Ensemble prediction and post-processing team reports to NGGPS

Tom Hamill

ESRL, Physical Sciences Division

tom.hamill@noaa.gov (303) 497-3060

"In public exegencies, there is hardly anything more prejudicial than excessive caution, timidity, and dilatoriness, as there is nothing more beneficial than vigour, enterprise, and expedition."

~~Alexander Hamilton~~
Fred Toepfer

Proposed team members

Ensemble system development	Post-processing
Tom Hamill (ESRL/PSD)	Tom Hamill (ESRL/PSD)
Yuejian Zhu (EMC)	Kathy Gilbert (WPC)
Dave Novak (WPC)	Matthew Peroutka (MDL)
Carolyn Reynolds (NRL)	Jeff Craven (NWS/Central Region)
Malaquias Peña (EMC)	Trevor Alcott (ESRL/GSD)
Walt Kolczynski (EMC)	Dan Collins (CPC)
Isidora Jankov (ESRL/GSD)	Michael Scheuerer (ESRL/PSD)
Phil Pegion (ESRL/PSD)	Dan Hodyss (NRL)
Zoltan Toth (ESRL/GSD, pending)	Yuejian Zhu (EMC)
	Zoltan Toth (ESRL/GSD, pending)

Draft NGGPS implementation plan for ensemble development and post-processing here.

Part 1:
Ensemble prediction system development

Ensemble prediction system development: NGGPS major objectives

(1) Develop and implement improved methods for **initializing ensemble predictions** (see also the DA component of NGGPS plan), including the initialization of the coupled environmental state (ocean, atmosphere, land, sea ice, and so forth).

(2) Develop methods to accurately **quantify model uncertainty** in ensemble prediction systems.

(3) Develop ensemble prediction system improvements that will facilitate the generation of reliable and maximally skillful guidance to **lead times of + 30 days and beyond**.

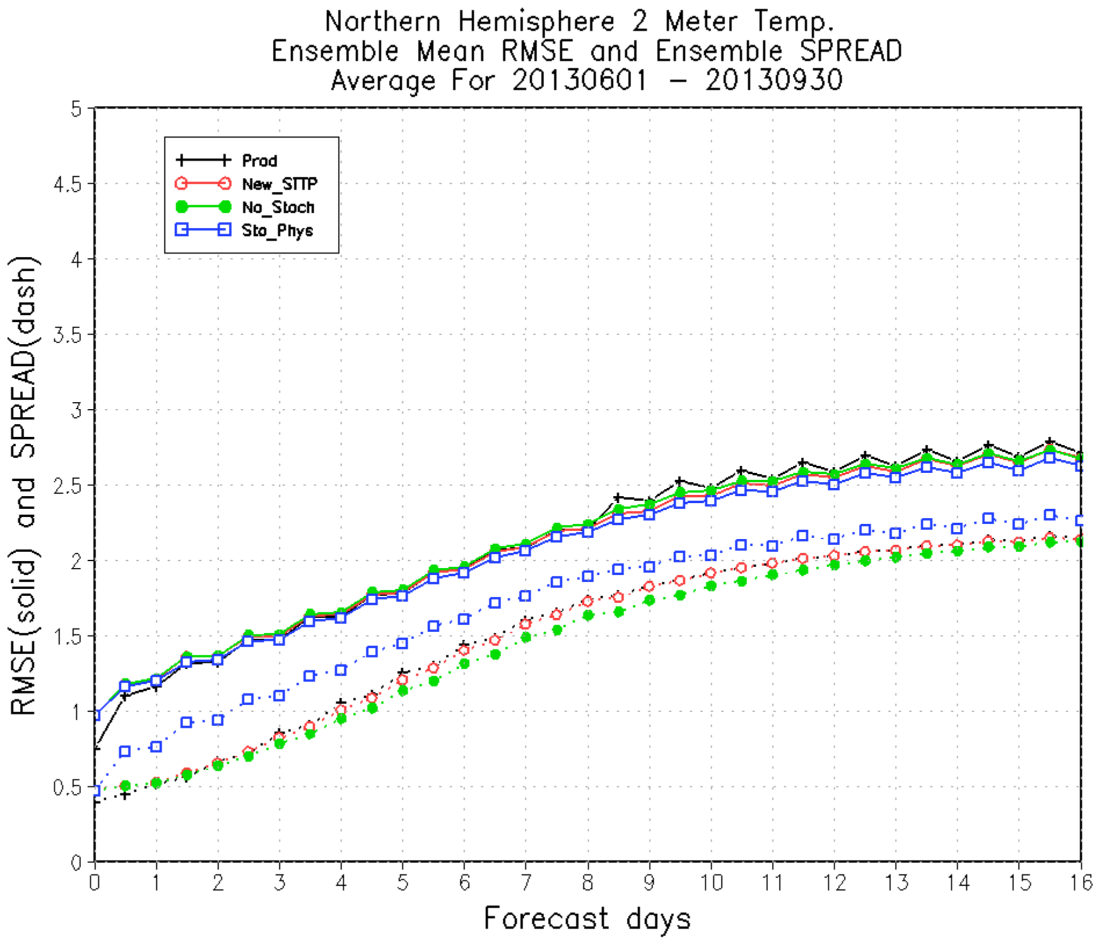
These all contribute to making more skillful and reliable probabilistic forecasts for high-impact weather at lead times of concern to NOAA and its customers.

Objective 1: initialization of ensembles

- 4D-En-Var or other related method is likely to be operational by early 2016, providing automatic *atmospheric* ensemble initialization. While more adjustments are contemplated, technology is **comparatively mature**, except:
- Methods for ocean, land, and sea-ice ensemble initialization.
  - Ensemble spread of T_{2m}, u,v @ 10m, precipitation are much more under-spread than, say, Z500.
  - Current GEFS doesn't provide realistic initial range of soil moisture
  - Initialization procedure should provide realistic covariances (e.g., no member with super-dry analyzed 2-m dewpoint over sopping-wet analyzed soil).
  - This technology is **adolescent**; some work at other centers to guide us.
- Position errors of coherent features (**immature**).
- Minimizing noise in analyzed state – this limits spread growth.
  - from small sample sizes.
  - from sub-optimal model uncertainty treatments.
  - **adolescent**; existing methods like Lynch filter sub-optimal.

Interactions with NGGPS data assimilation, land-surface, ocean teams expected.

Why is ocean, land, and sea-ice ensemble initialization a priority? In part because surface fields (and precipitation) are under-spread.



Even with modernized stochastic physics suite (discussion to come), surface temperature is under-spread, leading to unreliable probabilistic forecasts.

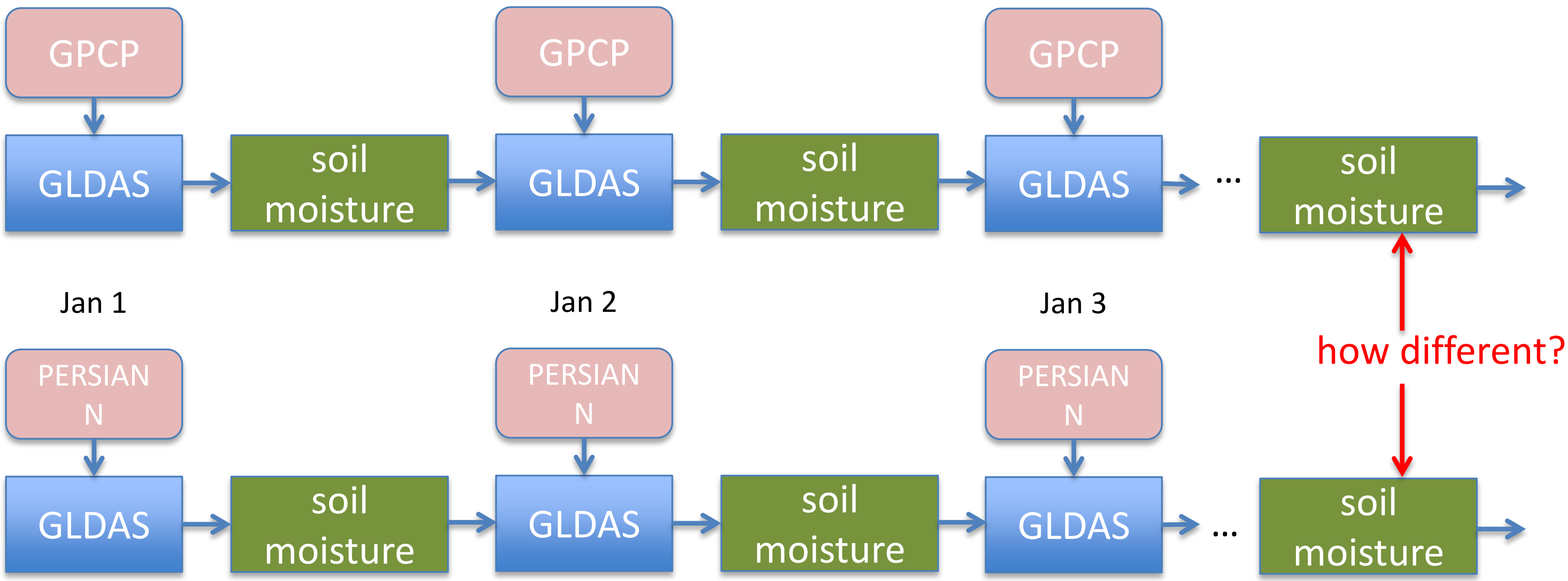
Are attempting to remedy this somewhat by:

(1) perturbations to soil moisture.
(2) land-surface parameter perturbation (discussed later).

figure c/o Walt Kolczynski, EMC.

Work in progress (via Sandy Supplemental): initialization of soil moisture.

- Determine what are realistic soil moisture perturbations by driving land-surface analyses with different precipitation data sets (cycled over many years).



The differences in soil moisture (variance, covariance, etc.) will be used to determine a reasonable perturbation methodology for initial soil moistures.

experiments by Maria Gehne, ESRL/PSD, with help from EMC land-surface team.

Ensemble prediction system objective 2: **treatment of model-related uncertainty**

- Ensemble prediction systems contain imperfections (e.g., finite resolution, sub-optimal numerics) and inappropriate deterministic assumptions (in parameterizations, model constants).
- These contribute to bias and limit spread growth of ensembles, resulting in biased, over-confident predictions.
- Methods for dealing with model uncertainty *in physically realistic ways* is **relatively immature** but greatly needed.

Envision collaborative work with NGGPS physics team.

Some approaches to dealing with model uncertainty

Approach	Benefits	Drawbacks
Multi-model / multi-parameterization ensemble	Shown many times over to increase spreads. "No-brainer" if sharing data between centers (NAEFS, National Blend of Models).	Members with different errors, biases. Difficult for one center to maintain and update a suite of models, parameterizations. Difficult to provide lengthy reforecasts.
"Simple" stochastic prediction methods (SPPT, SKEB)	Established, shown to provide benefit, comparatively easy	In some circumstances, not physically based. Can cause unexpected problems (e.g., SPPT introduces bias in climate simulations).
Physically based stochastic parameterizations	Ideally, one gets right answer (increased spread) *for the right reason*.	More in the realm of basic research; few methods are ready right now.
Post-processing	Major improvements in skill, reliability possible. Comparatively easy to implement.	Large reforecast and reanalysis data sets needed; doesn't fix underlying model problems.

