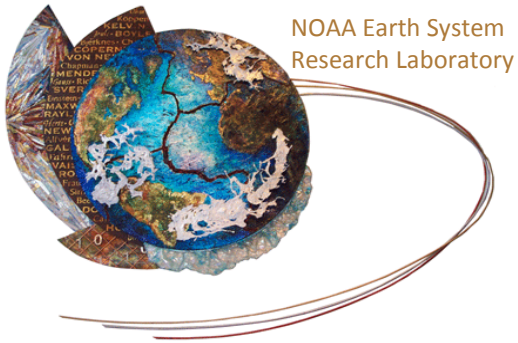


Reforecasts: what are they good for?

Using multi-decadal reforecasts from the NCEP Global Ensemble Forecast System.

Tom Hamill

NOAA ESRL, Physical Sciences Division

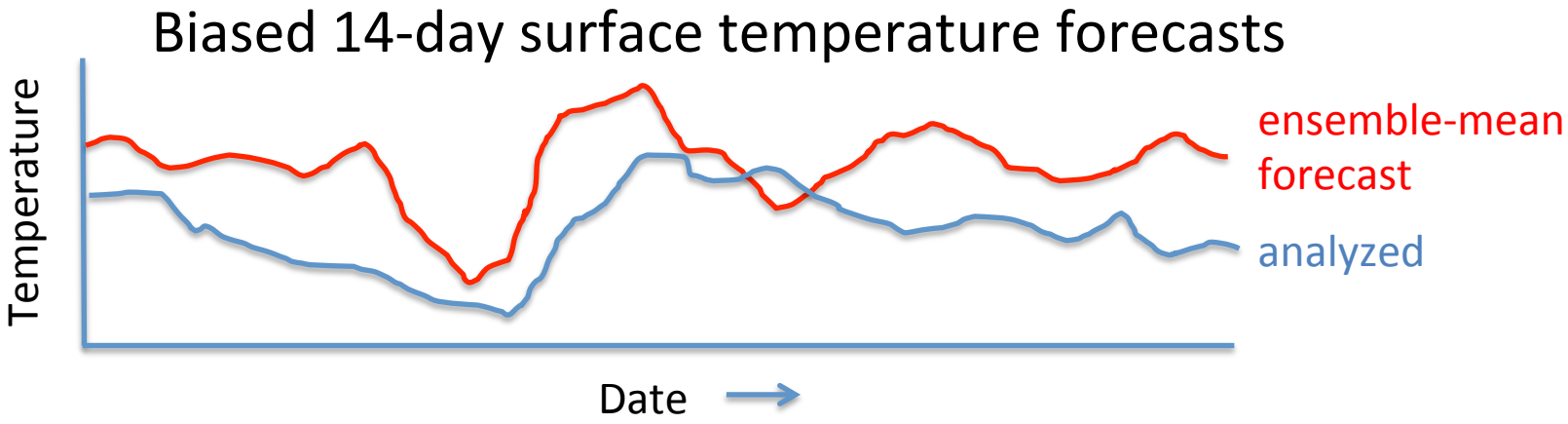
tom.hamill@noaa.gov

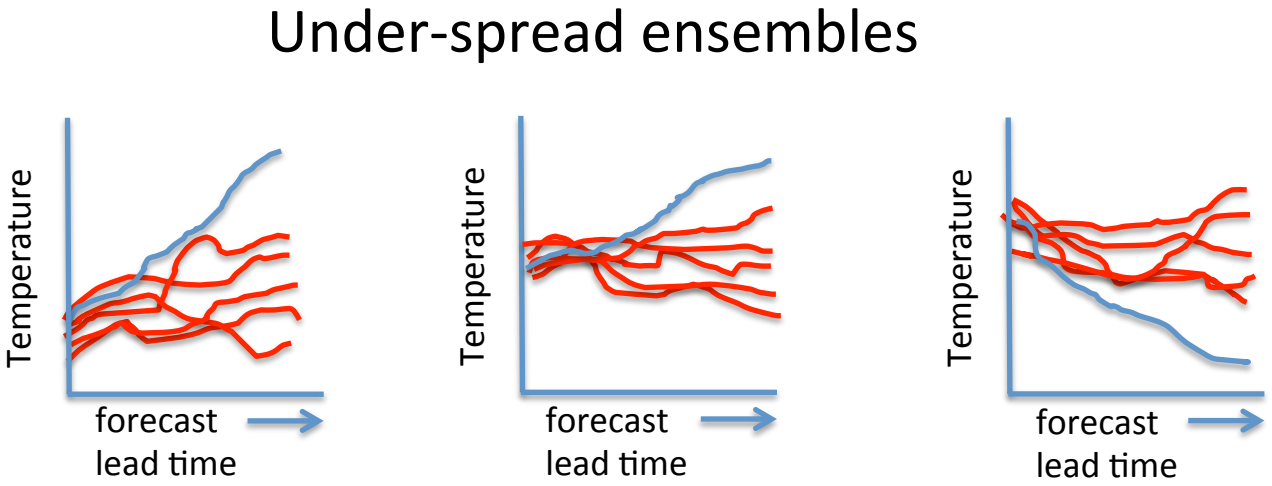
also: Jeff Whitaker, Gary Bates, Don Murray, Francisco Alvarez, Mike Fiorino, Tom Galarneau

What's to be covered here

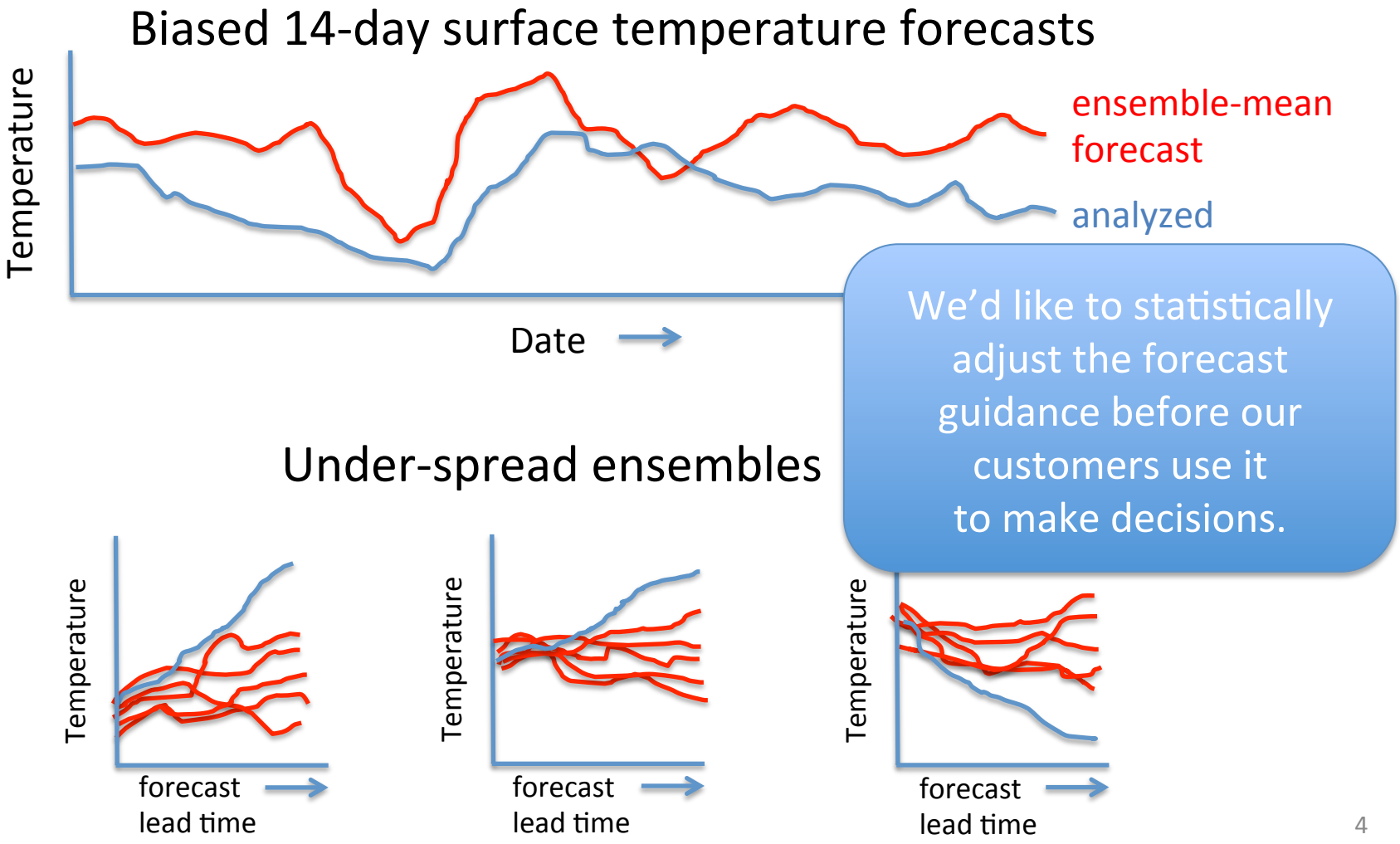
- A quick descriptions of the why / what of this reforecast database.
- Skill of raw reforecasts
- Post-processing method and examples
- Using reforecasts to diagnose forecastability of MJO and blocking.
- How you can access the reforecast data for your own applications.

Ensemble simulations of the weather and climate commonly have systematic errors.

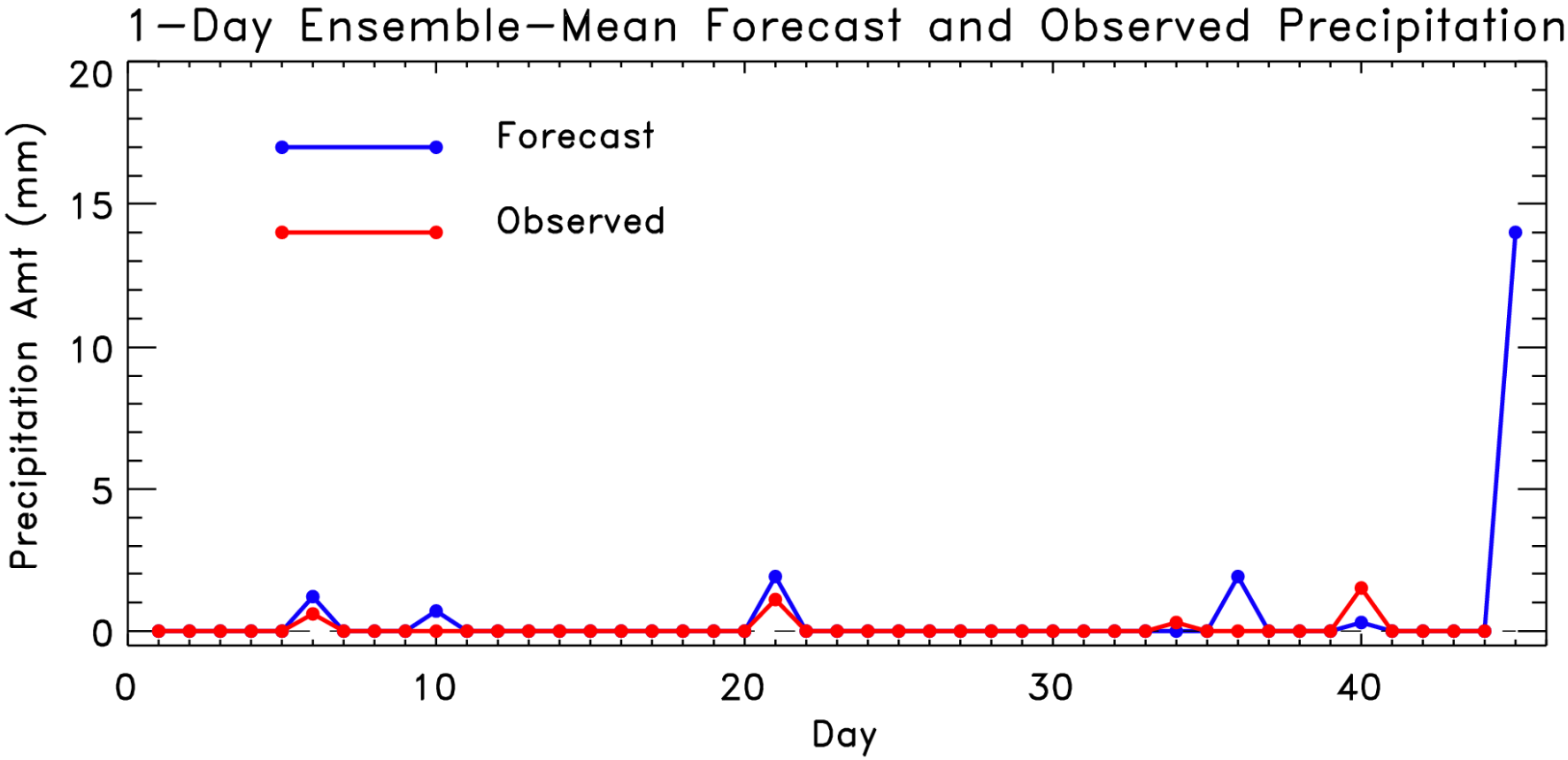




Ensemble simulations of the weather and climate commonly have systematic errors.

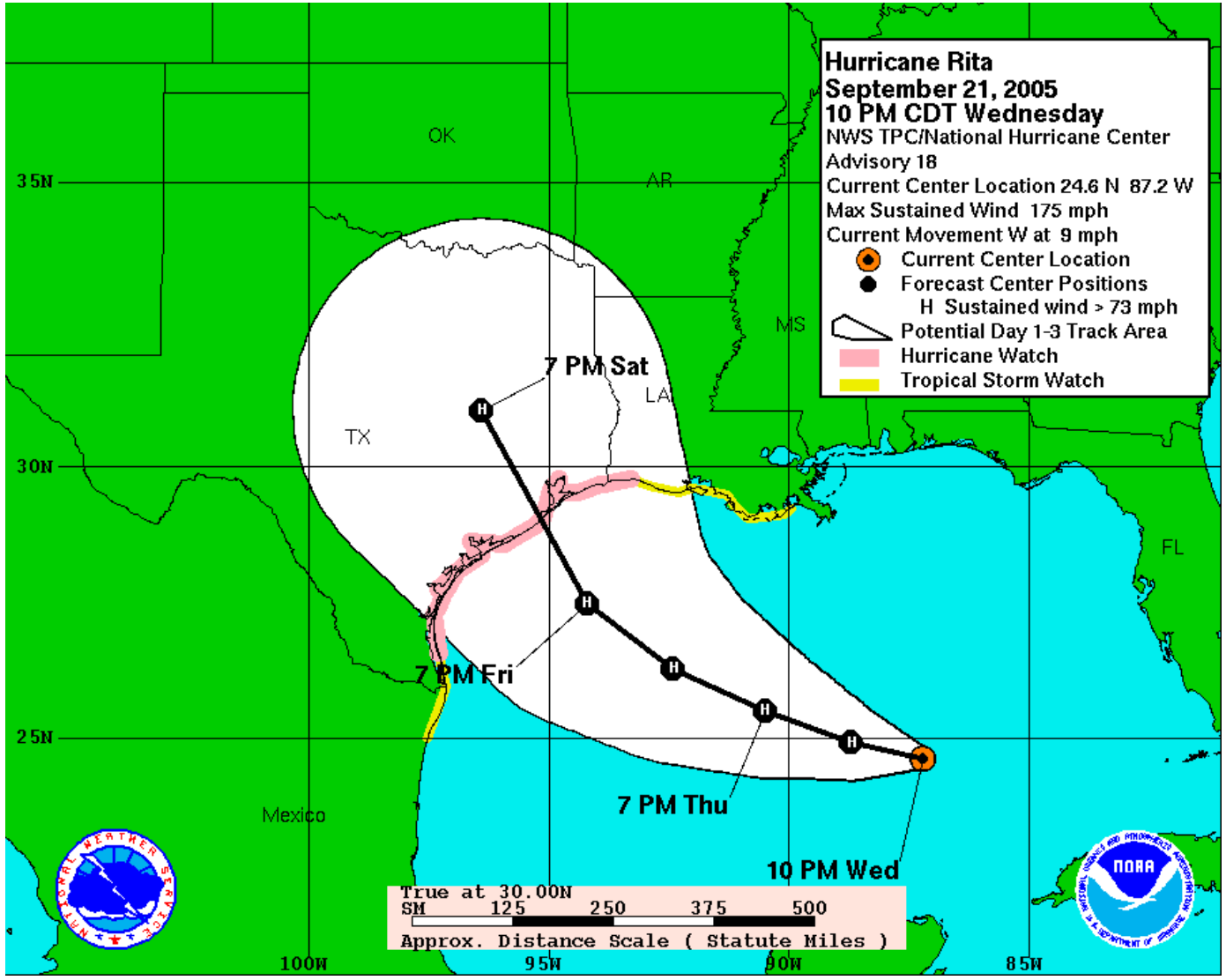


Statistical post-processing for rare events is challenging without a large training sample

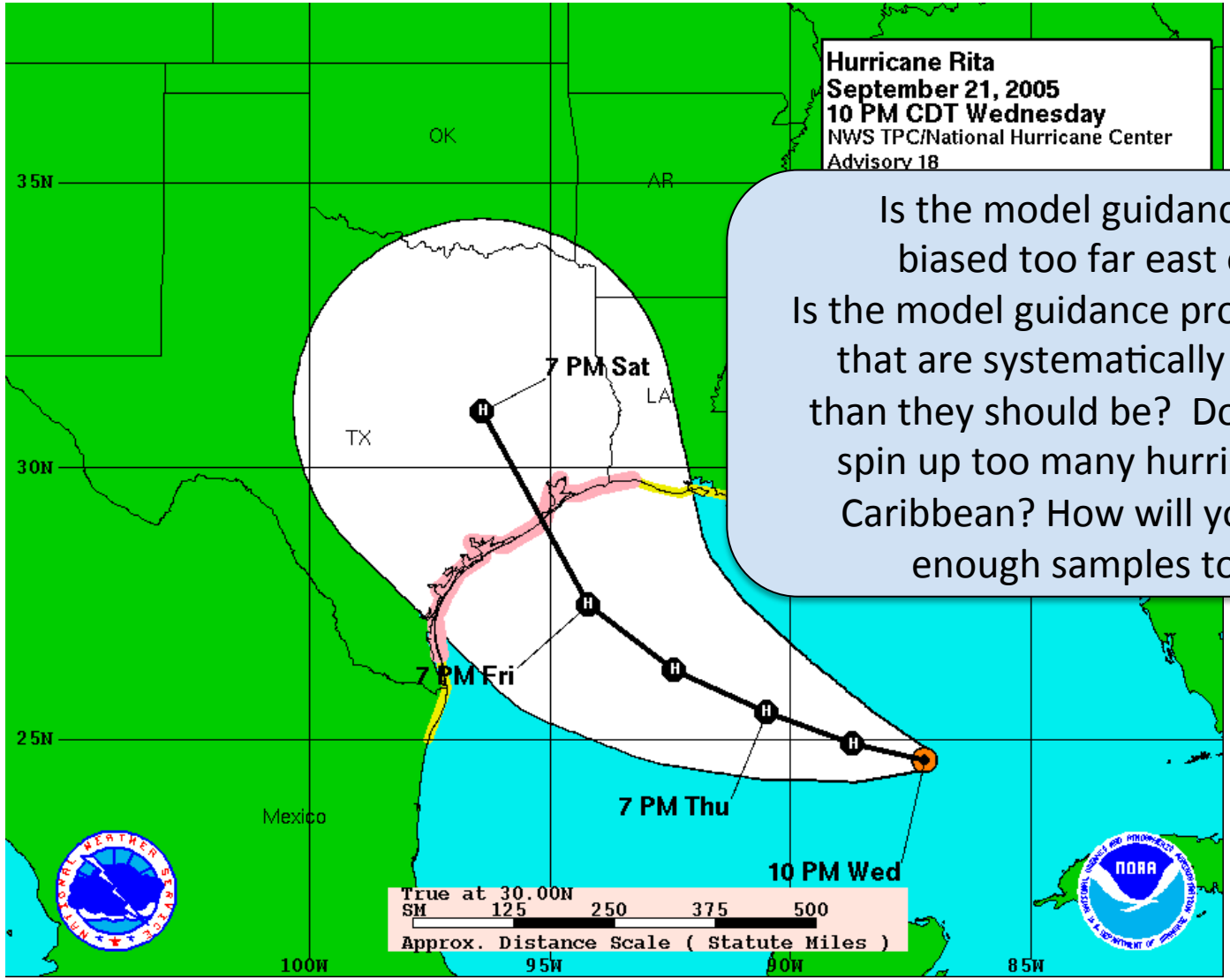


Say you want to statistically post-process your model precipitation forecast to improve it. Heavy precipitation events like the one today are the ones you care about the most. How do you calibrate today's forecast given past short sample of forecasts and observations?

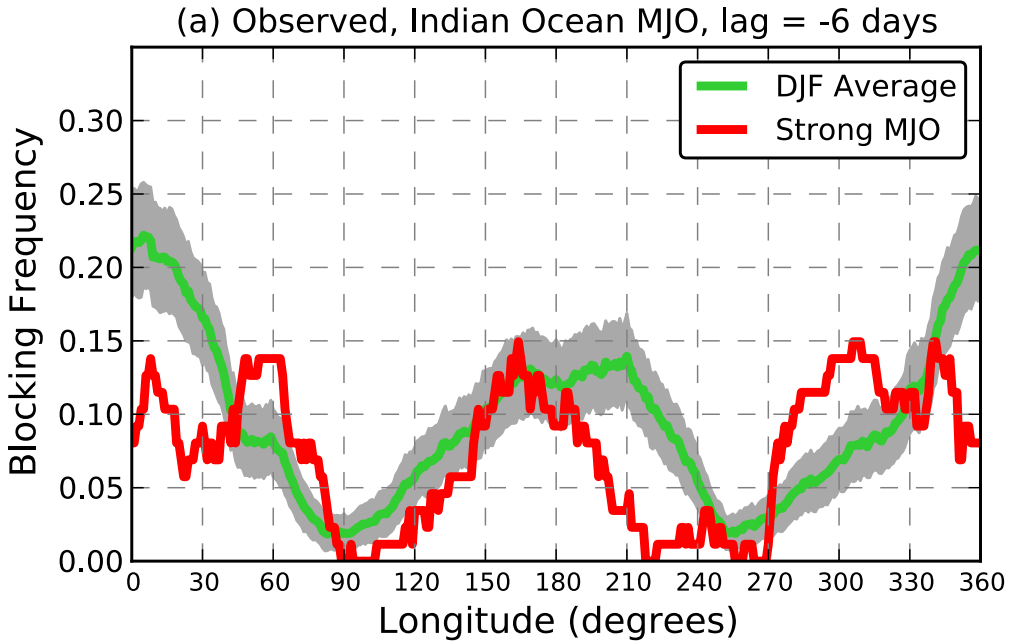
2005 Rita official forecast (Houston, TX evacuated)

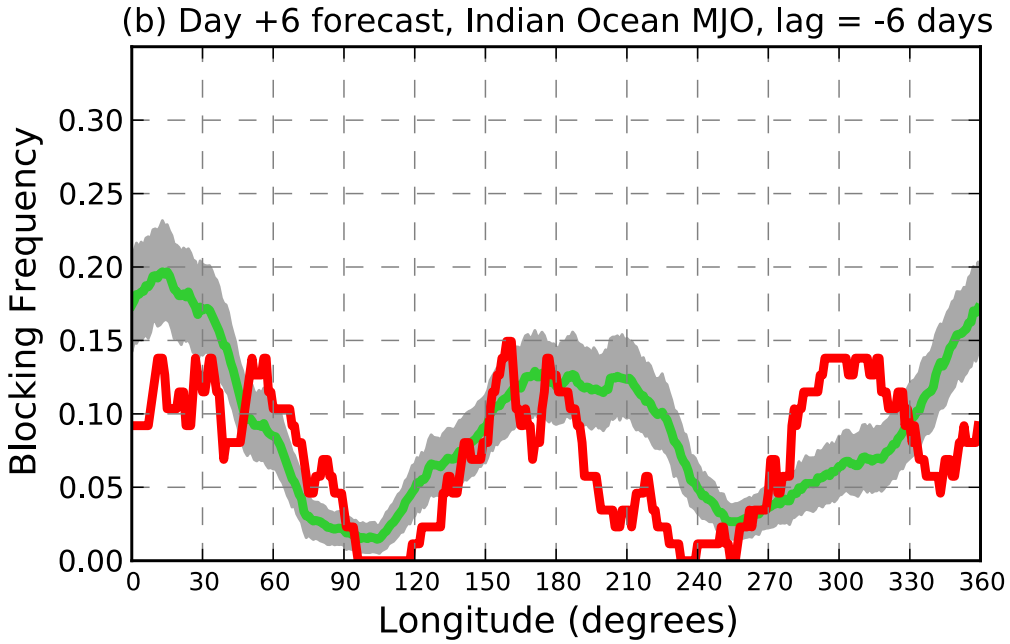


2005 Rita official forecast (Houston, TX evacuated)



Is the model guidance for Rita biased too far east or west?
Is the model guidance producing storms that are systematically less intense than they should be? Does the model spin up too many hurricanes in the Caribbean? How will you generate enough samples to know?





