PROPAGATING UNCERTAINTY INFORMATION IN THE FORECAST PROCESS

Zoltan Toth

Environmental Modeling Center
NOAA/NWS/NCEP

Acknowledgements:
Louis Uccellini, Steve Lord, Ken Mitchell, Dingchen Hou,
Yuejian Zhu, Joo-Hyung Son

http://wwwt.emc.ncep.noaa.gov/gmb/ens/index.html

OUTLINE / SUMMARY

- **WHY DO WE NEED AN ENSEMBLE?**
  - Science – For statistical consistency with verifying observations
  - Applications - For maximum value

- **WHEN DO WE NEED IT?**
  - Case dependent variations & strong non-linearity
  - Bias correction critical

- **ISSUES**
  - Statistical post-processing in 2 steps
    - Bias correction on model grid;
    - Downscaling to finer grid
  - Hind-casting
    - Must be done with latest model version, serious loss otherwise
  - Ensemble river flow forecasting experiments
    - Global application, true ensemble approach
    - Limitations due to lack of statistical corrections

WHY ENSEMBLES?

TRADITIONAL PARADIGM

- Single value forecast incomplete from viewpoints of
  - Science – Inherently statistically inconsistent with observations
  - Applications – Significantly fewer users, with less value
- Probabilistic forecasts needed – Generate them through
  - Single forecast integration
    - Accumulate error statistics over many cases (“bias correction”, eg, MOS)
    - Pro: Maximum possible fidelity in forecast - all comp. resources go into one solution
      - Improved statistical reliability; Slight increase in statistical resolution
    - Cons: Aggregate statistics - no case dependent variations in uncertainty captured

      As errors become nonlinear, single solution becomes unrepresentative
      - Loss of statistical resolution
  - Liouville equations
    - Theoretically proper solution in perfect model framework
      - Pdf of initial state integrated in time
        - » Impractical, enormous computational costs
  - Ensemble forecasts
    - Multiple integrations started with sample from estimated initial pdf
      - Provides multiple trajectories for critical downstream applications
    - Time evolution of pdf captured in truncated form (how many members needed?)
    - Ad-hoc methods aimed at capturing model related uncertainty

ENSEMBLE APPROACH

ENSEMBLES: WHEN?

- Single forecast approach favored when
  - ***Case-dependent variations*** are weak in
    - Level of linear error growth at short lead times
    - Pdf evolution at short lead times (ie, quasi-linear behaviour)
    - Model-related error behaviour (at any lead time)
  - Aggregate bias-correction algorithms adequate
- Use ensembles otherwise
  - Review criteria above for each application
  - Bias-correct both single value & ensemble forecasts (ie, pdf)
    - Decide on forecast configuration based on results
- "Generic" configuration
  - Higher resolution control for short lead time if beneficial
  - Lower resolution ensemble out to longer lead times
    - Benefits from combining hi-re control & lo-res ensemble at shorter leads?
- Considerations
  - Integrations must resolve phenomena of interest
    - Unless sophisticated statistical down-scaling techniques can be developed

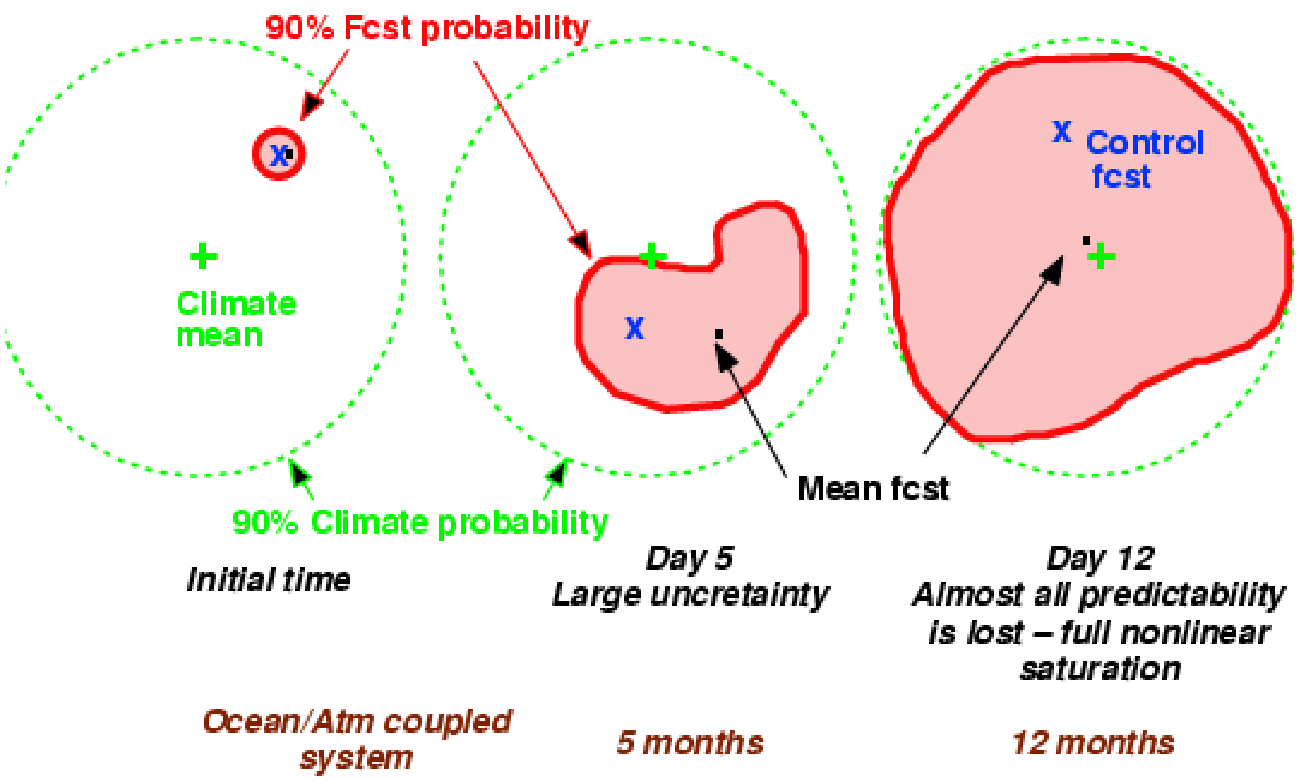
SCIENTIFIC BACKGROUND:
WEATHER FORECASTS ARE UNCERTAIN

ORIGIN OF FORECAST UNCERTAINTY

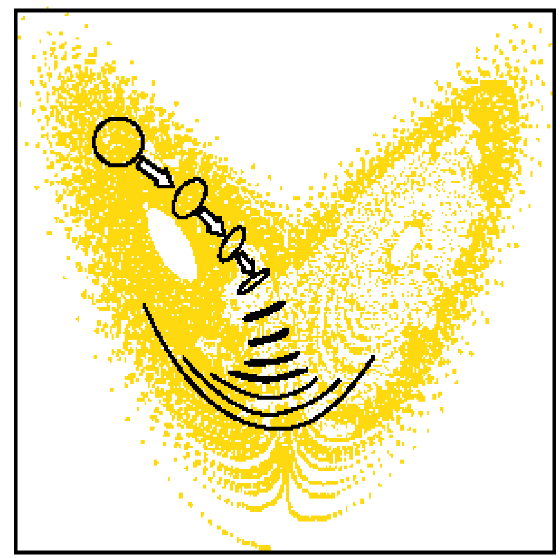
1) The atmosphere is a **deterministic system** *AND* has at least one direction in which **perturbations grow**

2) **Initial** state (and model) has **error** in it ==>

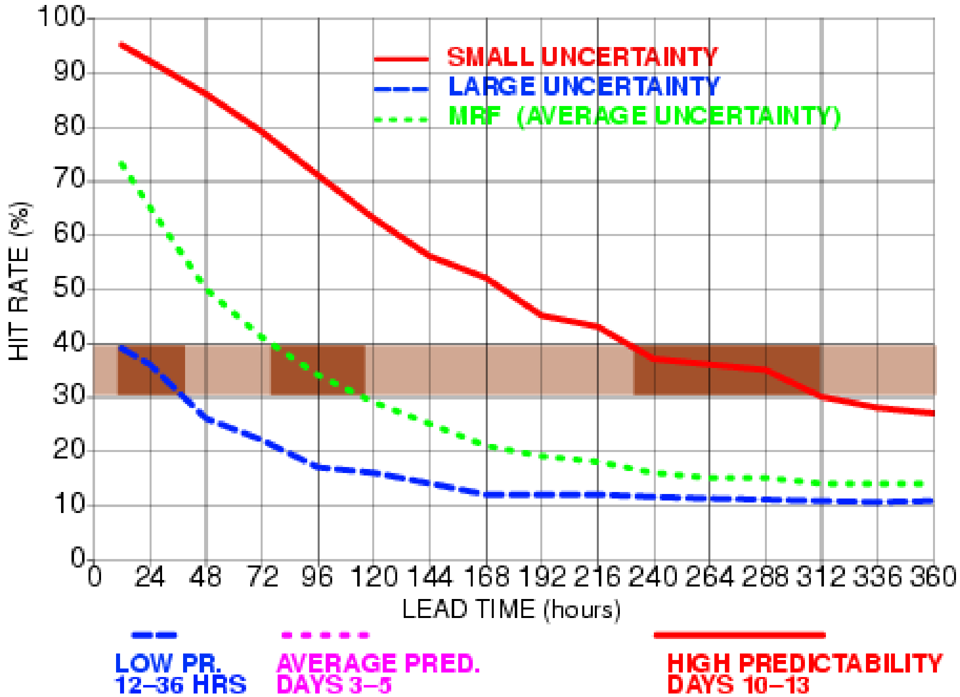
Chaotic system + Initial error =(Loss of) Predictability

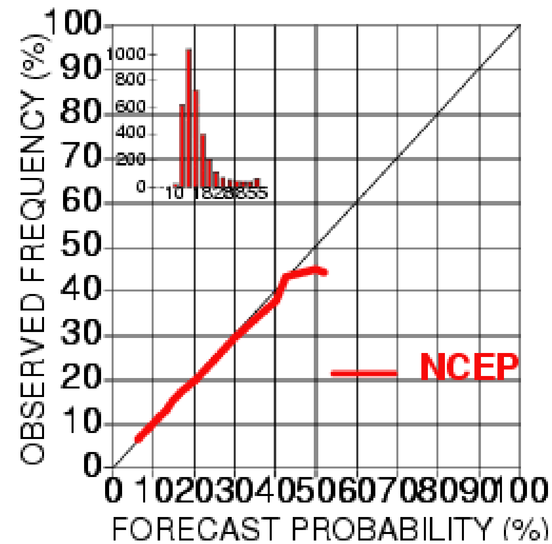


Buizza 2002



SEPARATING HIGH VS. LOW UNCERTAINTY FCSTS





Reliability diagram for 240–hour lead time 500 hPa height NH extratropics forecasts between March and May 1997. Forecast probabilities are based on how many ensemble members fell in any of 10 climatologically equally likely bins at each gridpoint, and are calibrated using verification statistics from the winter of 1995–96. Insert in upper left corner shows in how many events a particular forecast probability was used for the most likely bin (ensemble mode).

*THE **UNCERTAINTY OF FCSTS** CAN BE **QUANTIFIED IN ADVANCE***

HIT RATES FOR 1–DAY FCSTS

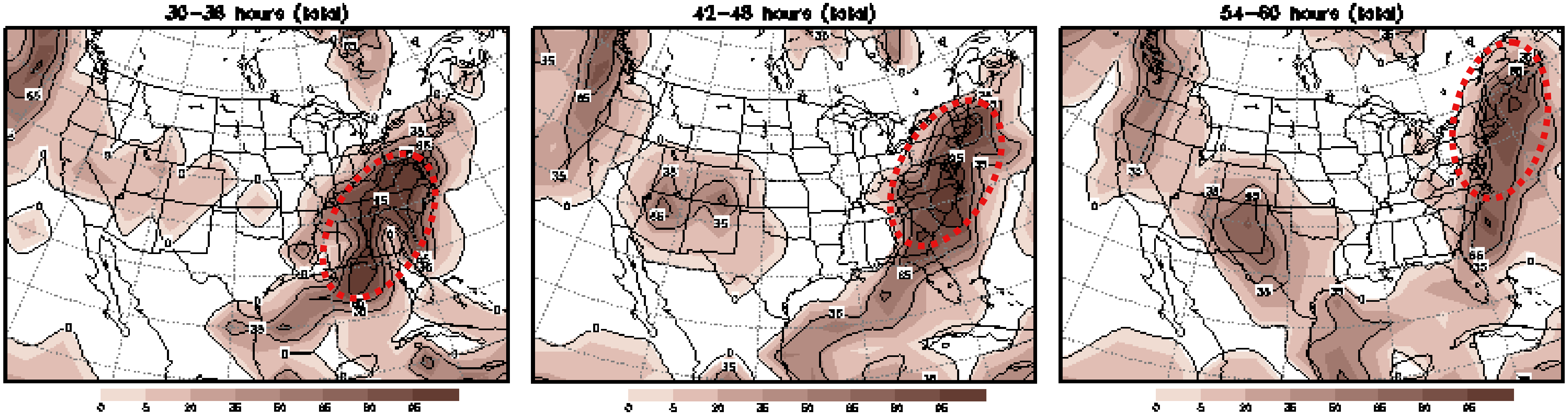
*CAN BE **AS LOW AS 36%**, OR **AS HIGH AS 92%***

10–15% OF THE TIME A 12–DAY FCST CAN BE AS GOOD, OR A 1–DAY FCST CAN BE AS POOR AS AN AVERAGE 4–DAY FCAST

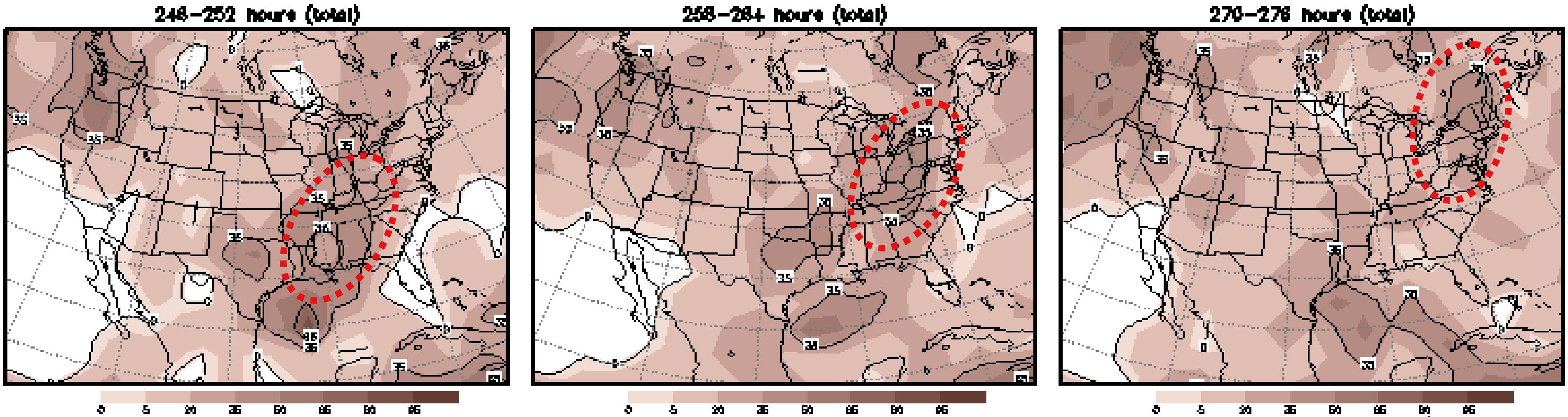
1–2% OF ALL DAYS THE 12–DAY FCST CAN BE MADE WITH MORE CONFIDENCE THAN THE 1–DAY FCST

AVERAGE HIT RATE FOR EXTENDED–RANGE FCSTS IS LOW –
VALUE IS IN KNOWING WHEN FCST IS RELIABLE

Ensemble Based Probabilistic Quantitative Precipitation Type Forecast
Ini:2004111100 Valid:2004111206-2004111212 Amount>=1.000mm/6hrs



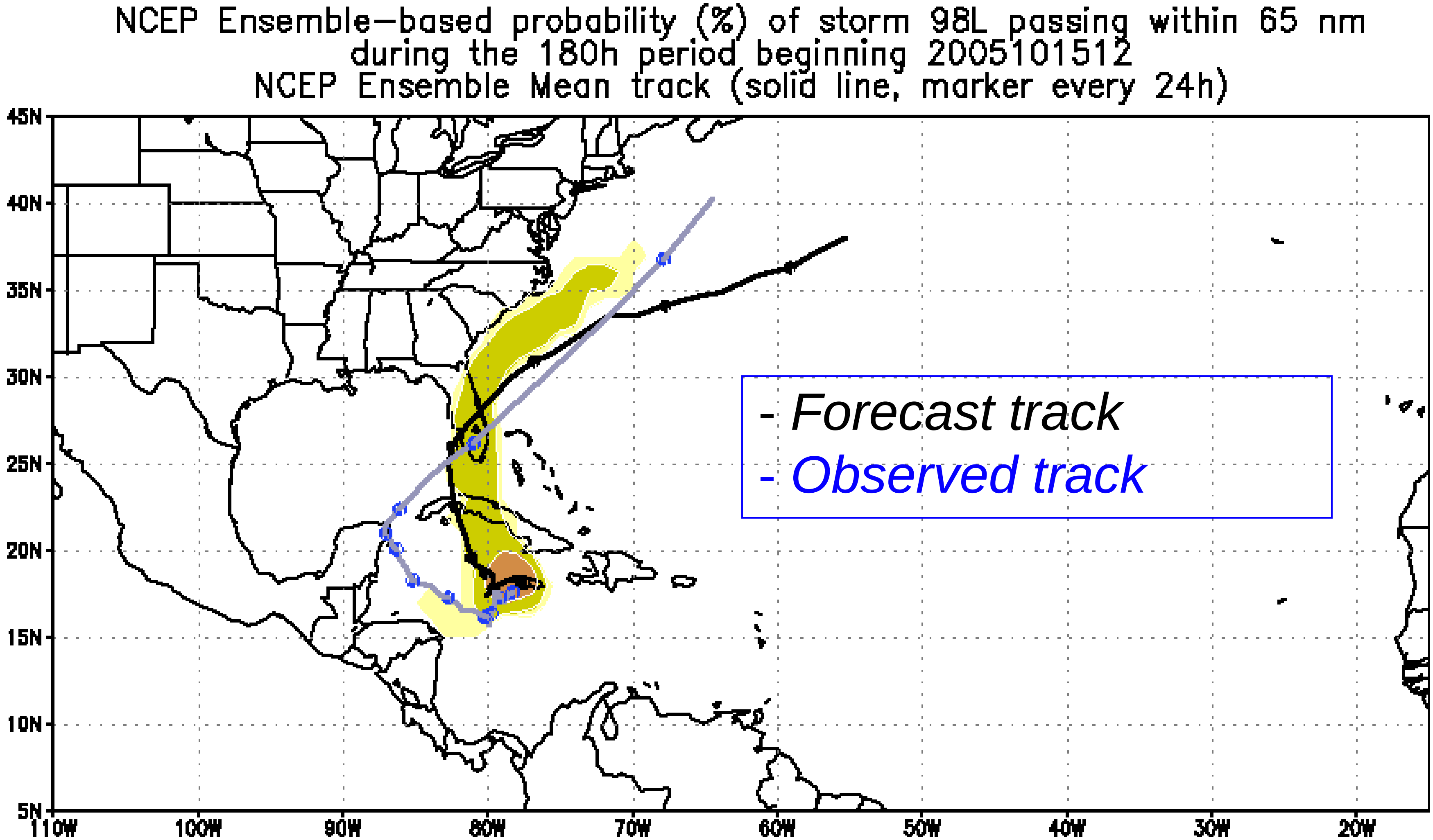
Ensemble Based Probabilistic Quantitative Precipitation Type Forecast
Ini:2004110200 Valid:2004111206-2004111212 Amount>=1.000mm/6hrs