Model uncertainty activities, Sandy Supplemental (SS) and NGGPS proposed

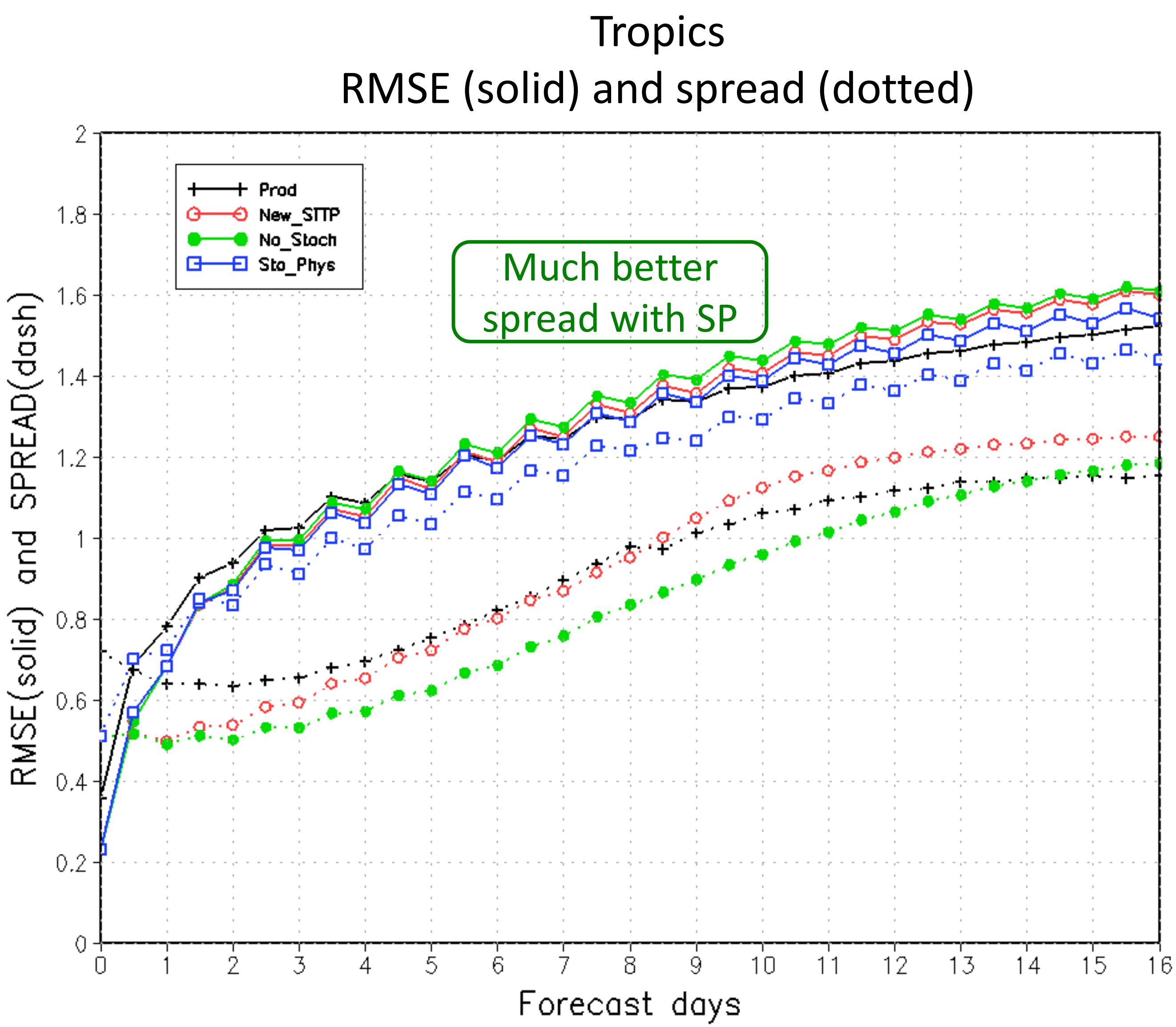
- Preparing an implementation of simple methods, many tested at other centers (SPPT, SKEB, SHUM; underway now via SS). Gratifying results. [**fairly mature**]
- Estimating parameter uncertainties associated with the land surface (underway now, not funded through to implementation, though). [**adolescent**] {see supplemental slides}.
- Stochastic parameterization (proposed).
  - Stochastic backscatter from convection (low-hanging fruit, demonstrated at UK Met Office). [**adolescent**]
  - Others harder, more basic research needed [**immature**].

Testing of simple and existing model uncertainty parameterizations

- Currently in GEFS: **STTP** (Stochastic Total Tendency Perturbations)
- Planned for next (≥ 2016) GEFS implementation (we're at knob-twiddling stage now):
  - **SPPT** (Stochastically Perturbed Physical Tendencies, from ECMWF)
    - tweaks by Phil Pegion to make precip. consistent with q tendency.
  - **SKEB** (Stochastic Kinetic-energy Backscatter, from ECMWF and Met Office)
  - **SHUM** (Stochastically perturbed boundary-layer RH, developed by Jeff Whitaker)

via USWRP, Sandy Supplemental, and HIWPP funds; c/o Phil Pegion (ESRL/PSD), Walt Kolczynski (EMC)

Summer results, 850-hPa temperature

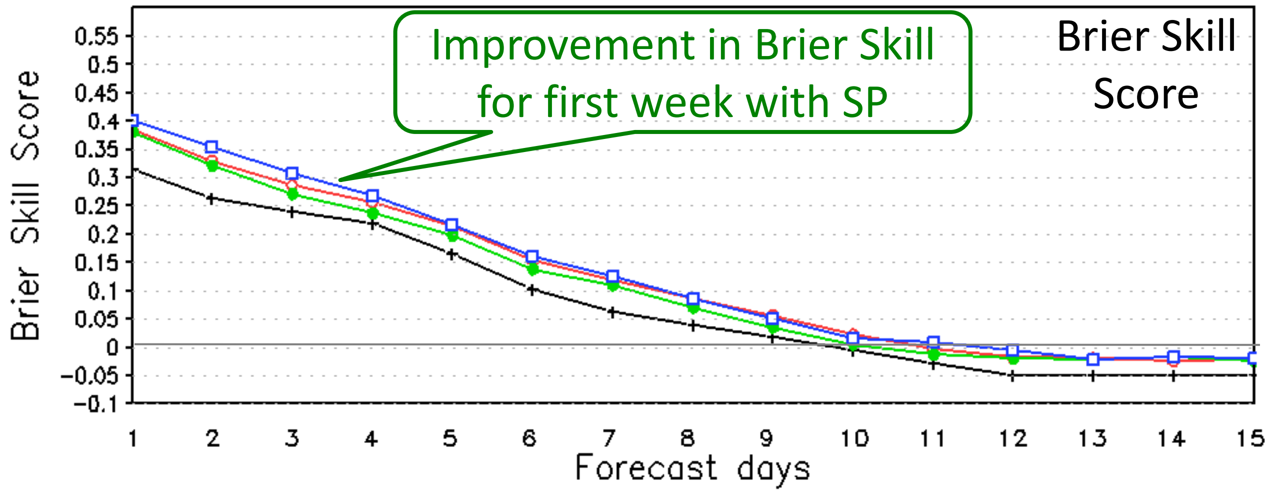


Production
New STTP
No Stoch
Sto. Phys
(SPPT, SKEB, SHUM)

(not all improvements are this impressive; figure c/o Walt Kolczynski, EMC).

Summer results, CONUS precipitation

verification against CCPA over CONUS on 1-degree grid.



> 1 mm threshold

Production

New STTP

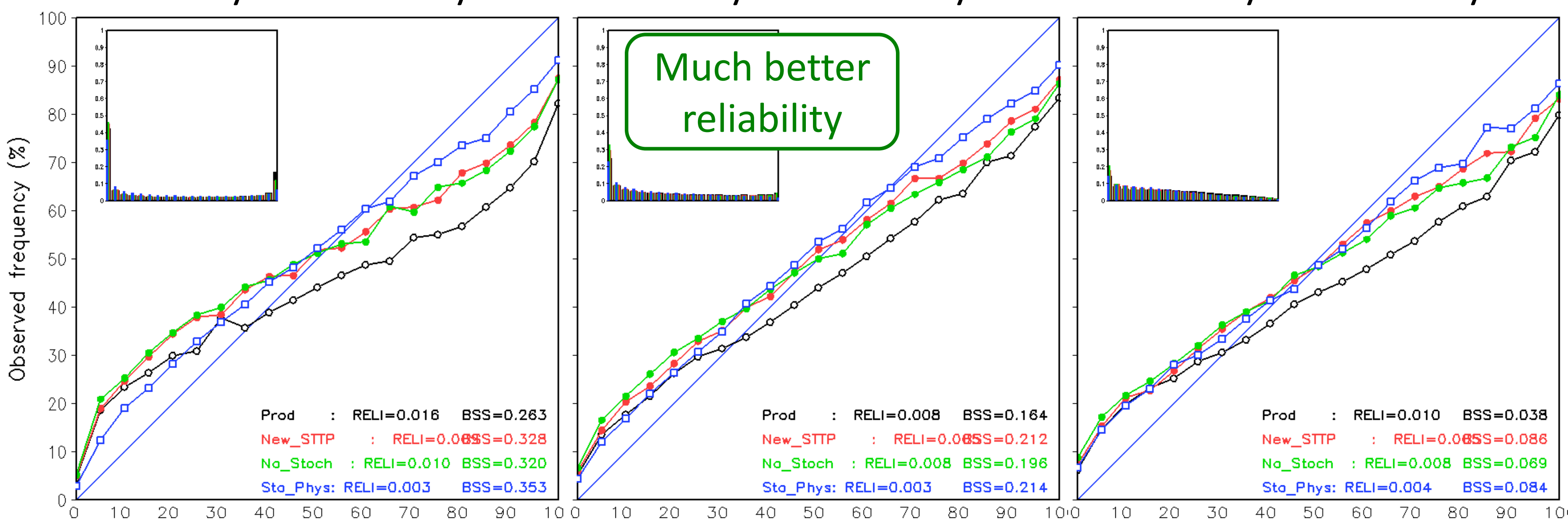
No Stoch

Sto Phys

Day +2 reliability

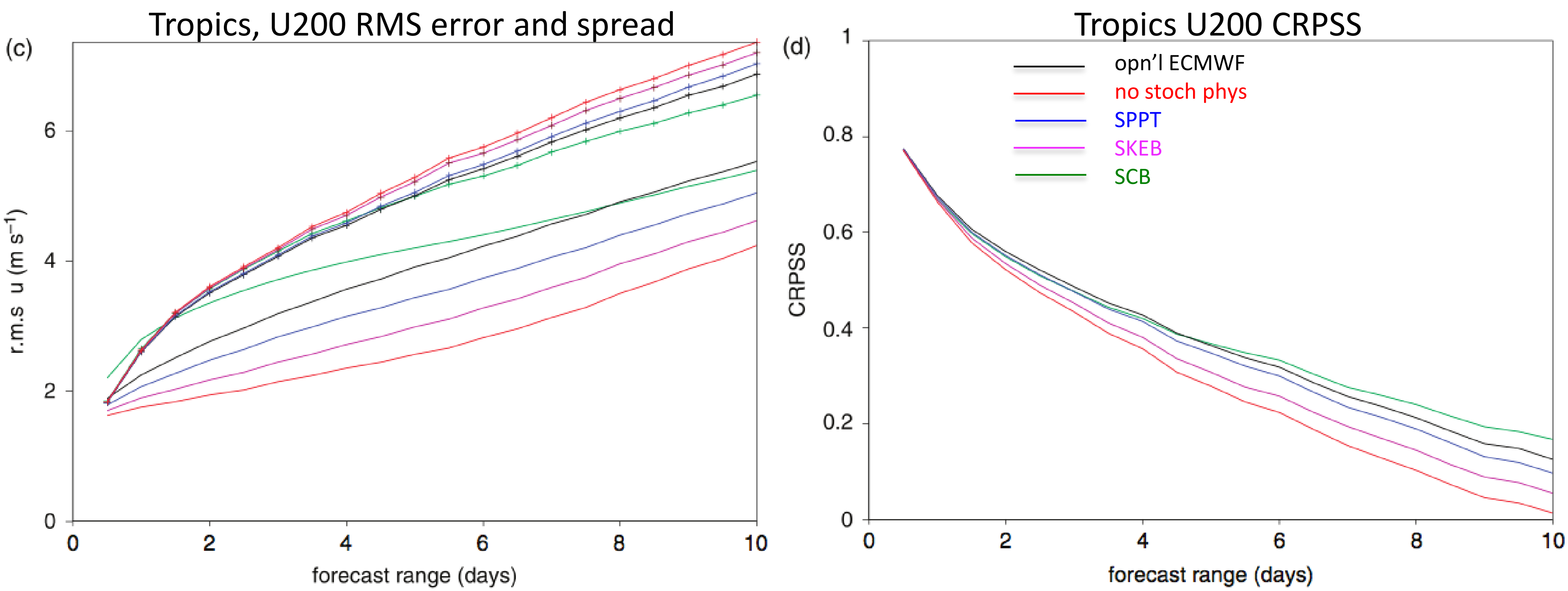
Day +5 reliability

Day +8 reliability



Physically based stochastic parameterization.

- Example of some “low-hanging fruit” in an otherwise basic-research area: recent Glenn Shutts (QJ, in press) stochastic backscatter for convection (SCB).