Need sufficient samples to understand whether the forecast model can predict low-frequency modes of variability.

The atmosphere suppresses blocking subsequent to an active Indian-Ocean Madden-Julian Oscillation (MJO) during the Northern Hemisphere winter. Does the forecast model suppress blocking as well? How can one detect that with only a season or so of past forecasts and with both blocking and strong MJOs happening infrequently?

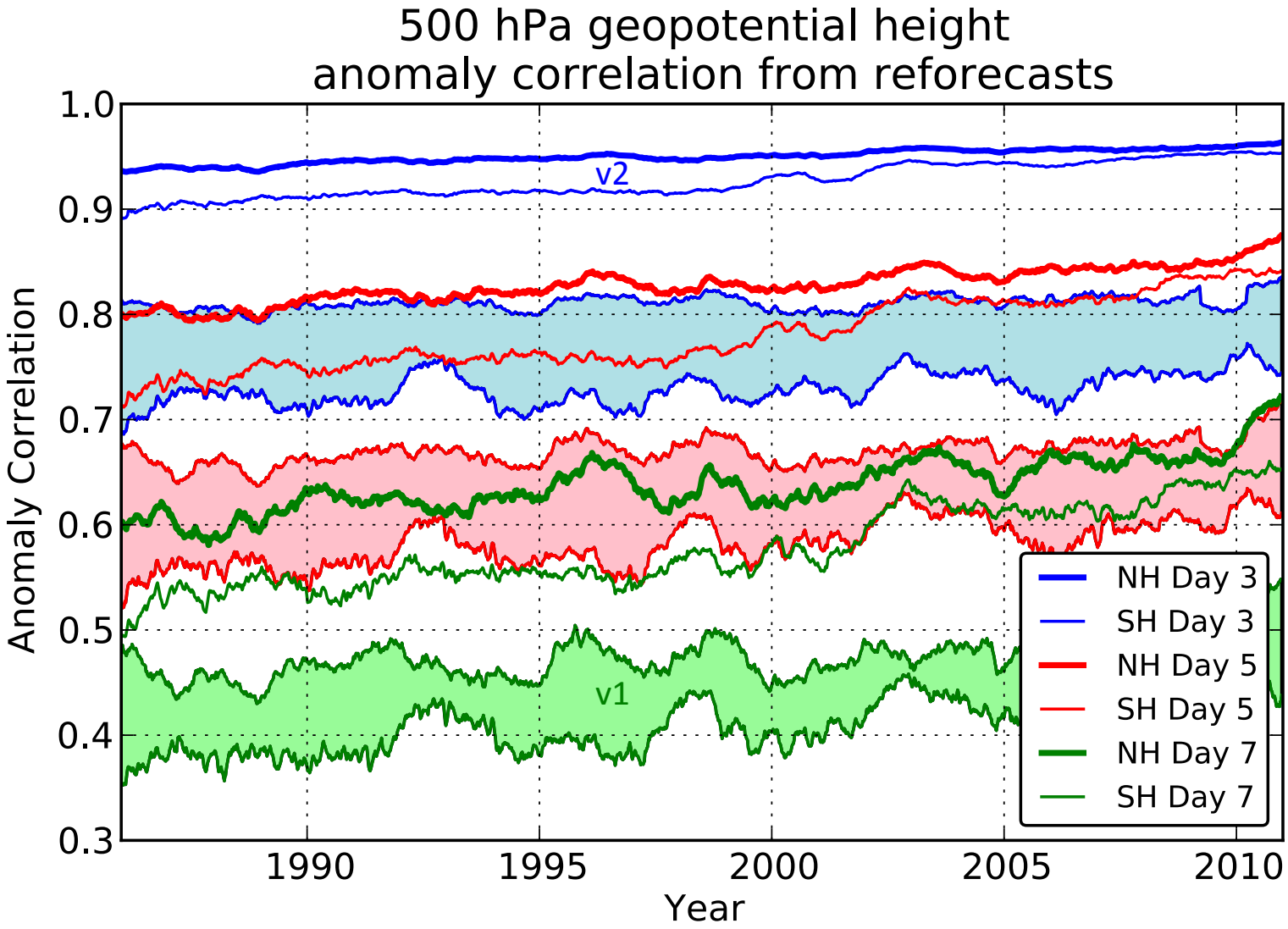
Reforecasts (hindcasts)

- Numerical simulations of the past weather (or climate) using the same forecast model and assimilation system that (ideally) is used operationally.
  - Common with climate, uncommon with weather models.

GEFS reforecast v2 details

- Seeks to mimic GEFS (NCEP Global Ensemble Forecast System) operational configuration as of February 2012.
- Once daily at 00 UTC, we produce an 11-member forecast, 1 control + 10 perturbed.
- These reforecasts were produced every day, for 1984120100 to current.
- CFSR (NCEP's Climate Forecast System Reanalysis) initial conditions (3D-Var) + ETR perturbations (cycled with 10 perturbed members). After ~ 22 May 2012, initial conditions from hybrid EnKF/3D-Var.
- Resolution: T254L42 to day 8 (~47-km grid spacing), T190L42 (~ 70 km) from days 7.5 to day 16.
- Fast data archive at ESRL of 99 variables, 28 of which stored at original ~1/2-degree resolution during week 1. All stored at 1 degree. Also: mean and spread to be stored.
- Full archive at DOE/Lawrence Berkeley Lab, where data set was created under DOE grant.

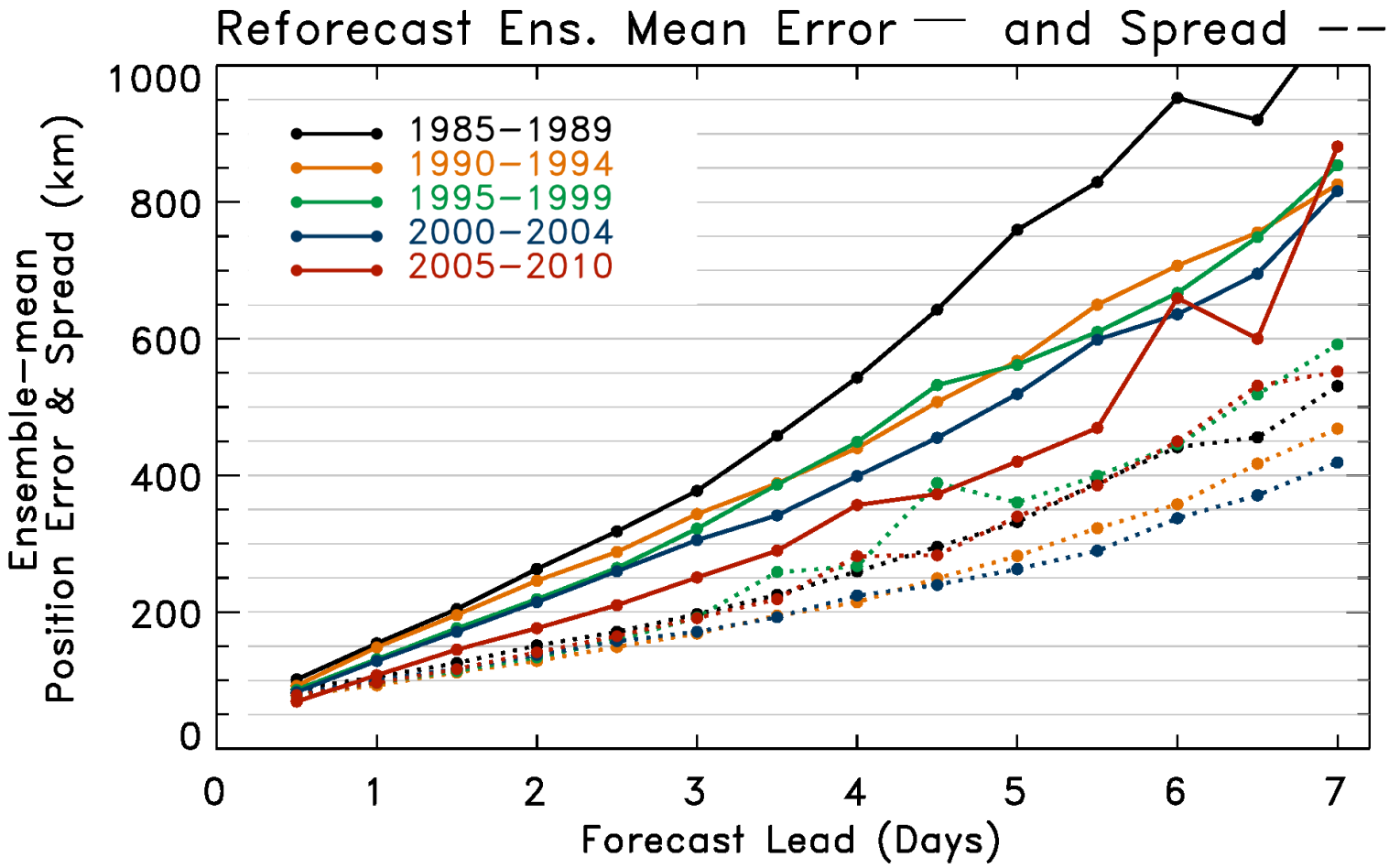
A few characteristics of the raw reforecasts



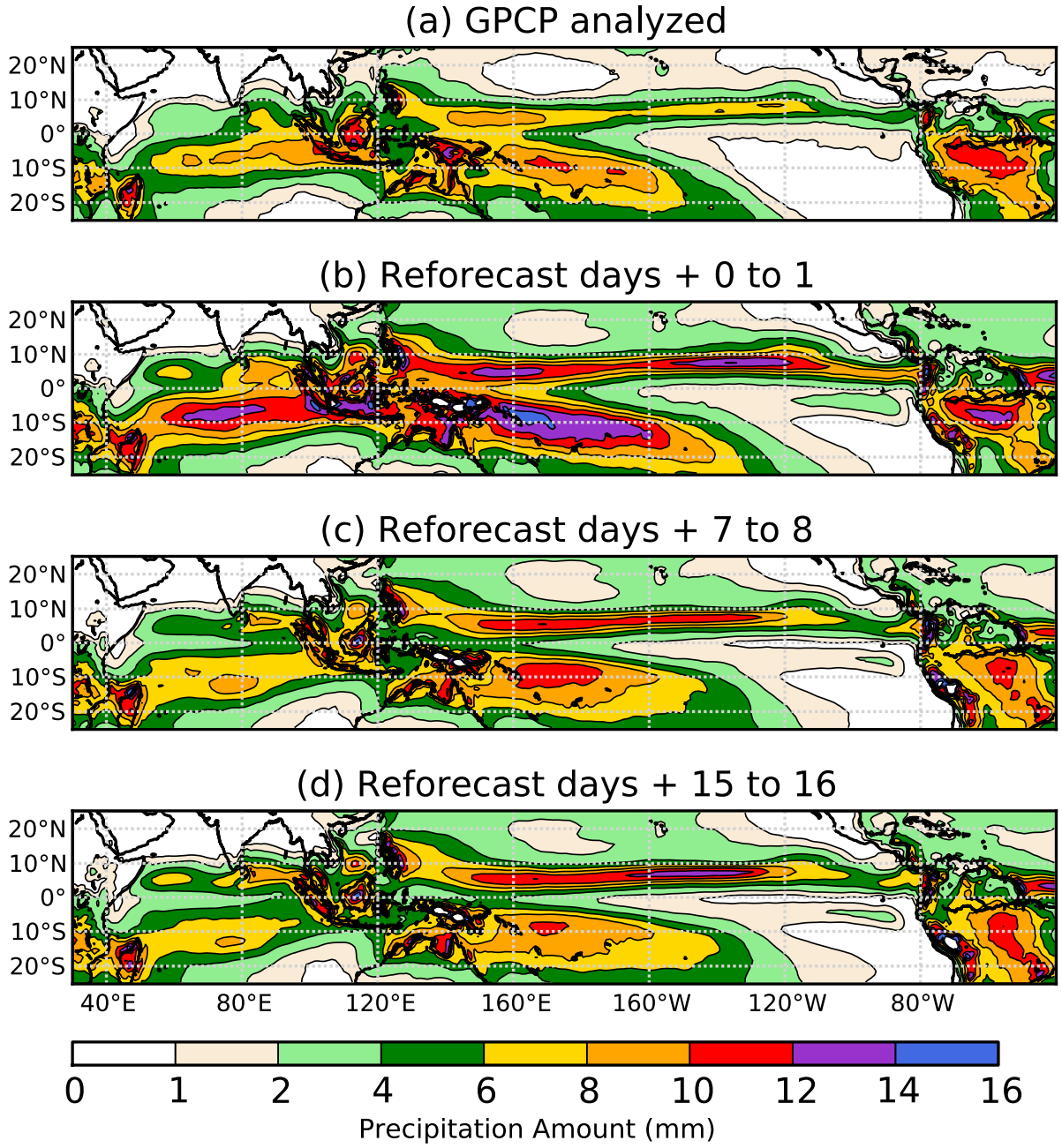
(1) CFSR initial conditions used in GEFS generally improve over the decades, leading to slight improvements in GEFS skill.

(2) About a +2 day improvement relative to 1998 GEFS T62 reforecasts.

Tropical cyclone track error and spread



A change in skill over time more evident with TCs, and in general with the tropical atmosphere relative to the extratropics.



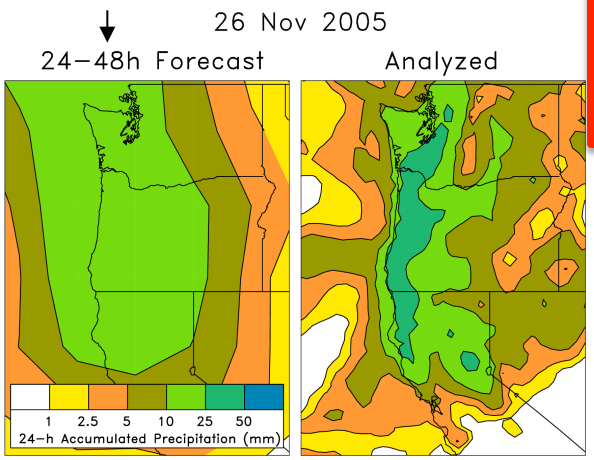
Daily rainfall climatology

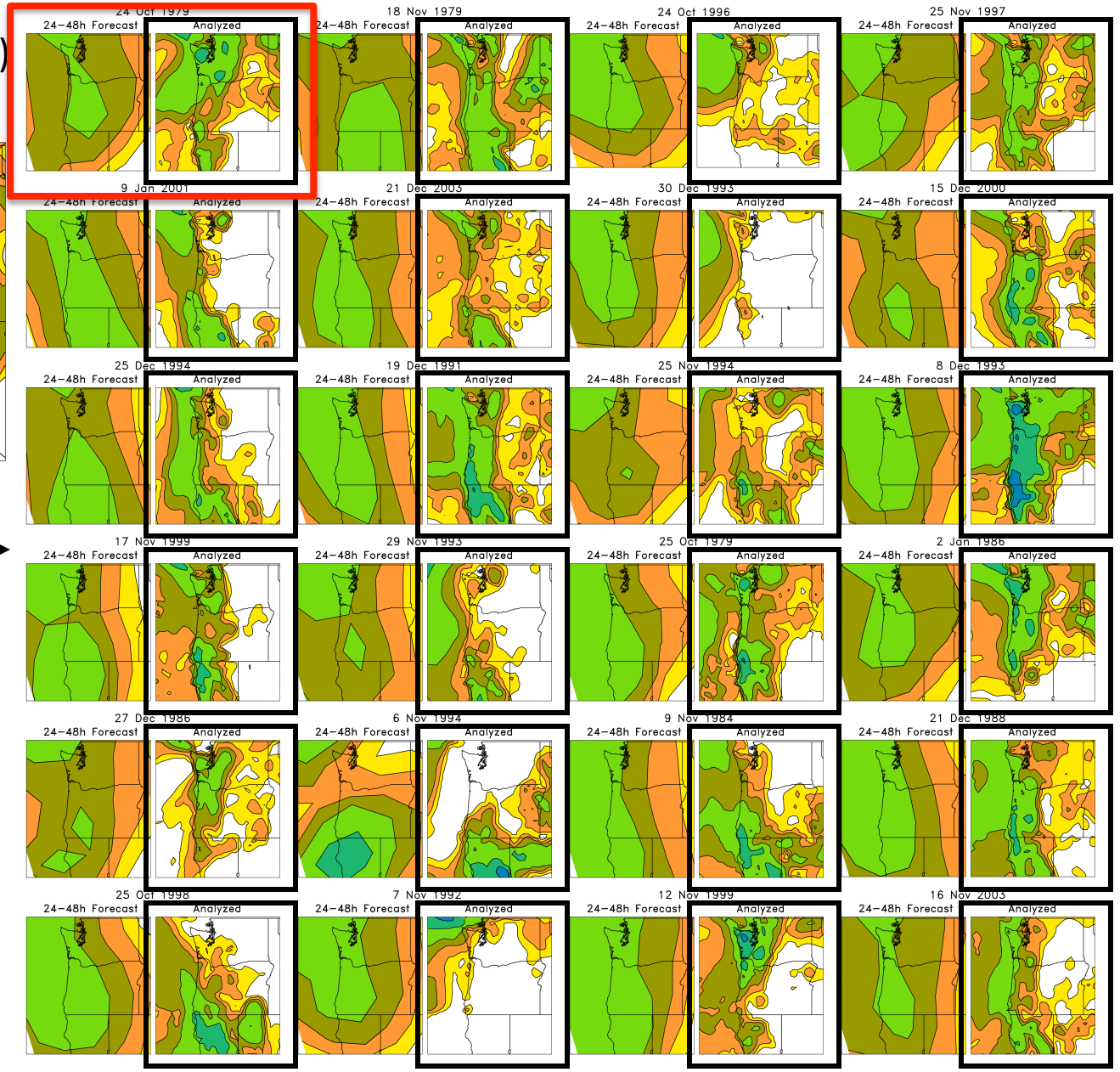
Large over-forecasting of precipitation on first day of forecast. Patterns of forecast precipitation at extended leads somewhat off, e.g., SPCZ not connected as well to ITCZ.

Application:
statistical post-processing
using reforecasts

An example of a statistical correction technique using those reforecasts

Today's forecast (& observed)





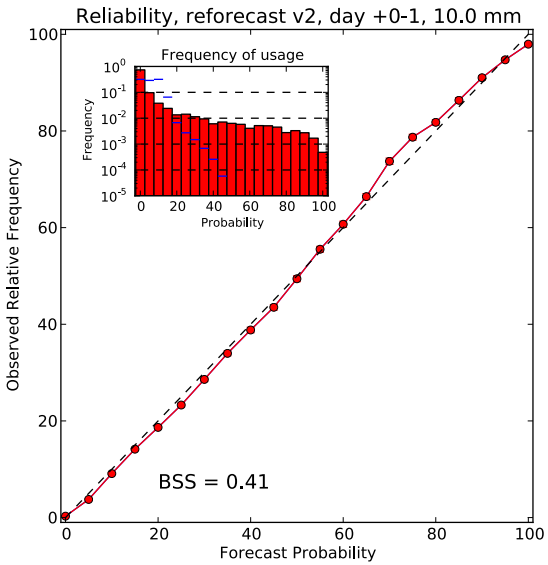
For each pair (e.g. red box), on the left are old forecasts that are somewhat similar to this day's ensemble-mean forecast. The boxed data on the right, the analyzed precipitation for the same dates as the chosen analog forecasts, can be used to statistically adjust and downscale the forecast.

Analog approaches like this may be particularly useful for hydrologic ensemble applications, where an ensemble of weather realizations is needed as inputs to a hydrologic ensemble streamflow system.

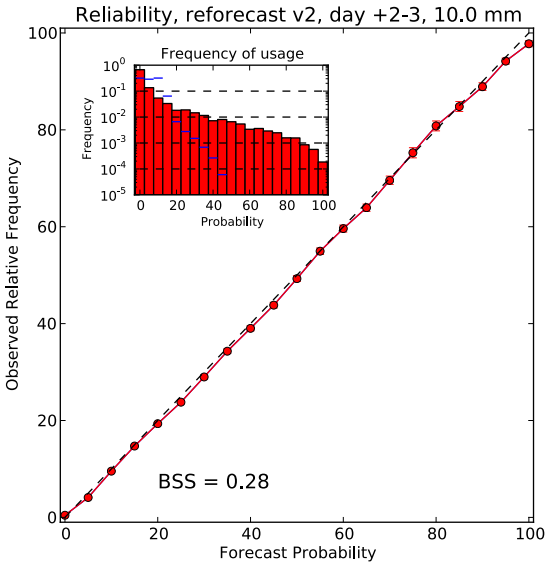
Reliability, > 10 mm precipitation 24 h^{-1}

Version 2 (2012 GEFS)

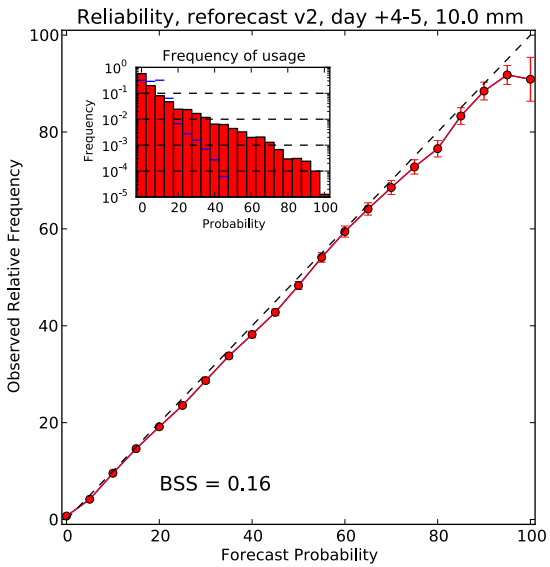
Day +0-1



Day +2-3



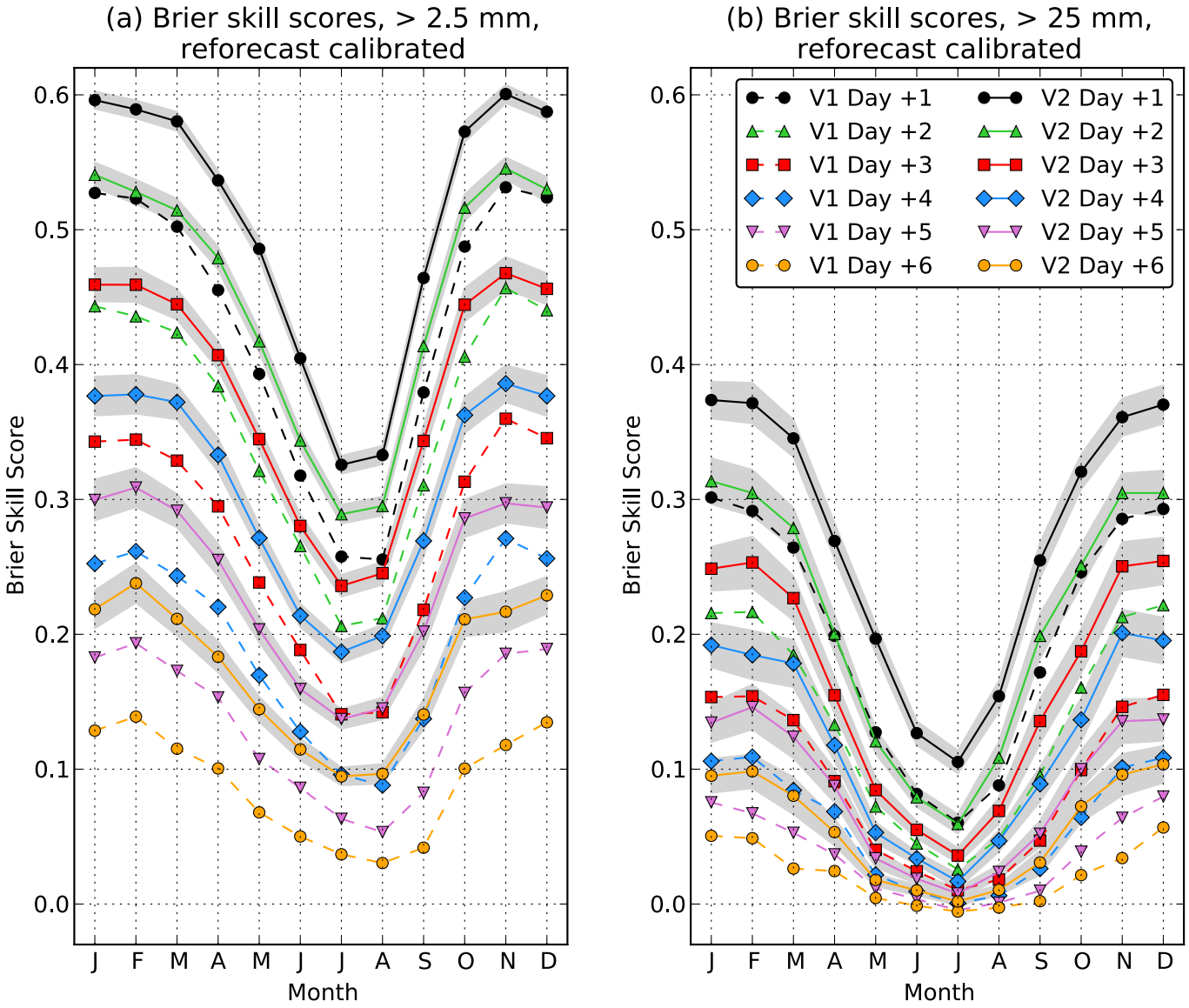
Day +4-5



Almost perfect reliability and improved skill is possible through analog-based calibration.

