Present and Future Computing Requirements for 20th Century Reanalysis Project and future Sparse Input Reanalyses for Climate Applications (SIRCA) 1830-2017

Gilbert P. Compo

University of Colorado CIRES and

NOAA Earth System Research Laboratory

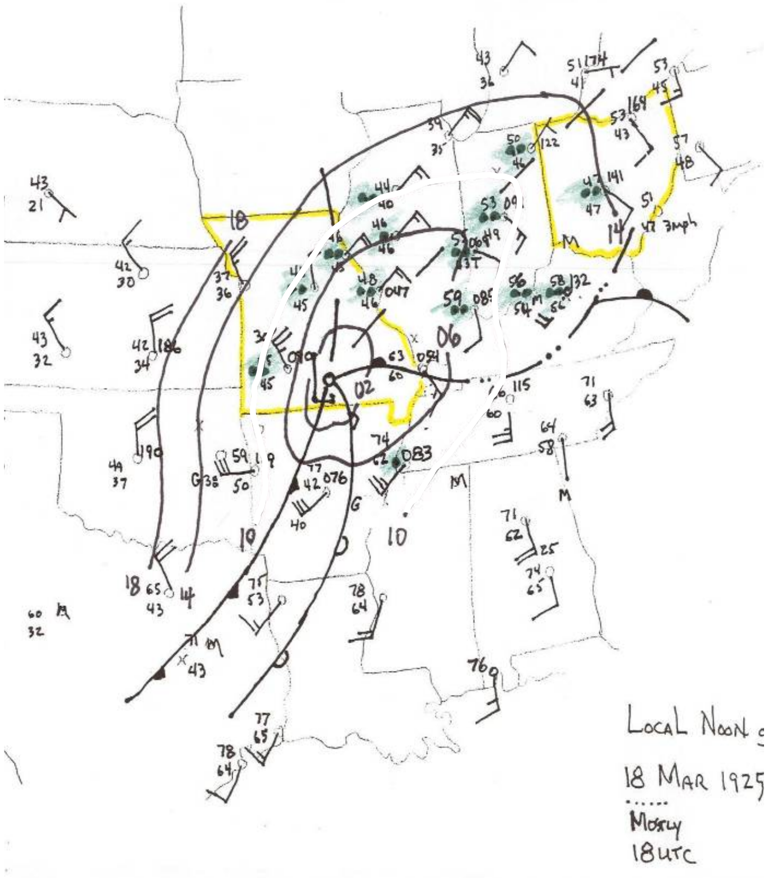
NERSC BER Requirements for 2017
September 11-12, 2012
Rockville, MD

1. Project Description

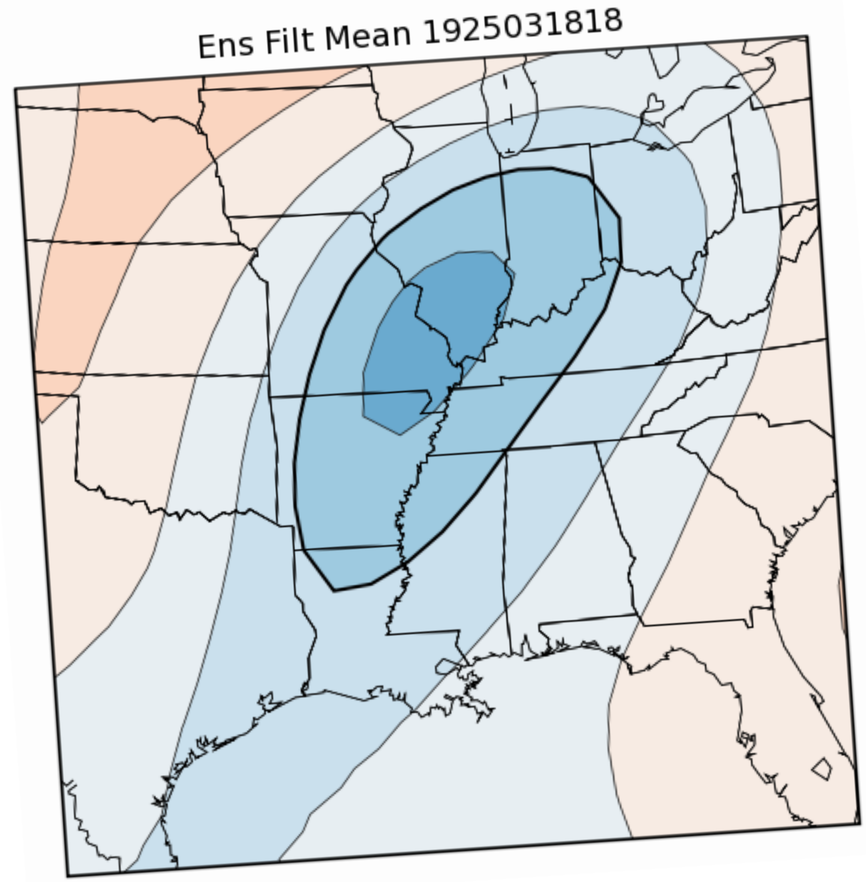
G. P. Compo, J.S. Whitaker , P. D. Sardeshmukh – U. of Colorado/NOAA ESRL
B. Giese, H. Seidel – Texas A&M