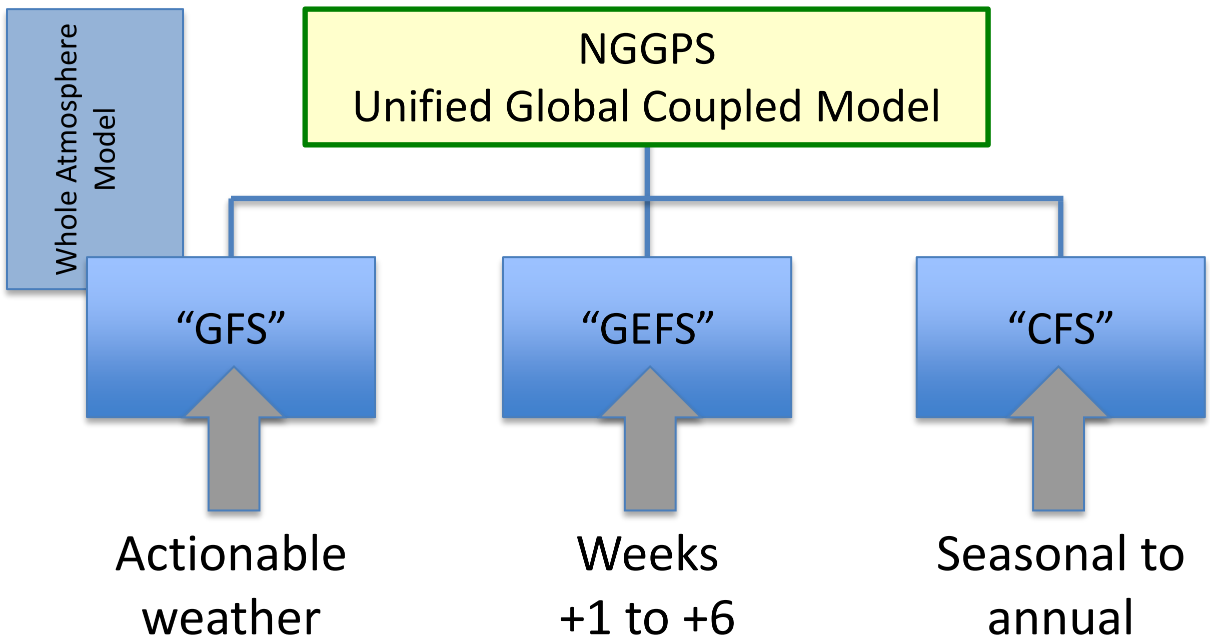


SCB increases spread, here at early leads, decreases RMS error, and improves probabilistic skill. Here SCB in isolation of other methods; more realistically, combined with them.

c/o Glenn Shutts: Ref: http://onlinelibrary.wiley.com/doi/10.1002/qj.2547/epdf

Ensemble prediction system objective 3: "GEFS" to infinity and beyond!

(well, at least to six weeks)



Update frequency	1 y	2 y	4 y
Length of Reanalysis	3 y	20-25 y	1979 - present
Cycles per day	4	1-4	TBD
Production machine	WCOSS	WCOSS	TBD

Extension of forecasts to week +6

- Experiment with GEFS system now, later in context of evolving unified coupled global model system
- Key science and technical questions
  - What sort of coupling is appropriate to +45 days lead? No direct ocean coupling, mixed layer, full coupling?
  - Develop methods for generating physically consistent atmosphere, ocean, land perturbations.
  - What configuration (ensemble size, resolution, reforecast duration/frequency) provides best use of available CPU resources?
  - Reanalysis / reforecast on WCOSS or other compute platform?
  - Does the prediction system faithfully model the (few) low-frequency modes of variability that may have predictable skill at 3-6 weeks?
    - MJO
    - Blocking / AO
    - ENSO

NGGPS external PI grants in 2015

- **Development and testing of a multi-model ensemble prediction system for sub-monthly forecasts.** Andrew W. Robertson, PI, Columbia University.
  - **Activities**: Develop and test a multi-model ensemble (MME) prediction system for sub-monthly forecasts (NCEP CFSv2, ECMWF and the Environment Canada model, and other models that become available).
- **Accelerating development of NOAA's next generation global coupled system for week-3 and week-4 weather prediction** Jim Kinter, PI George Mason University.
  - **Activities**: Conduct a series of model development and rigorous testing exercises designed to (1) correct systematic biases; (2) quantify the predictability and skill of weather forecasts for weeks 3-4.
- **An investigation of the skill of week-two extreme temperature and precipitation forecasts at the NCEP WPC.** Lance Bosart, PI, University at Albany, SUNY,
  - **Activities**: Evaluate newly proposed percentile forecast methods, persistent flow anomalies, and NH climate database in context of WPC's development of new forecast formats for Days 8-10. These forecast formats and methodologies for identifying EWEs will be tested in the WPC Hydrometeorological Testbed, and then will be implemented into WPC operations.
- **Exploitation of Ensemble Prediction System Information in support of Atlantic Tropical Cyclogenesis Prediction.** Chris Thorncroft, Pi, University at Albany, SUNY.
  - **Activities**: To ensure that recent and current research concerned with the variability of African easterly waves (AEW) structures and downstream tropical cyclogenesis probability is transferred into operational decision-making at NHC, and to develop and evaluate tools that exploit key information in dynamical ensemble prediction systems in support of tropical cyclogenesis prediction.

Management issues

- HPC and storage for research and development
- Is the NGGPS priority to fund:
  - Development of next-generation system, or
  - Incremental improvements to this generation's system, or
  - Both? In what mix?
- Related, what mix of high-risk/high-reward vs. low-risk/low reward in portfolio?
  - (Please don't expect low-risk/high-reward. If such things existed, we'd be doing them already).

Part 2:
post-processing.

NGGPS post-processing objectives

- Conduct post-processing "summit" [**mature**]
- Regularly generate supporting data sets, reanalysis/reforecast [**technology mostly mature; human, software, hardware infrastructure immature**].
  - necessary to support the advanced postprocessing development.
  - should include high-resolution reanalyses from a markedly improved RTMA or similar system.
- Improve post-processing algorithms for National Blend [**adolescent**]
  - Improve the post-processing and blending methods, allowing them to fully exploit the information in the improved ensembles (implicitly probabilistic)
  - Extend the post-processing and blending methods to include extra high-impact forecast variables and a wider range of forecast lead times.
- Develop post-processing techniques specific to the forecast problems of longer-lead forecasts (weeks 2-4) [**adolescent**].

For more, see recent white paper on major NWS changes in post-processing, here.

NGGPS objective #1: “Summit”

- **Major change #1**: Reanalysis/reforecast to be periodically generated; implies major changes to how we do post-processing. Need to sort these out.
- **Major change #2**: Post-processing to become increasingly oriented around National Blend and be probabilistic.
- **Build roadmap**: How can we organize and collaborate better?
  - Dispersed post-processing R&D now (MDL, EMC, ESRL, NSSL/SPC, AOML/NHC, WPC, CPC, etc.). Re-evaluate dispersed model, brainstorm ways of working together more effectively.
  - What is needed for more rapid technology transfer from OAR, NWS regions, academic sector?
  - What supporting tools/infrastructure needed? Libraries of common post-processing software, verification data sets, verification methods, etc.?
  - Lay out roadmap to manage the processes for these major changes.

NGGPS objective #2: regularly generate supporting data sets, reanalysis/reforecast (R/R).

- Global reanalysis procedures are **mature**, with exception of handling changes in observing systems.
- Infrastructure **immature**
  - Dedicated compute cycles;
  - Disk / cloud storage;
  - Re-usable, extendable observations database;
  - Diagnostic tools.
- **Procedures for freezing models immature** and runs contrary to EMC's past practices.
  - Especially difficult at times of major change.
- Hi-res surface reanalysis **adolescent** (two slides hence)

collaboration with NGGPS data assimilation group anticipated.

More on reanalysis/reforecast.

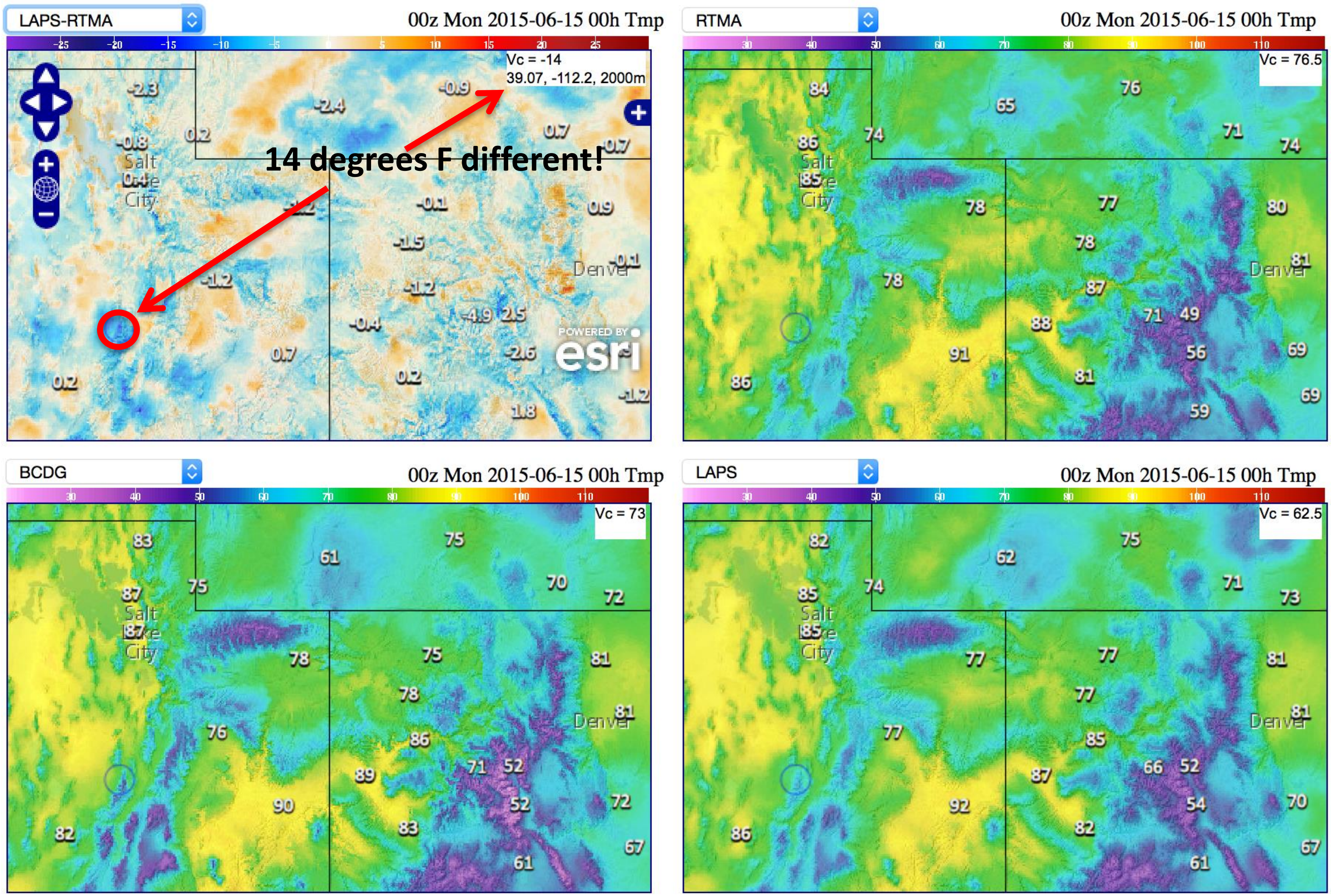
- Routine usage of reanalysis/reforecast (R/R) will dramatically improve post-processed guidance (see white paper).
- Hendrik Tolman at 2014 NCEP Production Suite review said EMC intends to move to a more ordered implementation and reanalysis/reforecast procedure:
  - **Seasonal**: ~ 4-year upgrade cycle, multi-decadal reanalysis/reforecast.
  - **Weekly**: ~ 2-year upgrade cycle, decadal reanalysis/reforecast.
  - **GFS or NGGPS replacement**: ~ 1-year upgrade cycle, a few years reanalysis/reforecast.
- Infrastructure (compute, storage, software, diagnostics) not in place yet. **Underlying technology to perform R/R is mostly mature, NWS infrastructure to do this regularly is immature**.
- Hence support to institutionalize reanalysis/reforecast/supporting infrastructure is a post-processing priority for NGGPS.
- Status: ESRL/PSD has funds to jump-start reanalysis system development, with anticipated additional funding for ESRL/PSD, EMC, CPC to support production. MDL seeking storage infrastructure funds.

High-resolution surface analysis /reanalysis (currently RTMA) challenges

- **Inherited biases from NWP model**: most DA systems correct model first guess to new observations.
  - If first guess is biased (usually is near surface), then analysis biased, especially in data-sparse regions.