I'd note that this "reforecast" ≠ "analog." You can apply whatever post-processing method you choose to design, and I suspect many could do even better.

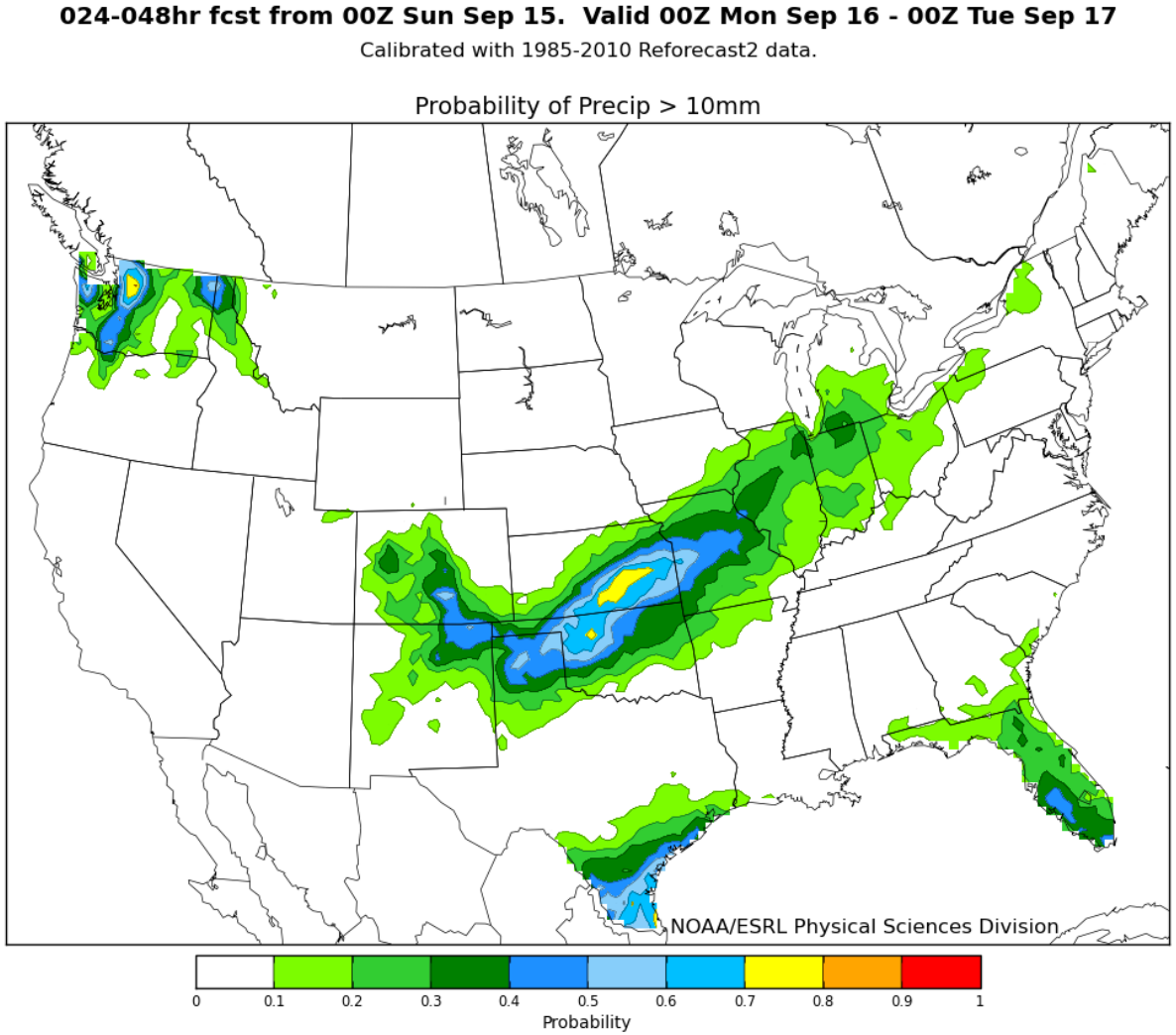
Precipitation reforecast skill



Reforecast skill of post-processed precipitation forecasts compared between version 1 (1998 T62 model) and version 2 (2012 T254 model).

Grey bars are 5/95 % confidence intervals.

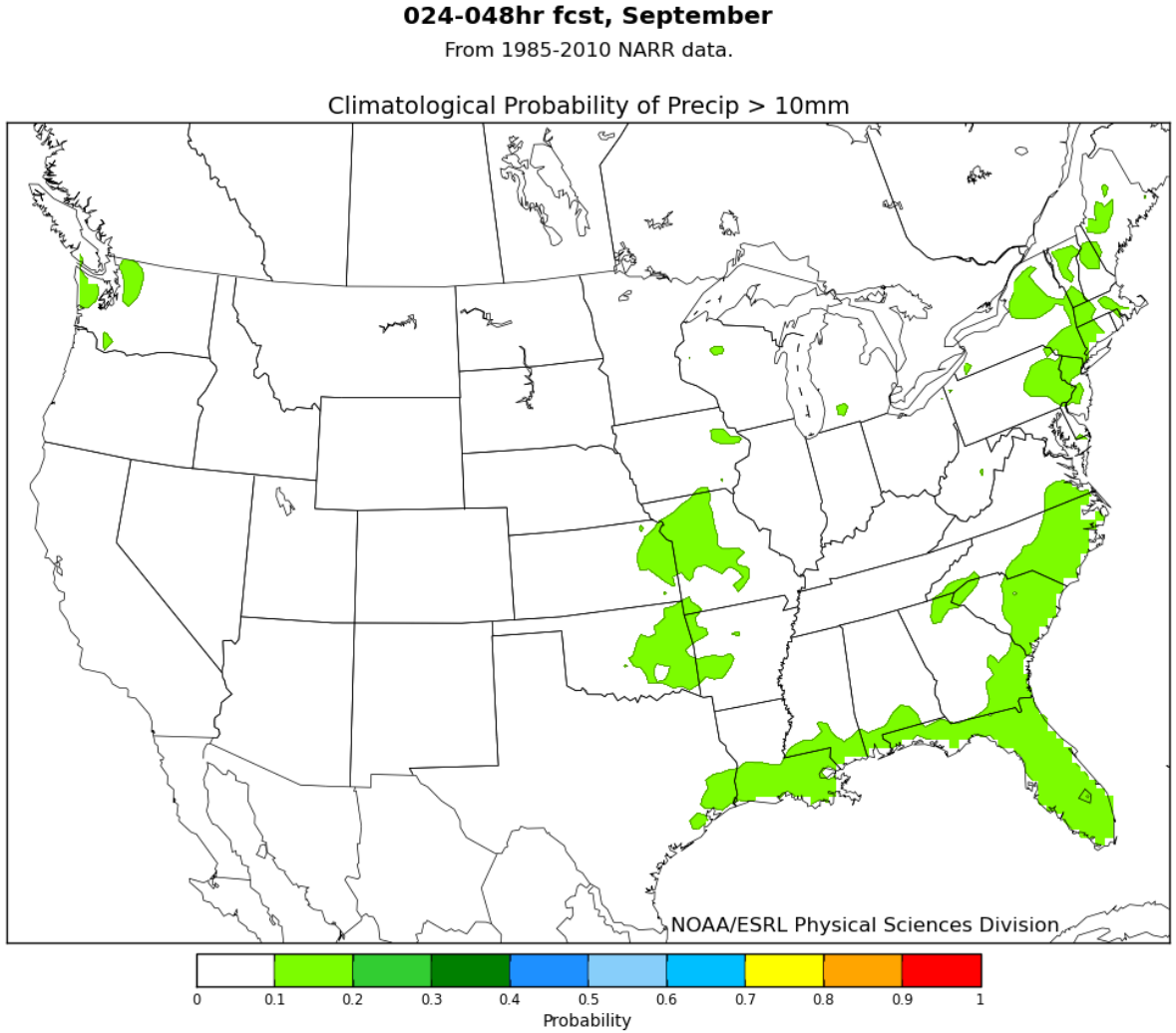
Example: recent forecast graphics



These forecasts were calibrated using 1985-2010 NARR 32-km precip. analyses and reforecasts.

http://www.esrl.noaa.gov/psd/forecasts/reforecast2/analogs/index.html

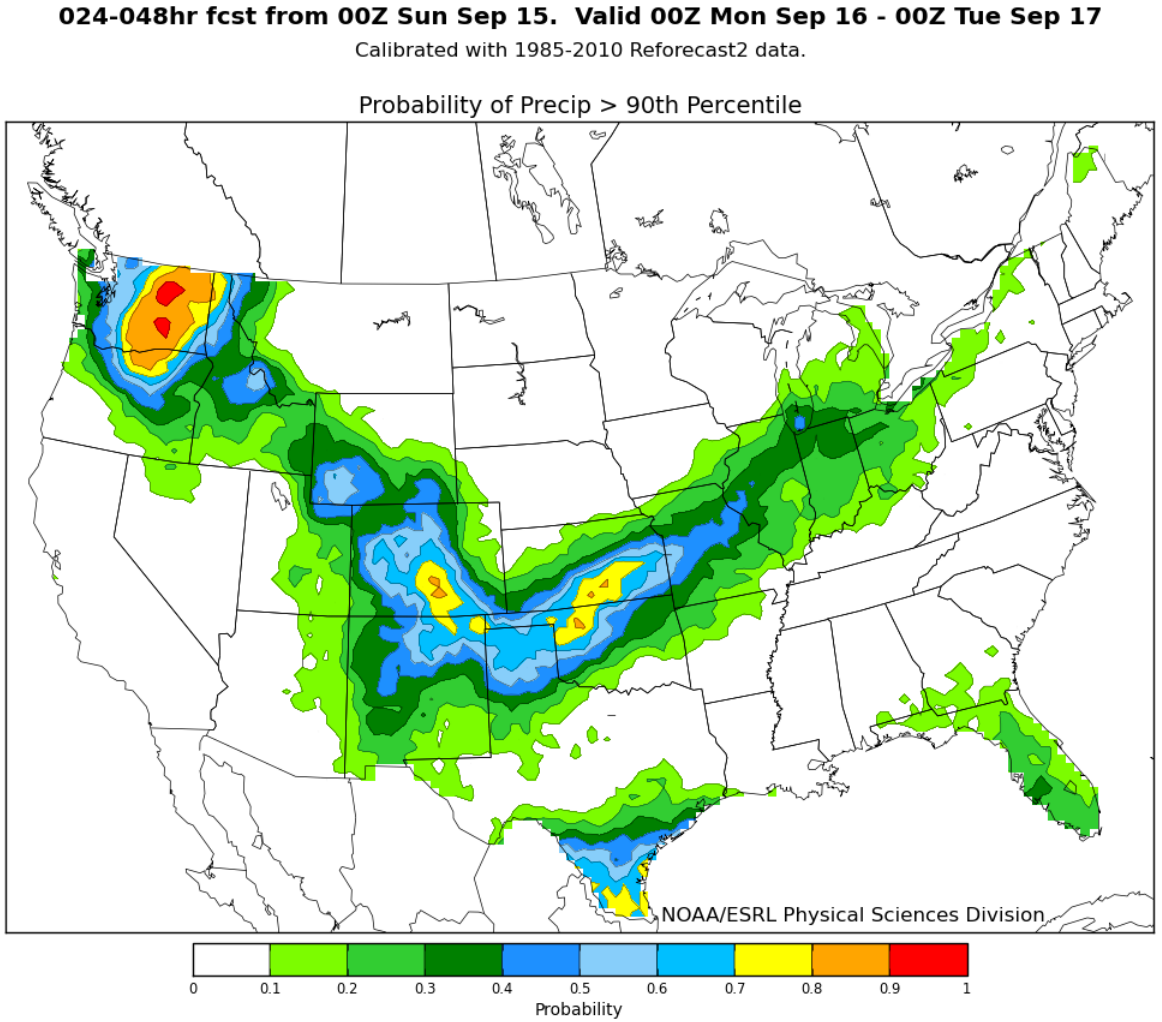
Example: recent forecast graphics



For comparison, on our web site we also provide climatological probabilities.

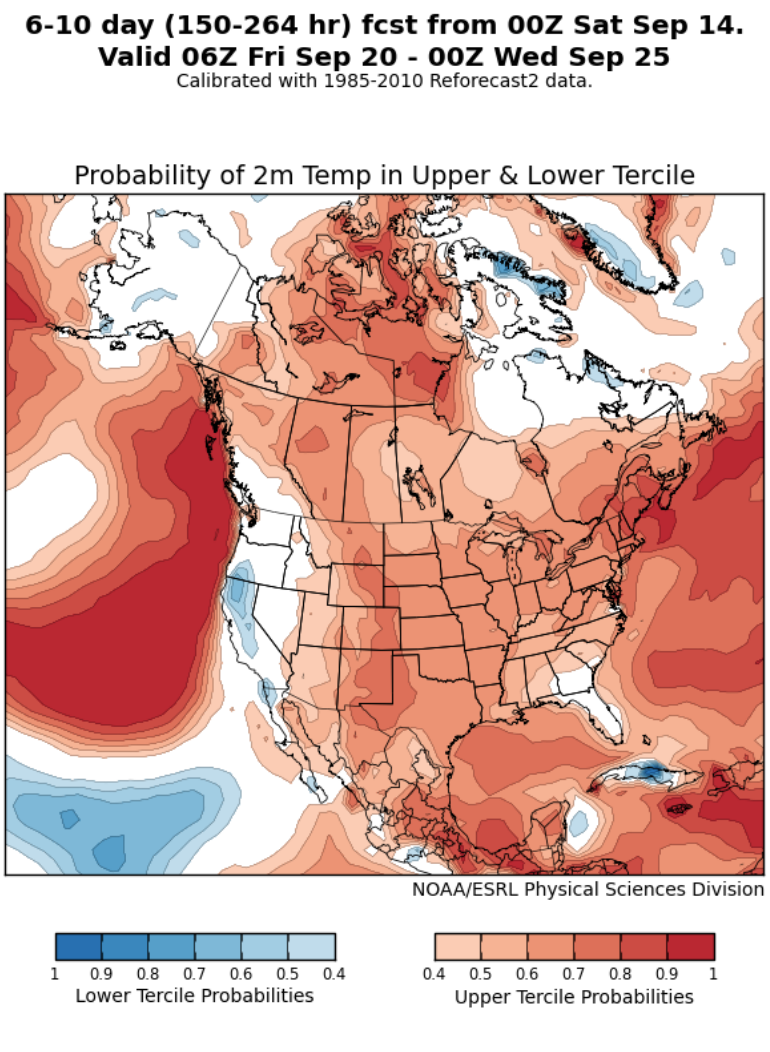
http://www.esrl.noaa.gov/psd/forecasts/reforecast2/analogs/index.html

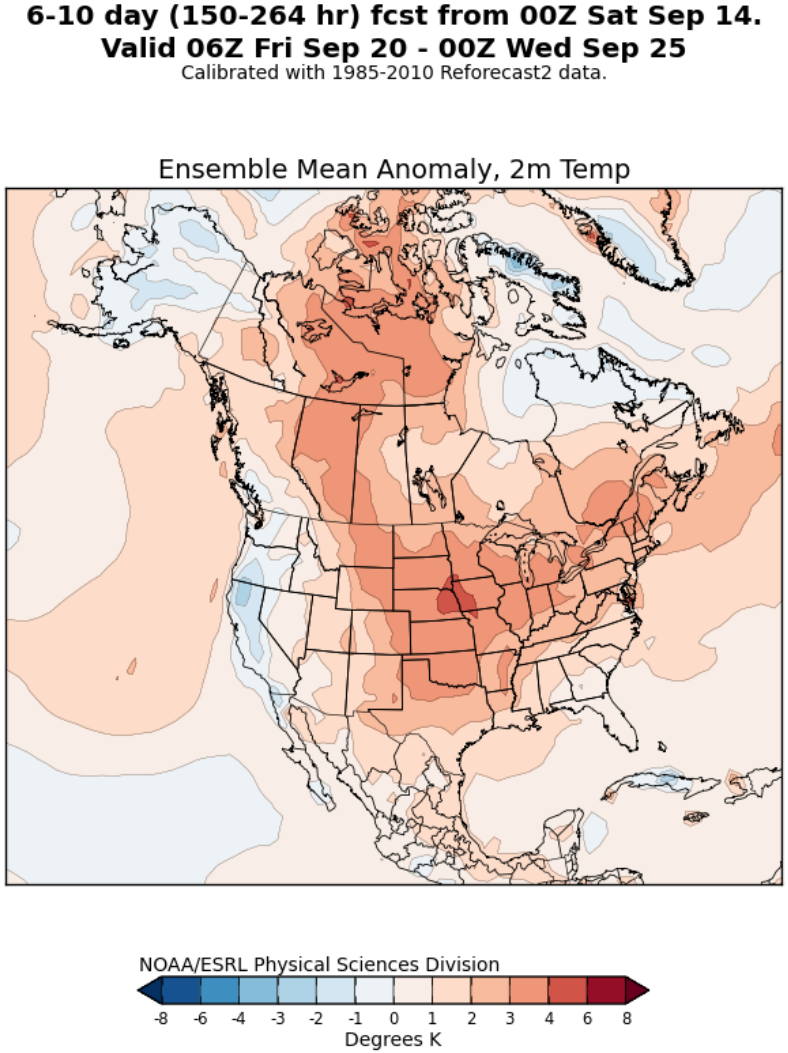
Example: recent forecast graphics



Here are probabilities of exceeding the 90th percentile of the climatological distribution, a way of normalizing for how dry each location is.

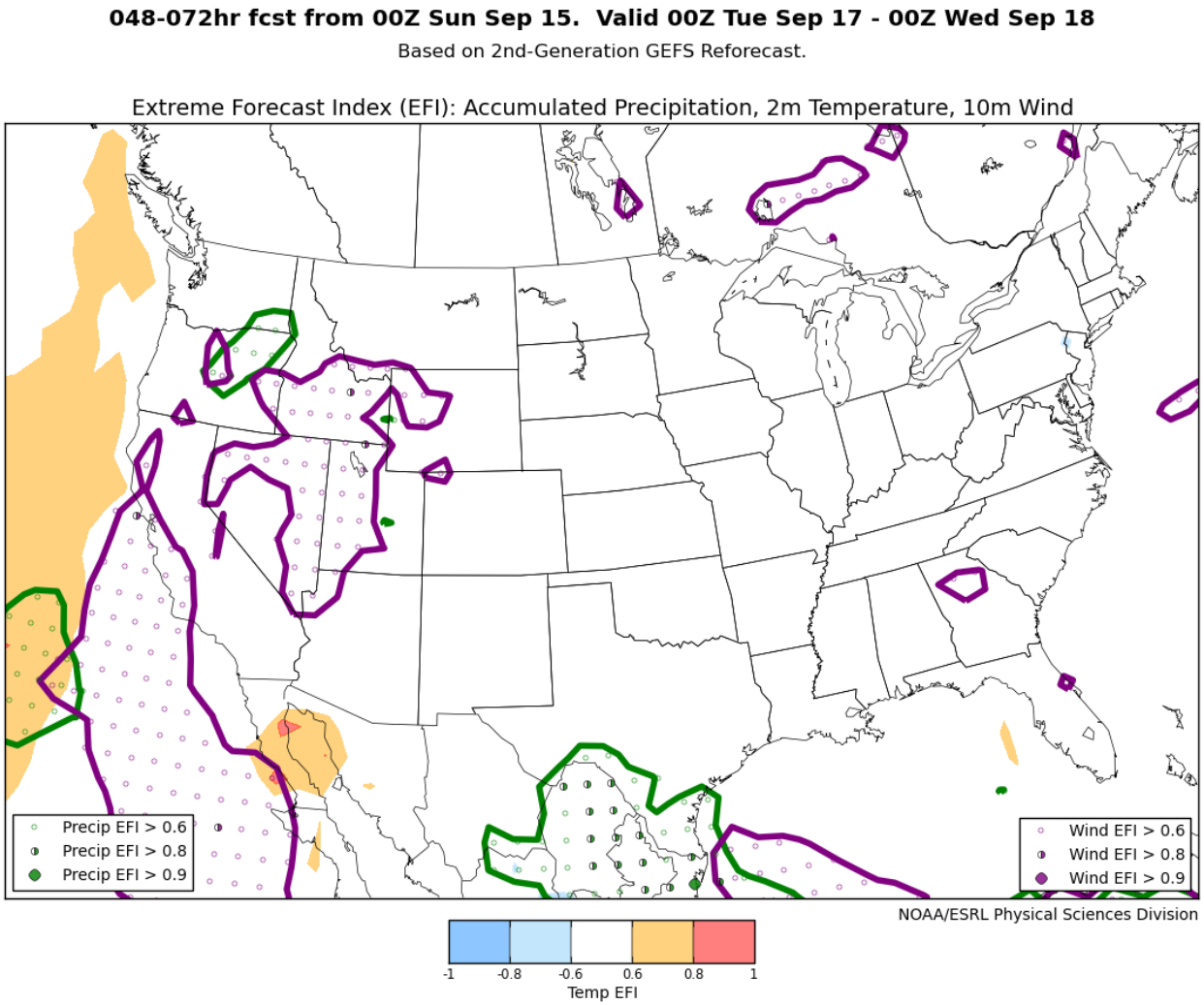
Example: post-processed 2-meter temperature forecasts





http://www.esrl.noaa.gov/psd/forecasts/reforecast2/medrange/index.html

Extreme forecast index

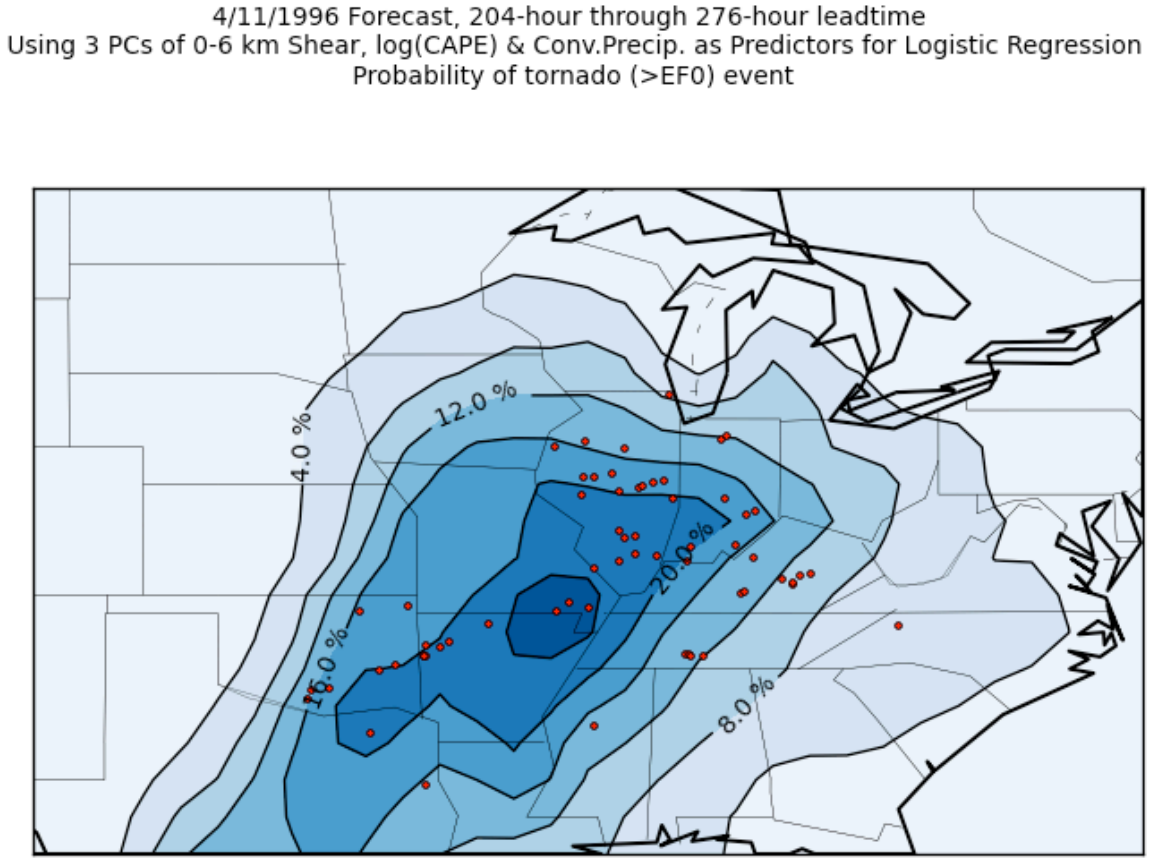


In the absence of reliable verification data for training a post-processing method, one can at least compare today's forecast and see how unusual it is relative to the model climatology.