HURRICANE WILMA

STRIKE PROBABILITY

Probability of storm within 65 nm vicinity of any point on map



Strike probability =>

NOAA/GFDL

PARADIGM SHIFT IN FORECASTING

	TRADITIONAL	NEW
PRIMARY OBJECTIVE	Single forecast	Probabilistic forecast
DERIVED QUANTITY	Forecast uncertainty	Single value or other

- Traditional paradigm sub-optimal in non-linear forecast situations
  - Ensemble mean is better estimate (because)
  - Probability distribution is captured by ensemble
- Numerical ensemble forecasting is favored approach
- Systematic errors to be mitigated via statistical bias correction & downscaling
- Unified approaches across scales / applications
- Link with CONOPS for high impact event forecasting
  - Automated ensemble-based forecast process
  - New, empowered role for forecasters
    - Directs & quality controls adaptive forecast process

PROPAGATING FORECAST UNCERTAINTY

OLD PARADIGM: Reduce Uncertainty	***FORECAST PROCESS***	**NEW PARADIGM:** Reduce & Assess Uncertainty
Misconstrued determinism	NATURE	Critical sensitivity to initial conditions - Chaos
Reduce obs. uncertainty	OBSERVING SYSTEM	Quantify obs. uncertainty
Estimate expected value	DATA ASSIMILATION	Estimate distribution
Reduce model errors	NWP MODELING	Reduce & represent model errors
Ad hoc opportunities	**ENSEMBLE FORECASTING**	Systematic approach
Reduce systematic error	STATISTICAL POST-PROCESSING	Calibrate uncertainty
Single value	BASIC PRODUCTS	Distributional characteristics
Yes or No forecasts tailored for decisions	USER SUPPORT SYSTEMS	Incorporate forecast uncertainty info
Limited forecast info - Restricted usage	SOCIETY	All forecast info – Optimal user decisions

Single value

Distribution

Ensemble Forecasting:
Central role – bringing the pieces together

ENSEMBLE DEVELOPMENT CONSIDERATIONS

- Common scientific principles - Chaos affects all spatial/temporal scales
  - Quantify all forecast uncertainty - Inseparable from forecasting in general
    - Links with observing system, data assimilation, numerical modeling, user applications
  - Represent all forecast uncertainty at their source - Otherwise poor reliability
    - Only chance to propagate true uncertainty through forecast process
- Unified approach
  - Common techniques across applications wherever appropriate / possible
- Ensemble team members
  - Work in implementation teams, coordinated with rest of EMC & NCO
  - Interact with broader research and user communities

COMPONENT		Adaptive Observations	Initial Perturbations	Model Perturbations	Statistical Post-Proc.	Product Generation	Verification
FORECAST SYSTEM LINK		*Obs. System Design*	*Data Assimilation*	*Numerical Modeling*			
APPLICATION	PEOPLE	Masutani, Song,	Wei	Hou, Du	Cui, Pena	Zhou, Zhu	Zhu, Zhou, Hou
Coupled	Pena						
Global	Zhu, Wobus						
Regional	Du						
High-Impact							
Ocean wave	Chen						
Sea Ice	Grumbine						
Riverflow/ Land-surface	Hou						

STATISTICAL POST-PROCESSING

Distinguish between

- Bias correction on model grid
  - Eliminate lead time dependency
  - Coarse to coarse resolution mapping - Cheap
- Downscaling to (much) higher resolution grid (NDFD)
  - Needed if effective model resolution is below desired output resolution
  - Coarse to fine resolution mapping – Expensive
  - Sub-grid variability must be added for ensembles
  - Best done by dynamical methods
    - LAM or variable resolution global model – Very expensive

Background

- Based on sample of forecast – truth pairs
  - Model, nature must be stationary
- Quality depends on signal (“systematic error”) to noise (random error) ratio
  - Improved when
    - Random error smaller (short range)
    - Sample size larger (hind-casts important for longer leads with larger random errors)
  - Degraded when more details sought
    - Need for larger sample

STATISTICAL POST-PROCESSING –PROPOSED SOLUTION

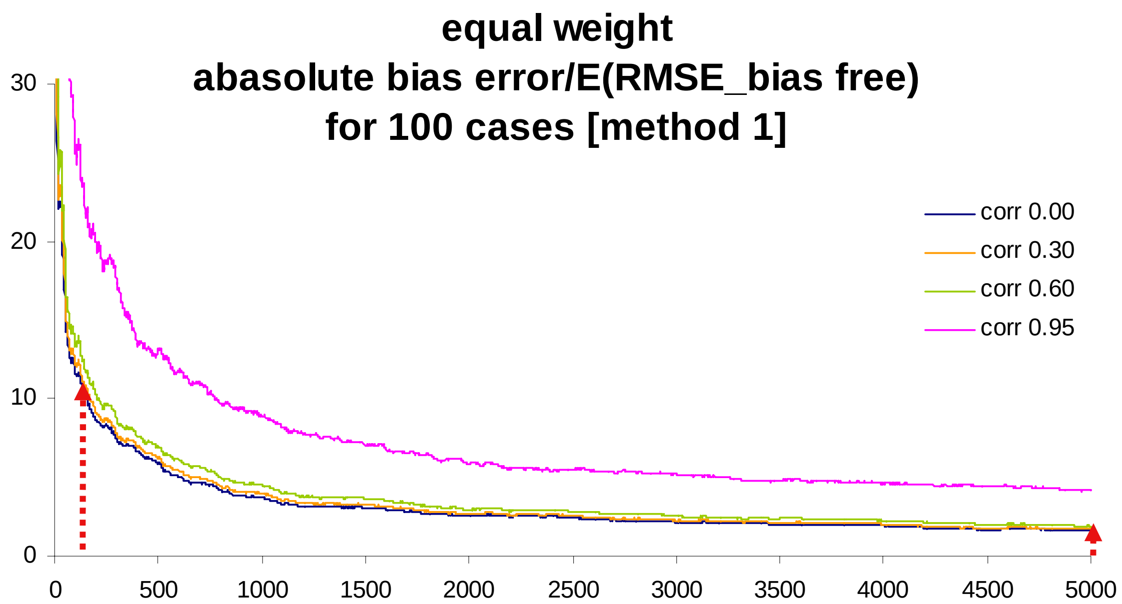
PROPOSED SOLUTION: Follow 2-step approach, develop centrally applied

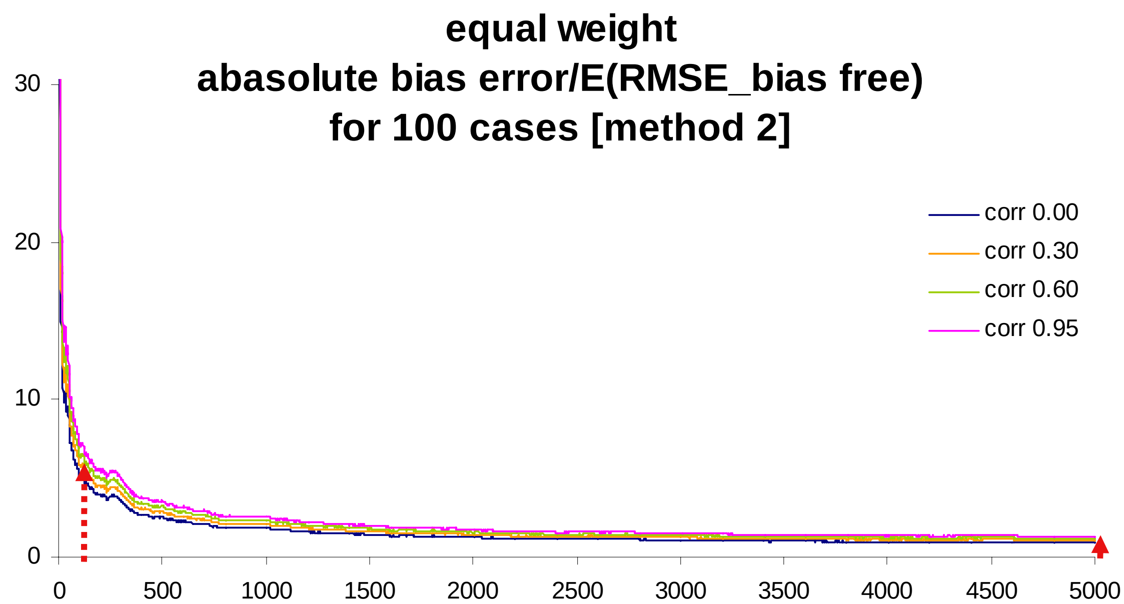
- Bias correction on model grid
  - Bayesian approach – can handle all non-Gaussian, non-linear situations
    - Can optimally merge hires control, lores ensemble, and climate information
    - Bias-corrected ensemble trajectories
- Downscaling to NDFD grid
  - Current methods – no sub-grid processes considered/added
    - Linear function of grid-scale info – limited utility
      - Climate anomaly
      - Downscaling vector
  - Alternative methods – sub-grid processes considered/added
    - More information / larger sample needed – *MDL, other collaborators*
      - Local analogs
        - » Must mosaic together independent patches
        - » Cannot well control what information is retained vs. stochastically replaced in ensemble
      - Stochastic generator
        - » More general solution
        - » Difficult to construct?

Forecast configuration evaluation

- Eventually must be done on user relevant bias-corrected and downscaled information
  - Selection criteria
    - System that produces best downscaled information at given computational cost

BIAS REDUCTION AS A FUNCTION OF SAMPLE SIZE

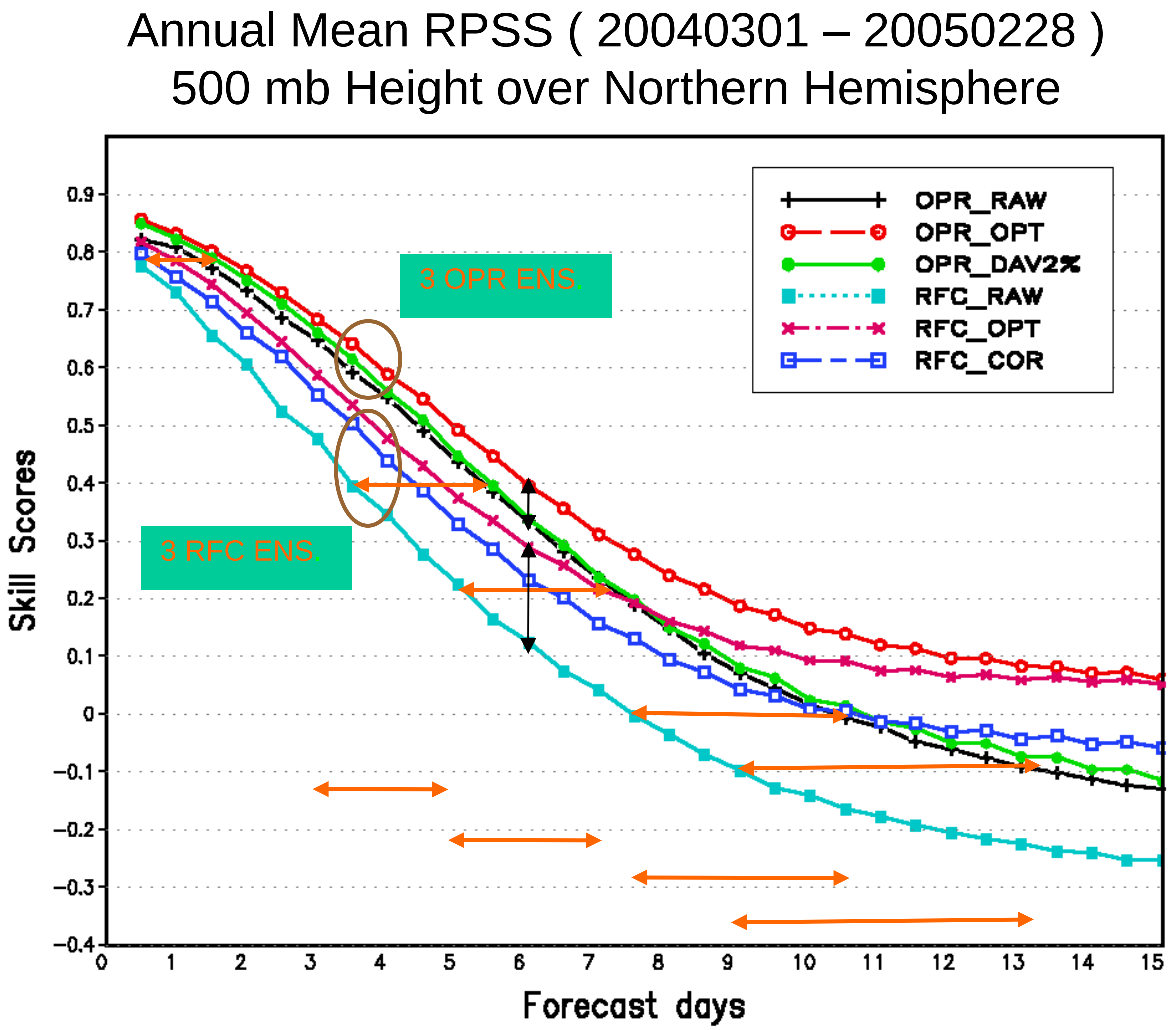




HINDCASTS WITH A FROZEN SYSTEM

- How much skill we lose by using an 8-10 year old forecast system?
  - 2 days skill at D3, 3 days at D7.5, 4+ days at 9D
    - Estimate 2% rms error reduction per year in recent yrs
- How much skill we gain due to use of larger sample?
  - Reduction depends on level of bias in forecast
    - CDC hind-casts have much larger bias
      - Large bias reduction observed
        - » Still does not compensate for loss of skill until 10+ day lead time
    - Operational forecast system
      - Not known
        - » Estimate via synthetic data
  - Use synthetic data to estimate bias error reduction with more data & current fcst
    - What is current level of NHE time mean error (estimated bias)?
      - ~6-14% of climate standard deviation (2m temp)
    - Gain from expanding sample
      - From 100 days (Kalman Filter method) to 5,000 days (50 yrs hind-cast)
        - » 5-10% (depending on method) of random rms error
        - » This is equivalent to 2.5-5 yrs of model development
- If DA/model has to be frozen for 3-5 yrs
  - Gain in reliability due to larger sample &
  - Loss in resolution and reliability due to use of frozen system
    - Are of similar magnitude
  - Is it worth it?

Raw, Optimal & Actual Bias Corrected Ensembles



Decaying average

bias correction improves RPSS for ll lead time vs. raw oper. ens.

- Climate error removed

  bias corrected reforecast gains significant improvement for all lead time vs. raw reforecast

- Operational vs. reforecast ens.

  oper. fcst is better than the bias-corrected reforecast out to 9-10 days. Beyond 10 days, bias-corrected reforecast becomes competitive to or better than oper. fcst

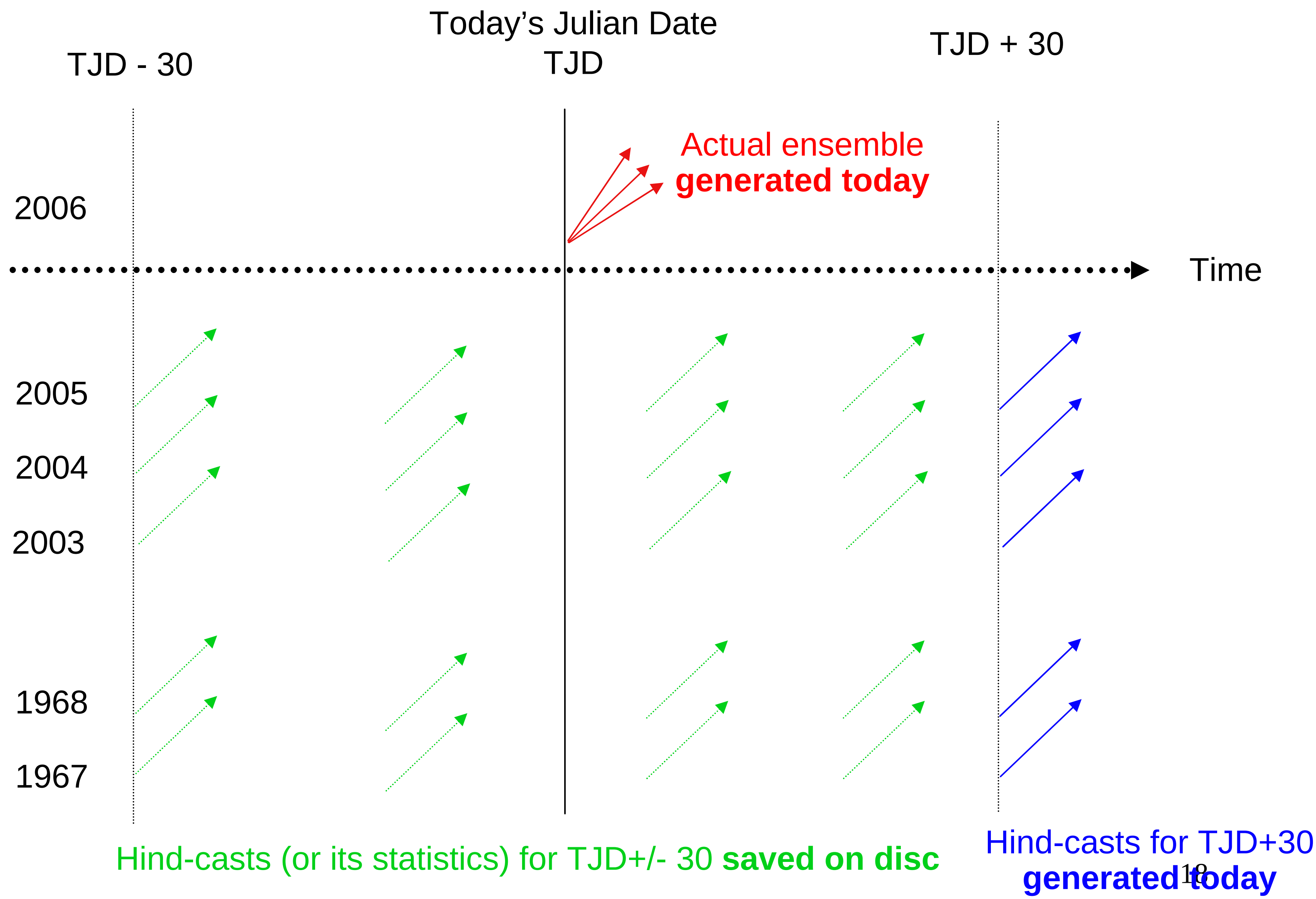
- Sign of improving

  larger for CDC reforecast

HIND-CASTING – PROPOSED SOLUTION

- Consider real-time generation arrangement as part of operations
- Use forecast process for hind-casts identical to operations
  - Cannot share hind-casts across applications if operational forecasts are not shared
  - Assumption: bias at longer lead does not depend on analysis
- Critical for longer range applications
  - Global ensemble
    - Highly non-linear applications such as river flow forecasting
    - Assume bias at long lead independent of analysis technique (use reanalysis)
  - Coupled ensemble
- Needed for refined, regime dependent bias correction / downscaling for
  - Regional ensemble, etc
  - Short-range bias depends on analysis technique
    - Must regenerate reanalysis with current DA system
      - Can reanalysis also be done in real time on next machine (2010)?