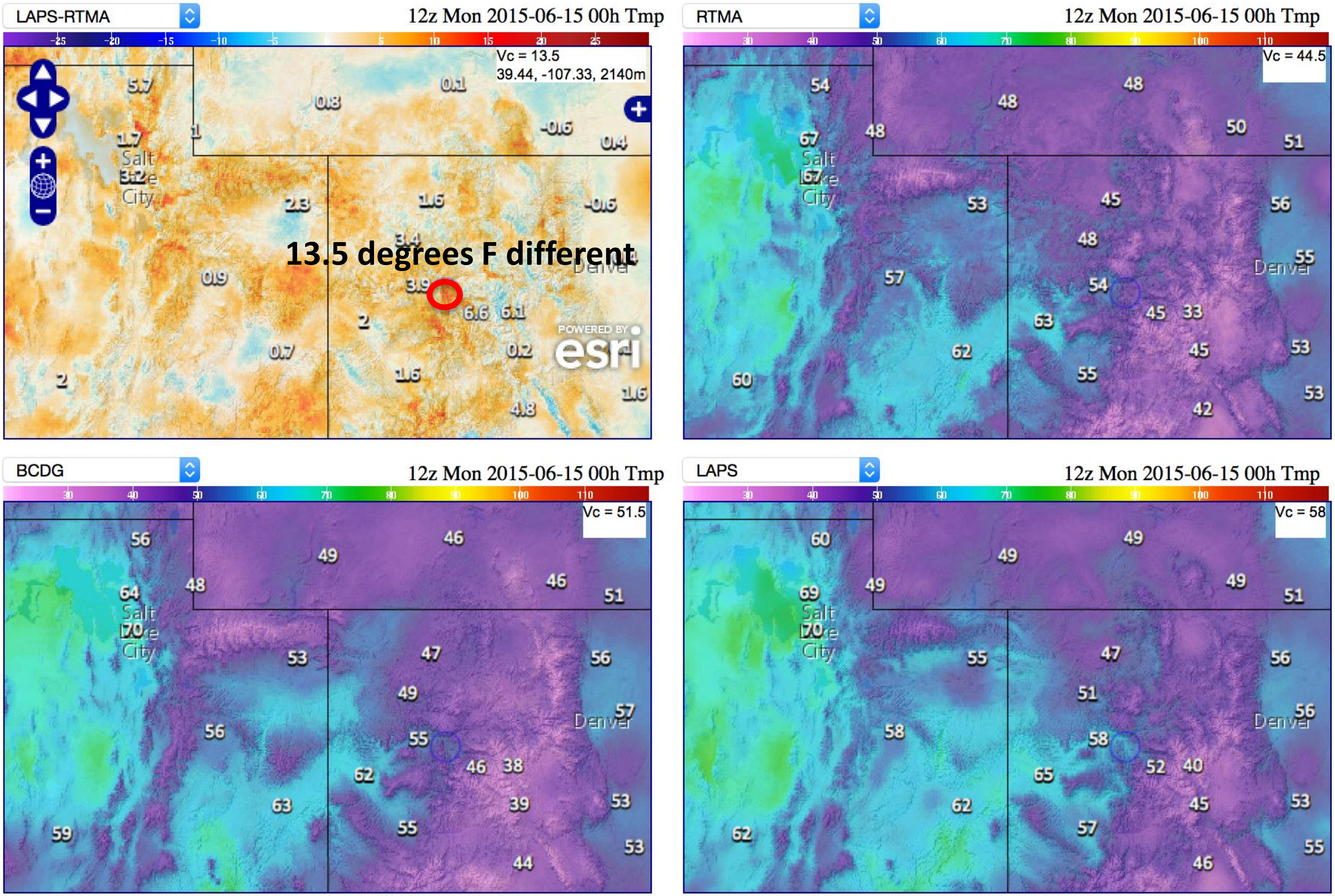
- **Procedures are relatively costly to run**; not sure EMC has set aside the time to run a retrospective RTMA back several decades.

Example of T_{2m} analysis differences

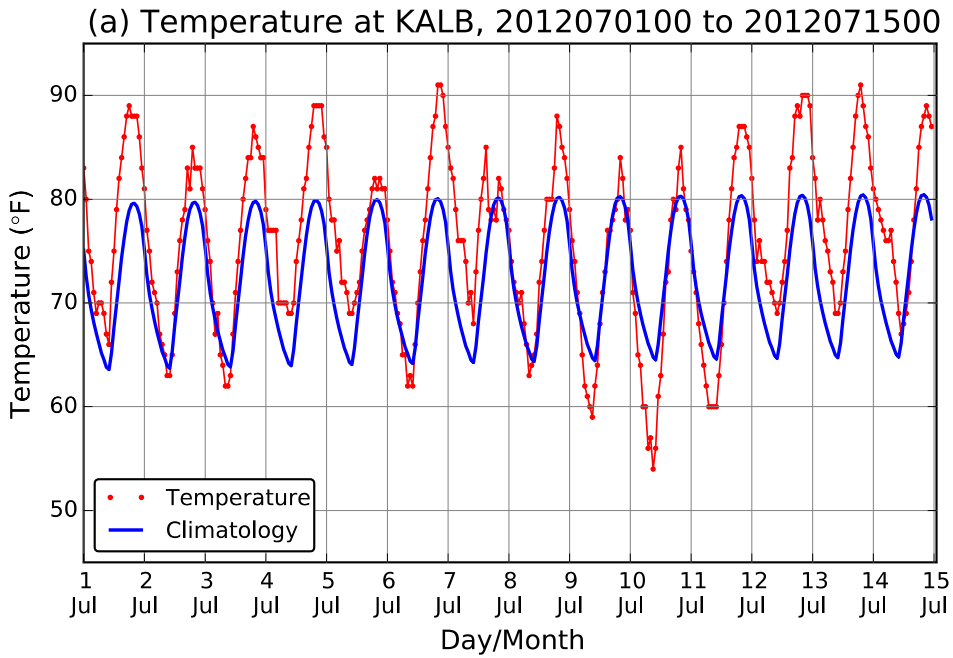


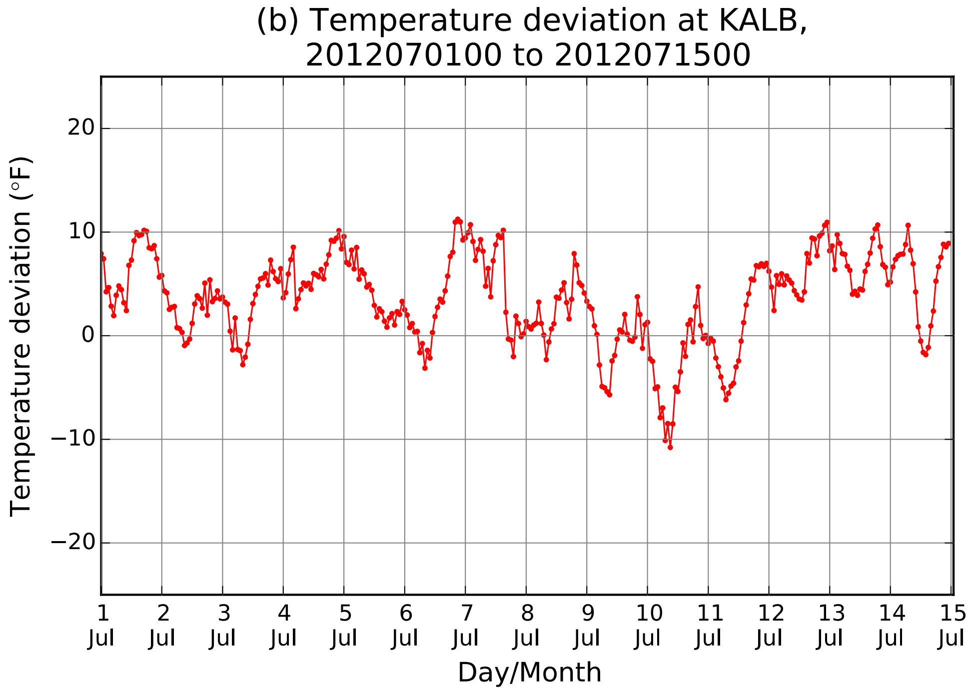
From (internal NOAA) http://www.mdl.nws.noaa.gov/~blend/blender.prototype.php

Another example of T_{2m} analysis differences



From (internal NOAA) http://www.mdl.nws.noaa.gov/~blend/blender.prototype.php





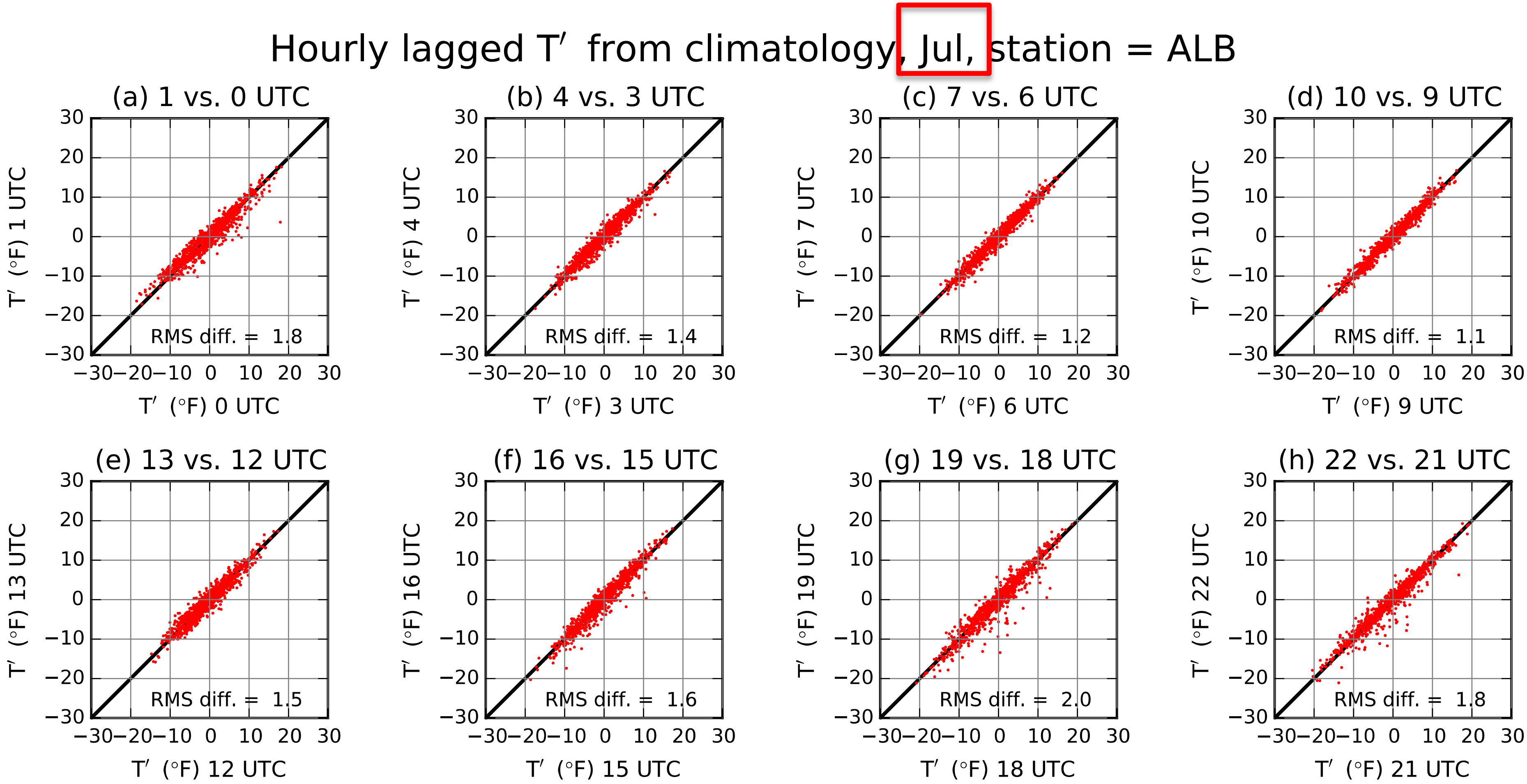
Time series of observed temperatures and deviations at Albany, NY.

A new, old idea:

What about *statistical* model for the first guess, based on deviation from climatology? Cheap, potentially unbiased.

Here, note that last hour's temperature deviation from climatology is frequently a decent first guess approximation for this hour's temperature deviation (and thus temperature, by adding back climatology).

How does last hour's temperature deviation from climatology predict this hour's temperature deviation?



Before the analysis, I had developed a climatology for Albany for each hour of the day and each day of the year. Hence, 10 AM's Local climatological temp is a bit warmer than 9 AM's temp. Persistence of last hour's deviation is pretty good.