This EFI plot is calculated following ECMWF's methodology.

http://www.esrl.noaa.gov/psd/forecasts/reforecast2/analogs/index.html

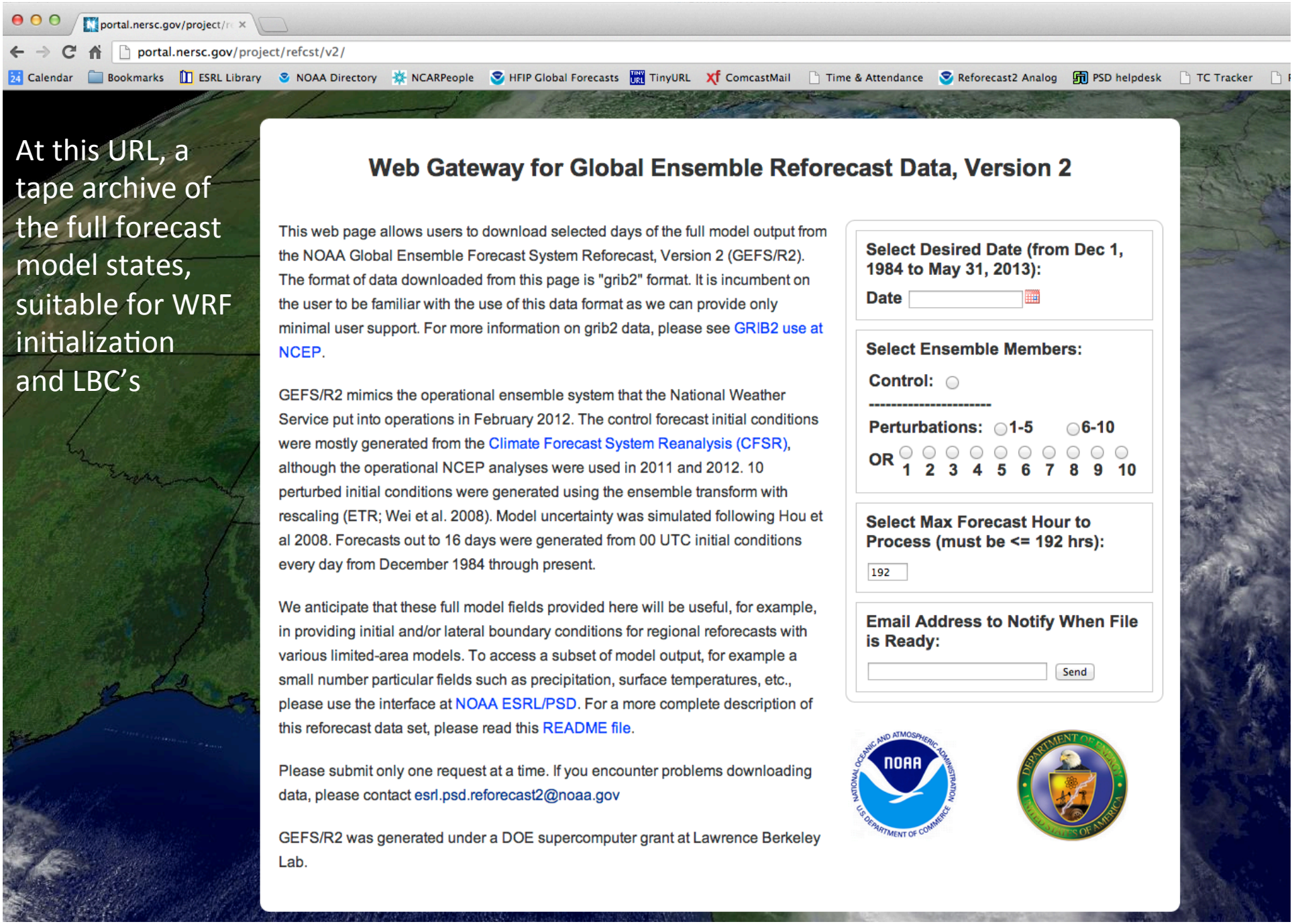
Application: extended-range tornado forecasting



Francisco Alvarez at St. Louis University, is working with me and others on using the reforecasts to make extended-range predictions of tornado probabilities.

Ph.D. work, in progress.

Application: providing initial and lateral boundary conditions to permit WRF retrospective forecasts


portal.nersc.gov/project/
portal.nersc.gov/project/refcst/v2/
Calendar
Bookmarks
ESRL Library
NOAA Directory
NCARPeople
HFIP Global Forecasts
TinyURL
ComcastMail
Time & Attendance
Reforecast2 Analog
PSD helpdesk
TC Tracker
At this URL, a tape archive of the full forecast model states, suitable for WRF initialization and LBC's
Web Gateway for Global Ensemble Reforecast Data, Version 2
This web page allows users to download selected days of the full model output from the NOAA Global Ensemble Forecast System Reforecast, Version 2 (GEFS/R2). The format of data downloaded from this page is "grib2" format. It is incumbent on the user to be familiar with the use of this data format as we can provide only minimal user support. For more information on grib2 data, please see GRIB2 use at NCEP.
GEFS/R2 mimics the operational ensemble system that the National Weather Service put into operations in February 2012. The control forecast initial conditions were mostly generated from the Climate Forecast System Reanalysis (CFSR), although the operational NCEP analyses were used in 2011 and 2012. 10 perturbed initial conditions were generated using the ensemble transform with rescaling (ETR; Wei et al. 2008). Model uncertainty was simulated following Hou et al 2008. Forecasts out to 16 days were generated from 00 UTC initial conditions every day from December 1984 through present.
We anticipate that these full model fields provided here will be useful, for example, in providing initial and/or lateral boundary conditions for regional reforecasts with various limited-area models. To access a subset of model output, for example a small number particular fields such as precipitation, surface temperatures, etc., please use the interface at NOAA ESRL/PSD. For a more complete description of this reforecast data set, please read this README file.
Please submit only one request at a time. If you encounter problems downloading data, please contact esrl.psd.reforecast2@noaa.gov
GEFS/R2 was generated under a DOE supercomputer grant at Lawrence Berkeley Lab.
Select Desired Date (from Dec 1, 1984 to May 31, 2013):
Date
Select Ensemble Members:
Control:
Perturbations: 1-5 6-10
OR 1 2 3 4 5 6 7 8 9 10
Select Max Forecast Hour to Process (must be <= 192 hrs):
192
Email Address to Notify When File is Ready:
Send
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
DEPARTMENT OF ENERGY

Some notes running WRF using data from this archive

(from http://www.esrl.noaa.gov/psd/forecasts/reforecast2/README.GEFS_Reforecast2.pdf)

- The proper Vtable file must be created prior to preprocessing the GEFS reforecast data. To do this, copy Vtable.GFS to your working directory for preprocessing the GRIB2 GEFS reforecast data. Rename the Vtable.GFS file as Vtable.reforecast, and modify the file as follows. First, add a line for specific humidity on pressure levels and at 2 m. The specifications for specific humidity should be as follows:

  metgrid Description: Specific Humidity
  metgrid units: kg kg-1
  metgrid Name: SPECHUMD
  GRIB2 Discp=0, Catgy=1, Param=0, Level=100
  GRIB1 Param=52, Level Type=100, From Level1=*
  metgrid Description: Specific Humidity at 2 m
  metgrid units: kg kg-1
  metgrid Name: SPECHUMD
  GRIB2 Discp=0, Catgy=1, Param=0, Level=103
  GRIB1 Param=52, Level Type=105, From Level1=2

- Second, remove the GRIB2 parameter number for relative humidity on pressure levels and at 2m. Note that ungrib.exe and metgrid.exe will calculate relative humidity for you if you have specific humidity. Finally, change the GRIB2 parameter number for PMSL from 1 to 0. No other known modifications to the Vtable are needed. Now that the Vtable.reforecast file is properly created, follow the instructions for running WRF-ARW on the WRF Users' Page [http://www.mmm.ucar.edu/wrf/users/].

- The files downloaded from DOE will have the fields for the different forecast lead times merged into one grib file. It will be your responsibility to break up that into separate grib files for each lead time.

- Multi-level soil moisture wasn't saved. README file indicates how to deal with this.

Demo: WRF ARW Regional Reforecast of Hurricane Rita (2005)

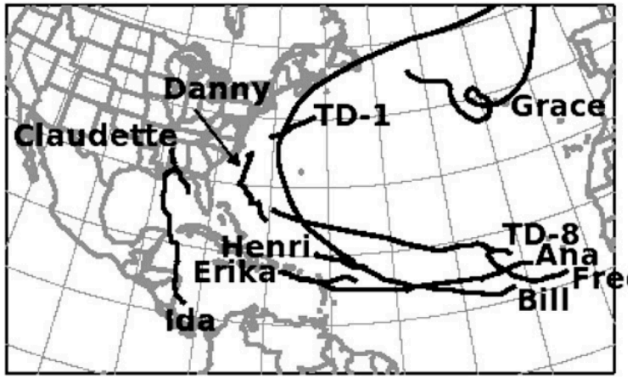
- An example of generating a WRF high-resolution regional reforecast ensemble using GEFS initial and LBCs
- Will show in this case that regional reforecast WRF ensemble provided value-added guidance to global ensemble forecast.

c/o Tom Galarneau, NCAR

Details of reforecast with WRF ARW v3.4 using GEFS for initial, boundary conditions

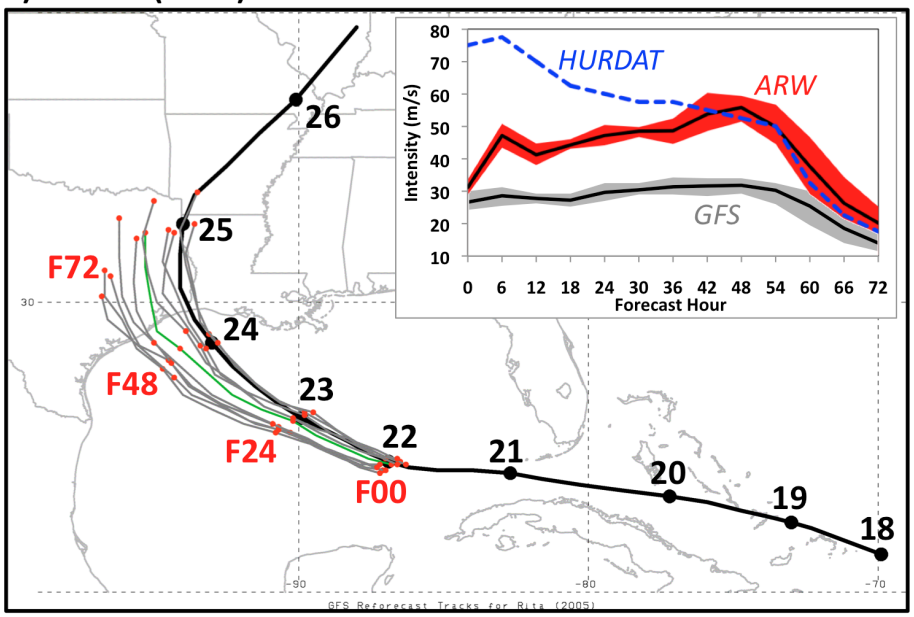
- Nested simulation 36-, 12- and 4-km with 36 vertical levels
  - 12- and 4-km moving nests
- Time steps: 180, 60, and 20 s
- Initial and boundary conditions from GFS reforecast ensemble members
- Tiedtke cumulus scheme on 36 and 12 km; explicit on 4 km
- YSU PBL scheme
- HYCOM ocean analysis
- WSM6 microphysics
- Noah land surface
- 2D Smagorinsky turbulence scheme
- Goddard shortwave radiation
- RRTM longwave radiation
- Second order diffusion
- Positive definite scalar advection
- Donelan wind-dependent drag formulation
- Garratt wind-dependent enthalpy surface fluxes

outer WRF domain

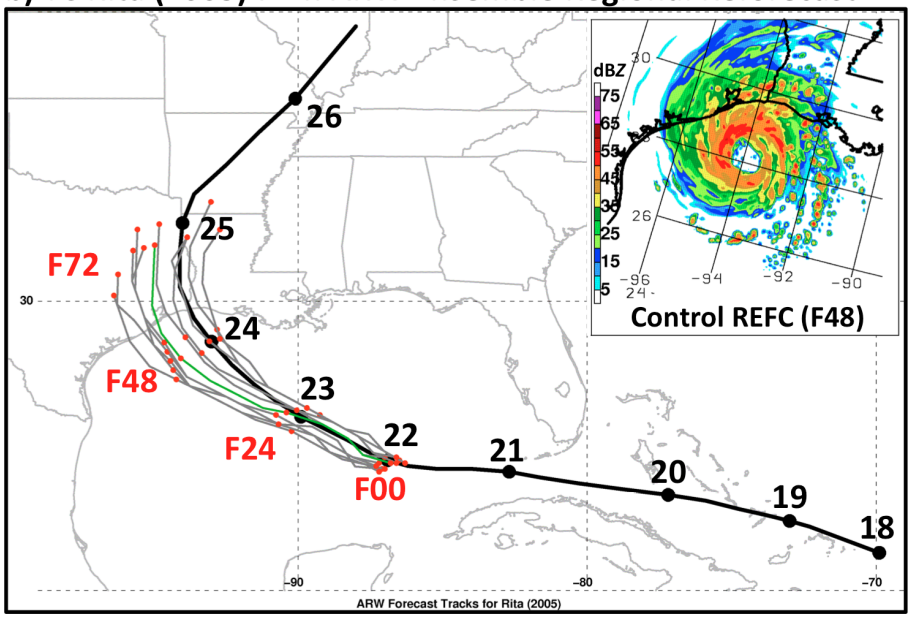


WRF-ARW reforecast ensemble results

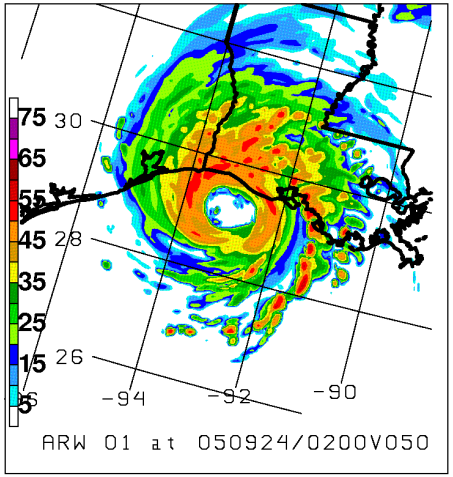
a) TC Rita (2005) 72-h GFS Ensemble Reforecast

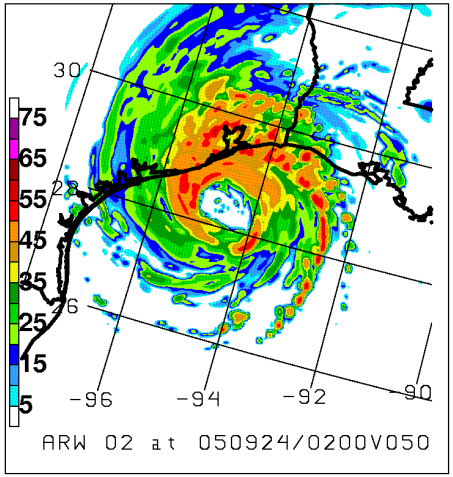


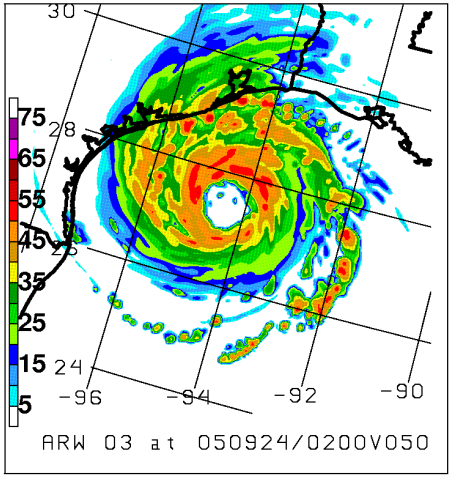
b) TC Rita (2005) 72-h ARW Ensemble Regional Reforecast

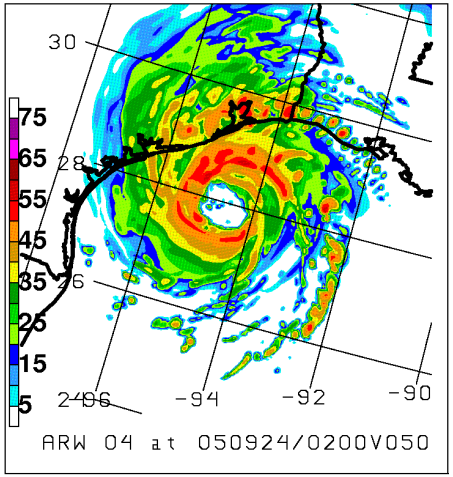


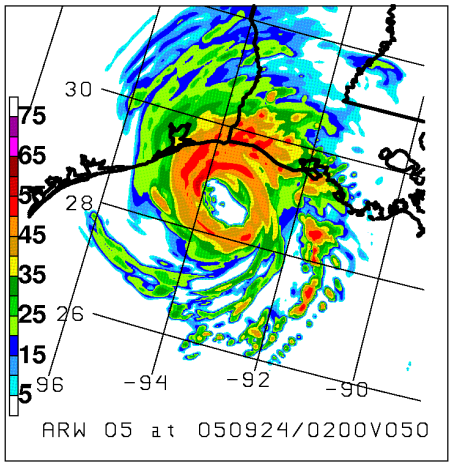
- Global reforecast ensemble is consistent with NHC forecast; indicating potential impact on Houston
- Significant left-of-track error and intensity was underestimated
- Rita vortex intensified in ARW regional reforecast despite terrible initial vortex
- Similar left-of-track error in ARW; suggests large-scale control on TC motion