REAL-TIME GENERATION OF HIND-CAST DATASET?



GLOBAL STREAMFLOW PREDICTION EXPERIMENTS

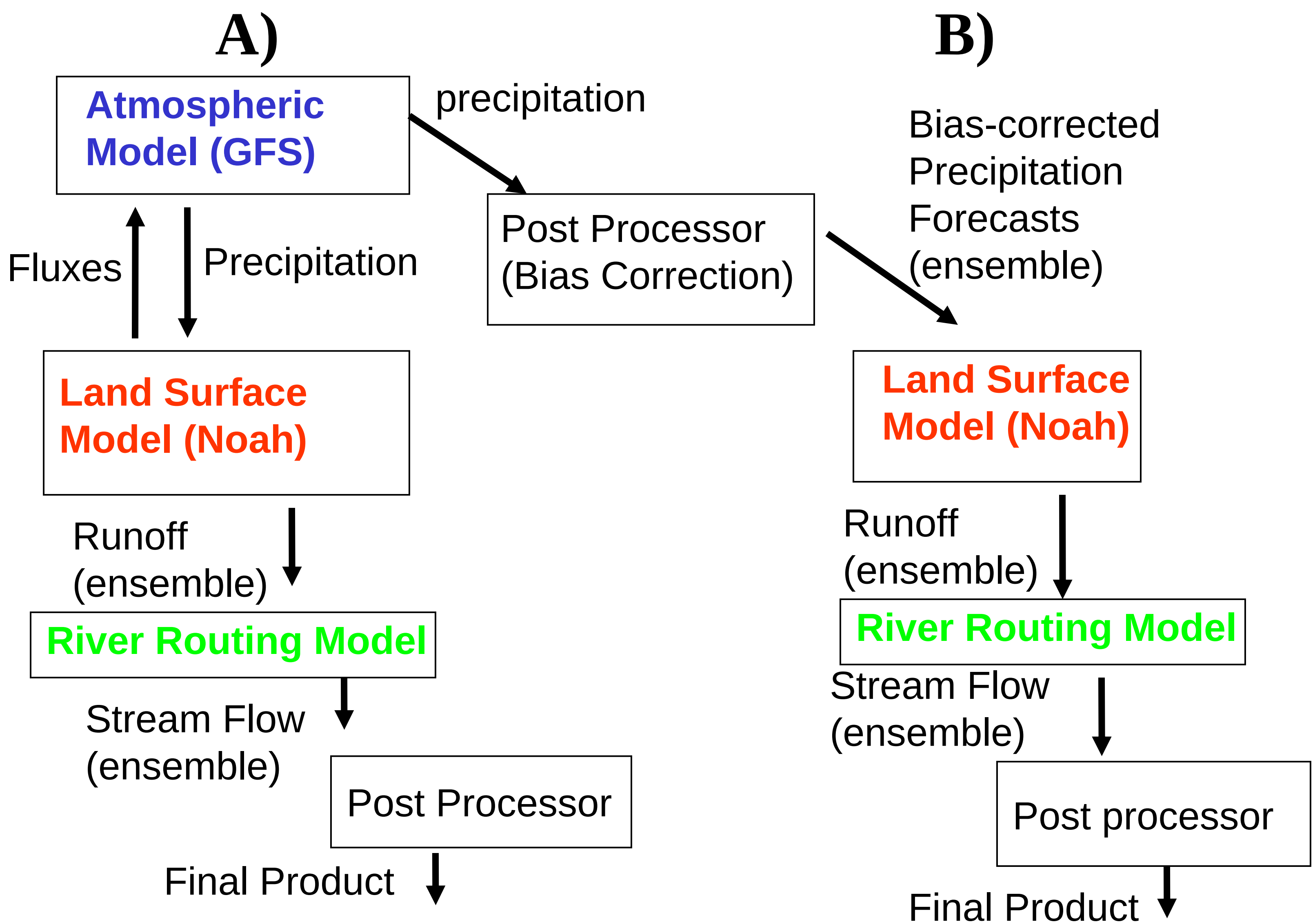
Purpose:

- Demonstrate feasibility of gridded river flow forecast in operational GEFS.
- Establish optimal configuration for river routing.
- Develop suitable strategy to account for uncertainties.
- Develop suitable methods of post processing and generating final products.
- Validate model output of stream flow.

General Strategy:

- NLDAS stream flow analysis used as initial condition and verification;
- Water management issues avoided by focusing on natural flow forecast;
- Global domain in mind with domestic and international users;
- River flow forecast capacity as a component of the ESMF system;
- Hind cast data set to be generated for post pressing.

Ensemble Stream Flow Forecast
Two Possible Approaches



Configuration

Design of Preliminary Experiment (Approach A)

- **River Routing Model**: linear program, distributed approach, same as used in NLDAS (Lohmann et al., 1998, 2004).
- **CONUS domain**, 1/8 degree grid size (same as NLDAS).
- **River Flow Direction Mask**: A D8 model, river stream in each grid point is discharged to 1 of the eight main directions (Lohmann, et al, 2004).
- **Forcing**: Runoff from global ensemble forecasts (GEFS) and the high resolution control forecast (GFS), interpolated to NLDAS grid.
  - Downscaling not considered yet
- **Uncertainty** considered in river routing: only in the forcing, added partially.
  - No initial uncertainty is included yet
  - Hydrological model error is ignored but systematic model errors can be corrected via post processing
- **Warming up period**: A dry start on Jan. 1st 2006 and a wetting up (warming up) period of 3 months
- **Evaluation**: For this preliminary test, using river routing products generated by the same model forced by NLDAS runoff, which is generated from observed precipitation (denoted as analysis). Observation will be used in follow up study.

Forecast Example (initiated April 1st, 2006) Stream Flow

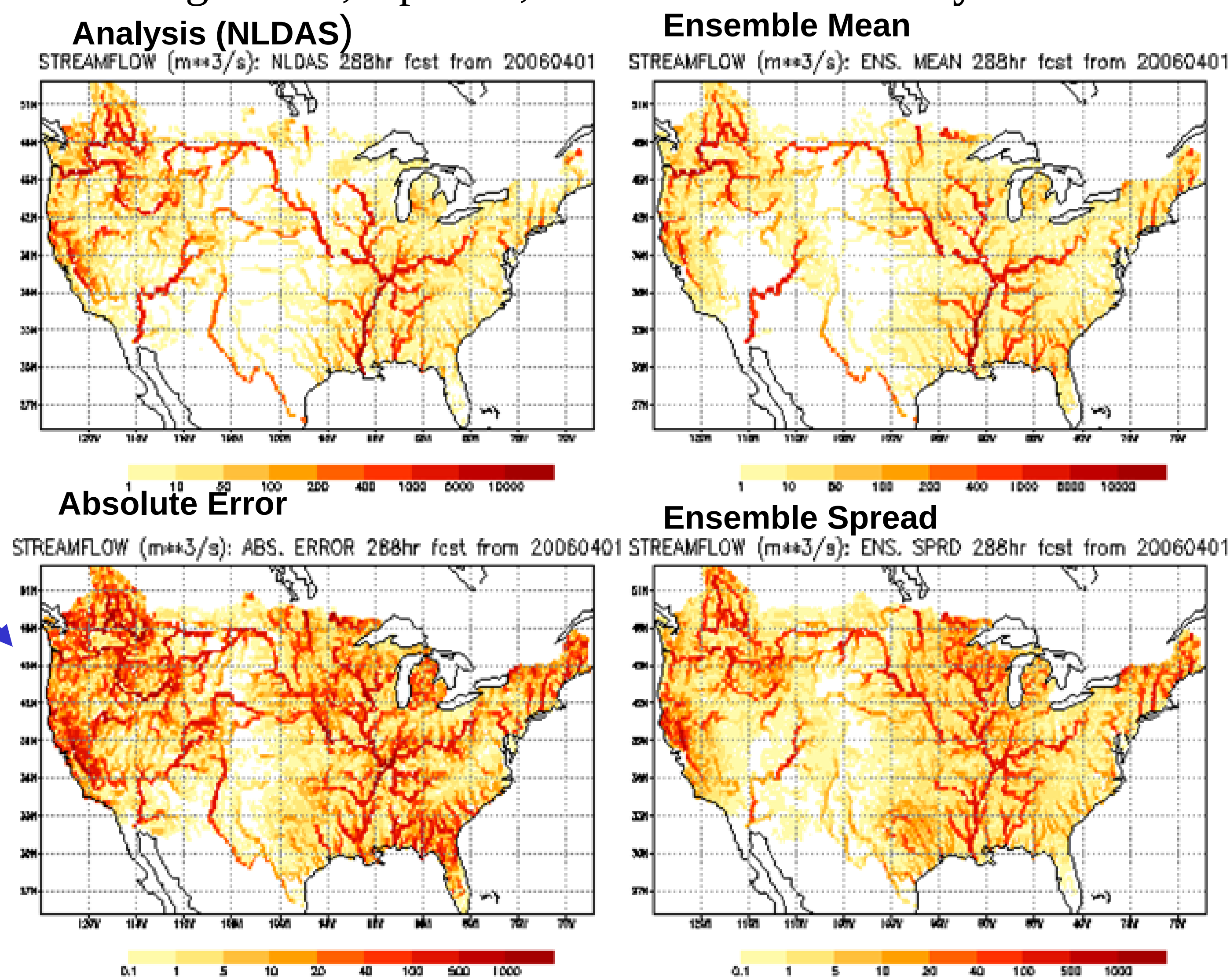
Stream Flow, Analysis and Ensemble Mean Forecast

Absolute Error of Ensemble Mean and Ensemble Spread

Forecast Starting at 00Z, April 1st, 2006. lead time **12 days**

Same as previous slide except for the error, where absolute value is plotted to compared with the spread.

Spread is comparable to error, but the value is smaller, especially in the West.



Mississippi, River
Vicksburg, MS
The Large Basin

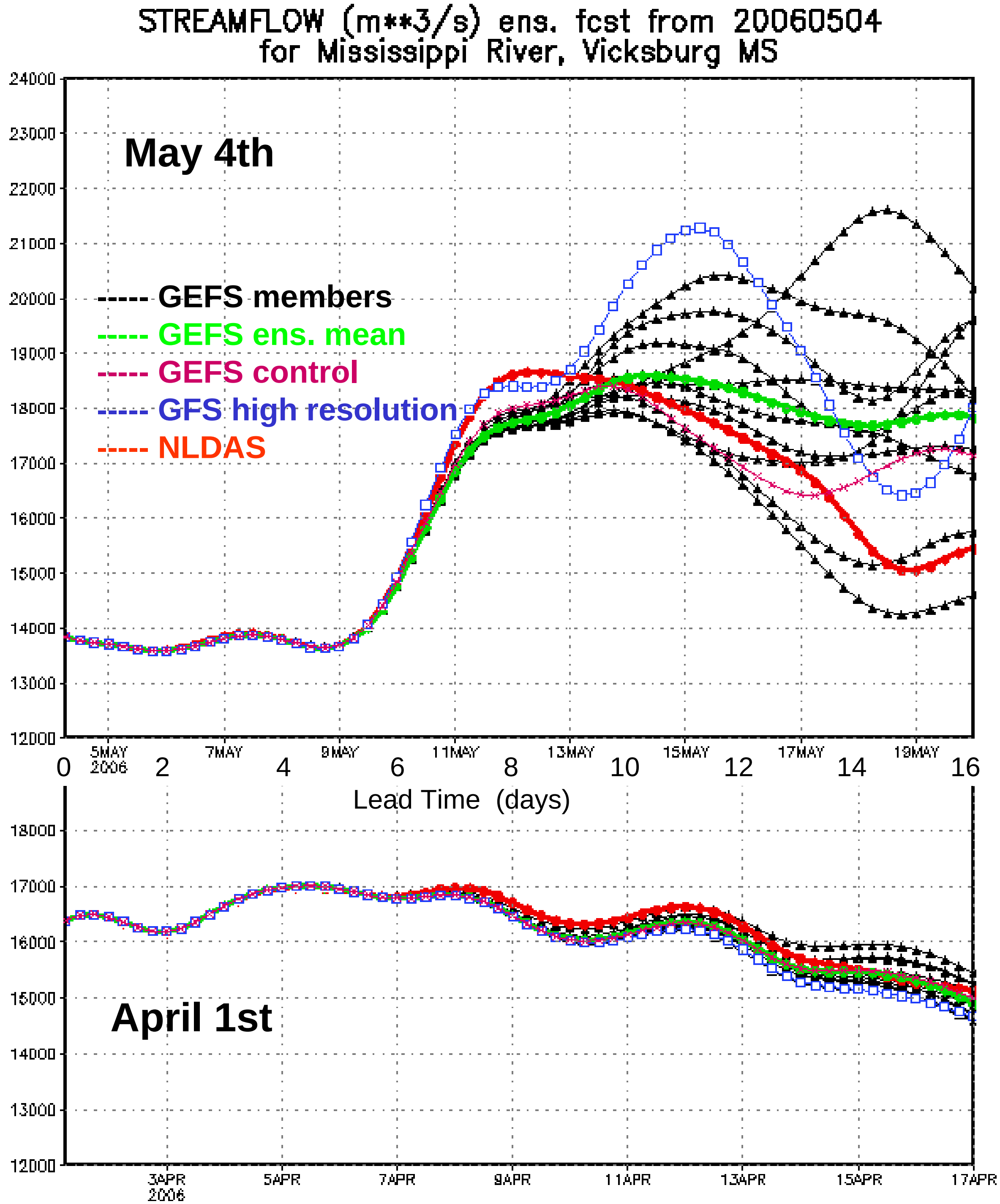
May 4th case

A major mid-range event well predicted
Significant spread in extended range

April 1st case

With out a major event,
all simulations are similar and spread is small.

Trend and events picked up.
Short lead time dominated
  by initial condition, showing
  little spread.
Spread Increases with time.

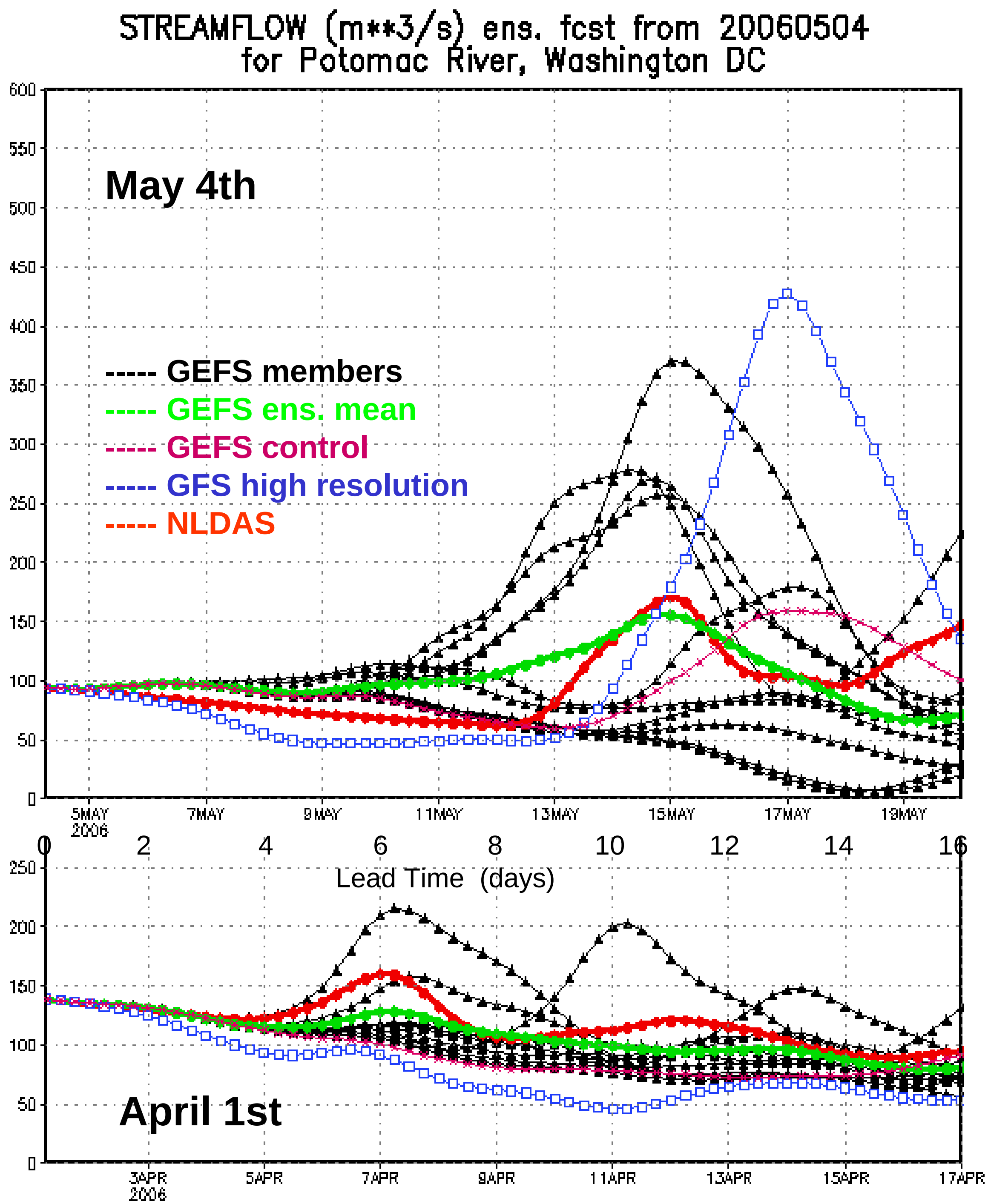


Potomac River
A Medium Sized Basin

In both cases

Single forecasts are insufficient.