Proposed statistical model to generate first guess

- T’(forecast next hour) =

  T’(analyzed this hour) +

  b_1*(analyzed cloud cover) +

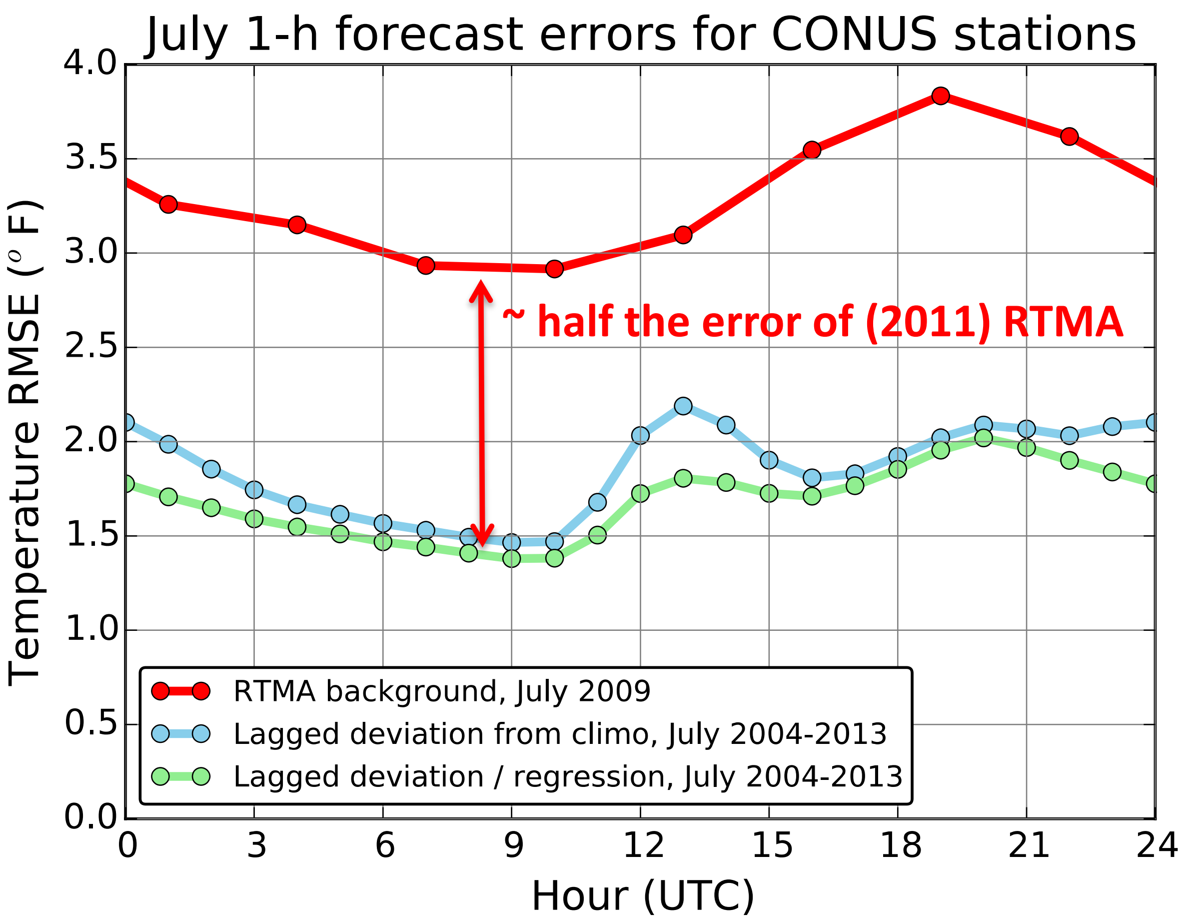
  b_2*(soil moisture) + … +

  b_{n-1}*(850 hPa vertical velocity)

  b_n*(forecast 925 hPa temperature change)

T’ is the deviation from the climatological mean *for this time of day, day of year.*

Statistical model first-guess errors *at stations* after regression



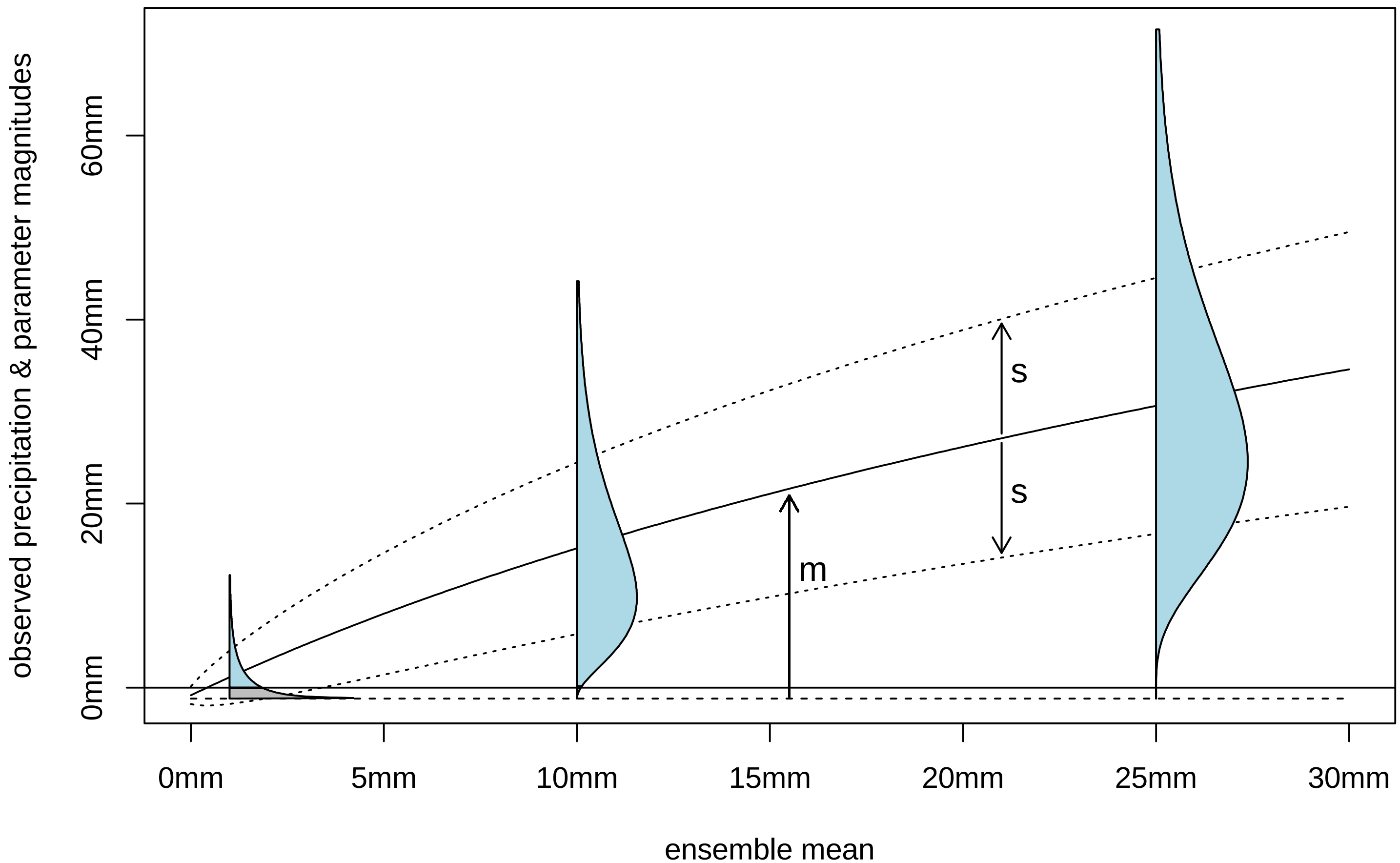
Lagged deviation from climo produces much lower error than RTMA **at stations**.

A small but noticeable further reduction of errors using predictors suggested in previous slides. (from de Pondeca et al., *W&F*, Oct 2011)

Results are not cross validated, but for a few selected stations, this appeared to have a negligible effect given the large training sample size (> 900).

Procedure for producing a gridded a first guess in progress.

Objective #3: improving post-processing algorithms



Advanced techniques are in development with Sandy Supplemental and NGGPS funds; more support will be needed for unusual, high-impact variables. **Adolescent.**

c/o Michael Scheuerer, ESRL/PSD.

Objective #4:
post-processing of weeks +3 to +4

- Initial-condition skill mostly gone, except episodically:
  - ENSO, MJO, blocking/AO.
- Small detectable signal buried in large amount of chaotic error, model bias.
- Lengthy reforecasts, stable models needed to tease out what skill there is.
- New post-processing techniques may be needed, tailored to unique challenges of these time scales.
- **Adolescent.**

NGGPS grants in 2015

- **Development of Ensemble Forecast Approaches to Downscale, Calibrate and Verify Precipitation Forecasts.** Dave Novak, WPC, PI
  - **Activities**: Enhance the skill of high-resolution quantitative precipitation forecasts (QPF) for detection of high-impact events via downscaling, quantile mapping
- **Calibration and Evaluation of GEFS Ensemble Forecasts at Weeks 2-4.** Ping Li, PI, SUNY Stony Brook.
  - **Activities**: Decompose GEFS extended range into a limited number of principal components to calibrate with observations.
- **Probabilistic Forecasts of Precipitation Type and Snowfall Amounts based on Global Ensemble Forecasts.** Tom Hamill, ESRL/PSD.
  - **Activities**: Develop novel experimental post-processing methods for precipitation type and snowfall amount.
- **An Investigation of Reforecasting Applications for Next Generation Aviation Weather Prediction: An Initial Study of Cloud and Visibility Prediction.** Dr. David Bright, NOAA/NWS/NCEP Aviation Weather Center, PI.
  - **Activities**: Utilize NOAA's second-generation Global Ensemble Forecast (GEFS) reforecast dataset, and be the first aviation-based GEFS reforecast study to construct a model climatology and downscaled calibrated prediction of instrument meteorological conditions (IMC).
- **Improved Statistical Post-Processing with the Bayesian Processor of Ensemble (BPE).** Zoltan Toth, PI, NOAA/OAR/ESRL/GSD.
  - **Activities**: Develop scientifically based, comprehensive algorithms and software for use in unified NWS statistical post-processing operations to address both the calibration of prognostic variables and the derivation of additional user variables. Test and demonstrate the algorithms for the calibration of prognostic variables.

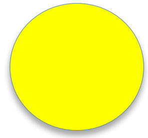
Post-processing management issues

- HPC and storage for research and development
- Big changes (R/R, National Blend) – is NWS prepared for them?
  - Dedicated compute cycles.
  - Process maturity; regular model implementations, frozen periods to conduct reforecasts.
  - Efficient storage/retrieval of the R/R and multi-model data sets needed.
  - Is post-processing truly regarded as part of the production cycle?
  - Leadership and coordination of this in NWS with diffuse activities; beyond part-time NGGPS committee.

Your feedback

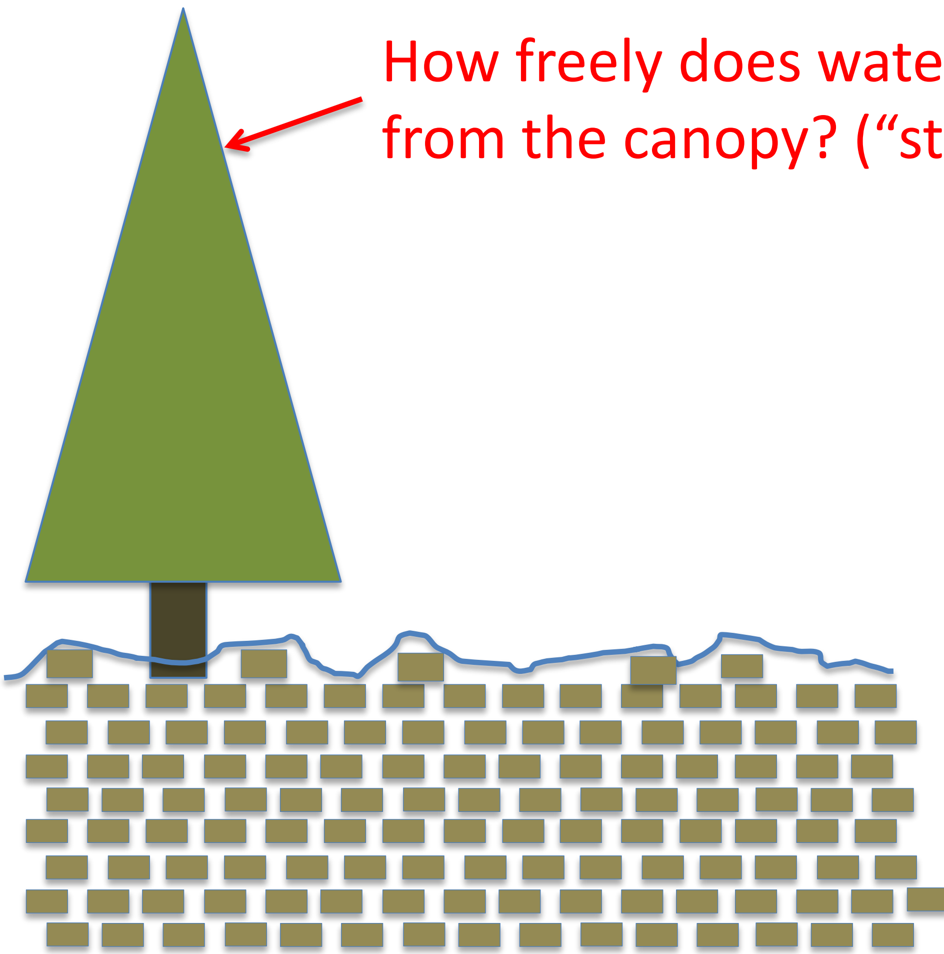
- Major objectives missed in ensemble or post-processing plans?
- Are we anticipating the right set of challenges?
- Are we consistent with NGGPS goals and other teams' plans?