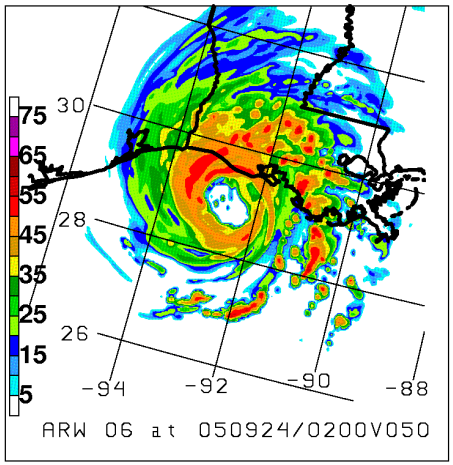


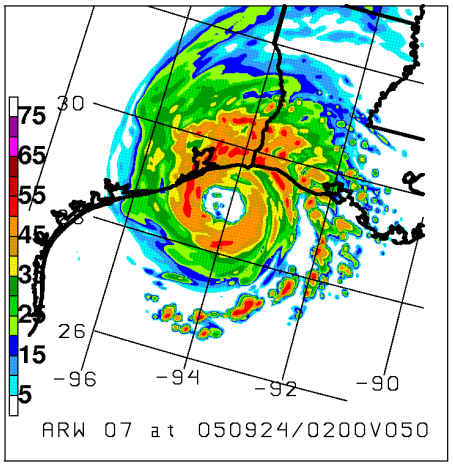


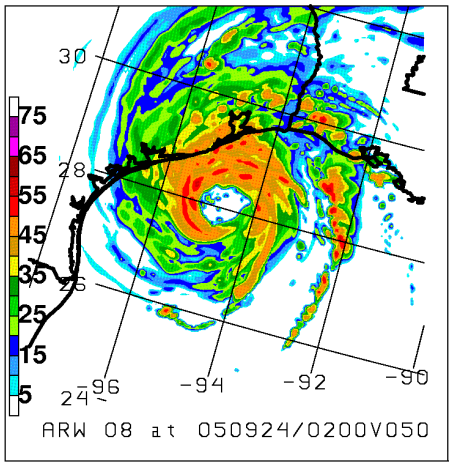


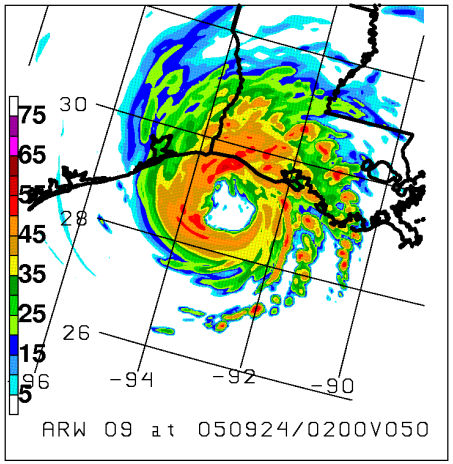


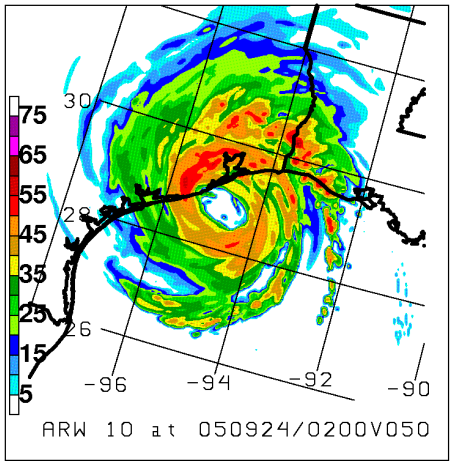


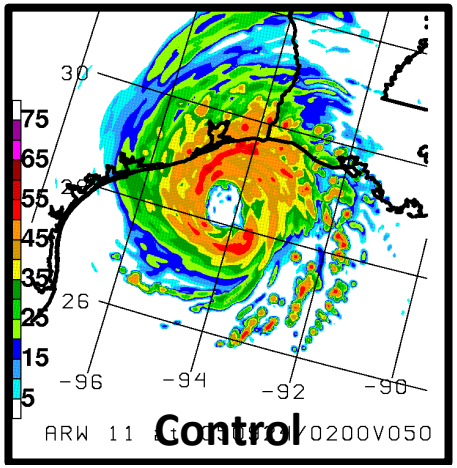










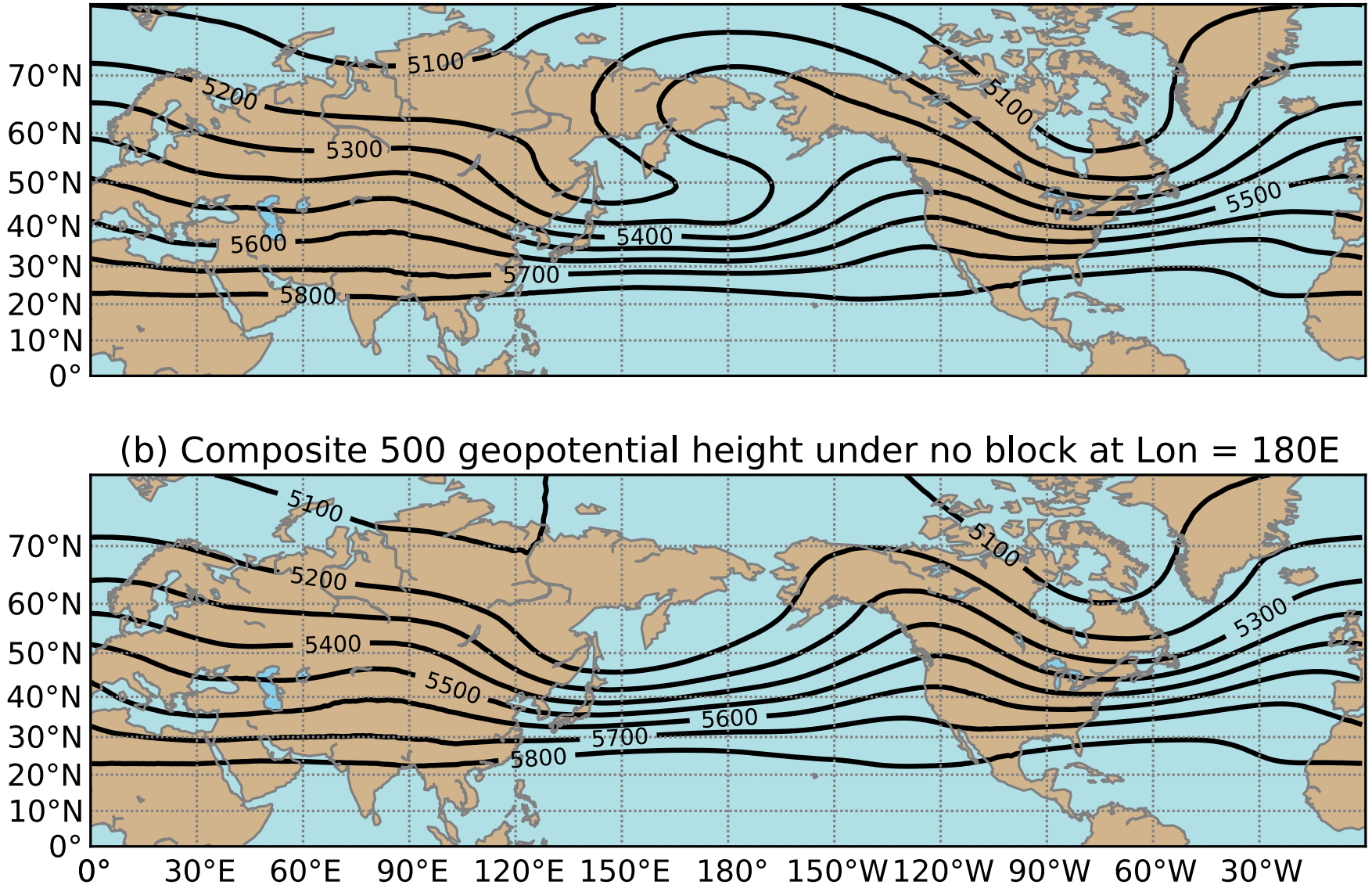


Maximum reflectivity (dBZ)
50-h ARW Forecast
Verifying 02Z/24 Sep '05

What else can one do with reforecasts?

Here we use them to examine the ability of the forecast model to skillfully predict low-frequency modes of variability and longer-lead forecasts (where large sample sizes are helpful). Specifically, the MJO and blocking, and their interactions.

(a) Composite 500 hPa geopotential height under block at Lon = 180E

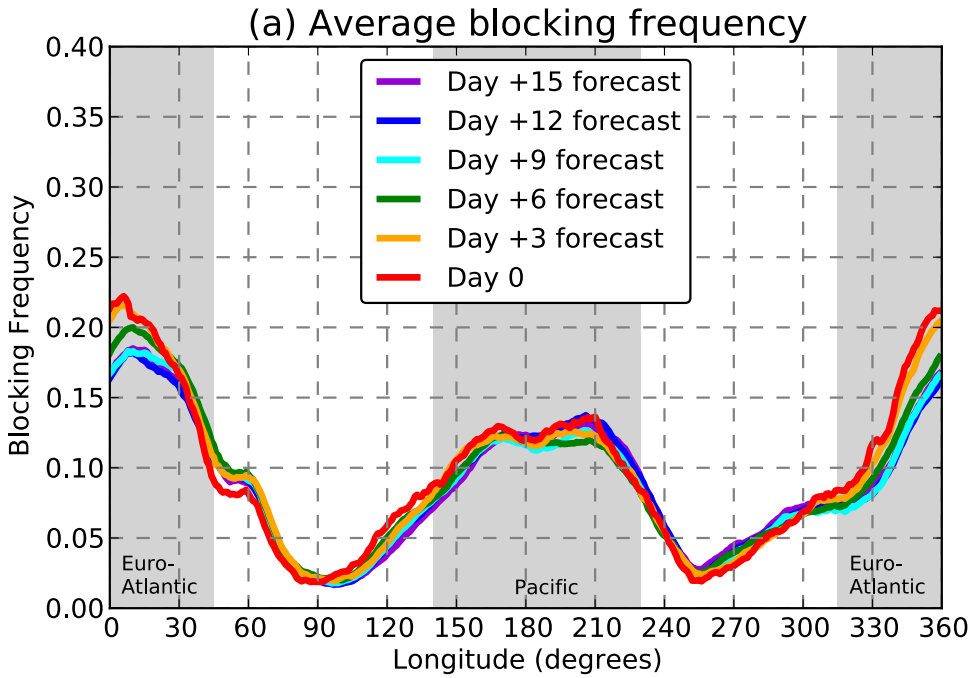


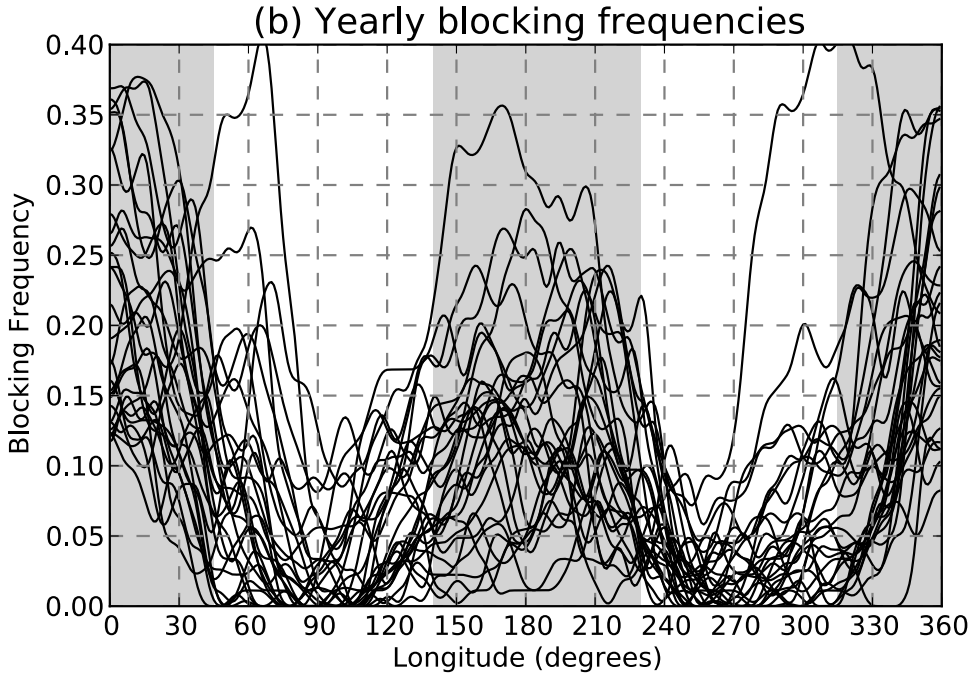
(b) Composite 500 geopotential height under no block at Lon = 180E

Dec-Jan-Feb 1985-2012 CFSR data. Blocks defined here by Tibaldi & Molteni algorithm.

Questions

- How well are blocking and MJO predicted in GEFS (reforecasts) at the medium range?
- Is blocking modulated by MJO, and is this modulation correctly forecast?

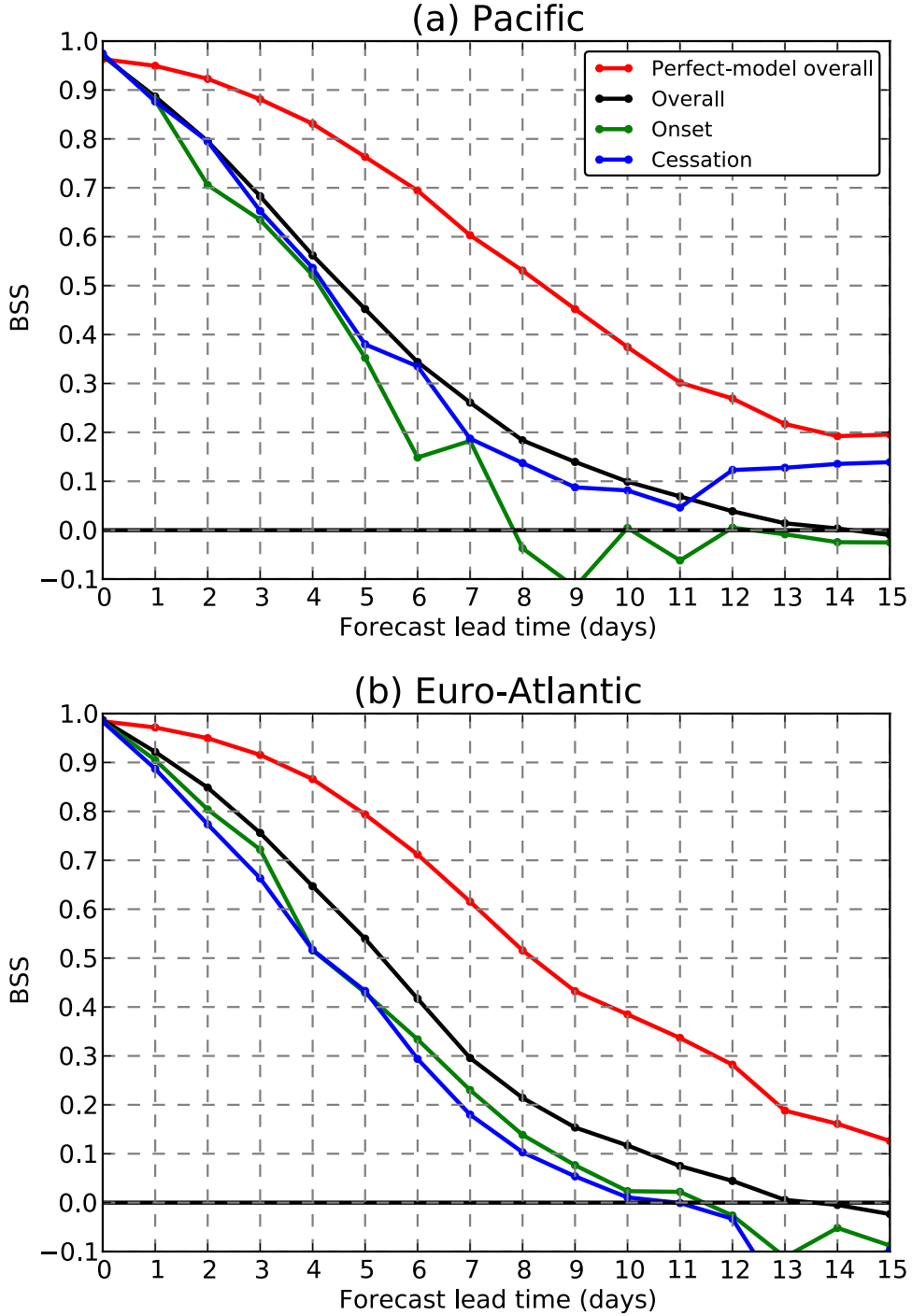




Blocking frequency and inter-annual variability

Blocking as defined in Tibaldi and Molteni (1990) using Z500. Grey bands defines Euro/Atlantic and Pacific blocking sectors in subsequent plots.

The large inter-annual variability of blocking may give you some sense of how misleading it may be to examine only a year or two of blocking forecasts.



Blocking skill in GEFS reforecasts

BSS (Brier skill score) as defined in supplementary slides.

Perfect model uses one member of ensemble as surrogate for analyzed.

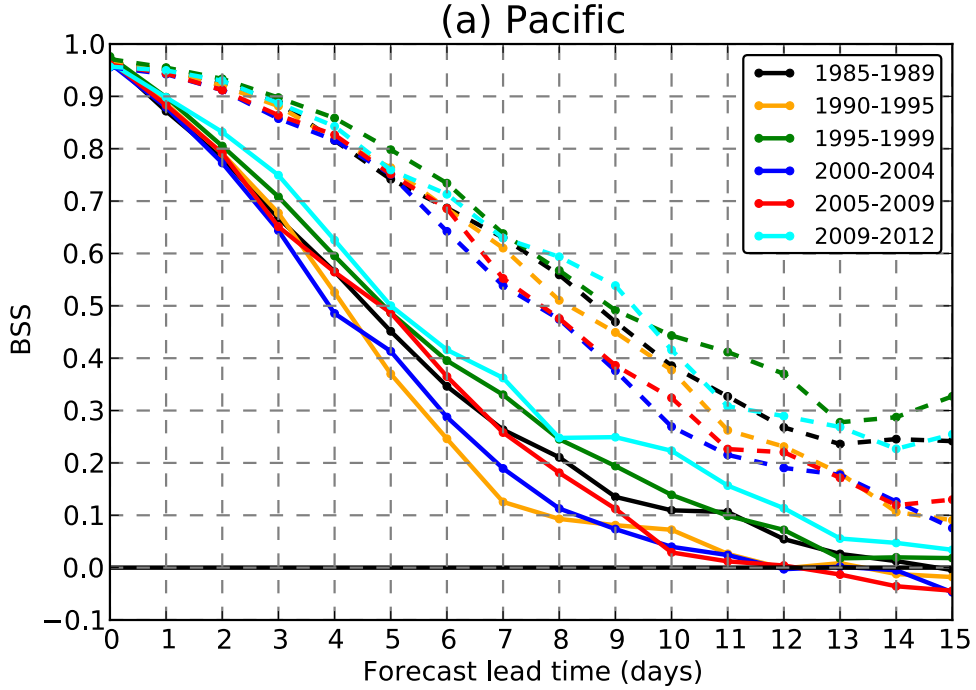
Real model: skill in blocking to ~13 days

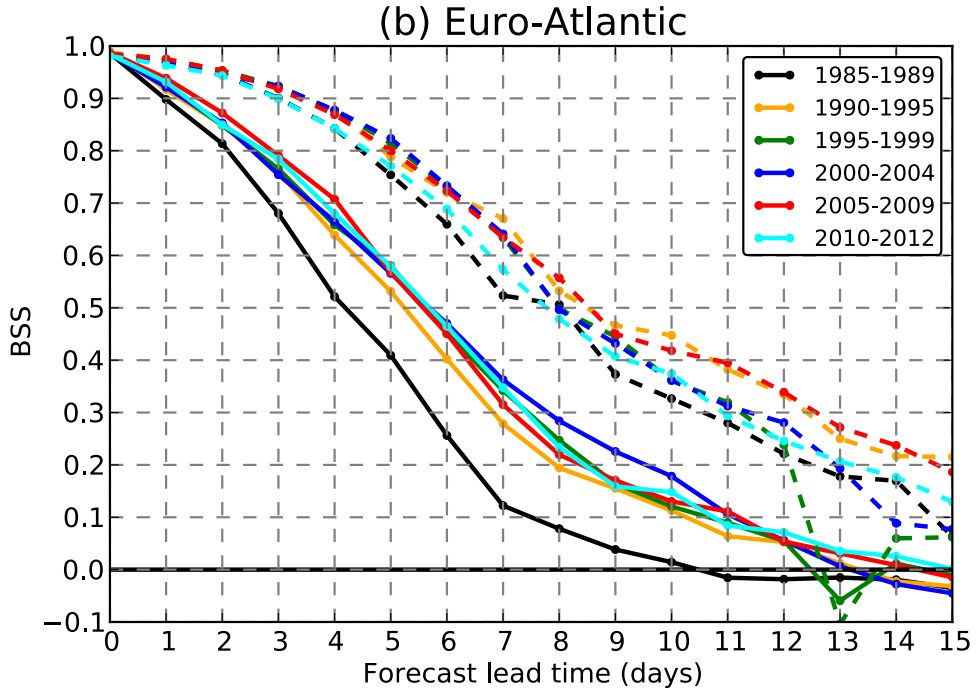
Perfect model: ~ 3-5 days longer skill.

Onset: date when there are more than 10 subsequent days where at least 20 degrees of longitude in a sector are blocked.

Cessation: date of end of that period.

Statistics include onset and previous 3 days, cessation and previous 3 days.





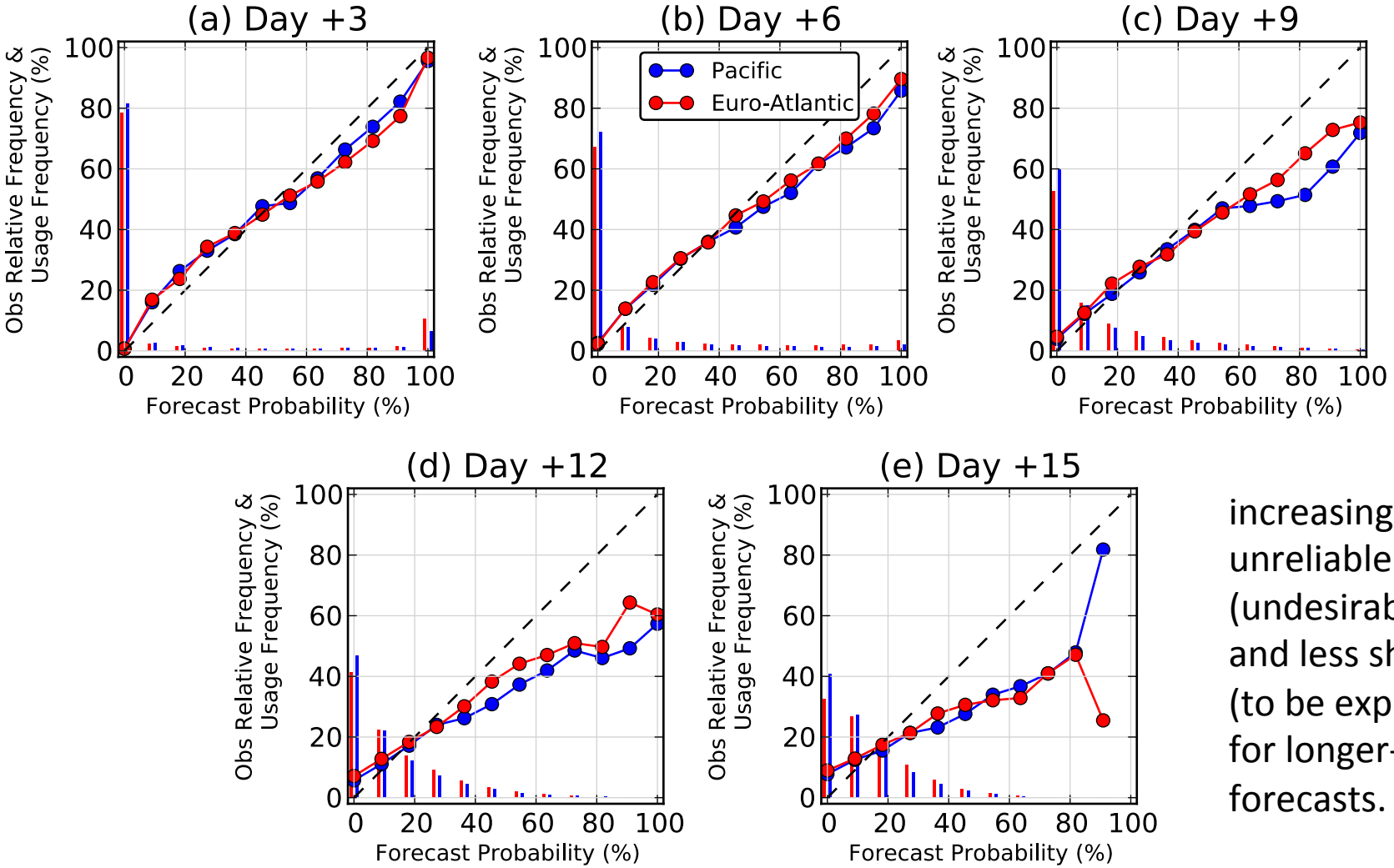
Blocking skill by half decade

—— actual skill.
--- perfect-model skill.

Decreased Atlantic sector skill in 1985-1989 period stands out.

Actual skill falls far short of perfect model skill. Some half-decadal variability due to natural variability, as seen in corresponding dashed lines.