Non-linear evolution of ensemble members help to improve forecast and catch major flood events.



Merrimack-Concord River
Lowell, MA
A Medium Sized Basin

Major Problem

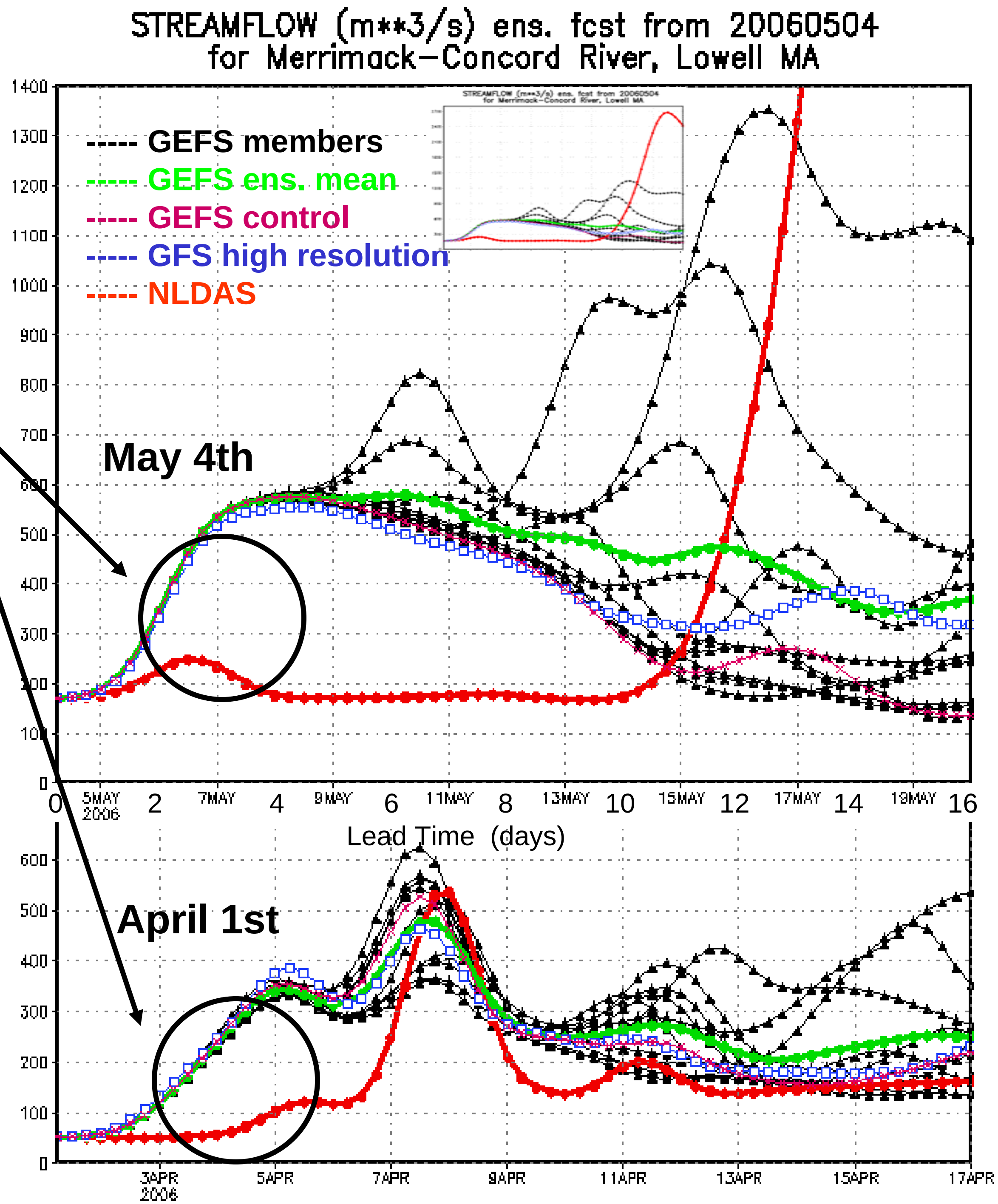
Underdispersive ensemble in grid and subgrid scale precipitation.

Mid-May Flood Event

Compared with the Early-April event, the Mid-May event is harder for the model to simulate. Nevertheless, the ensemble shows some skill indicating a major event with 10+ day lead, various amplitude and timing.

Early April

Major event forecast despite short range over-forecast



BACKGROUND

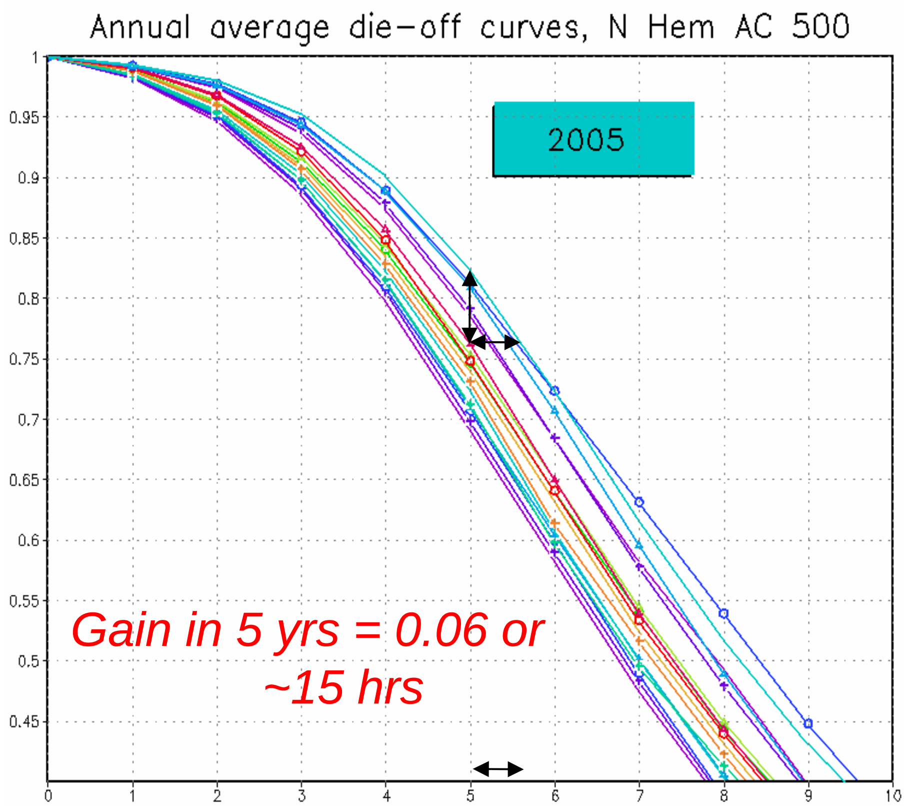
STATISTICAL POST-PROCESSING

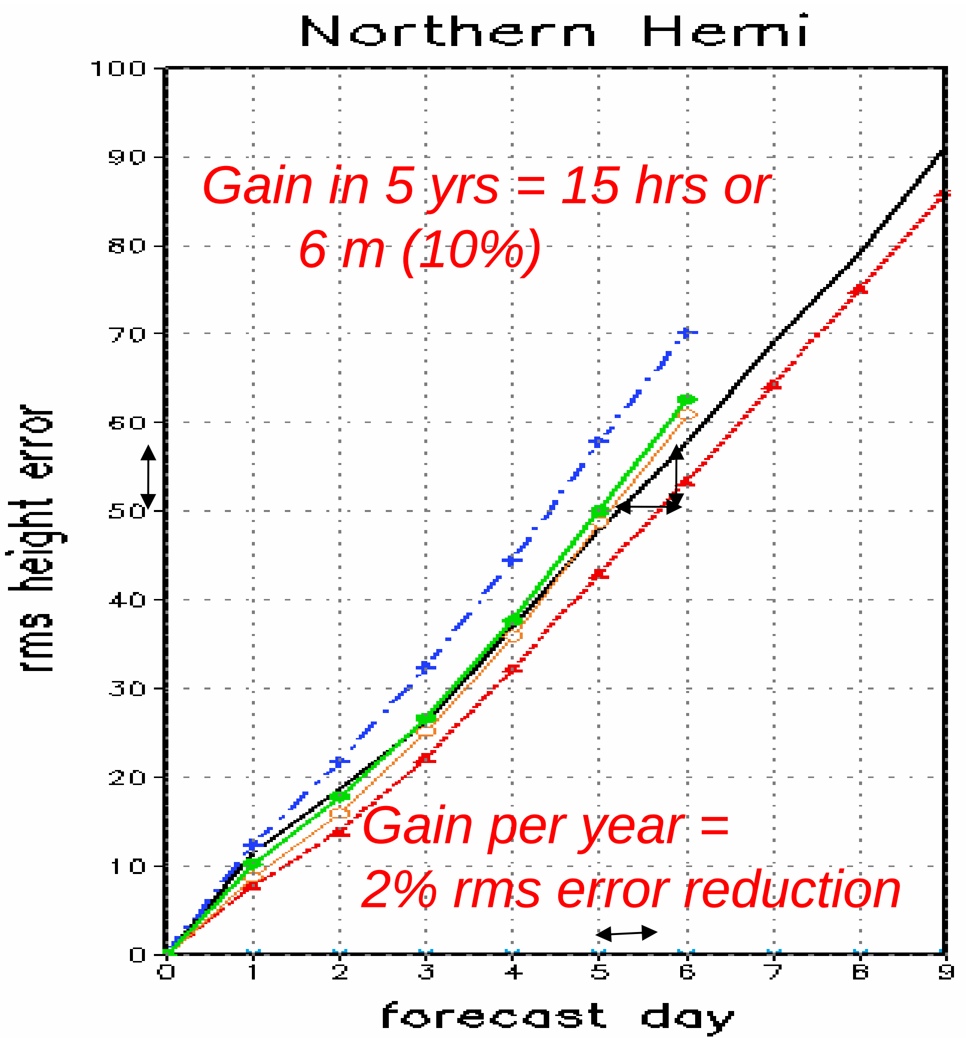
Approaches

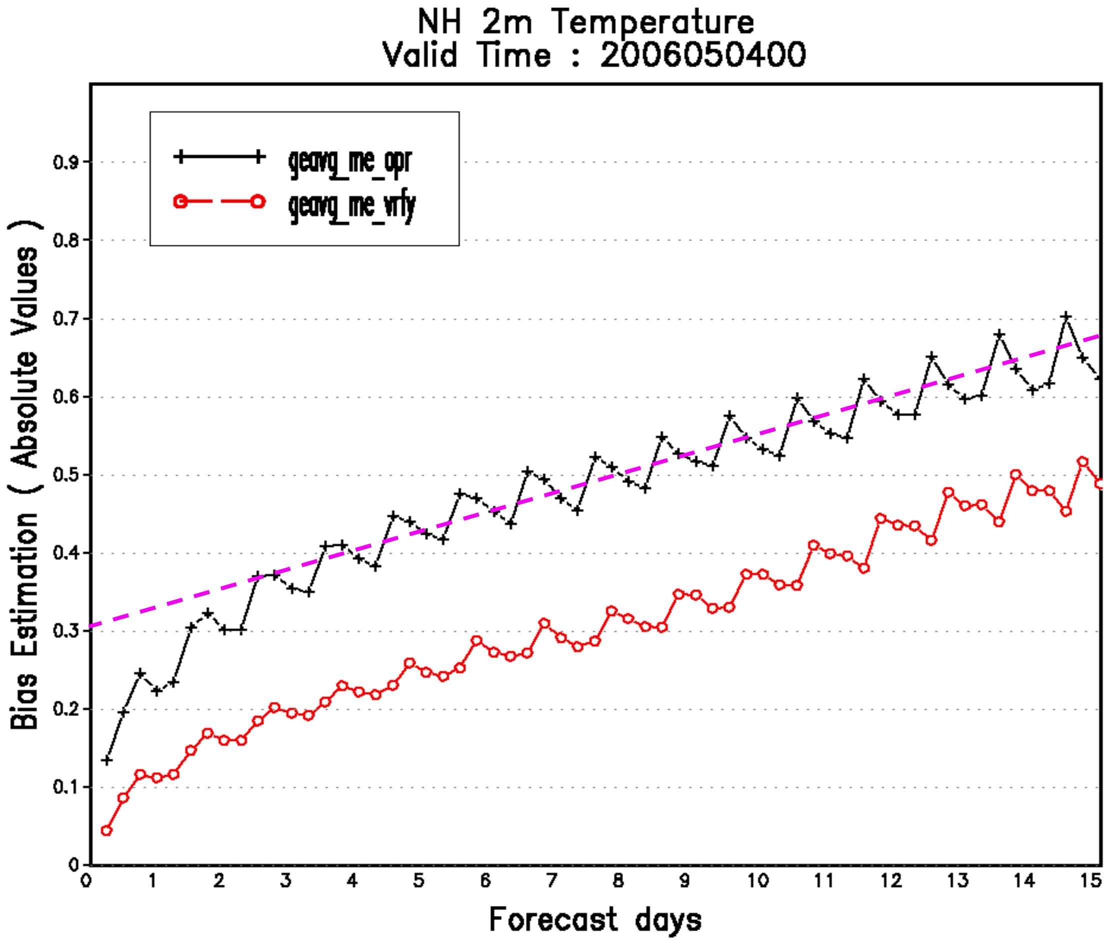
- 1-step "downscaling"
  - Potential advantage in reduced noise
- 2 steps
  - Lead-time dependent bias correction on model grid (cheap)
    - Diagnostic evaluation of model forecasts possible
  - Lead-time independent downscaling to finer grid (more expensive)
    - Applied on bias corrected forecasts - "Perfect prog" approach, not dependent on lead time
    - More flexibility

Applications differ in

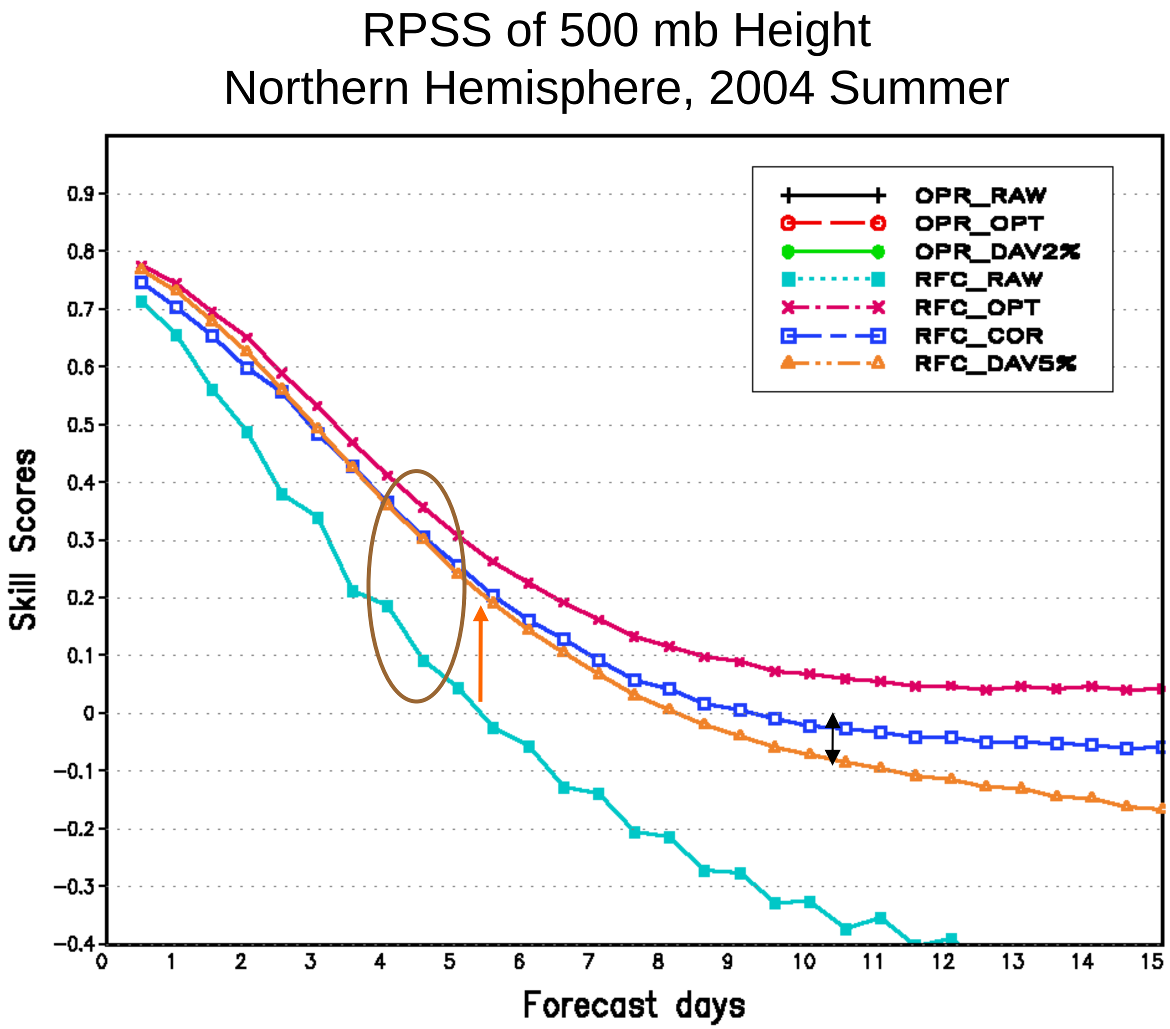
- Statistical method for extracting info from forecast – truth pairs
  - Linear regression, Bayesian, analog method
- What they bias correct statistically (skilful and not reliable statistically)
  - Only 1st, or also additional moments?
- How rest of information from forecast treated
  - Retain from raw forecasts (if info statistically reliable)
  - Replace stochastically (if info not skilful, not reliable)
    - Analog approach - difficult to control retained vs. lost information
    - Stochastic generator – preferred solution
  - Filter out (not skilful)
    - MOS hedges toward climatology, not suitable for ensemble applications


