, organization
  - Data repository and cloud-like access
  - Development of tools to interrogate data (EC)
- Research and education using results of pilot data
- Through EC – possibly the data repository and tools to interrogate post-pilot activities

For your consideration

- Are there science questions that could be explored and educational benefits that could be realized with [creation of] this data set?
- Is there community interest in contributing collaboratively to this endeavor?
- Are there benefits to public and private sector operational communities that could entice them to contribute?
- What infrastructure is necessary to make this possible?