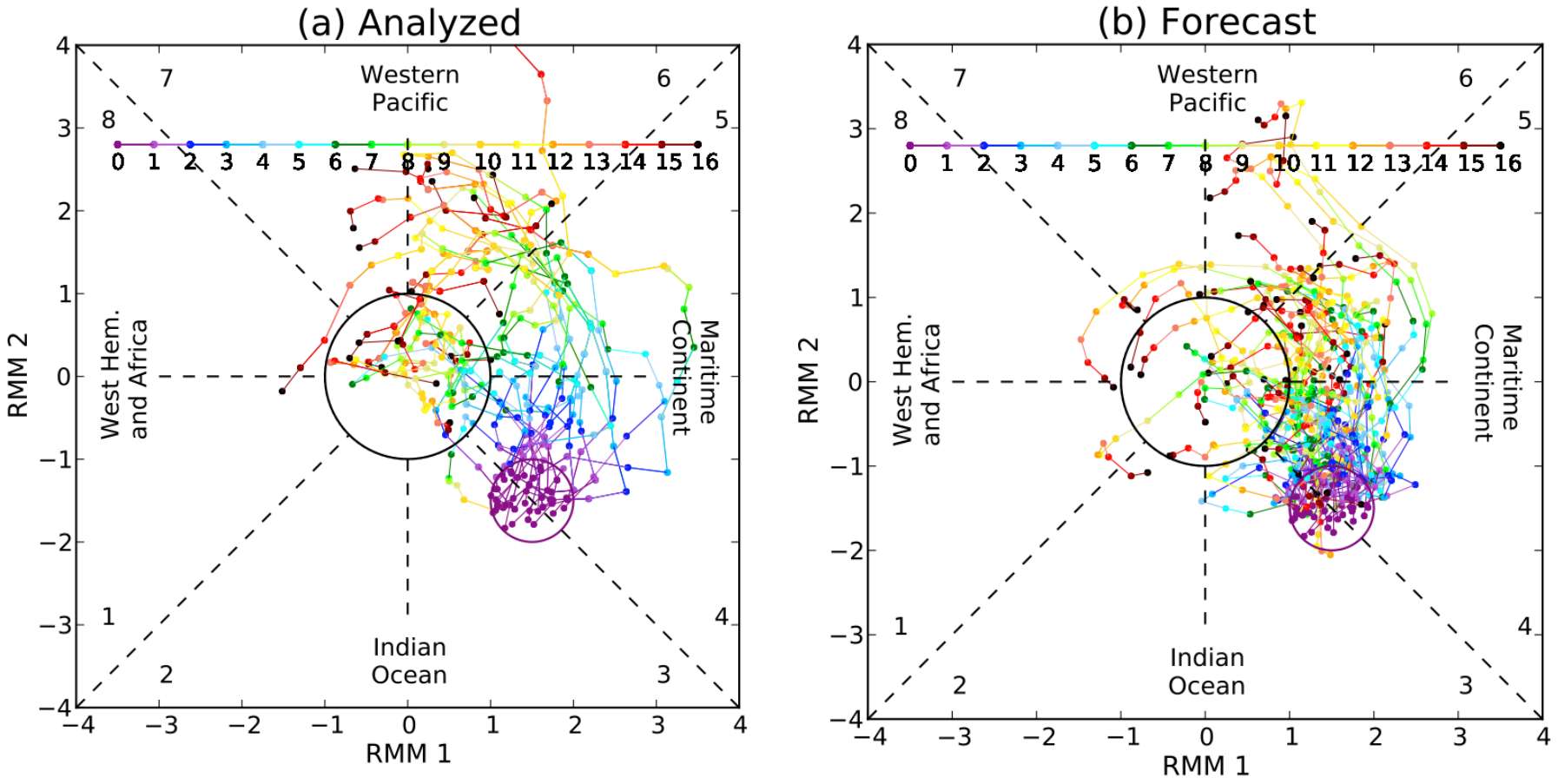
Reliability diagrams



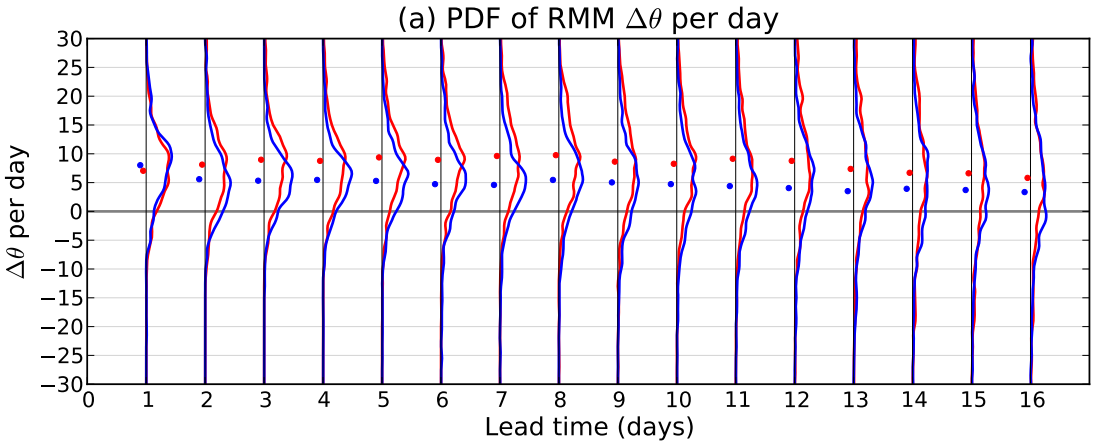
increasingly unreliable (undesirable) and less sharp (to be expected) for longer-lead forecasts.

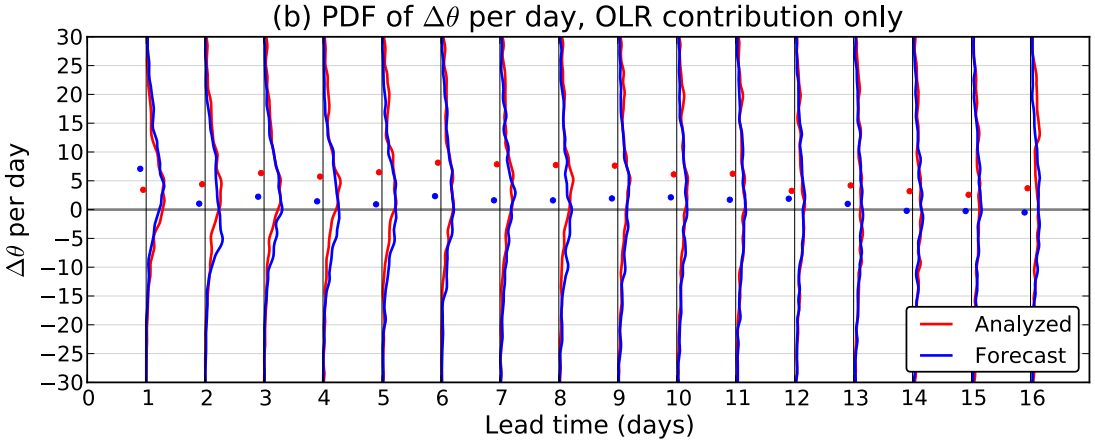
MJO diagnostics

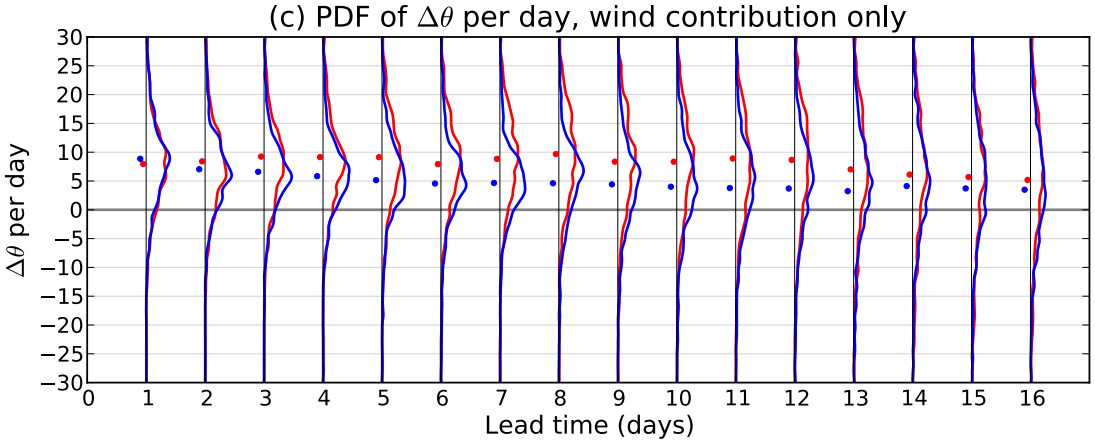
MJO, analyzed and deterministic forecast, E. Indian Ocean



Dates selected (Dec-Jan-Feb 1985-2012) where initial state was inside purple circle. Examine (RMM1, RMM2) phase plot for analyzed and forecast.



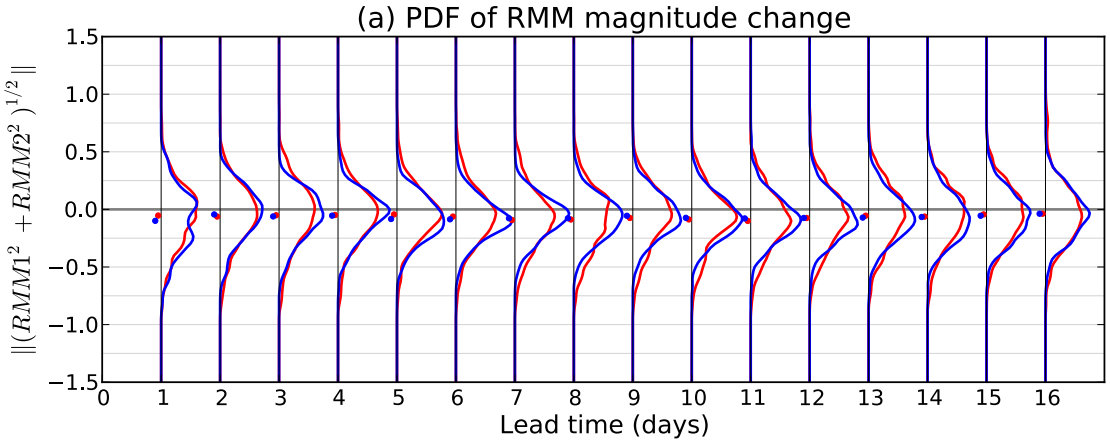


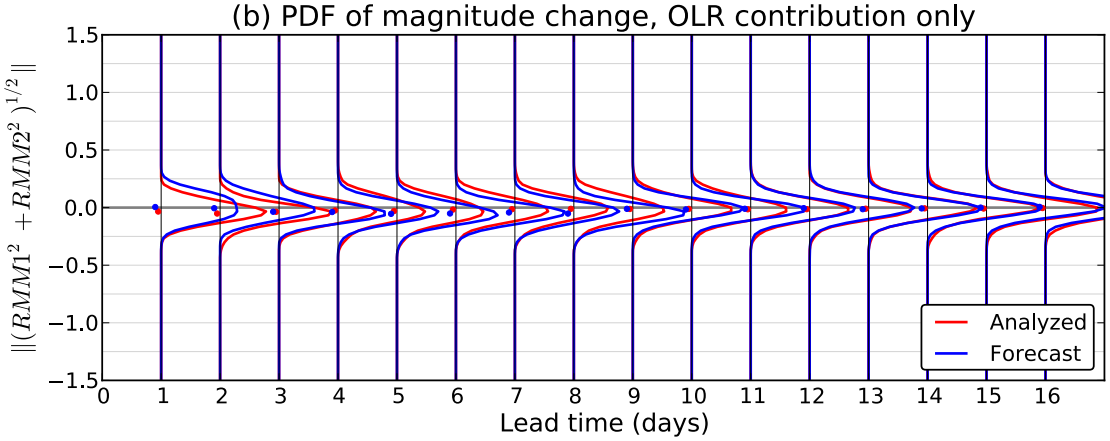


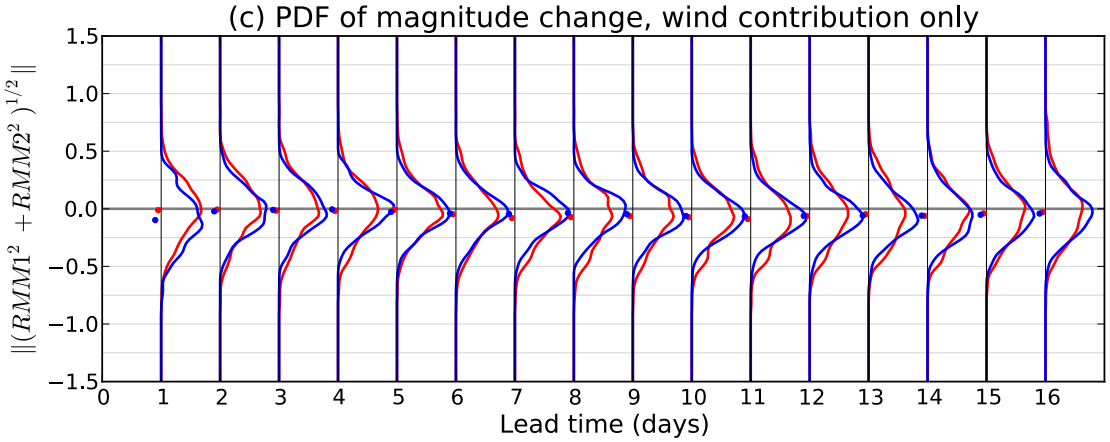
Propagation characteristics of MJO phase for strong MJO (magnitude > 1.0, OLR magnitude > 0.5, RMM2 < 0)

PDF shows change relative to data from a day prior, analyzed and ensemble members.

MJO forecasts propagate more slowly than analyzed, and more so for OLR component than wind components.



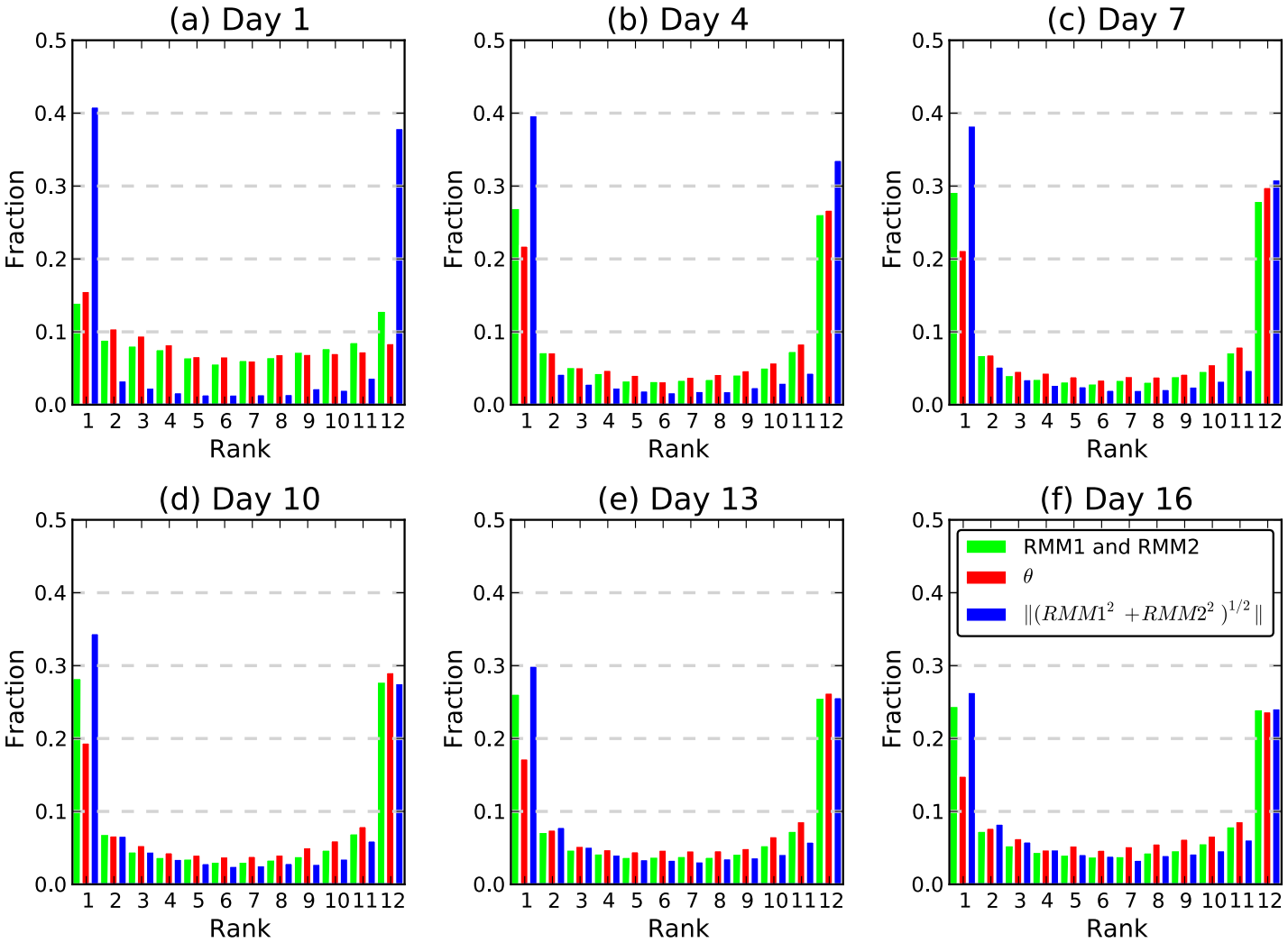




Propagation characteristics of MJO magnitude for strong MJO (magnitude > 1.0, OLR magnitude > 0.5, RMM2 < 0)

Forecast distribution is a bit more peaked, indicating that forecasts are more consistent in time in their magnitude than with analyzed conditions.

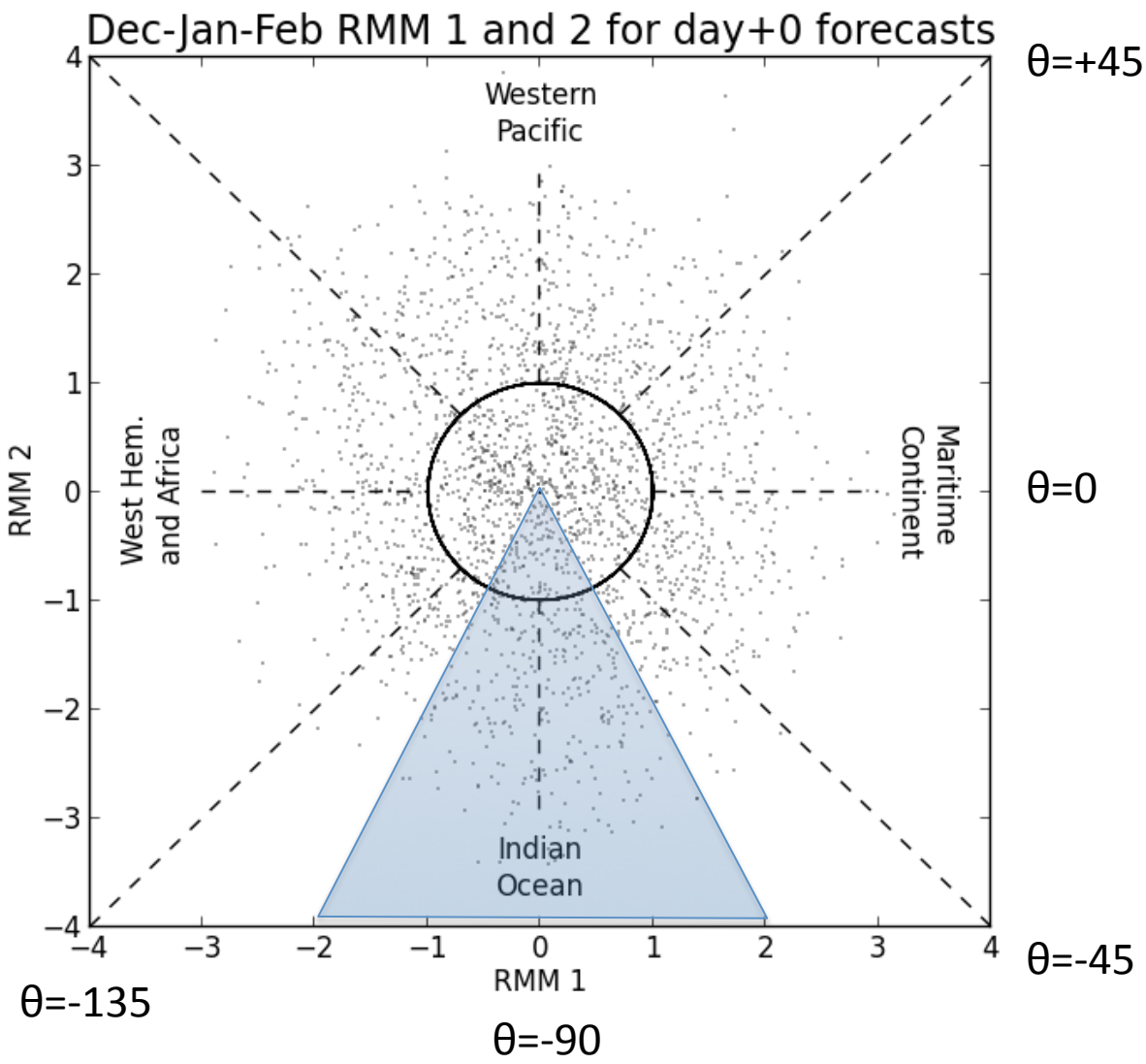
Rank histograms, RMM1 and RMM2 composited



GEFS reforecasts quickly develop insufficient spread and/or biased mean.
RMM most under-spread and/or biased at the medium range.
Especially insufficient variety of forecasts of MJO magnitude, even at short leads.

MJO – blocking interactions

My method of quantifying MJO phase



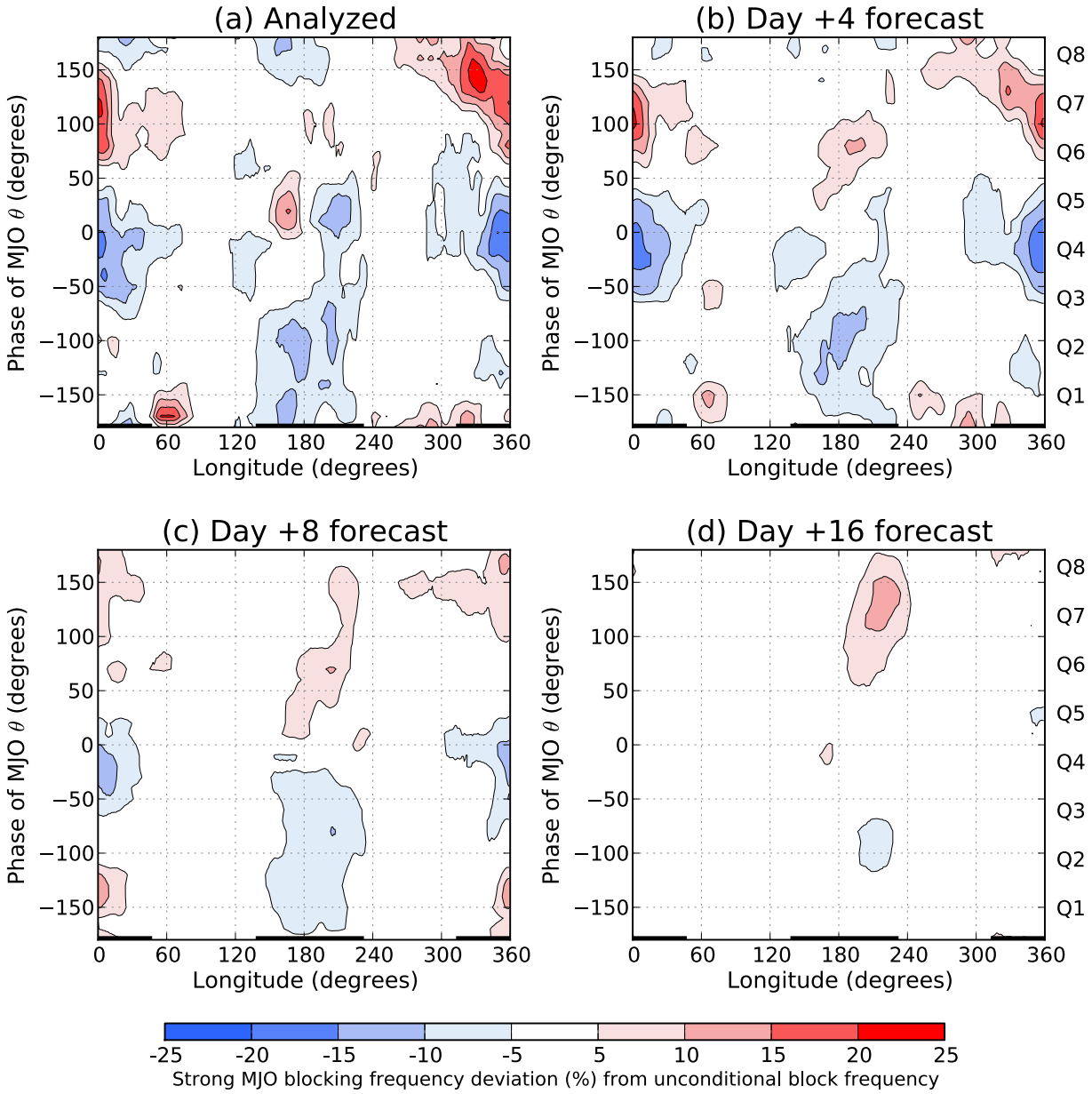
In subsequent plot you'll see I refer to the phase of MJO by its angle from x axis, a θ in conventional polar coordinates.

When examining statistics for $\theta=\theta_0$, I use RMM 1/2 samples with associated θ_0 +/- 22.5 degrees.

Example below for θ_0=-90 uses samples in blue cone.

A "strong" MJO is in the top 25% of RMM 1/2 amplitudes within the cone.