Annual average die-off curves, N Hem AC 500
2005
Gain in 5 yrs = 0.06 or
~15 hrs


Northern Hemi
rms height error
forecast day
Gain in 5 yrs = 15 hrs or
6 m (10%)
Gain per year =
2% rms error reduction


NH 2m Temperature
Valid Time : 2006050400
geavg_me_opr
geavg_me_vrfy
Bias Estimation (Absolute Values)
Forecast days
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Raw, Optimal & Actual Bias Corrected Ensembles

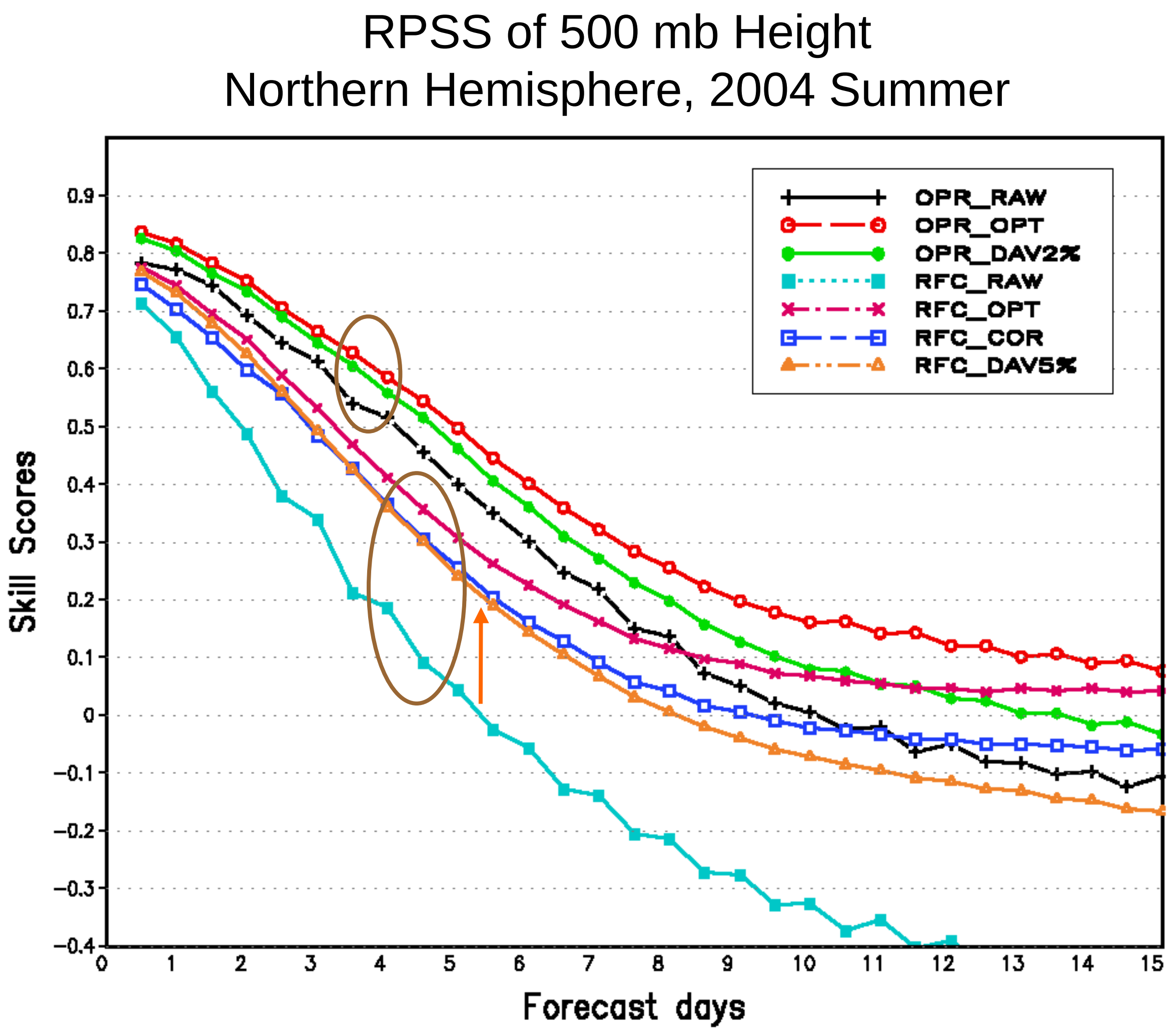


ecaying average applied to
DC reforecast

decaying method gives better results
ıan climate mean bias estimation for
ıort range (~day 5), value of regime
ependent correction

some gain from climate mean bias
orrection after _5 days

Raw, Optimal & Actual Bias Corrected Ensembles

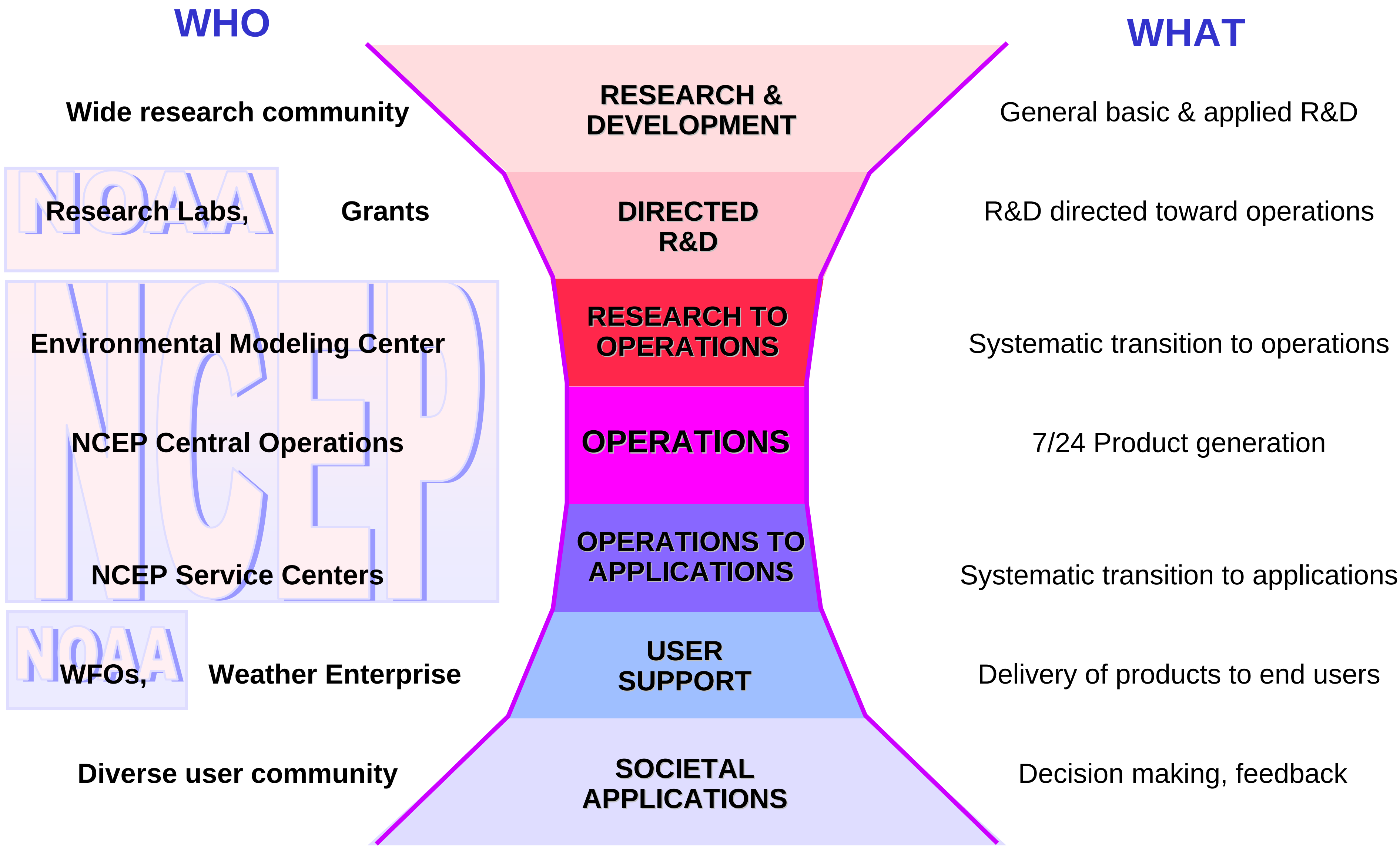


ecaying average applied to
DC reforecast

decaying method gives better results
an climate mean bias estimation for
ort range (~day 5), value of regime
ependent correction

some gain from climate mean bias
rrection after _5 days

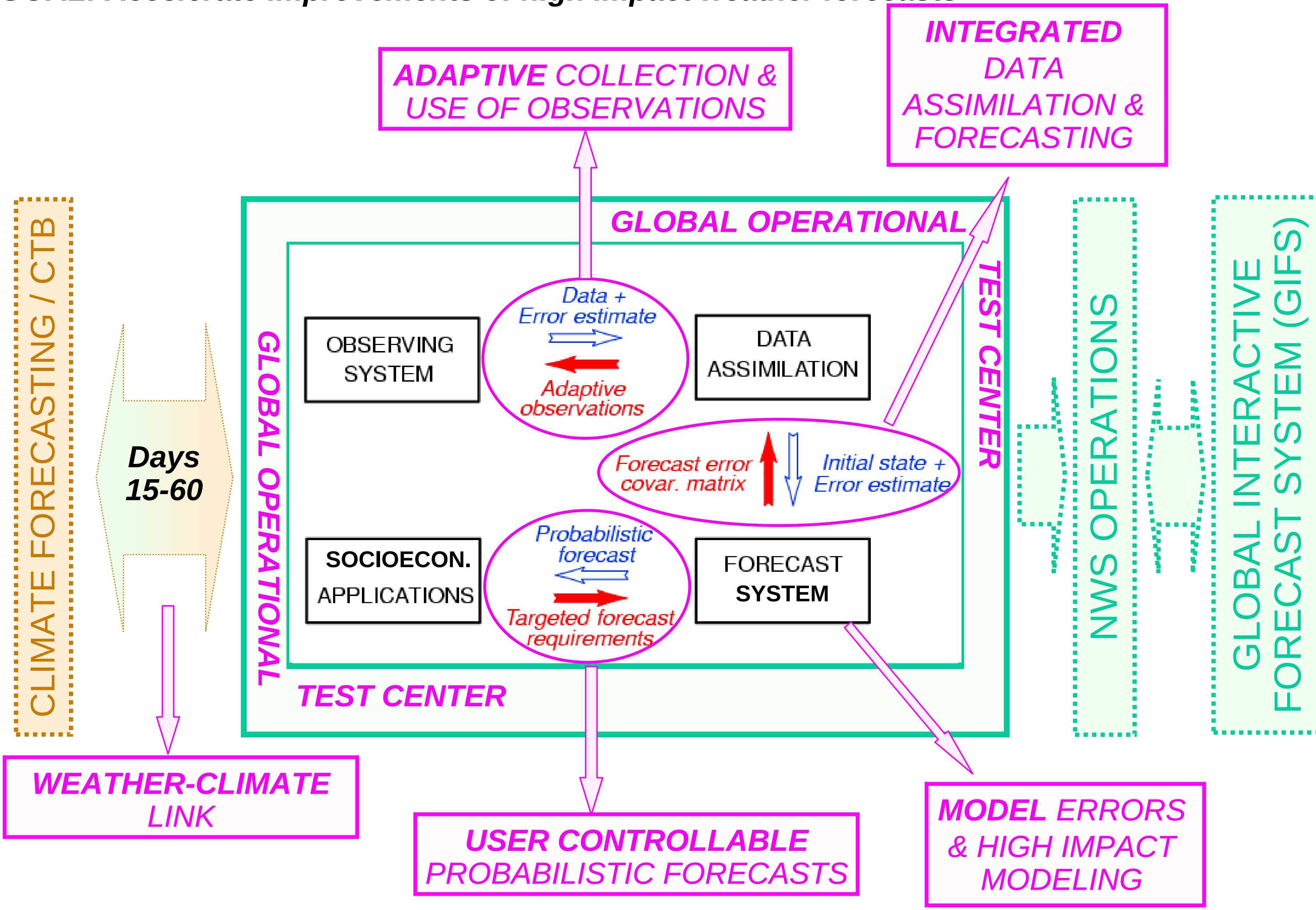
RESEARCH TO OPERATIONS TO APPLCIATIONS FUNNEL



ENSEMBLES AND THE RESEARCH COMMUNITY

LINKED THROUGH THORPEX – MAJOR INTERNATIONAL RESEARCH PROGRAM

GOAL: Accelerate improvements of high impact weather forecasts



ENSEMBLES AND NOAA SERVICES

- NWS requirements must be redefined
  - NWS operations is strictly requirement driven
    - Culture must change to support evolution in operations

- New NWS CONcept of OPerationS (CONOPS)
  - W&W Goal Theme involvement
- High Impact Events Theme
  - Adaptive and event driven
  - Integrated across the spectrum of services
  - Probabilistic approach
  - Enhanced automated guidance
  - New role for forecasters
  - Environmental Information Repository

- "Establish comprehensive suite of ensemble forecast systems ("forecast engine") that will facilitate the generation of automated forecast guidance products in the framework of the new NOAA CONOPS as the basis ("forecast engine") for NOAA operations regarding high impact events:
  - New automated "forecast engine that adapts to high impact events
    - Adaptive observations
    - Adaptive ensemble suite
    - Statistical post-processing

CONSIDERATIONS FOR OPERATIONAL IMPLEMENTATIONS

- Performance
  - Offline research, parallel development, pre-implementation testing
    - **User relevant verification statistics** (ie, bias corrected forecasts)

- Economy
  - Operations is narrowest point in Research-Operations-Applications funnel
    - Lots of research/development, **one system in operations**
    - **Computational efficiency**

- Maintenance
  - Minimize work needed for transfer (R2O, O2A, from machine to machine, etc)
    - **Unified approaches** preferred if performance not sacrificed

- Interconnectedness
  - Each piece of operations intimately connected with rest of system
    - **Incremental improvements** to existing system OR
    - Very careful **long-term planning** for major upgrades

2^{nd} Ensemble User Workshop

ACTIONS ON SUMMARY RECOMMENDATIONS

- **OVERALL** - Enhance coordination of ensemble-related efforts
  - Establish ensemble product working group *Done, NAWIPS success story*
  - Continue with monthly Predictability meetings *Done, 25 meetings*
  - Hold Ensemble User Workshops (part of reestablished SOO workshops) *In progress*
- **CONFIGURATION**

Global ensemble:	Implement hurricane relocation for perturbed initial conditions	***Done***
	Continue efforts to build multi-center ensemble	***Done, NAEFS***
Regional (SREF) ensemble:	Closer coupling with hires control (same initial time?)	***Resource needed***
	Run 4 cycles per day	***Done***

- **DATA ACCESS**
  - Provide access to all ensemble data (including members) *Resource needed*
  - Facilitate user controlled access to data (e.g. NOMAD, on demand, no rigid schedule) *Under planning*
- **STATISTICAL POST-PROCESSING (BIAS CORRECTION)**
  - Develop techniques for two-stage statistical post-processing *Done, downscaling under dev.*
  - Operationally implement post-processing techniques *Done*
- **PRODUCTS**
  - Develop a software toolbox for interrogating ensemble data *Done, NAWIPS*
  - Establish central/local operational product generation suites *In progress*
- **VERIFICATION**
  - Design & develop unified and modular ensemble/probabilistic verification framework *Resource needed*
- **TRAINING**
  - Establish NWS formal ensemble training requirements *OST?*
  - Contribute to Ensemble Training Workshops, international activities (AMS, WMO) *Done (Grumm, etc)*

3rd Ensemble User Workshop
EXPECT RECOMMENDATIONS ON

- Ensemble configuration
  - Relative resources/priorities for ensemble generation
  - Unified ensemble system for high impact applications
    - Hurricane, severe weather (storms), fire weather, air quality, dispersion, rapid update / now-casting
- Statistical post-processing
  - Unified bias correction on model grid to remove drift related errors
  - Downscaling to NDFD grid for high fidelity forecasts
- Products
  - Data depository
  - Interrogation / product generation tools
  - Data access
  - Product distribution
  - Verification
- Outreach
  - Support of / feedback from “weather enterprise”
    - Link with Decision Support Systems
  - Training
  - Societal value

HIND-CASTING

What

- With operational process, generate forecasts for past cases
  - Must use operational procedures otherwise lost purpose
  - Resource intensive

Purpose

- Increase sample size for bias correction / downscaling

Required for

- Longer lead bias correction
- Shorter lead bias correction with more details (regime dependent)

Options

- Freeze operational system
  - Generate hind-cast data set prior to use in operations
    - Labor intensive
    - Any improvements must wait until next hind-cast dataset can be prepared
- Generate hind-casts in real time, on continuous basis
  - Can upgrade forecast system any time following a 2-month parallel experiment
  - Computationally more expensive – Re-computes hind-casts with new system every year
  - Logistically simpler, institutionalized process
  - Cheaper in terms of human resources

BACKGROUND

THE START OF A NEW ERA IN WEATHER FORECASTING

Confluence of several developments

- Science readiness
  - Lorenz' discovery of chaos and what followed
- Technological revolution
  - Computing, telecommunication, display
- THORPEX Research Program
  - Aimed at 1-14 day high impact weather forecasting
    - Forecast uncertainty, adaptive techniques
- New NWS Concept of Operations (CONOPS)
  - Adaptive, high impact driven operations
- North American Ensemble Forecast System
  - Operational multi-center ensemble system
- National Research Council "Completing the Forecast" Report
  - Forecast uncertainty for better decision making
- Formation of NCEP Ensemble Team
  - Recognition of gravity of ensemble related issues

NATIONAL RESEARCH COUNCIL REPORT

COMPLETING THE FORECAST: CHARACTERIZING AND COMMUNICATING UNCERTAINTY FOR BETTER DECISIONS USING WEATHER AND CLIMATE FORECASTS

- “…no forecast is complete without a description of its uncertainty”
  - Need to act now
- Provide ensembles at various scales and applications
  - Contingent upon major upgrade of computational capabilities
- Engage/educate users, partners, social science in product development and usage
  - THORPEX, TPARC/IPY, NAEFS, Test-beds, etc
    - This workshop
- Provide access to all forecast data / verification information
  - Contingent upon major upgrade of telecommunication and storage facilities
- “NWS should take a lead role”
  - Major research and development enterprise (THORPEX)
    - NOAA THORPEX Program
    - Coordinate within NWS on transition to operations issues

NRC RECOMMENDATIONS

"NWS should take a leadership role"