Supplemental material

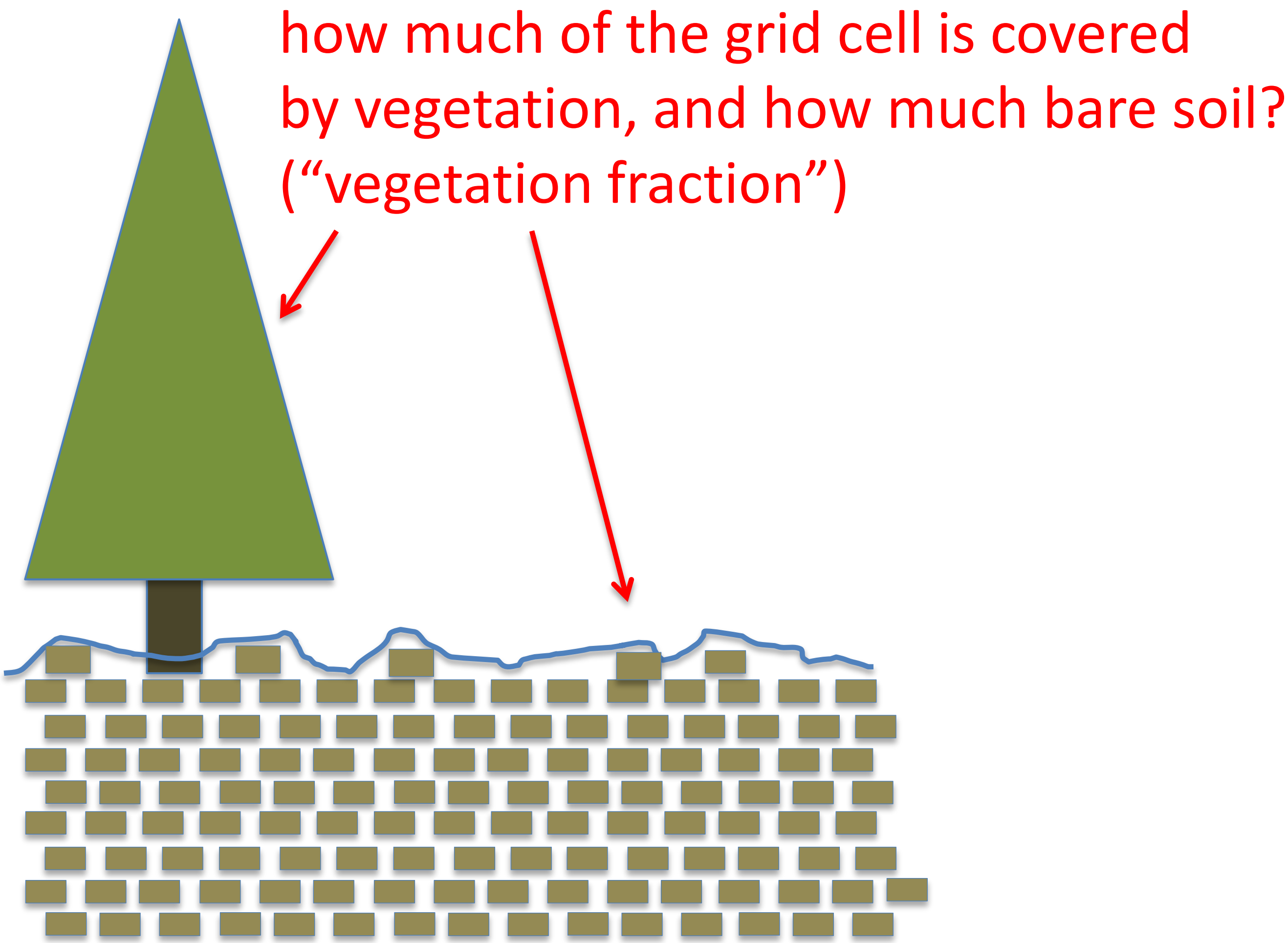


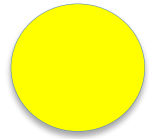
Land-surface parameter uncertainty also affects energy balance (and surface temp, moisture)

How freely does water evaporate from the canopy? (“stomatal resistance”)

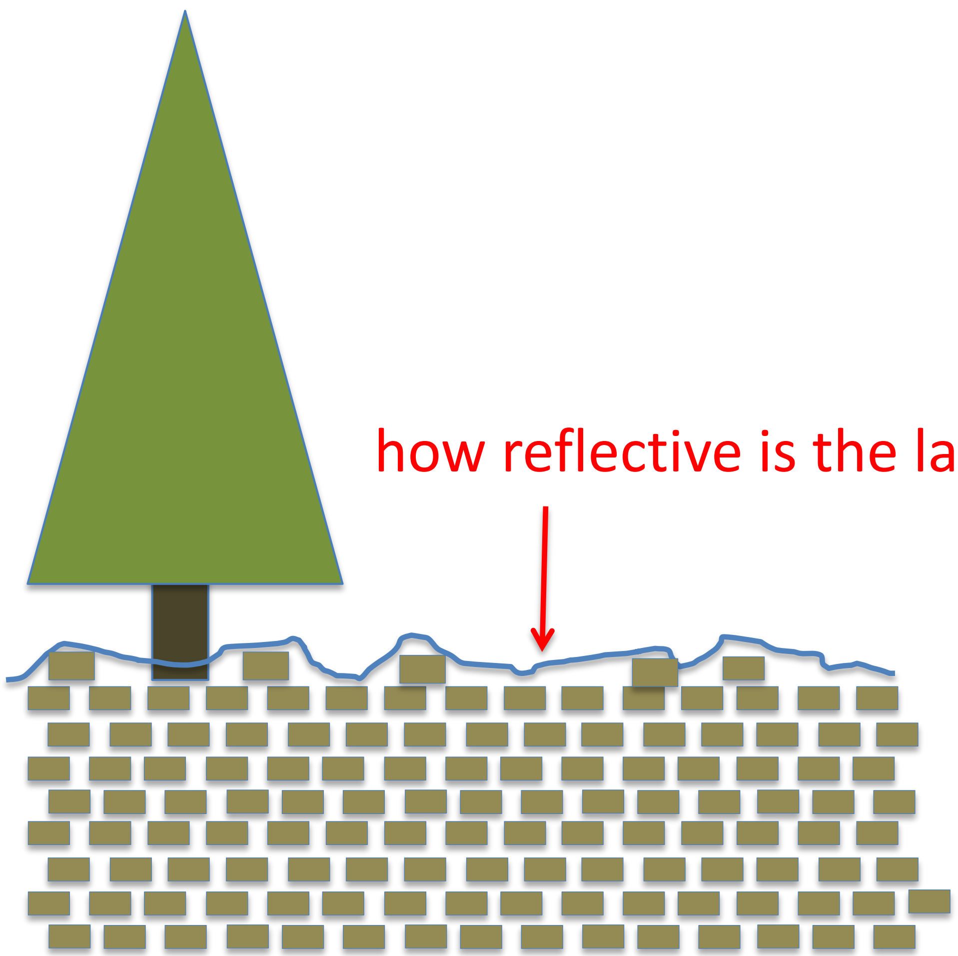


Land-surface parameter uncertainty also affects energy balance (and surface temp, moisture)

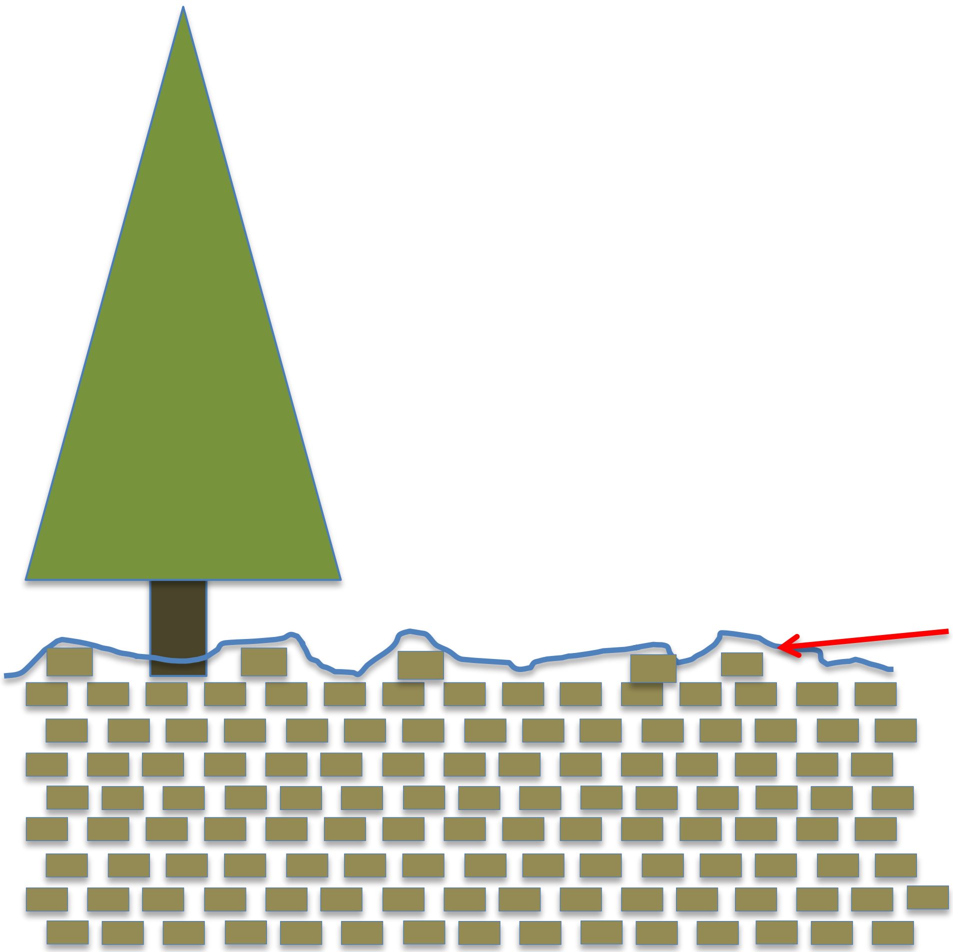




Land-surface parameter uncertainty also affects energy balance (and surface temp, moisture)

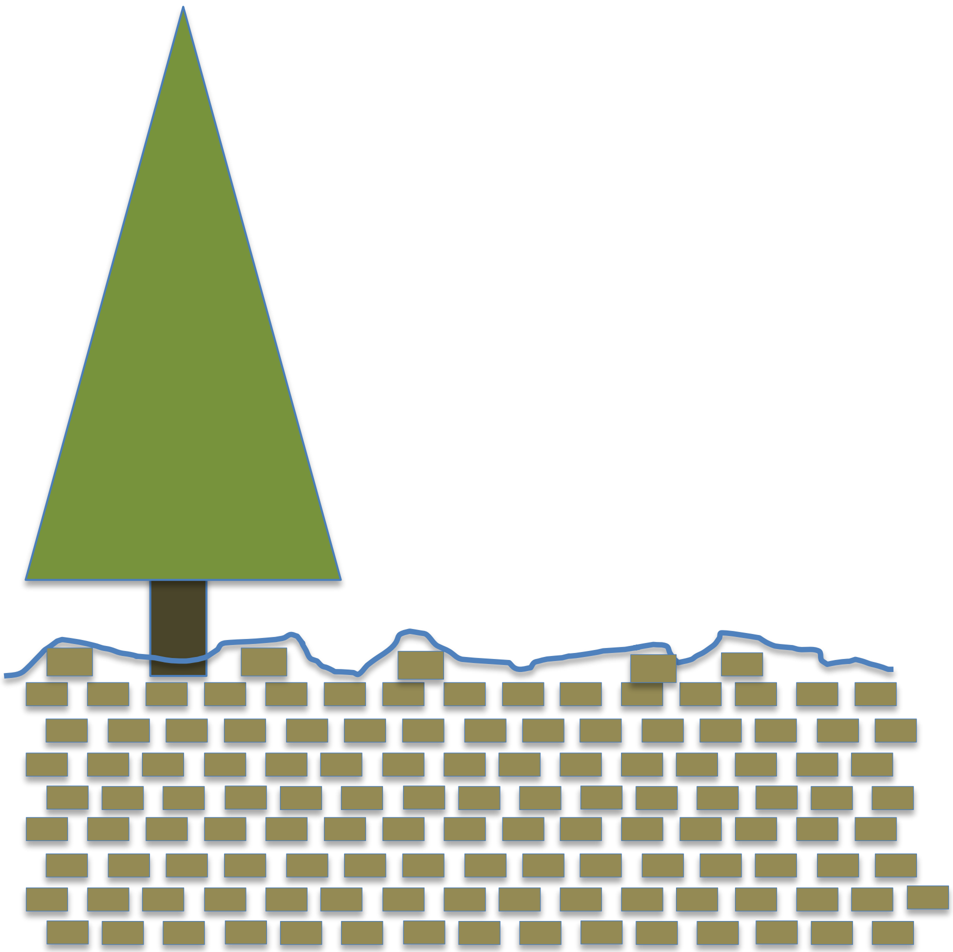


Land-surface parameter uncertainty also affects energy balance (and surface temp, moisture)