Blocking MJO relationship with strong MJOs



Top left panel shows, as the phase of a strong MJO varies along y axis, the change in blocking frequency from the overall climatological frequency for each given longitude (x axis). Note, for example, that as the phase of the MJO goes from 0° (Maritime continent) to 120°, Atlantic blocking frequency changes from suppressed to much more active than normal. This pattern is largely replicable in day + 4 forecasts, but much less so in day + 8 and day +16 forecasts. The GEFS is not replicating the blocking response to MJO forcing.

Accessing reforecast data

Data that is readily available from ESRL

Table 1: Reforecast variables available for selected mandatory and other vertical levels. Φ indicates geopotential height, and an X indicates that this variable is available from the reforecast data set at 1-degree resolution; a Y indicates that the variable is available at the native ~0.5 degree resolution. AGL indicates "above ground level."

Vertical Level	U	V	T	Φ	q	Wind Power
10 hPa	X	X	X	X		
50 hPa	X	X	X	X		
100 hPa	X	X	X	X		
200 hPa	X	X	X	X		
250 hPa	X	X	X	X		
300 hPa	X	X	X	X	X	
500 hPa	X	X	X	X	X	
700 hPa	X	X	X	X	X	
850 hPa	X	X	X	X	X	
925 hPa	X	X	X	X	X	
1000 hPa	X	X	X	X	X	
$\sigma \cong 0.996$	X	X		X		
$\sigma \cong 0.987$	X	X		X		
$\sigma \cong 0.977$	X	X		X		
$\sigma \cong 0.965$	X	X		X		
80m AGL	X,Y	X,Y				X,Y

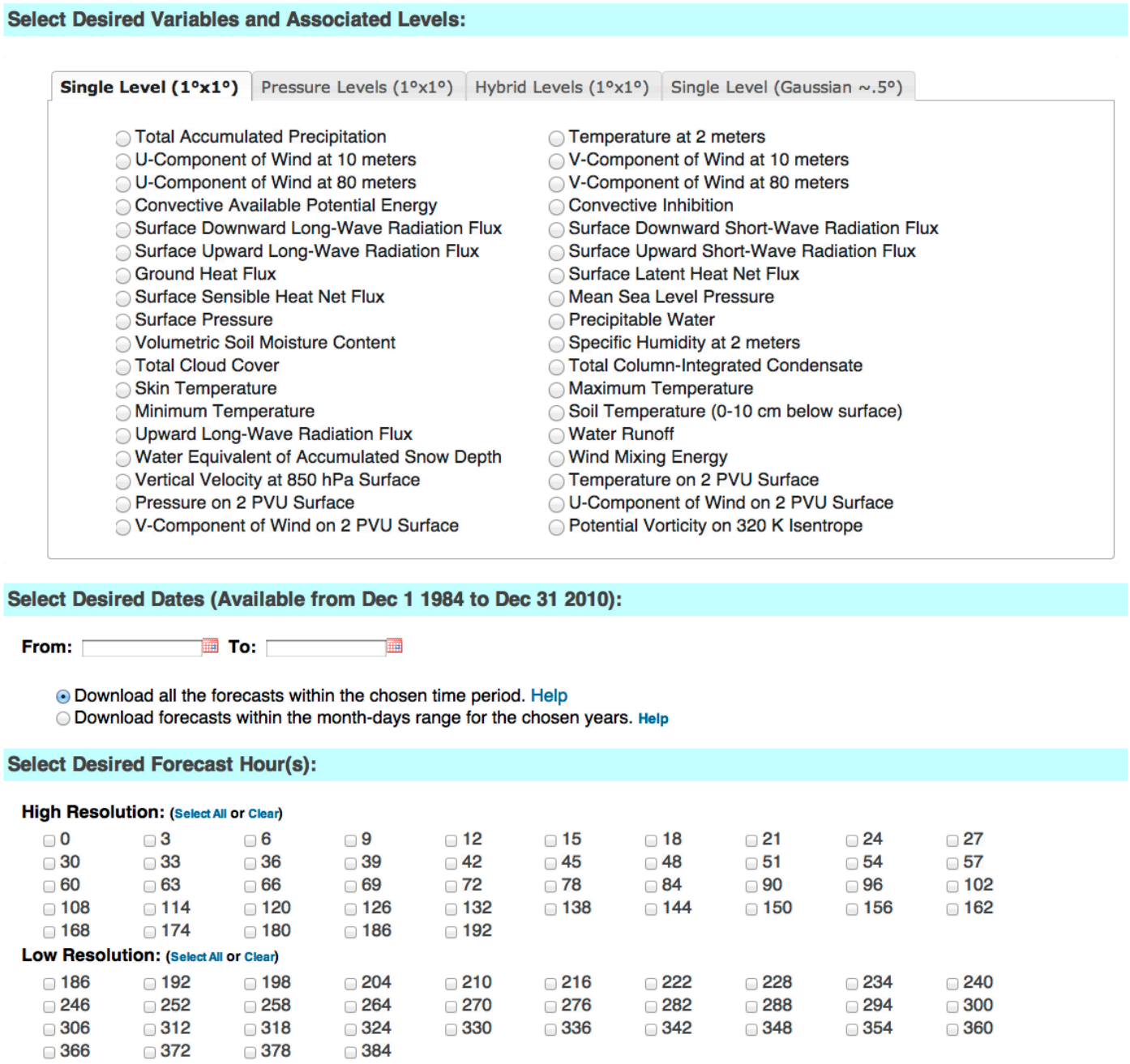
Also: hurricane track files

Data that is readily available from ESRL

Table 2: Single-level reforecast variables archived (and their units). Where an [Y] is displayed, this indicates that this variable is available at the native ~0.5-degree resolution as well as the 1-degree resolution.

Variable (units)
Mean sea-level pressure (Pa) [Y]
Skin temperature (K) [Y]
Soil temperature, 0.0 to 0.1 m depth (K) [Y]
Volumetric soil moisture content 0.0 to 0.1 m depth (fraction between wilting and saturation) [Y]
Water equivalent of accumulated snow depth (kg m^{-2}, i.e., mm) [Y]
2-meter temperature (K) [Y]
2-meter specific humidity (kg kg^{-1} dry air) [Y]
Maximum temperature (K) [Y]
Minimum temperature (K) [Y]
10-m u wind component (ms^{-1}) [Y]
10-m v wind component (ms^{-1}) [Y]
Total precipitation (kg m^{-2}, i.e., mm) [Y]
Water runoff (kg m^{-2}, i.e., mm) [Y]
Average surface latent heat net flux (W m^{-2}) [Y]
Average sensible heat net flux (W m^{-2}) [Y]
Average ground heat net flux (W m^{-2}) [Y]
Sunshine
Convective available potential energy (J kg^{-1}) [Y]
Convective inhibition (J kg^{-1}) [Y]
Precipitable water (kg m^{-2}, i.e., mm) [Y]
Total-column integrated condensate (kg m^{-2}, i.e., mm) [Y]
Total cloud cover (%)
Downward short-wave radiation flux at the surface (W m^{-2}) [Y]
Downward long-wave radiation flux at the surface (W m^{-2}) [Y]
Upward short-wave radiation flux at the surface (W m^{-2}) [Y]
Upward long-wave radiation flux at the surface (W m^{-2}) [Y]
Potential vorticity on θ = 320K isentropic surface (K m^2 kg^{-1} s^{-1})
U component on 2 PVU (1 PVU = 1 K m^2 kg^{-1} s^{-1}) isentropic surface (ms^{-1})
V component on 2 PVU isentropic surface (ms^{-1})
Temperature on 2 PVU isentropic surface
Pressure on 2 PVU isentropic surface

esrl.noaa.gov/psd/forecasts/reforecast2/download.html



Select Desired Variables and Associated Levels:

Single Level (1°x1°) | Pressure Levels (1°x1°) | Hybrid Levels (1°x1°) | Single Level (Gaussian ~.5°)

- Total Accumulated Precipitation
- U-Component of Wind at 10 meters
- U-Component of Wind at 80 meters
- Convective Available Potential Energy
- Surface Downward Long-Wave Radiation Flux
- Surface Upward Long-Wave Radiation Flux
- Ground Heat Flux
- Surface Sensible Heat Net Flux
- Surface Pressure
- Volumetric Soil Moisture Content
- Total Cloud Cover
- Skin Temperature
- Minimum Temperature
- Upward Long-Wave Radiation Flux
- Water Equivalent of Accumulated Snow Depth
- Vertical Velocity at 850 hPa Surface
- Pressure on 2 PVU Surface
- V-Component of Wind on 2 PVU Surface
- Temperature at 2 meters
- V-Component of Wind at 10 meters
- V-Component of Wind at 80 meters
- Convective Inhibition
- Surface Downward Short-Wave Radiation Flux
- Surface Upward Short-Wave Radiation Flux
- Surface Latent Heat Net Flux
- Mean Sea Level Pressure
- Precipitable Water
- Specific Humidity at 2 meters
- Total Column-Integrated Condensate
- Maximum Temperature
- Soil Temperature (0-10 cm below surface)
- Water Runoff
- Wind Mixing Energy
- Temperature on 2 PVU Surface
- U-Component of Wind on 2 PVU Surface
- Potential Vorticity on 320 K Isentrope

Select Desired Dates (Available from Dec 1 1984 to Dec 31 2010):

From: **To:**

- Download all the forecasts within the chosen time period. Help
- Download forecasts within the month-days range for the chosen years. Help

Select Desired Forecast Hour(s):

High Resolution: (Select All or Clear)

0	3	6	9	12	15	18	21	24	27
30	33	36	39	42	45	48	51	54	57
60	63	66	69	72	78	84	90	96	102
108	114	120	126	132	138	144	150	156	162
168	174	180	186	192					

Low Resolution: (Select All or Clear)

186	192	198	204	210	216	222	228	234	240
246	252	258	264	270	276	282	288	294	300
306	312	318	324	330	336	342	348	354	360
366	372	378	384						

Produces netCDF files.

Also: direct ftp access to allow you to read the raw grib files.

ftp access is also easy.

- $ ftp –i ftp.cdc.noaa.gov
- [user anonymous, pass=email address]
- cd /Projects/Reforecast2
- cd to YYYY/YYYYMM/YYYYMMDDHH of interest, e.g.,
  - $ cd 2000
  - $ cd 200001
  - $ cd 2000010100
  - $ ls

directory listing

```
150 Here comes the directory listing.
drwxr-xr-x    4 99       1201           80 Mar 11  2012 c00
drwxrwxr-x    3 99       1201           80 Jun 27  2012 mean
drwxr-xr-x    4 99       1201           80 Mar 11  2012 p01
drwxr-xr-x    4 99       1201           80 Mar 11  2012 p02
drwxr-xr-x    4 99       1201           80 Mar 11  2012 p03
drwxr-xr-x    4 99       1201           80 Mar 11  2012 p04
drwxr-xr-x    4 99       1201           80 Mar 11  2012 p05
drwxr-xr-x    4 99       1201           80 Mar 11  2012 p06
drwxr-xr-x    4 99       1201           80 Mar 11  2012 p07
drwxr-xr-x    4 99       1201           80 Mar 11  2012 p08
drwxr-xr-x    4 99       1201           80 Mar 11  2012 p09
drwxr-xr-x    4 99       1201           80 Mar 11  2012 p10
drwxrwxr-x    3 99       1201           80 Jun 27  2012 sprd
226 Directory send OK.
ftp>
```

top level director listing, with subdirectories for each perturbation number, the mean, and the spread

more directory listings

```
ftp> cd c00
250 Directory successfully changed.
ftp> ls
229 Entering Extended Passive Mode (|||53711|).
150 Here comes the directory listing.
drwxr-xr-x    2 99       1201        13312 Aug 13  2012 gaussian
drwxr-xr-x    2 99       1201        21504 Aug 13  2012 latlon
226 Directory send OK.
ftp> cd latlon
250 Directory successfully changed.
ftp> ls
229 Entering Extended Passive Mode (|||36968|).
150 Here comes the directory listing.
-rw-rw-r--    1 99       1201      4236525 Mar 28  2012 apcp_sfc_2000010100_c00.grib2
-rw-rw-r--    1 99       1201         3277 Aug 13  2012 apcp_sfc_2000010100_c00.grib2.inv
-rw-rw-r--    1 99       1201         1547 Jun 11  2012 apcp_sfc_2000010100_c00.grib2.pyidx
-rw-rw-r--    1 99       1201      3419759 Mar 28  2012 apcp_sfc_2000010100_c00_t190.grib2
-rw-rw-r--    1 99       1201         2525 Aug 13  2012 apcp_sfc_2000010100_c00_t190.grib2.inv
-rw-rw-r--    1 99       1201         1261 Jun 11  2012 apcp_sfc_2000010100_c00_t190.grib2.pyidx
-rw-rw-r--    1 99       1201      4447215 Mar 28  2012 cape_sfc_2000010100_c00.grib2
-rw-rw-r--    1 99       1201         2948 Aug 13  2012 cape_sfc_2000010100_c00.grib2.inv
-rw-rw-r--    1 99       1201         1547 Jun 11  2012 cape_sfc_2000010100_c00.grib2.pyidx
-rw-rw-r--    1 99       1201      3394508 Mar 28  2012 cape_sfc_2000010100_c00_t190.grib2
-rw-rw-r--    1 99       1201         2252 Aug 13  2012 cape_sfc_2000010100_c00_t190.grib2.inv
-rw-rw-r--    1 99       1201         1261 Jun 11  2012 cape_sfc_2000010100_c00_t190.grib2.pyidx
-rw-rw-r--    1 99       1201      4683420 Mar 28  2012 cin_sfc_2000010100_c00.grib2
-rw-rw-r--    1 99       1201         2905 Aug 13  2012 cin_sfc_2000010100_c00.grib2.inv
```

a select set of fields available on original Gaussian grid

a larger set available on the 1-degree lat/lon grid

more directory listings

```
ftp> cd c00
250 Directory successfully changed.
ftp> ls
229 Entering Extended Passive Mode (|||53711|).
150 Here comes the directory listing.
drwxr-xr-x    2 99       1201        13312 Aug 13  2012 gaussian
drwxr-xr-x    2 99       1201        21504 Aug 13  2012 latlon
226 Directory send OK.
ftp> cd latlon
250 Directory successfully changed.
ftp> ls
229 Entering Extended Passive Mode (|||36968|).
150 Here comes the directory listing.
-rw-rw-r--    1 99       1201      4236525 Mar 28  2012 apcp_sfc_2000010100_c00.grib2
-rw-rw-r--    1 99       1201         3277 Aug 13  2012 apcp_sfc_2000010100_c00.grib2.inv
-rw-rw-r--    1 99       1201         1547 Jun 11  2012 apcp_sfc_2000010100_c00.grib2.pyidx
-rw-rw-r--    1 99       1201      3419759 Mar 28  2012 apcp_sfc_2000010100_c00_t190.grib2
-rw-rw-r--    1 99       1201         2525 Aug 13  2012 apcp_sfc_2000010100_c00_t190.grib2.inv
-rw-rw-r--    1 99       1201         1261 Jun 11  2012 apcp_sfc_2000010100_c00_t190.grib2.pyidx
-rw-rw-r--    1 99       1201      4447215 Mar 28  2012 cape_sfc_2000010100_c00.grib2
-rw-rw-r--    1 99       1201         2948 Aug 13  2012 cape_sfc_2000010100_c00.grib2.inv
-rw-rw-r--    1 99       1201         1547 Jun 11  2012 cape_sfc_2000010100_c00.grib2.pyidx
-rw-rw-r--    1 99       1201      3394508 Mar 28  2012 cape_sfc_2000010100_c00_t190.grib2
-rw-rw-r--    1 99       1201         2252 Aug 13  2012 cape_sfc_2000010100_c00_t190.grib2.inv
-rw-rw-r--    1 99       1201         1261 Jun 11  2012 cape_sfc_2000010100_c00_t190.grib2.pyidx
-rw-rw-r--    1 99       1201      4683420 Mar 28  2012 cin_sfc_2000010100_c00.grib2
-rw-rw-r--    1 99       1201         2905 Aug 13  2012 cin_sfc_2000010100_c00.grib2.inv
```

these are the week +1 and week +2 grib files for accumulated precip.
other variables have other (hopefully obvious) names.

more directory listings

```
ftp> cd c00
250 Directory successfully changed.
ftp> ls
229 Entering Extended Passive Mode (|||53711|).
150 Here comes the directory listing.
drwxr-xr-x    2 99       1201        13312 Aug 13  2012 gaussian
drwxr-xr-x    2 99       1201        21504 Aug 13  2012 latlon
226 Directory send OK.
ftp> cd latlon
250 Directory successfully changed.
ftp> ls
229 Entering Extended Passive Mode (|||36968|).
150 Here comes the directory listing.
-rw-rw-r--    1 99       1201      4236525 Mar 28  2012 apcp_sfc_2000010100_c00.grib2
-rw-rw-r--    1 99       1201         3277 Aug 13  2012 apcp_sfc_2000010100_c00.grib2.inv
-rw-rw-r--    1 99       1201         1547 Jun 11  2012 apcp_sfc_2000010100_c00.grib2.pyidx
-rw-rw-r--    1 99       1201      3419759 Mar 28  2012 apcp_sfc_2000010100_c00_t190.grib2
-rw-rw-r--    1 99       1201         2525 Aug 13  2012 apcp_sfc_2000010100_c00_t190.grib2.inv
-rw-rw-r--    1 99       1201         1261 Jun 11  2012 apcp_sfc_2000010100_c00_t190.grib2.pyidx
-rw-rw-r--    1 99       1201      4447215 Mar 28  2012 cape_sfc_2000010100_c00.grib2
-rw-rw-r--    1 99       1201         2948 Aug 13  2012 cape_sfc_2000010100_c00.grib2.inv
-rw-rw-r--    1 99       1201         1547 Jun 11  2012 cape_sfc_2000010100_c00.grib2.pyidx
-rw-rw-r--    1 99       1201      3394508 Mar 28  2012 cape_sfc_2000010100_c00_t190.grib2
-rw-rw-r--    1 99       1201         2252 Aug 13  2012 cape_sfc_2000010100_c00_t190.grib2.inv
-rw-rw-r--    1 99       1201         1261 Jun 11  2012 cape_sfc_2000010100_c00_t190.grib2.pyidx
-rw-rw-r--    1 99       1201      4683420 Mar 28  2012 cin_sfc_2000010100_c00.grib2
-rw-rw-r--    1 99       1201         2905 Aug 13  2012 cin_sfc_2000010100_c00.grib2.inv
```

these precomputed inventory files allow *much* faster loading with wgrib2 or python.

Other files of interest

- in /Public/thamill you have netCDF files like

refcstv2_precip_ccpa_BBB_to_EEE.nc where BBB is the beginning lead time in hours, EEE is the end lead time in hours. These have 2002-2013 1/8-degree CCPA precipitation analyses and associated forecast information.

(note, if you downloaded these files earlier, please download again, as I corrected some bugs)

Some netCDF file detail

- [mac24:~/refcst2] thamill% ncdump -h /Rf2_tests/ccpa/netcdf/refcstv2_precip_ccpa_120_to_132.nc
- netcdf refcstv2_precip_ccpa_120_to_132 {
- dimensions:
- xa = 464 ;
- ya = 224 ;
- xf = 128 ; grid dimensions of analysis (xa, ya) and forecast (xf, yf)
- yf = 61 ;
- time = UNLIMITED ; // (4228 currently) number of dates stored
- ens = 11 ; number of ensemble members
- variables:
- float xa(xa) ;
- xa:long_name = "eastward grid point number, precip analysis" ;
- xa:units = "n/a" ;
- float ya(ya) ;
- ya:long_name = "northward grid point number, precip analysis" ;
- ya:units = "n/a" ;
- float xf(xf) ;
- xf:long_name = "eastward grid point number, precip forecast" ;
- xf:units = "n/a" ;
- float yf(yf) ;
- yf:long_name = "northward grid point number, precip forecast" ;
- yf:units = "n/a" ;
- float time(time) ;
- time:units = "hours since 1-1-1 00:00:0.0" ;
- int ensv(ens) ;
- ensv:long_name = "Ensemble member number (control, perts 1-10)" ;
- ensv:units = " " ;

more netCDF file detail

```
float lons_anal(ya, xa) ;
        lons_anal:long_name = "longitude" ;
        lons_anal:units = "degrees_east" ;
float lats_anal(ya, xa) ;
        lats_anal:long_name = "latitude" ;
        lats_anal:units = "degrees_north" ;
float lons_fcst(yf, xf) ;
        lons_fcst:long_name = "longitude" ;
        lons_fcst:units = "degrees_east" ;
float lats_fcst(yf, xf) ;
        lats_fcst:long_name = "latitude" ;
        lats_fcst:units = "degrees_north" ;
int yyyymmddhh_init(time) ;
        yyyymmddhh_init:longname = "Initial condition date/time in yyyymmddhh format" ;
int yyyymmddhh_fcstb(time) ;
        yyyymmddhh_fcstb:longname = "Forecast valid date/time in yyyymmddhh format" ;
int yyyymmddhh_fcste(time) ;
        yyyymmddhh_fcste:longname = "Forecast valid date/time in yyyymmddhh format" ;
float apcp_anal(time, ya, xa) ;
        apcp_anal:least_significant_digit = 2 ;
        apcp_anal:units = "mm" ;
        apcp_anal:long_name = "Analyzed Accumulated Precipitation from CCPA" ;
        apcp_anal:valid_range = 0., 1000. ;
        apcp_anal:missing_value = -99.99f ;
float apcp_fcst_ens(time, ens, yf, xf) ;
        apcp_fcst_ens:least_significant_digit = 2 ;
        apcp_fcst_ens:units = "mm" ;
        apcp_fcst_ens:long_name = "Ensemble member forecast accumulated precipitation" ;
        apcp_fcst_ens:valid_range = 0., 1000. ;
        apcp_fcst_ens:missing_value = -99.99f ;
```

longitudes, latitudes of analysis grid; note that when you read this in via fortran, the internal (y,x) storage gets flipped to (x,y), but not so when you read it in via python.

longitudes, latitudes of forecasts

precipitation analyses

precipitation forecast

more netCDF file detail

- float pwat_fcst_mean(time, yf, xf) ;
- pwat_fcst_mean:least_significant_digit = 2 ;
- pwat_fcst_mean:units = "mm" ;
- pwat_fcst_mean:long_name = "Ensemble precipitable water forecast mean, time averaged" ;
- pwat_fcst_mean:valid_range = 0., 1000. ;
- pwat_fcst_mean:missing_value = -99.99f ;
- float CAPE_fcst_mean(time, yf, xf) ;
- CAPE_fcst_mean:least_significant_digit = 2 ;
- CAPE_fcst_mean:units = "J/kg" ;
- CAPE_fcst_mean:long_name = "Ensemble CAPE (convective availble potential energy) forecast mean, time averaged" ;
- CAPE_fcst_mean:valid_range = 0., 7000. ;
- CAPE_fcst_mean:missing_value = -99.99f ;
- float CIN_fcst_mean(time, yf, xf) ;
- CIN_fcst_mean:least_significant_digit = 2 ;
- CIN_fcst_mean:units = "J/kg" ;
- CIN_fcst_mean:long_name = "Ensemble CIN (convective inhibition) forecast mean, time averaged" ;
- CIN_fcst_mean:valid_range = -3000, 3000 ;
- CIN_fcst_mean:missing_value = -9999.99f ;
- float T2m_fcst_mean(time, yf, xf) ;
- T2m_fcst_mean:least_significant_digit = 2 ;
- T2m_fcst_mean:units = "K" ;
- T2m_fcst_mean:long_name = "2-meter temperature forecast mean, time averaged" ;
- T2m_fcst_mean:valid_range = 200, 350 ;
- T2m_fcst_mean:missing_value = -9999.99f ;
- float MSLP_fcst_mean(time, yf, xf) ;
- MSLP_fcst_mean:least_significant_digit = 2 ;
- MSLP_fcst_mean:units = "Pa" ;
- MSLP_fcst_mean:long_name = "Mean sea-level pressure forecast mean, time averaged" ;
- MSLP_fcst_mean:valid_range = 85000., 108000. ;
- MSLP_fcst_mean:missing_value = -9999.99f ;

Other home-grown software

- python script for computing threat scores based on information in these files. Provide the beginning, ending lead times, a file name with a list of dates for end times of (24-h) forecast period, and it computes the threat scores.
- analog precip forecast software using this 1/8-degree CCPA precip. analysis information.