"Executive attention to coordinating…uncertainty information"

- Organizational issues – build on existing infrastructure
  - THORPEX is major research and development enterprise (THORPEX)
  - CONOPS is process of revamping NWS operations
  - Strengthen coordination between THORPEX & CONOPS
- Resource issues
  - NOAA THORPEX Program under ST&I / W&W
    - Used at ESRL and NCEP
  - Environmental Modeling Program
  - NCEP redirection if needed?
  - OST involvement?
  - OSIP process for NWS expenditures?
- Computational handicap
  - Major impediment to improved services
- Telecommunication gap
  - Serious bottleneck

"LEADING ROLE" VS. COMPUTATIONAL RESOURCES

- December 1992
  - First operational ensemble forecast system at NCEP
    - Shortly before routine forecasts initiated at ECMWF
    - Credits to Eugenia Kalnay, Joe Irwin, Zoltan Toth, Steve Tracton

- Current (October 2006) status
  - NCEP 14 members at T126L28
    - 4 times per day, out to 16 days
  - ECMWF 50 members at T399L62
    - Twice per day, out to 10 (soon 15) days

- Ratio of ECMWF vs. NCEP resources on global ensemble
  - 1-10 days 114 times less resources used
  - 1-16 days 35 times less resources used

- Inadequate level of general NWP computational resources
  - ~10 years behind leading center(s)
    - Serious handicap

PATH FROM THORPEX RESEARCH TO NOAA OPERATIONS

PHASE	***BASIC RESEARCH***	***APPLIED RESEARCH***	***TRANSITION TO OPERATIONS***	***NOAA OPERATIONS***
What?	***Answer Science Questions***	***Develop Methods***	***Prepare for Implementation***	***Generate Products***
Who?	***External investigators***	***NOAA Laboratories***	***Global Test Center / NCEP***	***NCEP Central Operations***

NSF, DOD, NASA

Financial Support? ***NOAA THORPEX PROGRAM*** ***NOAA NWS***

CONOPS

BACKGROUND

VALUE OF PROBABILISTIC FORECASTING

- Potential economic value of probabilistic forecasts
  - "…the value of reliable – and even moderately unreliable – probabilistic forecasts generally exceeds the value of … categorical forecasts" - Murphy 1977

- Potential economic value of ensemble forecasts
  - "… a winder range of potential users can benefit from the ensemble than from the control forecasts … the ensemble offers more economic value than the control forecasts" – Zhu el al. 2002

- Operational forecasting implications
  - "…important implications for operational forecasting … desirability of formulating and disseminating a wide variety of weather forecasts in probabilistic terms…" Murphy 1977
  - "A weather forecast is … not complete unless it is expressed in the form of probability distributions." - Zhu el al. 2002
  - "Uncertainty is thus a fundamental characteristic of weather, climate, and hydrological prediction, and no forecast is complete without a description of its uncertainty." NRC Report: "Completing the Forecast", Ban et al., 2006

USER REQUIREMENTS:
PROBABILISTIC FORECAST INFORMATION IS CRITICAL

ECONOMIC VALUE OF FORECASTS

Given a particular forecast, a user either does or does not take action (eg, protects its crop against frost) *Mylne & Harrison, 1999*

OBSERVATION	FORECAST: *YES*	FORECAST: *NO*
YES	H(its) ***Mitigated Loss***	M(isses) ***Loss***
NO	F(alse alarms) ***Cost***	C(orrect rejections) ***No Cost***

$$\text{Mean Expense}_{fc} = hML + mL + fC$$

$$\text{Mean Expense}_{perf} = oML$$

$$ME_{cl} = \min[oL,\ oML + (1-o)C]$$

$$\text{Value} = \frac{ME_{cl} - ME_{fc}}{ME_{cl} - ME_{perf}}$$

o=climatological frequency

Optimum decision criterion for user action: P(weather event)=C/L

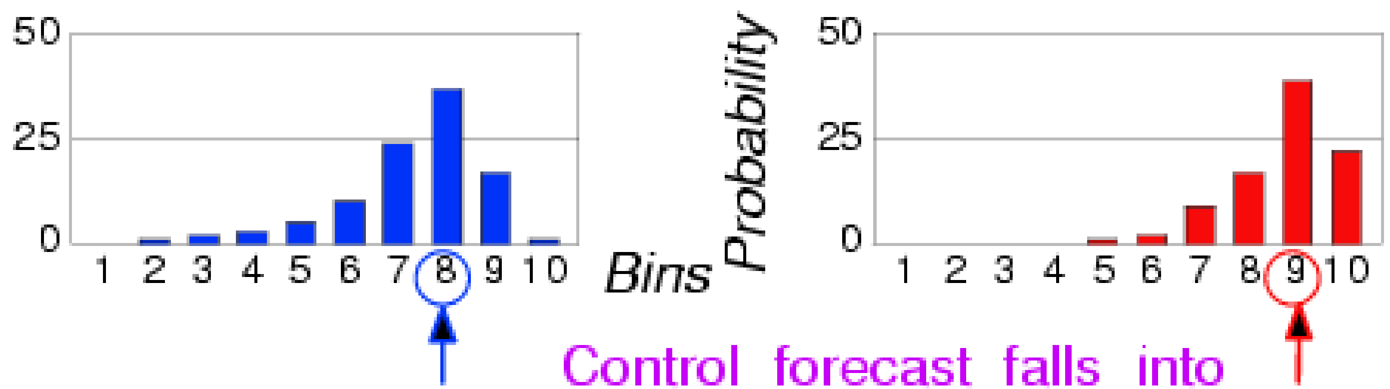
(Murphy 1977)

FORECASTING IN A CHAOTIC ENVIRONMENT –

PROBABILISTIC FORECASTING BASED A ON SINGLE FORECAST –

***One integration with an NWP model,* combined with past verification statistics**

DETERMINISTIC APPROACH - PROBABILISTIC FORMAT



•Does not contain all forecast information

•Not best estimate for future evolution of system

•***UNCERTAINTY CAPTURED IN TIME AVERAGE SENSE -***

•**NO ESTIMATE OF CASE DEPENDENT VARIATIONS IN FCST UNCERTAINTY**

BACKGROUND

THORPEX LINKS

ENSEMBLE-BASED DATA ASSIMILATION INTERCOMPARISON PROJECT

- GOAL:
  - Develop new ensemble-based DA methods and test against operational 3DVAR
- Collaborative project funded by NOAA THORPEX program
  - Coordinated by Zoltan Toth

PARTICIPANTS		*CONTRIBUTIONS*	*FUNDING*
ORGANIZATIONS	*SCIENTISTS*		
ESRL/PSD	Jeff Whitaker, Tom Hamill	EnSRF, model errors	NOAA THORPEX
University of MD	Istvan Szunyogh, Brian Hunt, Eric Kostelich, Eugenia Kalnay	LETKF, model bias	NOAA THORPEX
CIRA/CSU	Milija Zupanski, Arif Albayrak	MLEF, dual resolution, satellite bias	NOAA THORPEX
NRL	Craig Bishop, Daniel Hodyss	Background covariance representation	NOAA THORPEX
NCAR / DART	Jeff Anderson, Joe Tribbia	EAKF, parameter estimation	Other
NCEP/EMC	Mozheng Wei, Yucheng Song	Operational Benchmark, Evaluation/Implementation	NOAA THORPEX

THORPEX LINKS

ENSEMBLE-BASED DATA ASSIMILATION INTERCOMPARISON PROJECT

- Parallel development of two alternative paths:
  - Ensemble-based DA – led by Jeff Whitaker & Tom Hamill
  - Use of ensemble information within ¾ DVAR – led by Mozheng Wei & D. Parrish

TIMELINE	*TASKS*	
	ENSEMBLE-BASED DATA ASSIMILATION	VARIATIONAL APPROACH
Jan 2004	Planning started	
June2005	Build algorithms, conventional data tests	Develop Ensemble Transform (ET) perturbation scheme
June 2006	Expand algorithms, satellite data tests	Parallel testing, operational implementation of ET
June 2007	Complete algorithms, full inter-comparison	Analysis error variance estimates from 3DVAR
June2008	Prototype built	Operationally implement 3DVAR variance estimate into ET
June 2009	Off-line tests / enhancements	Develop methods for using ensemble covariance info in 3DVAR
June 2010	Parallel tests, Decision point	
Dec 2010	Operational implementation of better performing alternative	

THORPEX LINKS

ASSESSING / REPRESENTING MODEL-RELATED UNCERTAINTY

- Goals:
  - Improved understanding of model related uncertainties
  - Improved representation of model-related uncertainties in ensembles

- Participants / Contributions
  - Joao Teixeira (NURC) & Carolyn Reynolds (NRL)
    - Topic
  - & Eugenia Kalnay (Univ. Maryland
    - Topic
  - Jian-Wen Bao (ESRL/PSD)
    - Model perturbations (GFS physics?)
  - Dingchen Hou (NCEP)
    - Stochastic perturbations

- Funded by NOAA THORPEX program

THORPEX LINKS
OBSERVING SYSTEM DESIGN

- Goals:
  - Optimize global observing network
  - Develop adaptive data collection and processing algorithms

- Participants / Contributions
  - Bob Atlas (AOML), Dave Emmitt (SWA)
    - Observing System Simulation Experiment development
  - Chris Velden, Howard Berger (SIMSS)
    - Adaptive processing of satellite wind measurements
  - Terry Hock (NCAR)
    - Miniature dropsondes for use in drifsonde platforms
  - Yucheng Song, Michiko Masutani (NCEP)
    - Adaptive observational techniques, OSSE

- Funded mainly by NOAA THORPEX program

THORPEX LINKS
STATISTICAL POST-PROCESSING

- Goals:
  - Bias correct ensemble forecasts on model grid
  - Downscale bias-corrected ensemble onto finer (NDFD) grid

- Participants / Contributions
  - Richard Verret, Laurie Wilson et al (Meteorological Service of Canada)
    - NAEFS bias correction (BMA, etc)
  - Roman Krzisztofowicz (Univ. VA)
    - Bayesian Processor for Ensembles research and method development (NSF funding)
  - Yulia Gel (Univ. Waterloo)
    - Bias correction and downscaling research (self-supported)
  - David Unger (CPC)
    - Moment-based bias correction for NAEFS
  - Bo Cui (NCEP)
    - Moment-based bias correction, downscaling

- Funded partially by NOAA THORPEX program

THORPEX LINKS
PRODUCT DEVELOPMENT

- Goals:
  - Develop new numerical modeling applications
  - Develop new product generation tools and products

- Participants / Contributions
  - Scott Jacobs et al. (NCO)
    - NAWIPS ensemble functionalities
  - Richard Verret et al. (Meteorological Service of Canada, MSC)
    - NAEFS web-based products
  - David Unger et al. (CPC) and Richard Verret et al. (MSC)
    - Week-2 NAEFS products
  - Bob Grumbine (EMC)
    - Sea ice ensemble application
  - Dingchen Hou (EMC)
    - River flow ensemble application
  - Steve Silberberg, Binbin Zhou (NCEP)
    - Aviation weather guidance
  - Yuejian Zhu (NCEP)
    - NAEFS coordination

- Supported partially by NOAA THORPEX program

BACKGROUND

- **WHY WE NEED ENSEMBLES?**
  - Scientifically   Capture case dependent variations in uncertainty
  - Users   Downstream applicatns forced by nonlinear trajectories

- **WHEN ARE THEY CRITICAL?**
  - Case-dependent variations in chaotic and model error growth

- **HOW CAN THEY BE GENERATED?**
  - Multiple integrations
    - Initial perturbations
    - Model perturbations
  - Bias correction
  - Product generation

- **ISSUES IDENTIFIED**
  - Often no consensus solution, further research needed to refine issues
  - Apparently most promising paths recommended

ENSEMBLES: HOW?

- How to represent initial value related uncertainty?
  - Perturb initial conditions
- How to represent model related uncertainty?
  - Perturb model integration
- How many sample trajectories needed?
  - Ensemble size
- How to convey forecasts?
  - Trajectories and derived products

- **Unified approach across all applications** when practical
  - Based on general scientific principles
    - Choices based on / supported by experimental results when possible
  - Computational feasibility considered
  - Facilitated by ESMF framework – common interface for
    - Initial & model perturbations, bias correction, product generation
  - Applicable in most cases
    - Adjusted if/when necessary
  - Maintenance economic

HOW TO REPRESENT INITIAL VALUE RELATED UNCERTRAINTY?

Two distinct problems:

- Estimate analysis uncertainty
  - All estimates ultimately sample based => difficult to disentangle from sampling
    - Implicit solutions (ie, not explicit pdf)
- Choice among sampling strategies, given an estimate
  - Brute force (“Monte Carlo”) sampling – “Perturbed Observations” method
    - Run multiple analysis cycles with perturbed observations (Canadian approach)
    - Both growing and non-growing error space sampled with realistic amplitude
      - Very poor sampling of myriad non-growing directions
        - » Noise hurts analysis performance
  - Directed sampling
    - Singular vectors – fastest growth for pre-selected time period
      - Transient growth emphasized
      - Computationally expensive
      - No general solution: If transient growth is
        - » Important - Need different perturbations for various lead times
        - » Not important - No need for SVs
      - Most often norm used is uncoupled from analysis error estimates
        - » Relevant dynamics identified via growth optimization calculations
        - » No successful DA applications
    - Random sampling in growing sub-space
      - Based on principle of breeding: Cycle growing perturbations
        - » Capture dynamics of system responsible for error growth
        - » Ignore noise
        - » Successfully used in most ensemble-based DA efforts: eg, ETKF, etc

HOW TO REPRESENT INITIAL VALUE RELATED UNCERTRAINTY?

Proposed solution: Random sampling in growing sub-space – ET / ETKF

- Link with DA
  - GSI – ET
    - Take error variance from GSI to specify ensemble perturbation level
    - Feed back information from ensemble into background error covariance
  - Ensemble-based DA – ETKF
    - Same ensemble principles, except 2-way interactions tuned simultaneously
      - "Ensemble Filter" or
      - Variational solutions

- General applicability
  - As long as transient behaviour is not dominating
  - Ideal for "downstream" applications
    - ET provides series of perturbed analyses consistent in time
      - Important for wave, land surface, etc ensembles where
        - » Instantaneous error/perturbation dependent on *forcing history*

- Need for collaboration where DA & ensemble overlap
  - *Analysis is not complete until uncertainty estimate provided and assessed*
  - Consider ensemble-based DA alternative for revised product suite

HOW TO REPRESENT MODEL RELATED UNCERTRAINTY?

Theoretically not well understood problem

- Numerical model different from reality
  - Truncated model representation in resolved
    - Space and time scales (dynamics, physics)
    - Processes (physics)
    - Other approximations (numerics)
  - Other approximations/errors in representing nature due to
    - Lack of full understanding of nature
    - Mistakes (science and coding bugs)
    - Lack of accounting for model related uncertainties (for ensemble applications)
      - Deficiency in perturbation growth
- Performance metrics for modelling guidance differ depending on use
  - Traditional, single forecasts
    - Minimize single forecast error
  - Ensemble requirement
    - Account for case dependent model related uncertainty
- Systematic effort needed
  - Incorporate capability of simulating model related uncertainty
    - Strong collaboration between modelling and ensemble communities
  - *Model is incomplete until all model related uncertainty can be simulated*

HOW TO REPRESENT MODEL RELATED UNCERTRAINTY?

- **Current approaches**
  - Stochastic perturbations
    - “Catch-all” efforts to represent effect of unresolved scales of motion on resolved scales
      - Increase growth of spread (ie, properly simulate reaql level of predictability)
  - Multi-model (version) method
    - Pragmatic effort
      - Works to minimize effect of unidentified modelling errors
      - Possibly reduces case dependent biases (that cannot be removed statistically)
    - High development/maintenance costs
      - Effort given up by pioneering center (MSC)
    - Scientifically not appealing
      - Admits fractured nature of our knowledge
      - Must transcend
- **Proposed solution**
  - Continue development of stochastic perturbation method
    - Perturb resolved scales within ensemble subspace
  - Continue use of multi-model approach
    - Share development/maintenance costs with other centers
      - NAEFS: MSC, FNMOC, others
      - SREF / High impact modelling: MSC, FNMOC?
  - Embrace new modelling paradigm for ensemble era
    - Enable models to simulate all known modelling related uncertainties
    - Develop/evaluate models in ensemble (and not single forecast) framework

MEMBERSHIP VS. MODEL RESOLUTION?

TRADE-OFF BETWEEN MODEL VS. PDF RESOLUTION (LIMITED RESOURCES)

- Step-wise changes from single forecast to ensembles of increasing size: 1 => N
  - Decrease in model resolution
    - Degrading fidelity / statistical consistency
      - Bias correction / downscaling becomes more demanding
  - Increase in membership
    - Improving statistical resolution (case-dependent variations in pdf captured)
      - Potentially better forecasts
- Membership questions
  - Fewer members needed in phase of
    - Linear error growth (short lead time);
      - Bias correction to generate pdf
  - More members needed in phase of
    - Nonlinear error growth (longer lead times)
      - Highly nonlinear phenomena, eg, hurricane genesis
        - » Bias correction / downscaling to improve fidelity
    - To resolve higher moments of pdf

MEMBERSHIP VS. MODEL RESOLUTION?

PROPOSED SOLUTION

- Considerations
  - Cannot increase membership with lead time
    - Must compromise, considering entire time range of ensemble forecast
  - Integrations must resolve phenomena of interest
    - Unless sophisticated statistical down-scaling techniques can be developed
  - Potential gain from more members capped by level of refinement in other parts of forecast process
    - No point refining one aspect of forecast process – skill limited by weakest link
      - No use of very large ensemble with poor model
      - Bias correction / downscaling can interpolate/extrapolate pdf based on smaller ensemble
- "Generic" configuration guidelines for maximum overall benefits
  - Higher resolution control for short lead time if beneficial
  - Lower resolution ensemble out to longer lead times
    - Benefits from combining hi-re control & lo-res ensemble at shorter leads?
  - Ratio of ~1:2 horizontal resolution for ensemble vs. hires control
  - O(10) membership

WHAT PRODUCTS?

HIERARCHY OF PROGRESSIVELY MORE COMPLEX PRODUCTS

- First moment only – traditional approach
  - Best estimate of first moment (univariate mean or median)
  - “Most Likely Forecast” (multivariate)
    - Close to mean at short lead time
    - Display value only in highly non-linear phase (no unique solution)
- Second moment related information added
  - Two bounds of distribution (10 & 90 percentile of pdf)
- Likely scenarios (with likelihood), if they exist (based on statistical tests)
  - Multiple modes (univariate)
  - Clusters (multivariate)
    - Display value
- Probability of events that are defined by
  - Single variable (univariate)
  - Multiple variables (multivariate)
    - Quantitative use
- Full information – preferred approach
  - Pdf (univariate)
  - Ensemble member “trajectories” (multivariate)
    - All relevant information in explicit form
      - Ideal for downstream user applications

HOW TO DISSEMINATE PRODUCTS?