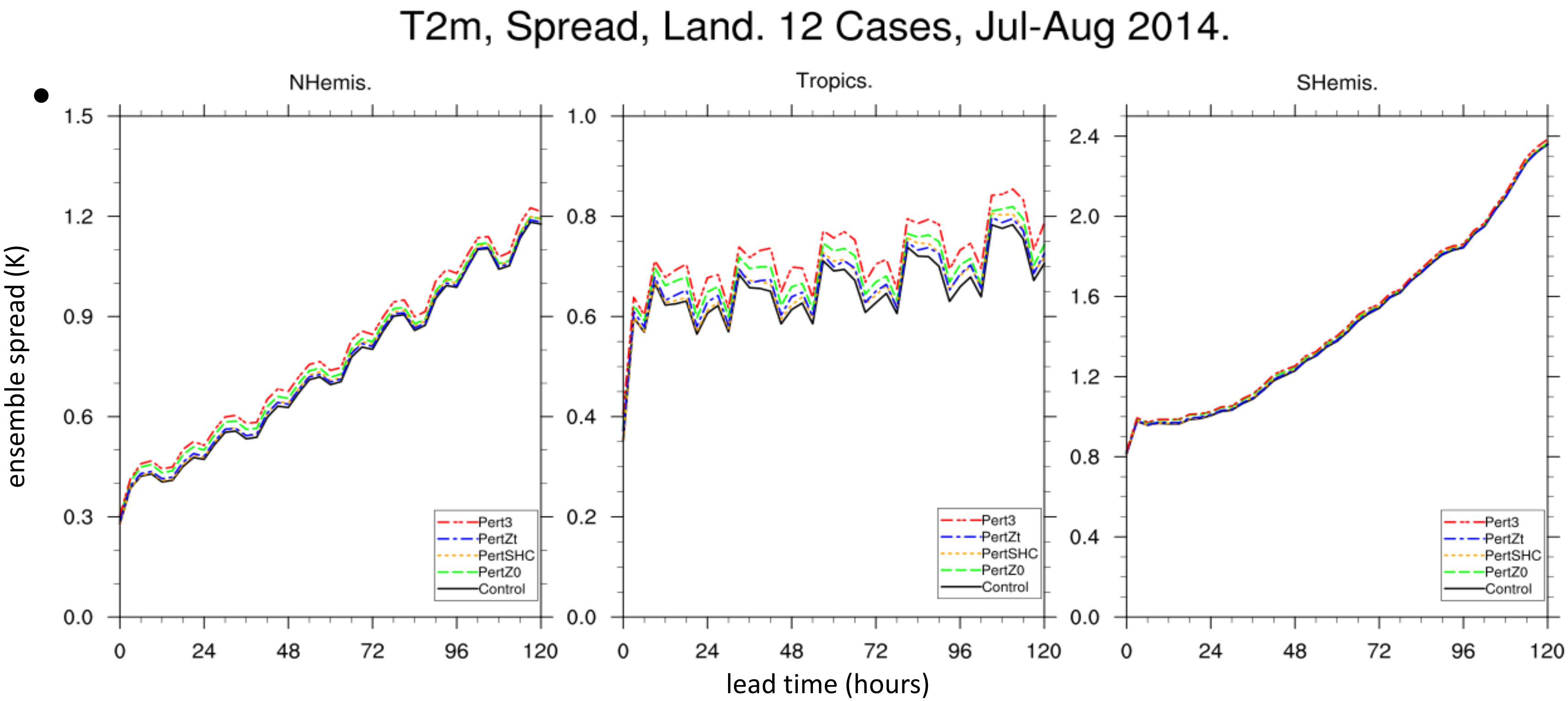
how rough is the surface?
(heat, momentum roughness lengths)

Land-surface parameter uncertainty also affects energy balance (and surface temp, moisture)



how permeable is the soil?
("hydraulic conductivity")

Work in progress (via Sandy Supplemental): initialization of land state

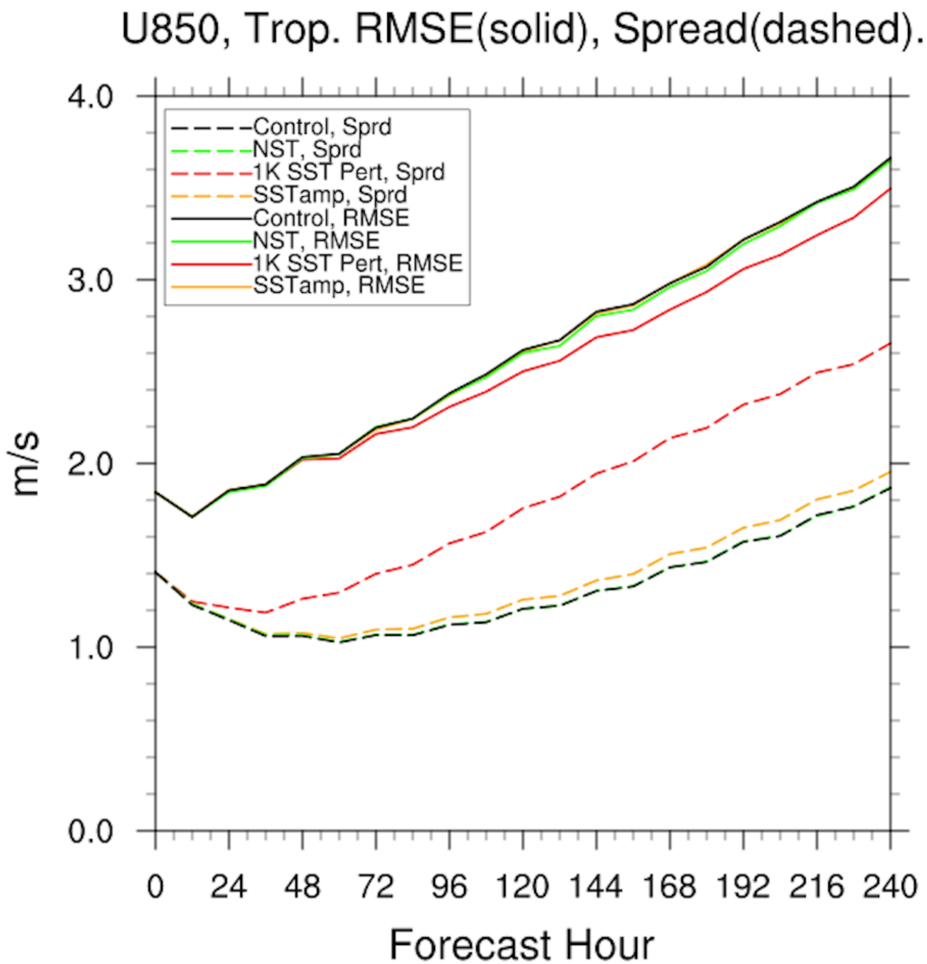


Some preliminary testing of perturbing soil hydraulic conductivity (SHC), momentum roughness length (z0), and heat/momentum roughness length ratios (zt) in reduced resolution GEFS (c/o Gary Bates, ESRL/PSD). Small increases in surface spread, more in tropics and summer hemisphere.

Would like to gather estimates of these parameters from various operational centers to set the uncertainty bounds. Also may explore perturbations to LAI, stomatal resistance, albedo.

Note: interesting paper on conceptual difficulties with LSMs: Best et al. June 2015, *J. Hydrometeor.*

Perturbing the ocean state and diurnal SST variation effect.



c/o Gary Bates, ESRL/PSD. See also recent Tennant and Beare, 2013 *QJRMS*, DOI: 10.1002/qj.2202. Also, McClay et al., 2012 *JGR*, DOI: 10.1029/2011JD016937

4 experiments:

1. Control (no SST perts)
2. NSST (ocean skin temps permitted to vary with weather, insolation)
3. O(1K) random perts applied over all oceans (larger than justifiable)
4. Ocean initial perts varying geographically based on estimated error; Std dev ~ 0.2-0.3 K except ~1 K in some Southern Ocean.

Other regions and variables, less impact.

Here, perturbations random and not designed to co-vary with atmospheric perturbations in a realistic way.

Major change # 2: “National Blend” product development

- National Blend concept:
  - Ingest multi-model, multi-center deterministic and ensemble forecasts.
  - Post-process and downscale them to 2.5-km NDFD grid.
  - Populate *all NDFD elements* with National Blend guidance as starting point for local WFO forecasters.
  - Possibly extend NDFD in future to include more probabilistic information.
- Anticipated result:
  - very high quality automated guidance.
  - less manual intervention by forecasters, thus greater consistency between WFOs.
  - forecasters ready and able to take on more decision-support roles.

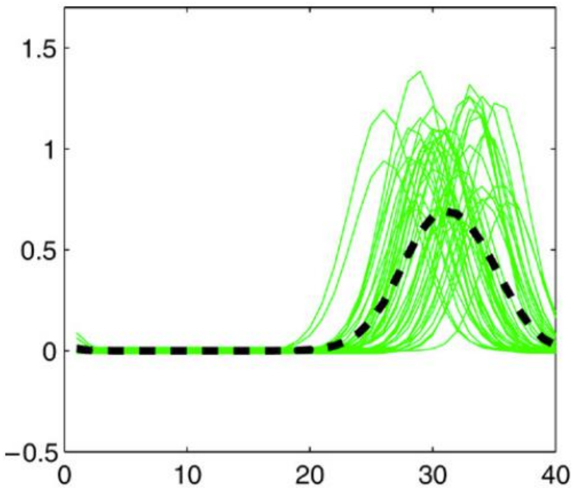
post-processing concepts adolescent, supporting infrastructure immature.

National Blend challenges.

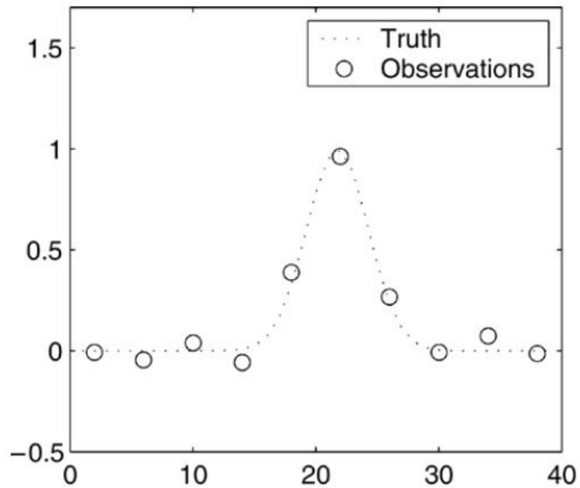
- Reaching data-sharing agreements with international partners.
- Ideally, need reforecasts to achieve consistently high-quality post-processed guidance.
- Need quality hi-res. surface reanalyses for training, validation (see next slide).
  - need them CONUS, AK, HI, Guam, PR.
- Developing new algorithms to exploit richness of ensemble data and potentially longer training data sets.
  - For all variables, including more difficult ones like snowfall amount, precipitation type, sky cover.
- Similar in concept to NAEFS; should they be integrated?

Position errors of coherent features

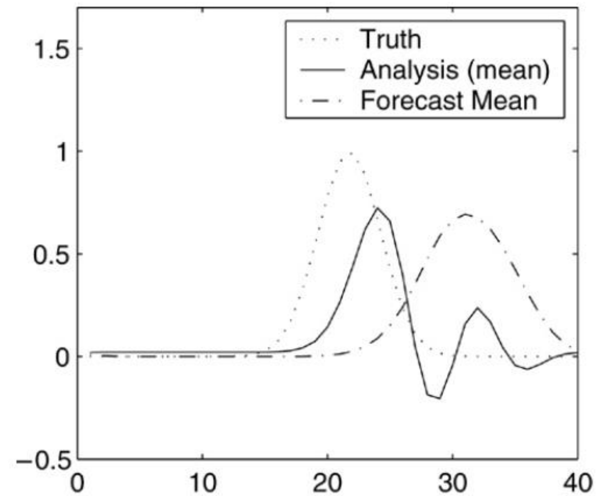
(“field alignment” or “feature calibration and alignment”)



(A) Forecast ensemble.



(B) Truth and observations.



(E) EnKF analysis.

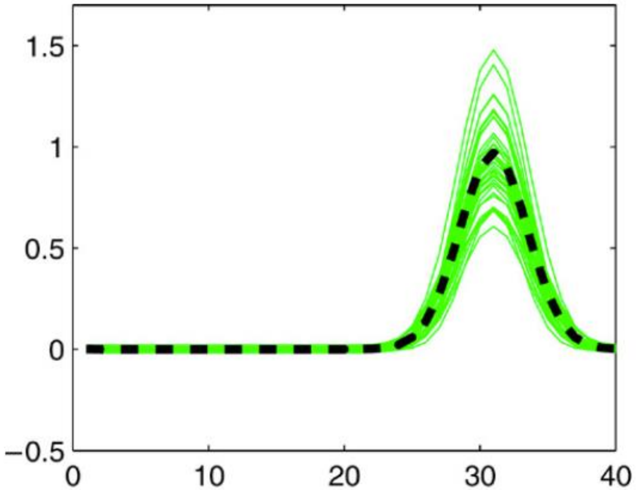
Above: first guess ensemble is mistaken about the position of a coherent feature. Resulting analysis has two small features, one at observation location, one at first-guess location; current *data assimilation not well set up to handle consistent position errors*.

We’ve seen this problem with tropical cyclones. Would like to get beyond more ad-hoc “vortex relocation.”