Define BSS for evaluating blocking skill

- The blocking Brier Skill score is calculated after summing forecast and climatological Brier scores over the relevant longitudes in either the Pacific or Atlantic basins, respectively, then averaged. For example (Pac):

$$BSS = 1.0 - \frac{BS_{forecast}}{BS_{climo}}$$

$$BS_{forecast} = \sum_{l_p=1}^{nlons} \sum_{i=1}^{ndates} \left(p_i^{forecast}\left(l_p\right) - o_i\left(l_p\right)\right)^2$$

$$BS_{climo} = \sum_{l_p=1}^{nlons} \sum_{i=1}^{ndates} \left(p_i^{climo}\left(l_p\right) - o_i\left(l_p\right)\right)^2$$

$$o_i\left(l_p\right) = \left\{ \begin{array}{ll} 1 & if\ blocked \\ 0 & if\ unblocked \end{array} \right\}$$

$$p_i^{forecast}\left(l_p\right) = ensemble-based\ probability\ of\ block\ for\ this\ longitude$$

$$p_i^{climo}\left(l_p\right) = climatological\ probability\ of\ block\ for\ this\ longitude$$

Supplementary slides

Computing the CRPSS of GEFS RMM1 and RMM2 forecasts

- CRPSS = 1 – CRPS(forecast) / CRPS(climatology)

$$CRPS(forecast) = \sum_{i=1}^{ndates} \sum_{j=1}^{ncats} \frac{1}{ncats} \left(\Phi_{forecast}\left(i, x(j)\right) - \Phi_{analyzed}\left(i, x(j)\right) \right)^2$$

$$CRPS(climo) = \sum_{i=1}^{ndates} \sum_{j=1}^{ncats} \frac{1}{ncats} \left(\Phi_{climo}\left(i, x(j)\right) - \Phi_{analyzed}\left(i, x(j)\right) \right)^2$$

$$x(1) = -5.0, \quad x(2) = -4.9, \quad \ldots \quad , \; x(ncats) = +5.0$$

$$\Phi(\,\cdot\,) = cumulative\ distribution\ function\ for\ either\ RMM1\ or\ RMM2$$

- $\Phi(\cdot)$ estimated from normal distribution fit to sample mean and standard deviation.

Blocking computation method: follows Tibaldi and Molteni, 1990 *Tellus*

(1950). The procedure we have applied is as follows: the 500 hPa field is firstly evaluated on a 4° by 4° regular latitude–longitude grid covering the Northern Hemisphere. Then the geopotential height gradients GHGS and GHGN (referring to middle and high latitudes respectively) are computed for each longitude point of the grid:

$$\mathrm{GHGS} = \frac{Z(\phi_o) - Z(\phi_s)}{(\phi_o - \phi_s)},$$

$$\mathrm{GHGN} = \frac{Z(\phi_n) - Z(\phi_o)}{(\phi_n - \phi_o)},$$

where

$\phi_n = 80°\,\mathrm{N} + \Delta,$
$\phi_o = 60°\,\mathrm{N} + \Delta,$
$\phi_s = 40°\,\mathrm{N} + \Delta,$
$\Delta = -4°, 0°$ or $4°.$

A given longitude is then defined as "blocked" at a specific instant in time if the following conditions are satisfied for at least one value of Δ:

(1) $\mathrm{GHGS} > 0,$

(2) $\mathrm{GHGN} < -10$ m/deg lat.

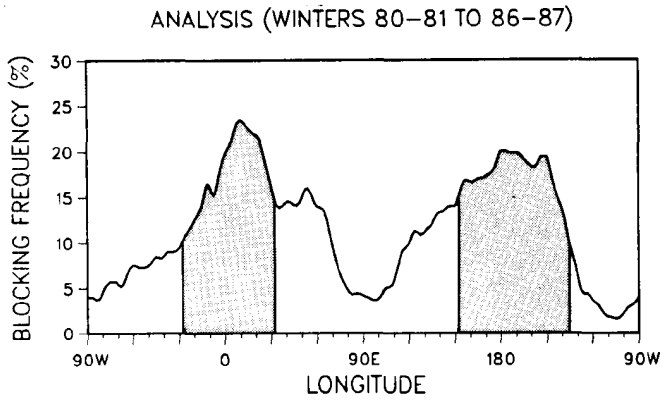


Fig. 1. Percentage frequency of blocking (objectively defined in Section 2) as a function of longitude and computed on all ECMWF daily objective analyses of our database.

There are alternatives, such as PV-based index by Pelly and Hoskins. While these may have some advantages, this old standard used hereafter.

MJO task force data

Center	Model	Data Stream ID	Ensemble Members	Forecasts Start	Length (Days)	Realtime Data FTP	Version 1 Plots	Model Climatology
NCEP	GFS EPS T126[1]	NCPE	21	Nov 2007	15	-----	Yes	No
NCEP	GFS T382[1]	NCPO	1	Jan 2008	15	-----	Yes	No
NCEP	CFS T62[1]	NCFS	4	Jan 2007	40	-----	Yes	Yes
CMC	GEMDM_400x200[2]	CANM	20	Jun 2008	16	Yes	Yes	No
UKMO	MOGREPS[3]	UKMA	1	Oct 2007	15	Yes	Yes	No
UKMO	MOGREPS[3]	UKME	23	Oct 2007	15	Yes	Yes	No
ABOM	GASP T239[4]	BOMA	1	Jan 2008	10	Yes	Yes	No
ABOM	GASP EPS T119[4]	BOME	32	Aug 2008	10	Yes	Yes	No
ABOM	POAMA1.5bT47[4]	BOMC	1	Jan 2008	40	Yes	Yes	No
ABOM	POAMA1.5T47[4]	BOMH	1	Jan 2008	40	Yes	No	Yes
ECMWF	VAREPS T299/T255[*]	ECMF	51	Jun 2008	15	Yes	Yes	No
ECMWF	VAREPS T299/T255[*]	ECMM	51	Jun 2008	15	Yes	Yes	Yes
ECMWF	SFSv3 T159[*]	EMON	51 (W)	Jun 2008	32	Yes	Yes	No
ECMWF	SFSv3 T159[*]	EMOM	51 (W)	Jun 2008	32	Yes	Yes	Yes
JMA	GSM WEPS T319[6]	JMAN	51	Nov 2008	9	Yes	Yes	No
CPTEC	GWEFS T126[7]	CPTC	15	Feb 2009	15	Yes	Yes	No
IMD	NCMRWF T254[8]	IMDO	1	Jun 2009	7	Yes	Yes	No
IMD	NCMRWF EPS T80[8]	IMDE	8	------	7	No	No	No
FNMOC	NOGAPS T119[9]	NGAP	10	------	10	No	No	No
TCWB	CWB EPS T119^	TCWB	1	Oct 2009	40	Yes	Yes	No

MJO deterministic verification metrics

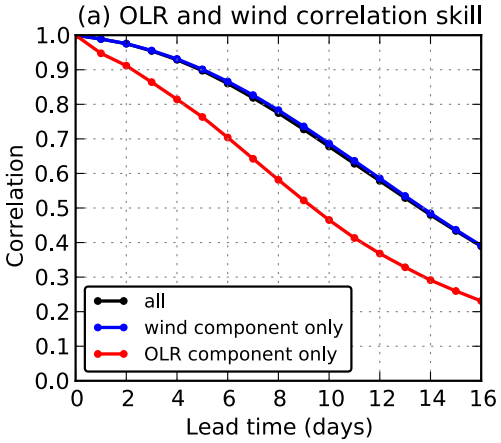
$$\mathrm{COR}(\tau) = \frac{\sum_{i=1}^{N} [a_{1i}(t)b_{1i}(t) + a_{2i}(t)b_{2i}(t)]}{\sqrt{\sum_{i=1}^{N} [a_{1i}^2(t) + a_{2i}^2(t)]}\sqrt{\sum_{i=1}^{N} [b_{1i}^2(t) + b_{2i}^2(t)]}},$$

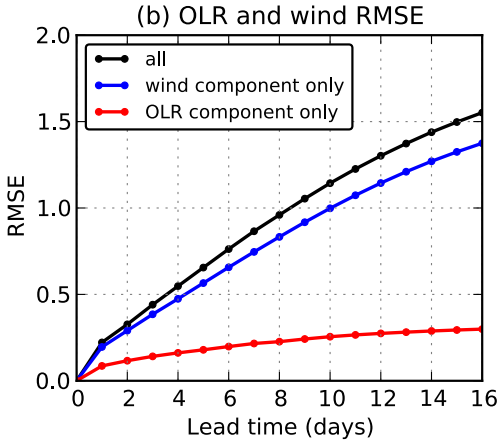
where $a_{1i}(t)$ and $a_{2i}(t)$ are the observed RMM1 and RMM2 at day t, and $b_{1i}(t)$ and $b_{2i}(t)$ are their respective forecasts, for the ith forecast with a τ-day lead. Here, N is the number of forecasts.

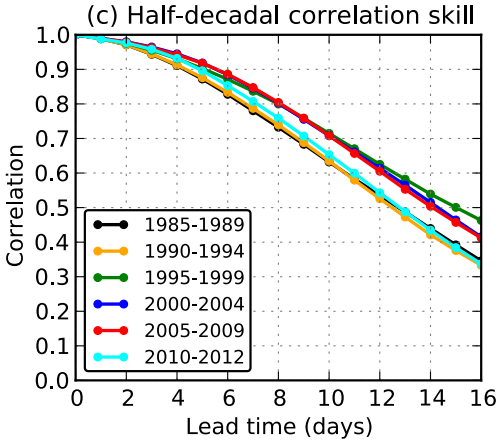
$\mathrm{COR}(\tau)$ measures the skill in forecasting the phase of the MJO, which is insensitive to amplitude errors. $\mathrm{COR}(\tau)$ is equivalent to a spatial pattern correlation between the observations and the forecasts when they are expressed by the two leading combined EOFs.

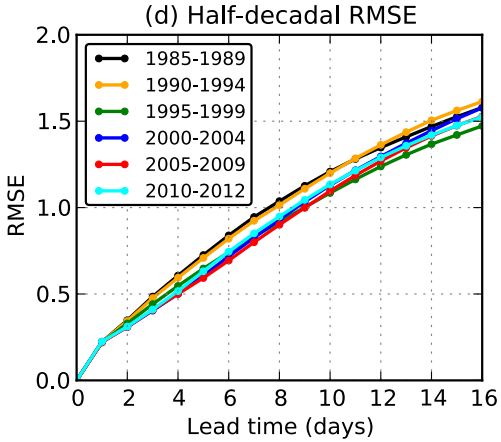
$$\mathrm{RMSE}(\tau) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \{[a_{1i}(t) - b_{1i}(t)]^2 + [a_{2i}(t) - b_{2i}(t)]^2\}}.$$

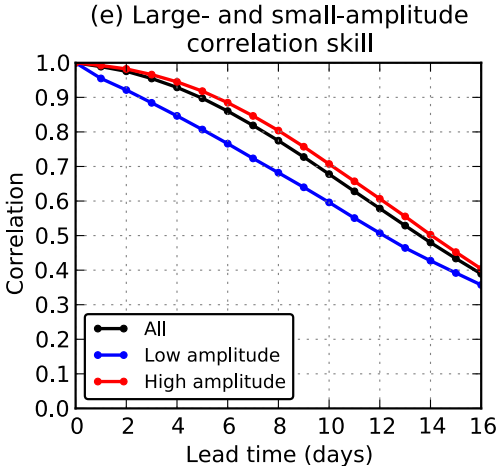
from Lin et al., Nov 2008 *MWR*.

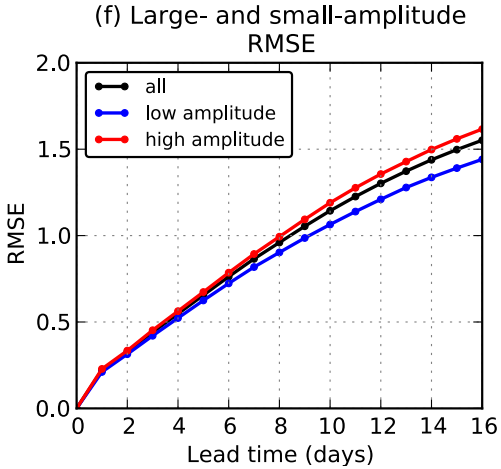








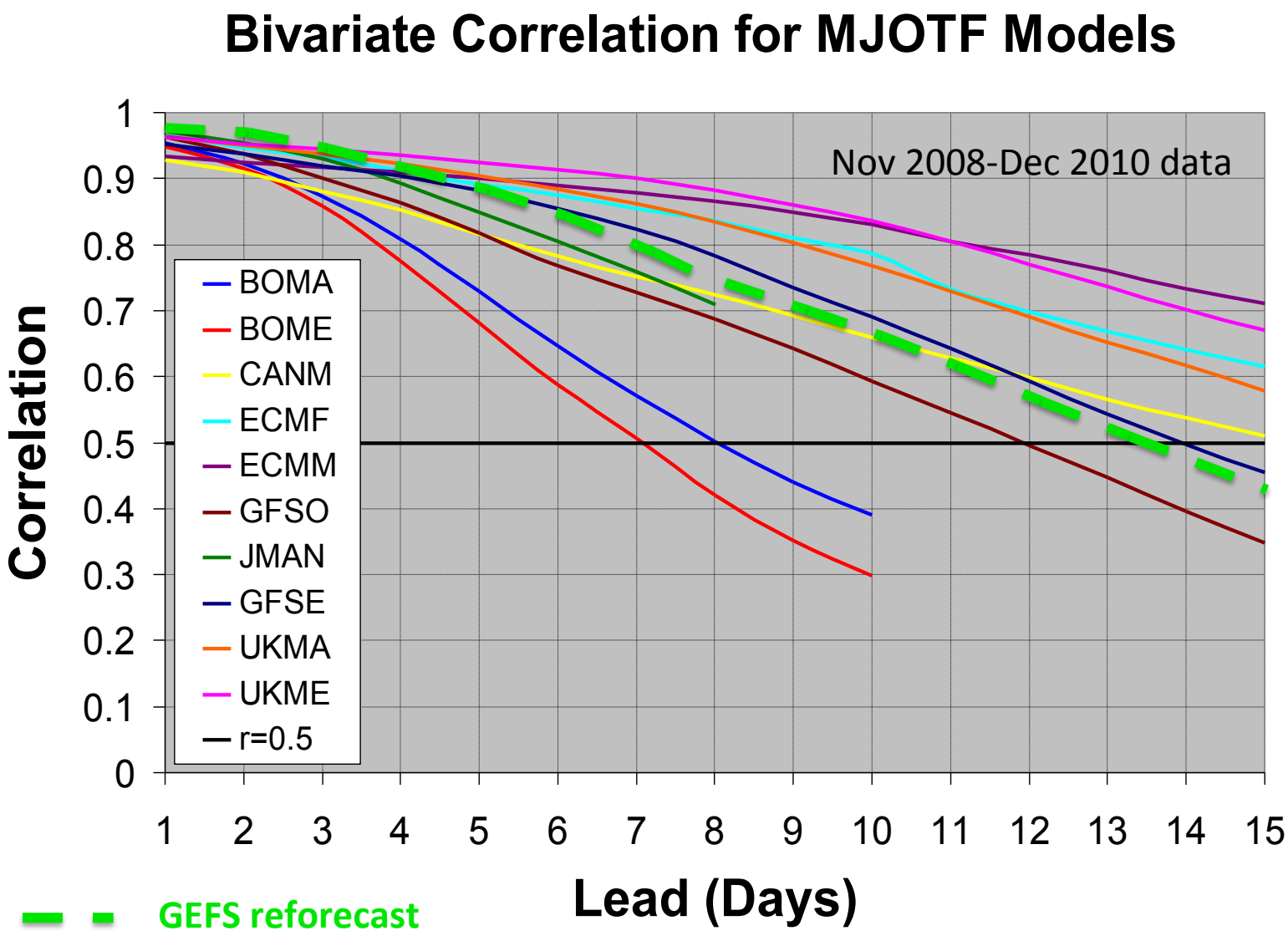




Bi-variate RMM1 and RMM2 correlation and RMSE

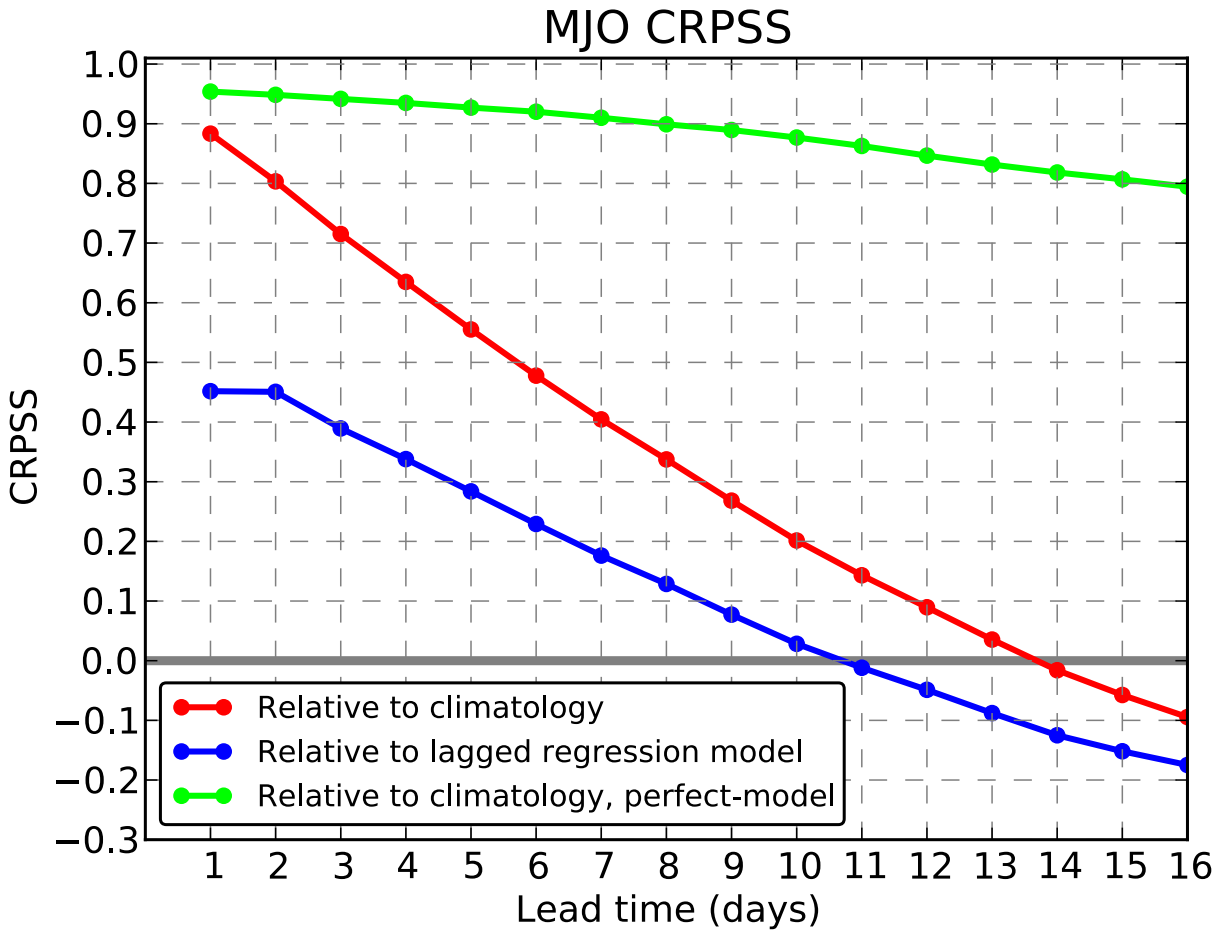
Low amplitude:
$\sqrt{(RMM1^2+RMM2^2)} < 1$
High amplitude:
$\sqrt{(RMM1^2+RMM2^2)} \geq 1$

Comparing against MJO task force data…

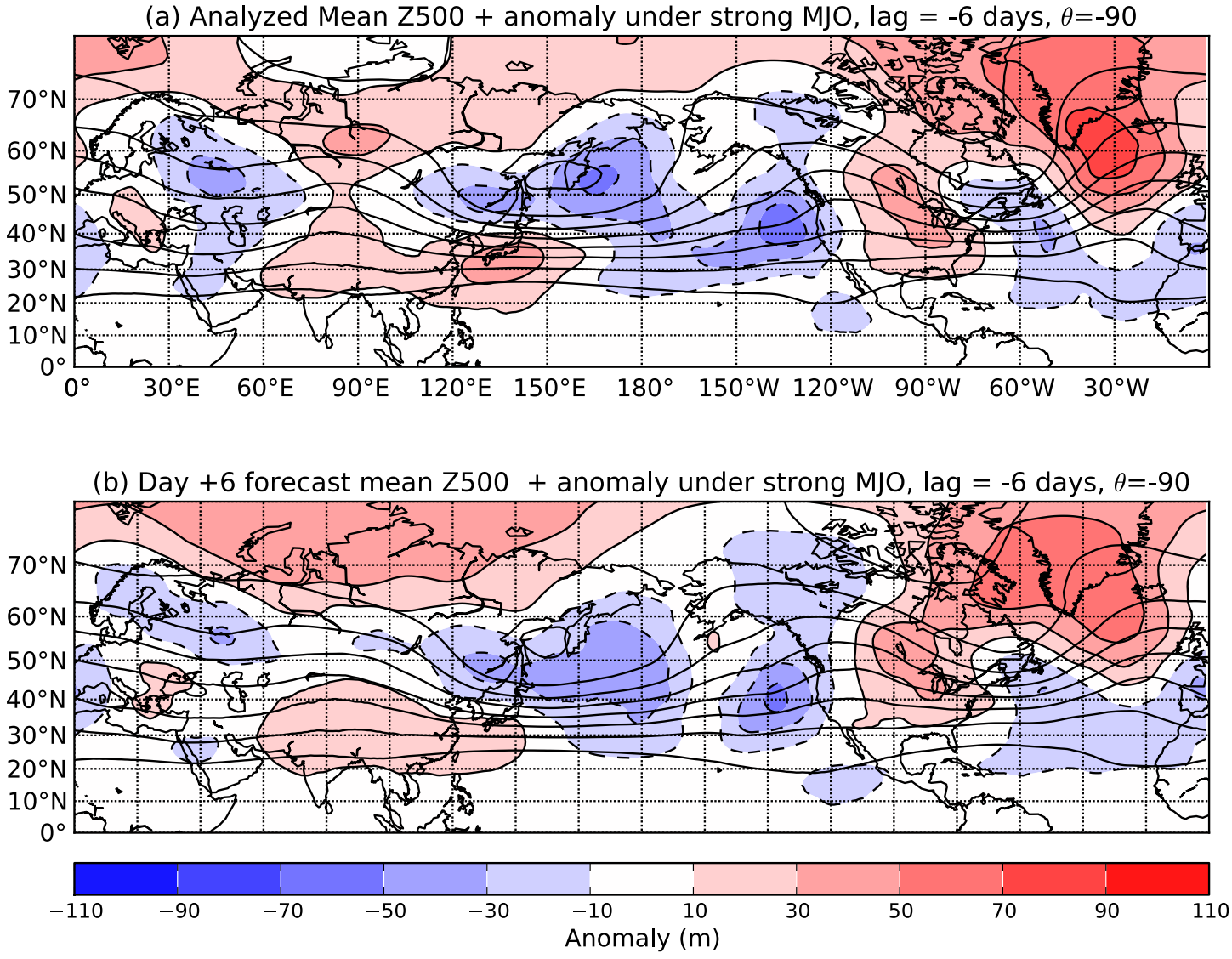


c/o Jon Gottschalk

Probabilistic forecast verification: CRPSS



Method for computing CRPSS discussed in supplementary slides. References are climatology and a regression model based on lagged persistence using day 0, -5, -10, -15, -20, and -25 RMM1 and RMM2 values.

The extremely high skill of the perfect-model scenario likely exaggerates the best case; as seen in rank histograms, there is undue consistency among ensemble members. This inflates perfect-model skill.

Z500 anomalies under strong (6-day lagged) Indian Ocean MJO