MULTIPLE CHANNELS

- Routinely prepare and actively distribute often-used basic products
  - Univariate ensemble mean, spread, 10-90 percentiles, PQPF, etc
    - AWIPS, NAWIPS, ftp, web
- Stage all raw and bias-corrected ensemble forecasts on ftp sites
  - Internally, for product generation engine
  - On ftp sites, for professional users
    - Include hind-casts to permit user specific bias correction
  - Downscaled information will require orders of magnitude more storage
- On-demand access to derived information – web/ftp access
  - Use product generation engine
    - Accesses bias-corrected ensemble database
    - Derives any desired product
    - NAWIPS software developed in collaboration with NCO
  - NOMADS functionalities serve User Support Systems
    - Strongly encouraged by NRC panel report
  - Conflict with some private sector partners?

INTERFACE OF UNIFIED ENSEMBLE APPROACH WITH DIFFERENT FORECAST SYSTEMS

ENSEMBLE ASPECT / FORECAST SYSTEM		Initial Perturbations	Model Perturbations	Statistical Post-Pro.	Product Generation	Verification
MAJOR SYSTEMS	Coupled					
	Global					
	Regional					
	High-Impact					
DOWNSTREAM APPLICATIONS	Ocean wave					
	Sea Ice					
	Riverflow/ Landsurface					

- Experts working on different aspects of unified ensemble approach (columns)
- Others responsible for pulling together all pieces for specific systems (rows)
- Critical area: High-Impact ensemble systems
  - Unified framework for very high resolution ensemble, embedded into regional ensemble
    - Tropical storm; Severe weather; Storm surge; Fire weather; Air Quality, etc applications
    - Must share basic infrastructure to prevent proliferation of systems
    - Can adapt basic structure as needed for special applications (eg, different model versions used)
- Suggested long term goal: Variable resolution modeling
  - Single framework to address multiscale processes, replaces current global and multitude of LAM integrations
    - Simplified modeling, DA, ensemble infrastructure
  - Scientifically challanging - 5+ years (global resolution higher by then)
  - Adaptively configure model to serve all high impact cases with pre-defined priorities
  - THORPEX research/development resources may be available

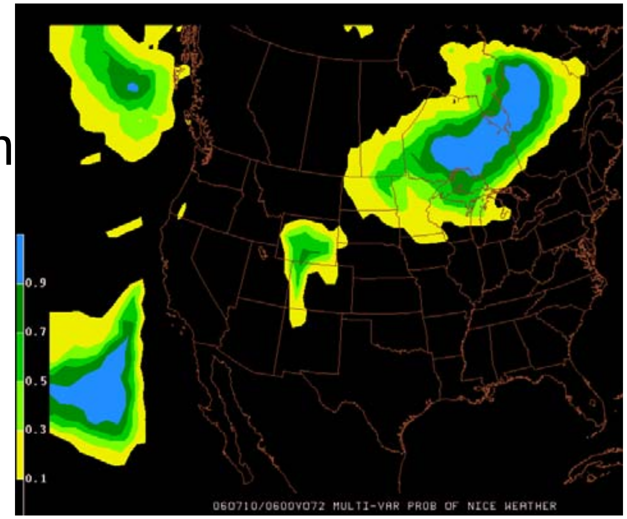
BACKGROUND

- Quality of forecast process depends on its
  - Statistical resolution
    - Ability to distinguish (provide unique signals before) future events
      - Temporal sequence foreseen - Inherent value of forecast process
        - » NWP methods used in 6 hr – 15 days range, statistical methods are not viable
    - Can be improved by NWP DA & model development
  - Statistical reliability
    - Ability to simulate (not predict) nature faithfully
      - Realism, fidelity - But no info on temporal sequence
    - Can be improved by
      - NWP model development
      - Use of statistical bias correction methods
        - » Better statistical methods
        - » Larger data sample - Can be perfectly corrected with large enough sample

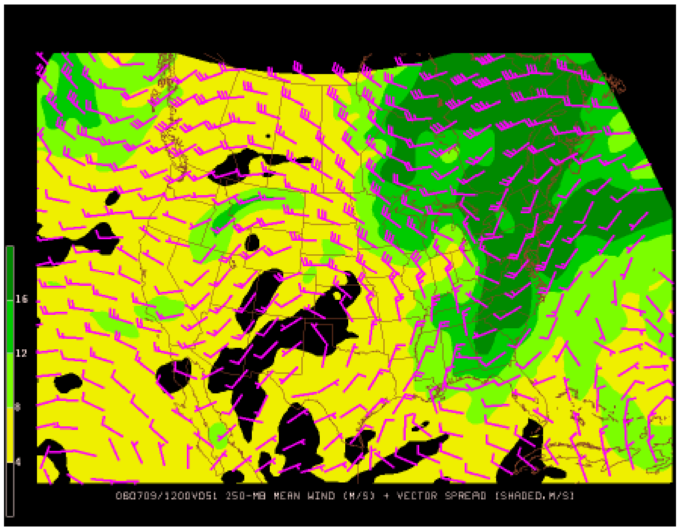
- Utility of forecasts depends on both resolution and reliability
  - Dual requirement of
    - Continuously improved DA & model

NAEFS PROJECT

- **Major accomplishments**
  - First operational multi-center ensemble system
  - NAWIPS ensemble functionalities
    - Integrated ensemble product generation suite

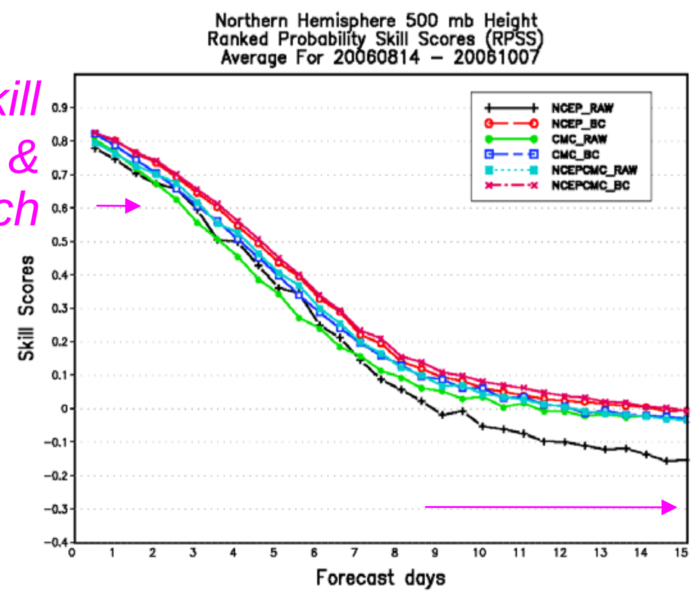


Probability of nice weather (precip/humidity/cape/mslp)



Ensemble mean wind and vector spread

1-6 days gain in skill due to bias correction & multi-center approach



Project participants

- EMC: Bo Cui, Yuejian Zhu, Richard Wobus, Dingchen Hou, Zoltan Toth
- NCO: David Michaud, Brent Gordon, Scott Jacobs, Steve Schotz, Luke Lin
- CPC: Ed Olenic, David Unger, Dan Collins

ENSEMBLES: HOW?

FOR EACH TIME SCALE / APPLICATION

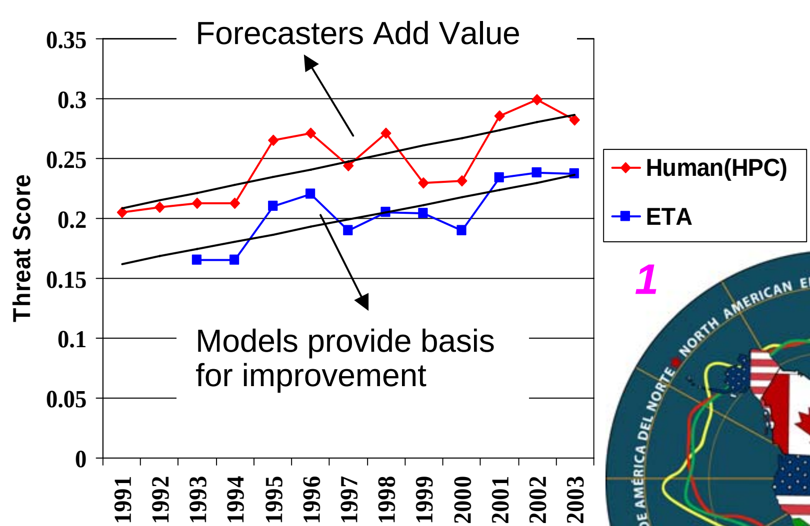
- How to represent initial value related uncertainty?
  - Perturb initial conditions
- How to represent model related uncertainty?
  - Perturb model integration
- How many sample trajectories needed?
  - Ensemble size
- How to convey forecasts?
  - Trajectories and derived products

- Unified approach when practical across all applications
  - Based on general science principles
    - Choices supported by experimental results when possible
  - Applicable in most cases
    - Adjusted if / when necessary
  - Maintainable with limited resources
  - Computationally feasible

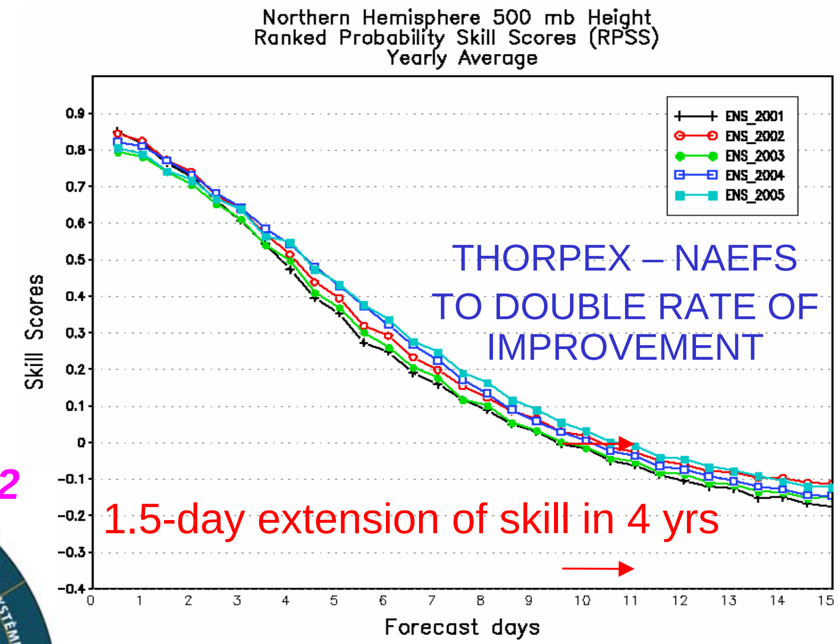
THORPEX GOAL

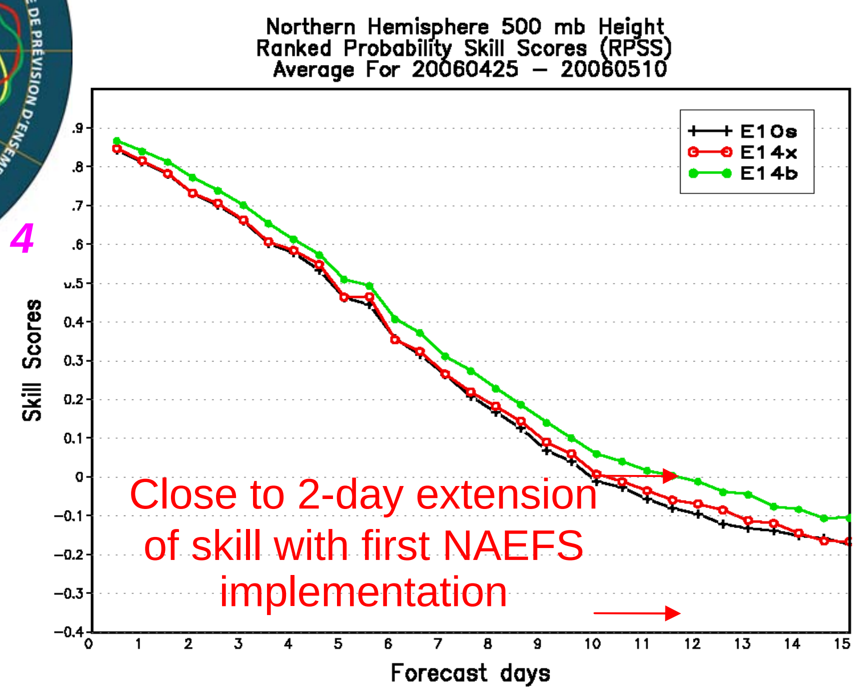
- Accelerate improvements in skill & utility of high impact forecasts
  - All improvements related to advances in NWP skill
    - We must accelerate improvements in NWP skill
    - Maintain/improve application procedures

Impact of Models on Day 1 Precipitation Scores



IMPROVEMENT IN PROBABILISTIC SKILL OVER PAST 4 YEARS





NORTH AMERICAN ENSEMBLE FORECAST SYSTEM (NAEFS)

1 2 3 4

- Operational multi-center ensemble system
- **Significant acceleration** in skill
  - Joint ensemble research
    - More achieved in one implementation than in previous 4 yrs
  - *Implementations at participating centers have immediate impact for all participants*
    - Shortcutting the typical 2-3 year development path that takes to adapt changes internally

NAEFS Performance Review

Appendix 6

KEY PERFORMANCE MEASURES

Improvement in Ensemble Forecasts					
Requirement		**Threshold**	**Actual 25Apr-10May06**	**Variance**
Ensemble Mean 3-14 Day Lead Time	Bias Reduction (%)	50%	30-70%	Met or exceeded in Tropics & up to D3 elsewhere; slightly below otherwise
	RMS Error Reduction (%)	10%	Up to 10%	Met up to D3, below expected D4 and beyond
Improvement in Ensemble-based Probabilistic Forecasts	3 Day	6 Hours	12 hrs	Exceeded
	7 Day	12 Hours	16 hrs	Exceeded
	10 – 14 Days	24 Hours	48 hrs	Exceeded

-

NOAA THORPEX IMPLEMENTATION PLAN

PROGRAM COMPONENT	ACTIVITIES	DELIVERABLES			PARTICIPANTS/ CONTRIBUTORS*	RECIPIENT
		2006-2008*	2009-2011	2012-2014		
Adaptive collection & use of observations	*Develop, test, and prepare for operational implementation adaptive methods for the collection and processing of observational data*					GEOSS
	In-situ platforms	Develop new adaptively usable platforms/sensors	Test new platforms/sensors	Assess new platforms / sensors / methods in optimal mix of observations	NCAR, UAS, ST&I	NOAA/GIFS operations
	Satellite platforms	Develop new methods for adaptive collection & use of observations	Test new methods		NESDIS, JCSDA, CIMMS, ETL	NPOESS operations
	Adaptive methodologies	Improved methods for global applications	Develop integrated adaptive observ. system		U. Miami, NRL	NWS operations
Integrated data assimilation & forecasting	*Develop, test, and prepare for operational implementation methods that integrate data assimilation and NWP forecasting procedures, including high impact modeling*					*EMP / NWS Operations*
	Ensemble-based DA project	Develop components of ensemble-based DA system	Develop & test ensemble-based DA system	Compare new system with "hybrid" DA system	CDC, NRL, CIRA, UM, NCAR	NCEP Data Assimilation System
	Model error representation	Develop components of model error representation	Integrate components of model error representation	Test new system in various applications, including DA	UM, NRL	NCEP Ensemble Forecast System
	High-impact modeling	Develop wave, ice, land surface model components	Develop adaptive grid-based global-to-regional modeling	Test new modeling system in global demo project	SAIC	NCEP GFS/NAM systems
User controllable probabilistic products	*Develop, test, and prepare for operational implementation new products and user procedures for probabilistic early warning of high impact weather*					*LFW, NCEP Service Cntrs*
	Statistical post-processing	*Develop basic methods for ensembles*	Add downscaling methods	Refine/integrate methodologies	MDL, CDC, IRI	Hydrology
	High impact applications	Develop application tools	Provide probabilistic guidance during field & demonstration programs		AMMA, IPY, IRI, HEPEX, GIFS	NAEFS
	Equitable use of resources	Study value & cost of forecasts	Develop equitable use plans at national level	Develop equitable use plans at international level	SIP, Stratus, IRI	GIFS
Weather-Climate link	*Develop, test, and prepare for operational implementation jointly with the climate community an observing, data assimilation, forecasting and application suite providing seamless forecasts for days 15-60*					*EMP, LFW*
	Observing system	Compare observing needs of weather & climate forecasting	Develop synergistic observing practices	Optimize observing network for weather & climate forecasting	ST&I, OGP, CTB	GEOSS
	Data assimilation & forecasting	Study coupled ocean-atmosphere model initialization & drift problems	Demonstrate improved skill via use of atmosphere/land/ocean initial conditions	Coupled atmosphere/land/ocean forecast model applications for day 1-60	CTB, OGP	NCEP Operations
	Applications	Compare weather & climate forecast applications	Develop tools suitable for both weather & climate forecasts	Demonstrate value of seamless weather-climate forecast applications	IRI, OGP, CTB	CPC, WFOs, GIFS
Global Operational Test Center	*Facilitate testing of successful candidate observing, DA, modeling & application tools*		*Integrate all research results by testing first in OSSE, then pre-implementation environment, prototypes of new THORPEX NWP forecast system*			*Research community*
	Test-bed for OSE, DA, NWP methods	Set up facility; Data impact studies; Targeting guidance	Serve research users; Data impact studies; Targeting guidance		JCSDA	AMMA, IPY, other demonstration / field projects
	Observing System Simulation Framework	Make system accessible to external researchers	Test THORPEX NWP components	Test integrated forecast system	JCSDA, SWA	NAEFS, GIFS
	Ensemble archive	Set up global archive	Augment global, add regional archive	Augment global/regional archives	NCDC, NCAR	NAEFS, TIGGE

*NOAA THORPEX Funding

*NOAA THORPEX Grantee

NORTH AMERICAN ENSEMBLE FORECAST SYSTEM (NAEFS)

- International project to produce operational multi-center ensemble products

- Combines global ensemble forecasts from Canada & USA
  - 30+ members per cycle, 2 cycles per day from MSC & NWS
    - 6-hourly output frequency out to 16 days
    - 1x1 lat/lon resolution

- Generates products for
  - Intermediate users
    - E.g., weather forecasters at NCEP Service Centers (US
  - Specialized users
    - E.g., hydrologic applications in all three countries
  - End users
    - E.g., forecasts for public distribution in Canada (MSC), Mexico (NMSM), Caribbean, South America, Africa (AMMA)