- Our present focus is to reconstruct the global weather and ocean states back to the 1850s at 2° latitude/longitude resolution for the atmosphere and 0.5 ° latitude/longitude for the ocean. This reconstruction includes quantified uncertainties for the first time.
- By 2017, we expect to reconstruct global weather and climate of the ocean and atmosphere back to the 1830s at 0.5 ° degree latitude/longitude for the atmosphere and 0.25 ° degree latitude/longitude to improve the representation of hurricanes, severe storms, and other extreme and high-impact weather.

Sea Level Pressure analyses from 20th Century Reanalysis for Tri-State Tornado Outbreak of 18 March 1925 (deadliest tornado in U.S. history)



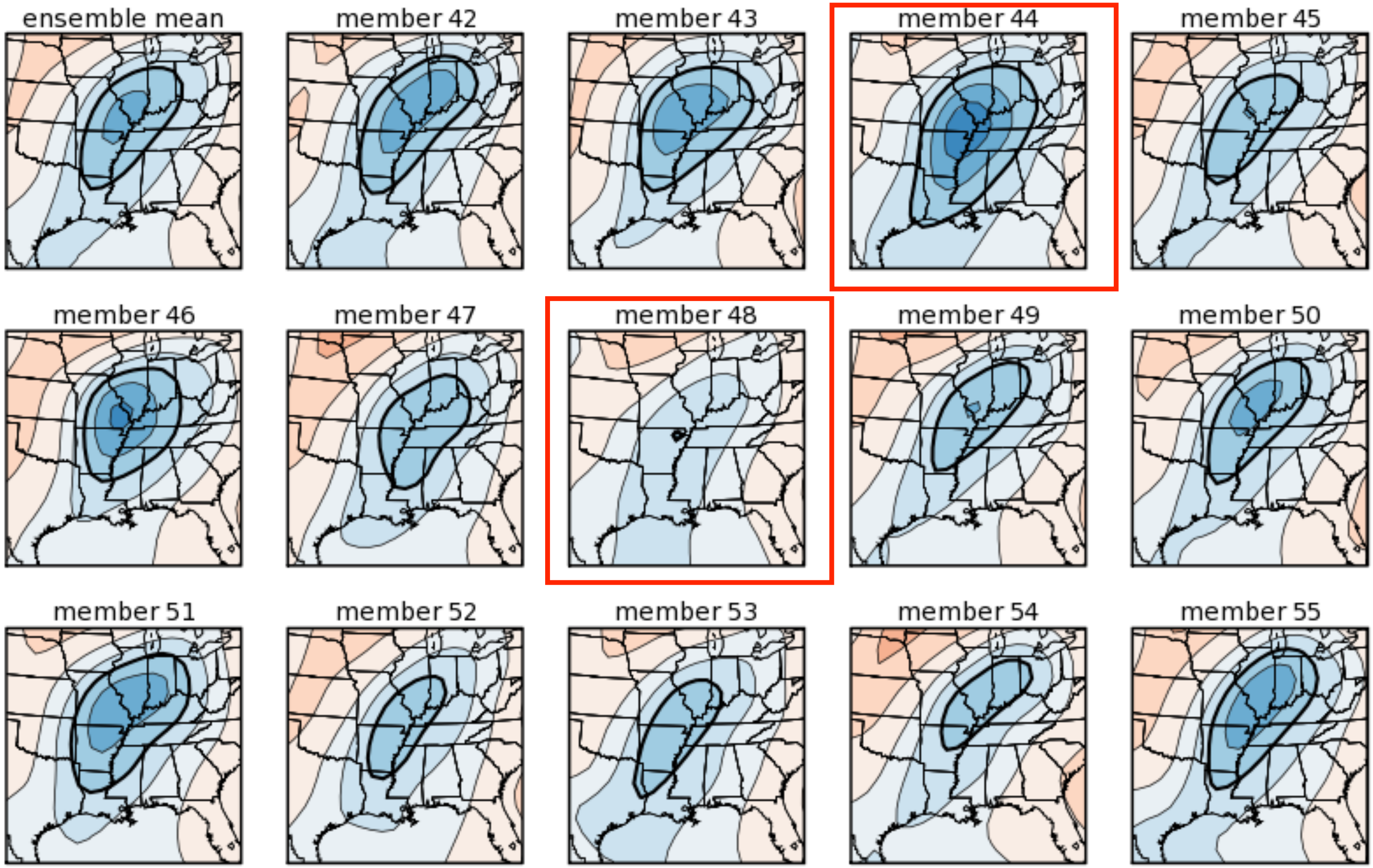
Manual Analysis, courtesy B. Maddox



Ensemble mean from Ensemble Filter (4 hPa interval, 1010 hPa thick)

NOTE!!! This analysis did not use ANY of the observations shown on the left.

Range of possibilities for Sea Level Pressure
18 March 1925 18Z using 14 (of 56) members