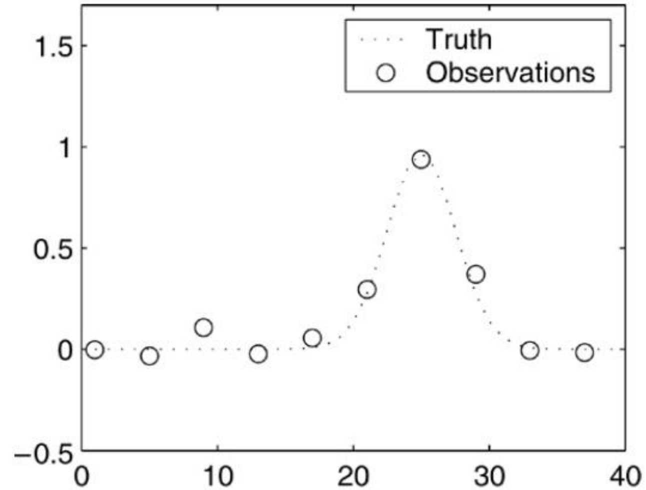
More rapid updates (hourly?) could ameliorate this, but there are other challenges to that, such as numerical noise with ensemble-based systems.

Ref: Ravela et al. 2007, http://dx.doi.org/10.1016/j.physd.2006.09.035. See also recent articles by Nehrkorn et al., MWR.

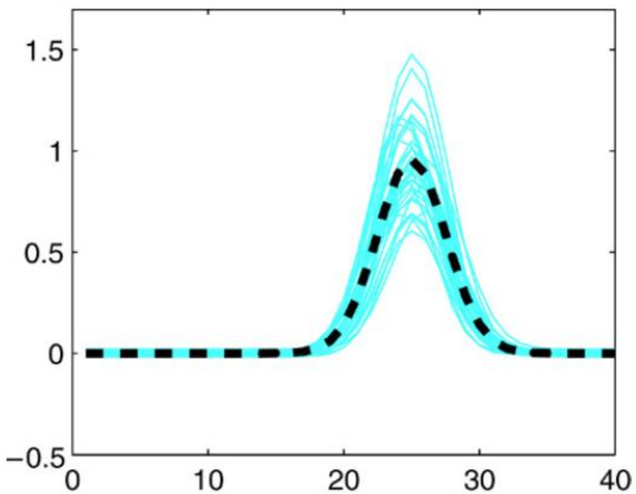
“Field alignment”



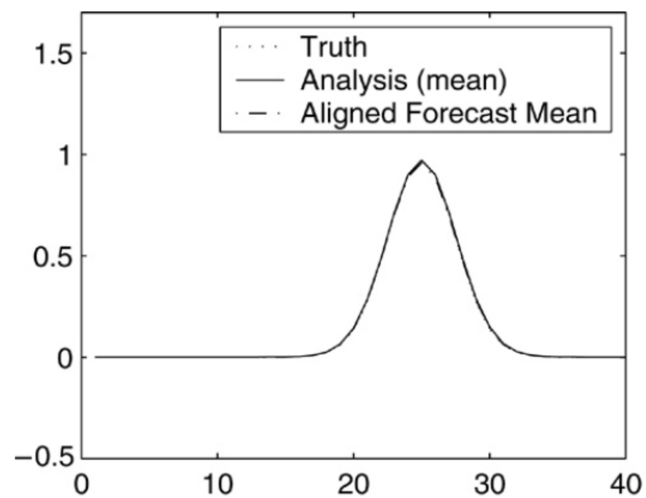
(A) Forecast ensemble.



(B) Truth and observations.



(C) Aligned forecast ensemble.



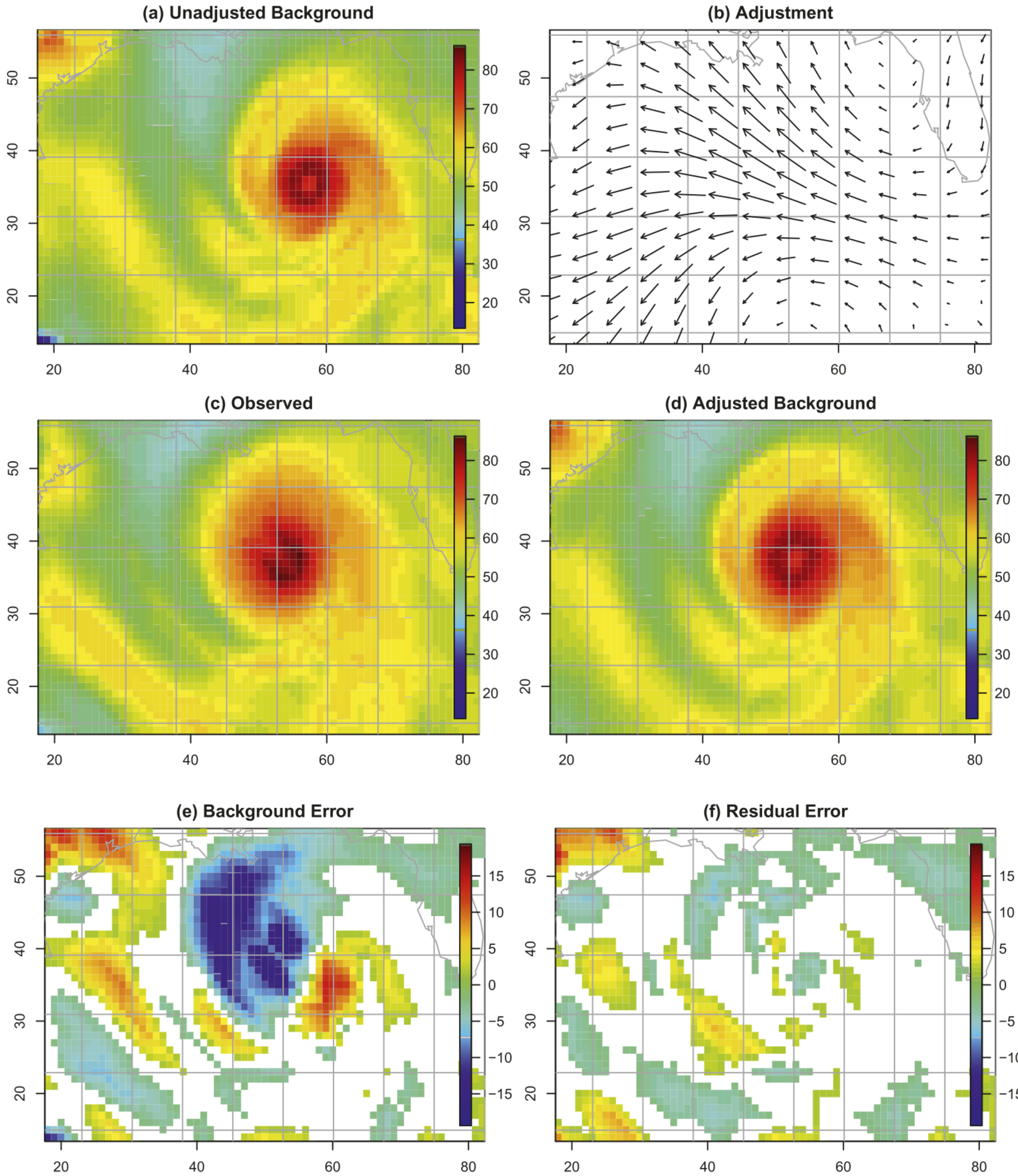
(E) EnKF analysis.

Under field alignment and related techniques, data assimilation is split into two steps:

(1) adjustment for position errors, and

(2) adjustment for amplitude errors.

Ref: ibid.



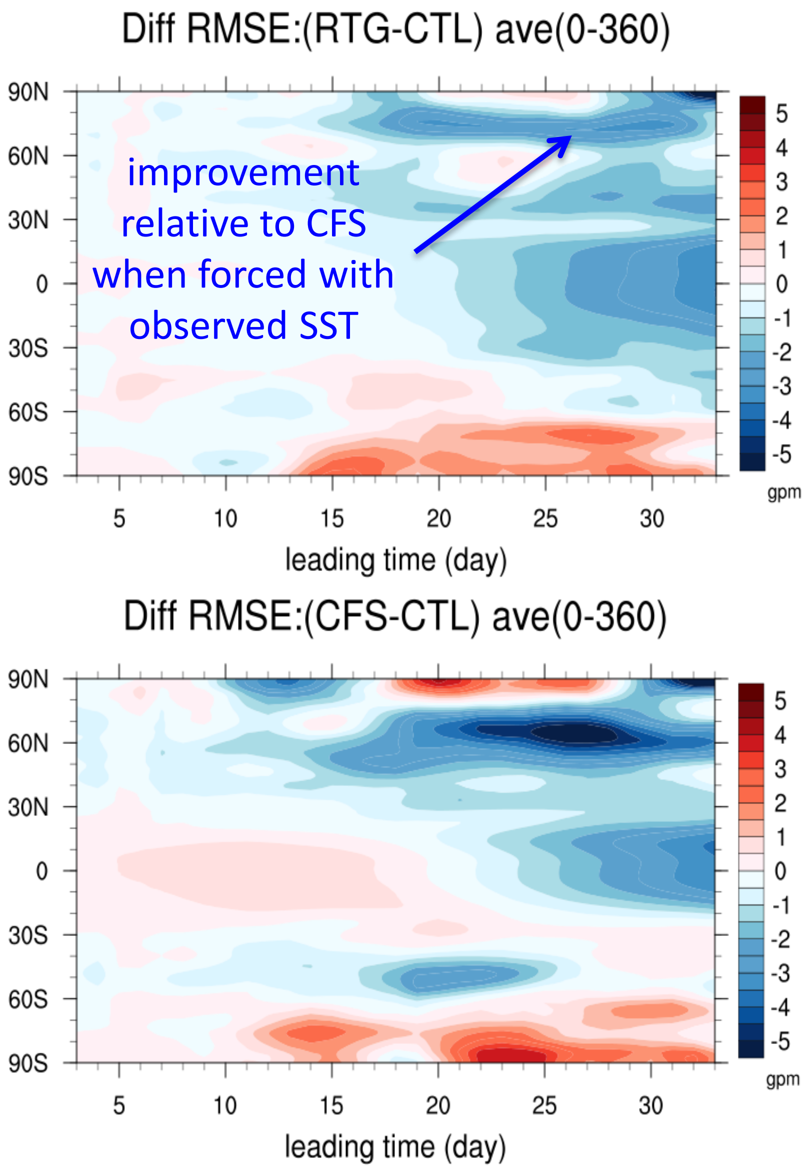
Example: hurricane position adjustment to water-vapor imagery

A field of adjustment vectors (panel b) warps the background forecast so it is more consistent with the observed. There is a smoothness penalty to avoid over-fitting.

Ref: Nehrkorn et al. 2014, MWR, DOI: 10.1175/MWR-D-13-00164.1

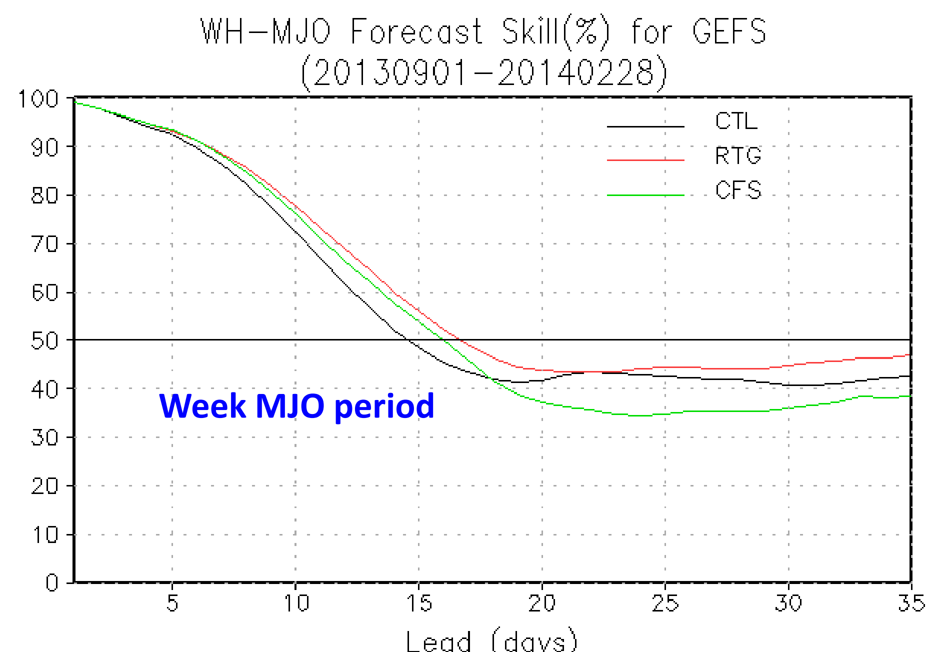
Skill potential with perfect SST forcing (AMIP)

Evolution of error Z500 difference
5-day running mean (09/01/2013 – 2/28/2014)



c/o of Wei Li; see also Barsugli et al, 1998 *BAMS*, **80**, 1399-142.

- Control (CTL or PARA): analysis SST relaxes to climatology
- Optimum (RTG): realistic SST forcing every 24 hours (AMIP like)
- Forcing (CFS): CFSv2 predicted SST forcing every 24 hours



c/o Qin Zhang following Ferranti et al. 1990, *JAS*, **125**, 2177 – 2199.