- Prototype ensemble component of THORPEX Global Interactive Forecast System (GIFS)
  - Operational outlet for THORPEX research using THORPEX Interactive Grand Global Ensemble (TIGGE) archive

MAJOR NOAA THORPEX TRANSITIONAL PROJECT

- North American Ensemble Forecast System (NAEFS)
  - Coordinated among National Weather Services of Canada, Mexico, US

- NAEFS provides
  - Operational requirements for THORPEX community
    - Interface to research community
  - Outlet for all THORPEX research
    - Must improve NAEFS performance to be valuable
  - Delivery mechanism for NA THORPEX
    - Forecast product distribution to less developed regions
    - Application procedures must connect with NAEFS for real-time use
  - Training opportunity at national & regional levels
  - Performance measure for NA THORPEX
    - Double rate of forecast improvement during 2005-2014 over 1995-2004
    - Detailed list of probabilistic measures to be developed
      - Mainly measures related to high impact events
        - » Extreme temperature (for extended periods?)
        - » PQPF
        - » High winds (tropical storms)
        - » Severe weather (based on CAPE & shear?)
      - Vanilla measure for long term performance reference
        - » 1000 & 500 hPa height

BACKGROUND

WORKING GROUP PARTICIPANTS (26)

CONFIGURATION

Co-leaders: Jun Du and Mozheng Wei

Participants: Rick Knabb, Richard Wobus, Ed O'Lenic, Dingchen Hou

DATA ACCESS

Co-leaders: Yuejian Zhu and David Michaud

Participants: David Bright, Minh Nguy, Kathryn Hughes

STATISTICAL POST-PROCESSING

Co-leaders: Paul Dallavalle & Zoltan Toth

Participants: Keith Brill, Andrew Loughe, DJ Seo, David Unger

PRODUCTS & TRAINING

Co-leaders: Jeff McQueen and Pete Manousos

Participants: Paul Stokols, Fred Mosher, Paul Janish, Linnae Neyman, Bill Bua, Joe Sienkiewicz, Binbin Zhou

ADDITIONAL WORKSHOP PARTICIPANTS (14)

Steve Tracton, Mike Halpert, Brian Gockel, Brent Gordon, Mark Antolik, Barbara Stunder, Michael Graf, Dave Plummer, Steve Schotz, Jon Mittelstadt, Malaquias Pena, Glen Zolph, Steve Lord, David Caldwell

OVERALL ISSUES / *RECOMMENDATIONS*

- **Enhance coordination of ensemble-related efforts**
  - Among NCO Service Center users
  - Between users and NCO / EMC developers
  - Between global and regional ensemble groups within EMC
    - Share research, development, and operational procedures where possible
  - *Establish NCO / EMC / Service Centers Ens. Products Working Group*
  - *Continue (expand via telecom?) monthly Predictability Meetings*

- **Optimize NCO operational job stream with user input**
  - For improved integrated forecast decision support
  - *Periodically reevaluate job stream from user and science perspectives*

- **Reestablish Annual NCEP SOO Workshop**
  - Rotate focus of workshop among various topics
  - *Hold Ensemble User Workshop every 3-4 years*

ENSEMBLE CONFIGURATION - CURRENT STATUS

CONFIGURATION	***Global***	***Regional***
Cycles per day	4	2
Membership	10	15
Resolution	T126L28 till 7.5 days T62L28 beyond	48km
Time range	16 days	63 hrs
Model version(s)	Single GFS	ETA (2 conv. schemes), RSM
Initial perturbation	Breeding	Breeding
Boundary perturbations	N/A	Global ensemble

ISSUES -

COLLABORATIVE PROJECTS MUST ENABLE OPERATIONAL IMPLEMENTATIONS

- **Global**
  - North American Ensemble Forecast System – with Met. Service of Canada
    - Post-processing & product development – Aimed at operational applications
  - THORPEX – NOAA, NA, & international collaborators
    - Projects on initial and model related perturbations – Path to operations
- **Regional**
  - Northeast Energy Project – OAR & Industry collaborators
    - Heat wave forecast related research – Should transition into operations
  - WRF – FSL, NCAR, USAF and other collaborators
    - Potential for rapid development of next generation operational system

ENSEMBLE CONFIGURATION - RECOMMENDATIONS

- **Global ensemble**
  - Implement hurricane relocation for perturbed initial conditions
    - Experiment with techniques used successfully with GFS system
  - Continue efforts to build multi-center ensemble
    - Combine NCEP, ECMWF, MSC, JMA, FNMOC ensembles
    - Best possible multi-model approach (with added benefits of initial condition variability)
- **Regional ensemble (SREF)**
  - Consider running ensemble & hires ETA (WRF) control from same initial time
    - Utility of off-cycle ensemble (9 & 21Z) is limited when used with 12Z & 00Z controls
      - Differences between ensemble & hires control from different cycles hard to interpret
    - Closer coupling between ensemble & hires control allows proper interpretation of both
    - Alternative suggestions for computer resource allocation:
      - Increase less the resolution for both ensemble & hires control in future implementation
      - Decrease resolution for hires forecast beyond, eg, 36 hrs (if skill is not degraded)
      - Run ~5 initial/model perturbation members along with hires control, finish rest of ensemble later
      - Run 5 members from early, hires from final analysis, finish ensemble, run hires window for dominant clusters
      - Study feasibility of combining information from older ensemble with newer hires forecast (J. Du's suggestion)
  - Introduce 4 cycles per day, out to 84 hrs if possible – Run ensemble at 00, 06, 12, & 18Z
    - Will allow comparison of hires control and lowres ensemble, enhancing utility of both
- **Additional suggestions for both systems**
  - **Membership**
    - Evaluate effect of increased membership in combination with post-processing gains
  - **Spatial resolution** – *As computational resources increase*,
    - Increase ensemble resolution (~50-50% resources for hires control & lores ensemble)
  - **Initial perturbations**
    - Continue research aimed at better quantification of initial uncertainty
  - **Model error representation**
    - Continue research on stochastic model perturbations & model diversity

ENSEMBLE DATA ACCESS - CURRENT STATUS

- **Global ensemble** – 1x1 grid, pgrib, enspost, Sager file types
  - NCEP Service Centers All data available - Limited NAWIPS access
  - AWIPS Limited data out to 84 hrs – need 180 hrs (WAFS?)
  - NCEP ftp servers All data available
  - NWS server 2 cycles only – Need to add 06 & 18Z cycles?
- **Regional ensemble** (SREF) – GRIB212 (40km)
  - NCEP SCs All data available – Limited NAWIPS access
  - AWIPS No access to data – Need selected variables
  - NCEP ftp servers Selected variables only – All data needed?
  - NWS server None – Need to post data
- **Issues:**
  - Disc space usage – Inefficient due to use of multiple file formats
    - Same data packaged in various formats for convenient access and historical reasons
  - Bandwidth limitations – Ftp overload due to data access limited to prepared files
    - Typical user needs only fraction of downloaded data
  - Increase in data volume – Need advance planning to facilitate future data access
    - Ensembles from other centers; Increased resolution, membership

ENSEMBLE DATA ACCESS - RECOMMENDATIONS

- Provide access to all ensemble data (including members)
  - Allows optimal use of ensemble information by diverse user base
  - Should be feasible given low cost of disc storage space
    - Lower resolution ensemble has similar data volume to hires control
    - Temporary disc space limitations should be mitigated by
      - Freezing output resolution (or list of available variables)
- Facilitate user controlled access to data (e. g., NOMAD)
  - Allow users to choose what they want to download by
    - Selecting members, variable, level, time and spatial domain of interest
    - Providing basic functionalities to manipulate data (eg, download derived statistics only – see Products Working Group recommendations)
  - Consider for NAWIPS, AWIPS, and ftp dissemination
  - Eliminates need for duplicate data files
  - Significantly reduces bandwidth requirements
  - Prototype system exists (NOMAD, all global ensemble data available)
    - As interim solution until system operational, introduce split pgrib files?
- Shift to use of GRIB2 format
  - WMO sanctioned standardized & uniform format for ensemble data
    - Need for international ensemble data exchange (see Configuration WG)
  - x3 (for global) to x5 (for regional) reduction in file size

ENSEMBLE STATISTICAL POSTPROCESSING - CURRENT STATUS

- **NWP models, ensemble formation are imperfect**
  - Known model/ensemble problems addressed at their source
    - No “perfect” solution exists, or is expected to emerge
  - Systematic errors remain and cause biases in
    - 1^{st}, 2^{nd} moments of ensemble distribution
    - Spatio-temporal variations in 2^{nd} moment
    - Tails of distributions
- **No comprehensive operational post-processing in place**
  - MOS applied on individual members (global ensemble, MDL)
  - QPF calibration of 1^{st} moment (global ensemble, EMC & CPC)
  - Week 2 calibration with frozen system (global ensemble, CDC)
- **Issues:**
  - Users need bias-free ensemble guidance products
    - Bias-corrected ensemble members must be consistent with verification data
  - Algorithms must be relatively cheap & flexible for operational applications
    - Post-process on model grid first, then “downscale” to NDFD grid / observs?
  - Level of “correctible” details depends on
    - Bias signal vs. random error noise ratio
    - Sample size of available forecast/observation training data pairs
      - Relatively small sample for short-med. ranges – Capture regime dependent bias?
      - Much larger for extended ranges – Capture climatological bias via frozen system?

ENSEMBLE STATISTICAL POSTPROCESSING - RECOMMENDATIONS

- Develop techniques for two-stage statistical post-processing:
  - 1) Assess and mitigate biases on model grid with respect to analysis fields
    - Feedback to model / ensemble development
    - 1st moment correction based on: Time mean error; Cumulative distributions
    - 2nd moment correction based on: Time mean ratio of ens mean error & spread
    - Post-processed forecasts bias corrected with respect to reanalysis fields
      - Generate anomaly forecasts using global/regional reanalysis climatology
  - 2) Downscale bias-corrected fcsts from model grid to NDFD/observatn locations
    - “Smart” interpolator for bias correction and variance generation on fine scales
      - Multiple regression (MOS); Bayesian methods; Kalman Filtering; Neural nets
    - Apply downscaling methods on bias-corrected fields (no lead time dependence)
      - Use large reanalysis and corresponding observational data base (&/or NDFD analysis fields)
  - To describe ensemble-based pdf forecasts, use 3-parameter distributions
    - Test two methods, find best fitting analytic distribution (out of ~25 candidates)
      - Simple method: Fit actual ensemble data
      - Kernel approach: Find best fit to climate data, then apply it on each member w/weight
- Operationally implement post-processing techniques
  - Apply basic bias-correction techniques centrally (NCO) to serve wide user base
    - Post-process all variables used from the ensemble (first model, then derived variables)
  - Disseminate bias-corrected forecasts on lowres ensemble model grid
    - Save disc and bandwidth resources
    - Keep raw forecast fields also accessible for special user processing needs
  - Use additional post-processing (if any) locally to address special needs, eg:
    - Hurricane forecasting

ENSEMBLE PRODUCTS - CURRENT STATUS

- **Product development software**
  - Some functionalities exist
    - Scattered around different developers/platforms/users
      - NCO operations
      - NAWIPS official build
      - NAWIPS development by NCEP SOOs
      - AWIPS
      - Other platforms
- **Products generated centrally by**
  - NCO Limited number of gridded products (operational)
  - EMC Additional set of gridded and web-based products (non-operational)
- **Issues:**
  - Lack of standard/common software toolbox for ensembles
    - Missing functionalities
    - Multiple software versions of existing functionalities
    - Duplication of efforts
  - Lack of comprehensive, well designed set of products
    - Non-standard set of products/displays (global vs. regional ensembles, etc)
    - NAWIPS, AWIPS requires access to products (web not enough)
    - Need for operationally generated and supported web product suite

ENSEMBLE PRODUCTS - RECOMMENDATIONS

- Develop a software toolbox for interrogating ensemble data
  - Establish development team - NCO, EMC, NCEP Service Center experts
  - Compile list of required functionalities – See attached list
  - Develop standard software package (subroutines) for each functionality
    - Work in NAWIPS framework
    - Ensure software (subroutines) are portable to different platforms
    - Ensure batch and on demand processing capabilities
    - Provide interactive processing/display capability where needed
    - Offer subroutines for use by AWIPS and broader inter/national community
    - Consider WRF, NAEFS, THORPEX applications
- Establish operational/local product generation suites
  - Use standard software toolbox for product generation
  - Identify list of products – See attached list for NCEP Service Center requests
  - Type of product generation based on typical usage:
    - Every day - Generate centrally (NCO), produce multiple file formats (NAWIPS/web)
    - Occasionally - On demand, locally (NCEP Service Centers)
    - Interactively - On screen manipulation/interrogation (NAWIPS)
  - Distribute centrally generated products within NAWIPS, AWIPS
  - Set up and maintain operational NCEP ensemble product web page
    - Post products on web page for use by broader community
    - Provide limited interactive query tools if desired (examples within NOMAD)
    - Contribution to THORPEX goals for use by less developed nations

ENSEMBLE PRODUCTS - FUNCTIONALITIES

List of centrally/locally/interactively generated products required by NCEP Service Centers for each functionality are provided in attached tables (eg., *MSLP, Z,T,U,V,RH, etc, at 925,850,700,500, 400, 300, 250, 100, etc hPa)*

	FUNCTIONALITY	***CENTRALLY GENERATED***	***LOCALLY GENERATED***	***INTERACTIVE ACCESS***
1	Mean of selected members			
2	Spread of selected members			
3	Median of selected values			
4	Lowest value in selected members			
5	Highest value in selected members			
6	Range between lowest and highest values			
7	Univariate exceedance probabilities for a selectable threshold value			
8	Multivariate (up to 5) exceedance probabilities for a selectable threshold value			
9	Forecast value associated with selected univariate percentile value			
10	Tracking center of maxima or minima in a gridded field (eg – low pressure centers)			
11	Objective grouping of members			
12	Plot Frequency / Fitted probability density function at selected location/time (lower priority)			
13	Plot Frequency / Fitted probability density as a function of forecast lead time, at selected location (lower priority)			

Additional basic GUI functionalities:

- Ability to manually select/identify members
- Ability to weight selected members

Potentially useful functionalities that need further development:

- Mean/Spread/Median/Ranges for amplitude of specific features
- Mean/Spread/Median/Ranges for phase of specific features

ENSEMBLE VERIFICATION – CURRENT STATUS

For lack of time, this topic was not discussed at the workshop

- Global ensemble verification package used since 1995
  - Comprehensive verification stats computed against analysis fields
  - Inter-comparison with other NWP centers
- Regional (SREF) verification package
  - Basic measures computed routinely since 1998
  - Probabilistic measures being developed independently from global ensemble
- Issues
  - Need to unify computation of global – regional ensemble verification measures
  - Unified framework must facilitate wide-scale national/international collaboration:
    - North American Ensemble Forecast System (collaboration with Met. Service Canada)
    - THORPEX International Research Program
    - WRF meso-scale ensemble developmental and operational activities
  - Facilitate wider community input in further development/enhancements
    - How to establish basis for collaboration with NCAR, statistical community, etc

ENSEMBLE VERIFICATION - RECOMMENDATIONS

- Design unified and modular ensemble/probabilistic verification framework
  - Data handling/storage
    - Use standard WMO file formats as ensemble data input
    - Allow non-standardized user/site specific procedures
  - Computation of statistics
    - Establish required software functionalities (scripts) and verification statistics (codes)
    - Jointly develop and share scripts/subroutines with standard input/output fields
    - Improvements to common infrastructure benefit all
    - Comparable scientific results, independent of investigators
  - Access/display of output statistics
    - Explore if standard output file format(s) feasible? Use text or FVSB-type files?
    - Develop/adapt display software for interactive interrogation of output statistics
      - Examples: FVS display system; FSL approach to WRF verification

- Develop and implement new verification framework
  - Utilize existing software and infrastructure where possible
  - Direct all internal ensemble-related verification efforts toward new framework
  - Share work with interested collaborators
    - Meteorological Service of Canada (subroutines, L. Wilson and colleagues)
    - FSL (display tools, A. Laugh)
  - Make new software available to national/international community
    - Coordinate further development with wider community (WMO, etc input)

ENSEMBLE VERIFICATION – DESIGN SPECIFICATIONS

Compute statistics selected from list of available

- Point-wise measures, including:
  - RMS, PAC for individual members, mean, median
  - Measures of reliability (Talagrand, spread vs. error, reliability components of Brier, RPSS, etc)
  - Measures of resolution (ROC, info content, resol. comps. of BSS, RPSS, potential econ.value, etc)
  - Combined measures of reliability/resolution (BSS, RPSS, etc)
- Multivariate statistics (e.g., PECA, etc)
- Variables & lead times – make all available that are used from ensemble

- **Aggregate statistics as chosen in time and space**
  - Select time periods
  - Select spatial domain (pre-designed or user specified areas)
- **Verify against observational data or analysis fields**
  - Scripts running verification codes should handle verification data issues
  - Use same subroutines to compute statistics in either case
  - Account for effect of observational/analysis uncertainty?
- **Define forecast/verification events by either**
  - Observed/analyzed climatology, e.g., 10 percentile thresholds in climate distribution
    - Automatically compute thresholds for forecast values
  - User specified thresholds – automatically compute corresponding climate percentiles
  - Ensemble members (like in Talagrand stats) – compute climate percentiles
- **Facilitate the use of benchmarks**:
  - Climatology, persistence, or specified prior forecast data set
- **Prioritize and find balance between**
  - Flexibility vs. complexity; operational vs. research use, etc

ENSEMBLE TRAINING

- CURRENT STATUS:
  - NCEP Training Material available since 2002 (P. Manousos)
  - COMET professional training module to be released soon (B. Bua)
    - Includes winter weather, severe weather, and general weather forecasting problems
  - Webcast module based on COMAP presentation by D. Bright (by 09/2004, B. Bua)
  - NWS WFO teletraining using VISITView (B. Bua, proposed)
    - Practical use of ensembles

- ISSUES:
  - Official NWS training opportunities/requirements not established
  - Training for professional national, international, and lay users needed
  - Share training resources nationally/internationally

- RECOMMENDATIONS:
  - Establish NWS formal ensemble training requirements
  - Consider organizing AMS Ensemble Training Workshops
    - Practicing broadcast etc meteorologists
    - Emergency managers
  - Share training material on national/international level
    - Establish NWS OS FAQ on ensembles
    - NCEP/HPC International Desk – Spanish/Portugese translations of existing material
    - Exchange ideas/material with WMO Ensemble Training initiative
    - Contribute to Socio-Economic Applications part of intl. THORPEX research program