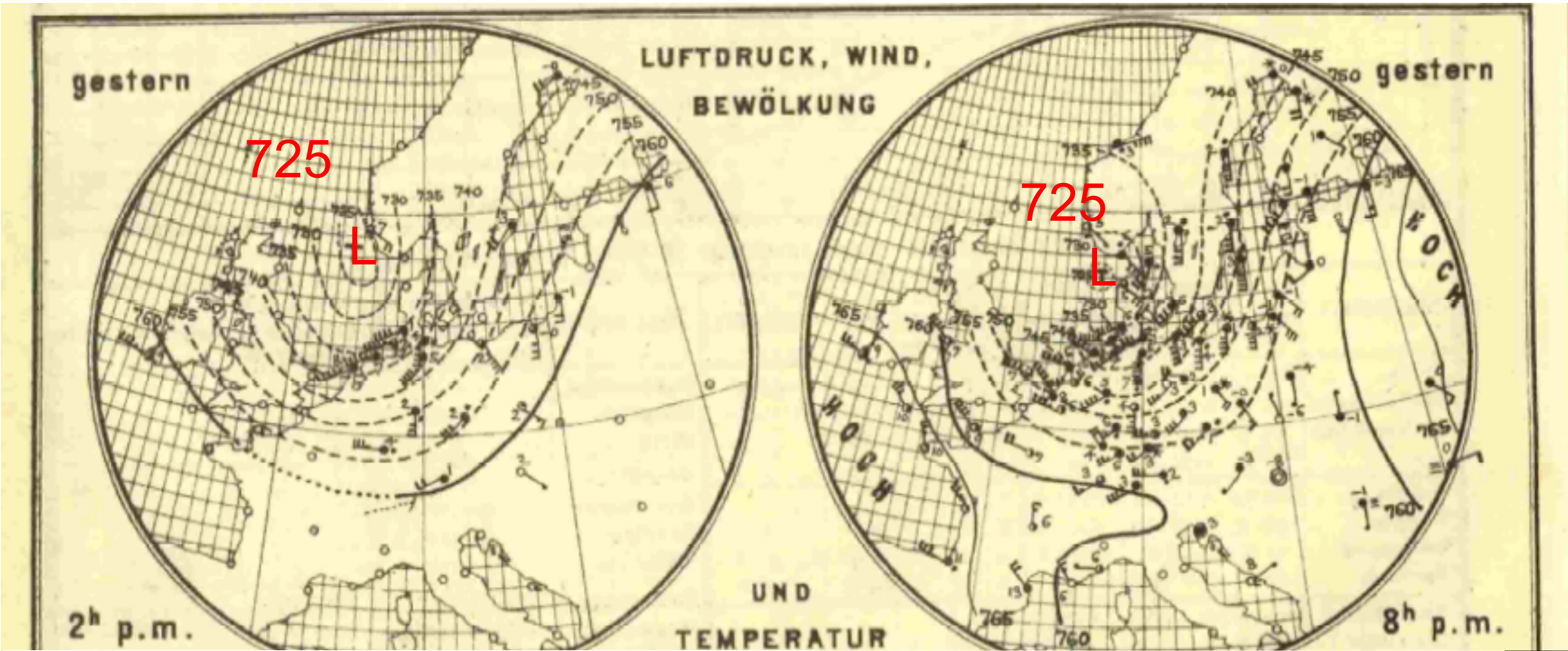


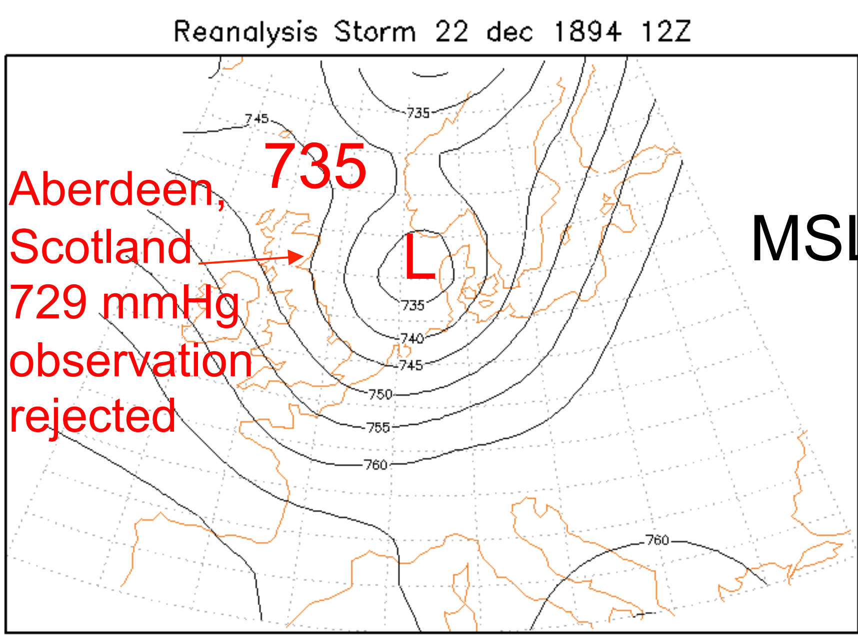
Ensemble of 56 possible realizations consistent with the observations

De Storm van 1894 (Zenit. 2010)

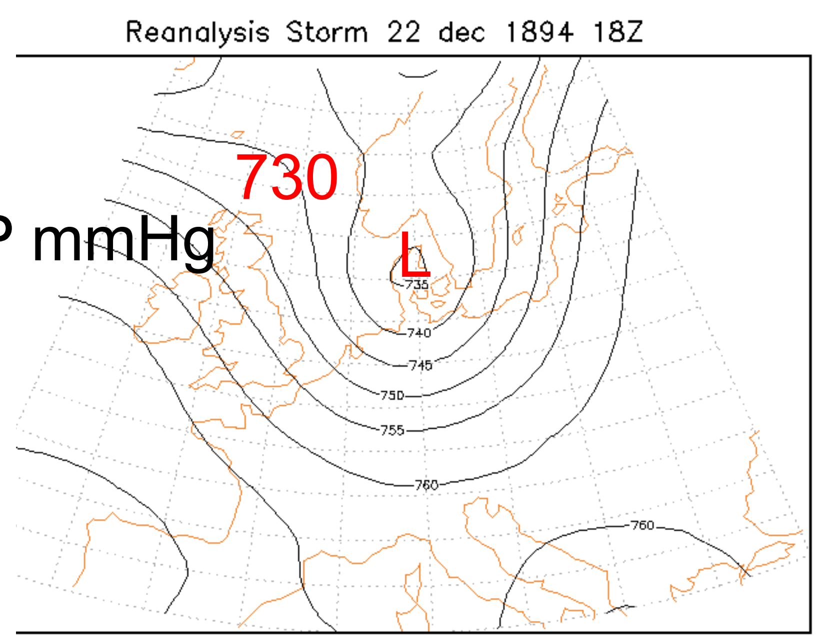
Henk de Bruin and Huug van den Dool

Frank Beyrich and **Britta Bolzmann** (DWD) provided 1894 weather maps of the ***Seewarte Hamburg***





MSLP mmHg



De Bruin and van den Dool (2010)

2. Computational Strategies

- We approach this problem computationally at a high level for **20CR**:1. generate 56 six-hour weather forecasts at 2 degree and 28 levels

  2. update these forecasts with the available surface pressure observations using a Monte-Carlo approximation to the Kalman Filter called an ensemble Kalman Filter.
- **Simple Ocean Data Assimilation system**:

  1. generate ocean climate model state of 10 days
  2. determine difference with sea surface temperature observations
  3. re-run model with the difference as forcing.

  The codes we use are the NCEP Global Forecast System, ensemble data assimilation system, Parallel Ocean Program, Simple Ocean Data Assimilation
- Our biggest computational challenges are efficient use of MPI and have a larger ensemble in the face of considerable I/O
- Our parallel scaling is limited by I/O

- These codes are characterized by these algorithms:

NCEP Global Forecast System: discretized approximation to compressible Navier-Stokes fluid equations on a sphere and laws of thermodynamics. Using spherical harmonics for dynamical terms and parameterized representation of nonlinear terms including moist processes, turbulence, and radiation.

Ensemble Kalman Filter: Monte-Carlo approximation to optimal solution for combining a dynamical model and observations when errors are Gaussian. Includes representation of model error via co-variance inflation.

- We hope to unify the atmosphere (20CR) and ocean (SODA) assimilation into a single coupled or loosely coupled system. Because the ocean timescales are long, the system may have to be run in sequence, rather processing many years in parallel, as done in 20CR.

- Compute at NERSC because easy to get lots done efficiently.

3. Current HPC Usage (see slide notes)

- Machines currently using: (was) Franklin, Hopper
  - Carver (was euclid) for data pre and post-processing
  - dtn's for data transfer
- Hours used in 2012
  - Franklin: 3.9 million
  - Hopper: 6 million
  - Carver:? , DTNs: ?
  - Euclid: 3 thousand
- Typical parallel concurrency and run time, number of runs per year
  - 20CR: 1344 cores, 1 hour generates 2 days of assimilated data, 4500 runs in a year
  - SODA: 2400 cores, 5.5 hours generates 1 year of assimilated data , 161 runs in a year for each experiment * 2-3 experiments
- Data read/written per run
  - 20CR: 150GB
  - SODA: 4.5TB
- Memory used globally: 6GB

- Necessary software, services or infrastructure
  netCDF4, python, climate data operators, NCL, IDL, intel compiler fortran90 and C,
  NERSC science gateway, data transfer nodes, HPSS

- Data resources used
  - HPSS: 250 TB
  - /project : 80TB
  - /scratch (s) : 100TB

4. HPC Requirements for 2017

(Key point is to directly link NERSC requirements to science goals)

- Compute hours needed (in units of Hopper hours)
  - 7 X 10^8 (700 million) hours to generate 64 global atmosphere states every six hours from 1830 to 2017 at 0.5 degree latitude/longitude resolution and 64 vertical levels to compare variations in hurricanes and severe storms with those simulated by climate models. Every 5th year is generated in parallel (32 streams).
- Changes to parallel concurrency, run time, number of runs per year
  - 20CR: 262144 cores, 1 hour, 1826 runs
  - SODA: 4800 cores, 4 hours, 150 per experiment, 2 -3 experiments
- Changes to data read/written
  - 20CR: 1.6TB per run = 3PB
  - SODA: 9.0TB per run = 1.35 PB
- Changes to memory needed per node | globally
  - 25 GB on at least one node | 36 GB globally

5. Strategies for New Architectures

- Our strategy for running on new many-core architectures (GPUs or MIC) is *to wait for compiler directives*

- To date we have prepared for many core by …

- We are already planning to do …

- To be successful on many-core systems we will need help with *implementation in the absence of compiler directives*.

5. Summary

- What new science results might be afforded by improvements in NERSC computing hardware, software and services? **Determining the observed variability and change in hurricanes, severe storms and floods, polar lows , and other extreme and high impact weather events that require higher spatial resolution than used in 20CR**
- Recommendations on NERSC architecture, system configuration and the associated service requirements needed for your science – **need to improve I/O**
-

  NERSC generally refreshes systems to provide on average a 2X performance increase every year. What significant scientific progress could you achieve over the next 5 years with access to 32X your current NERSC allocation?

  Reconstruct global ocean and atmosphere weather and climate back to 1830 at 4 x the resolution currently used and include objective estimates of uncertainty in both.
- What "expanded HPC resources" are important for your project? – **More processors and shorter queue times**

Thank you