Characterization of Forecast Uncertainty by Means of Ensemble Techniques

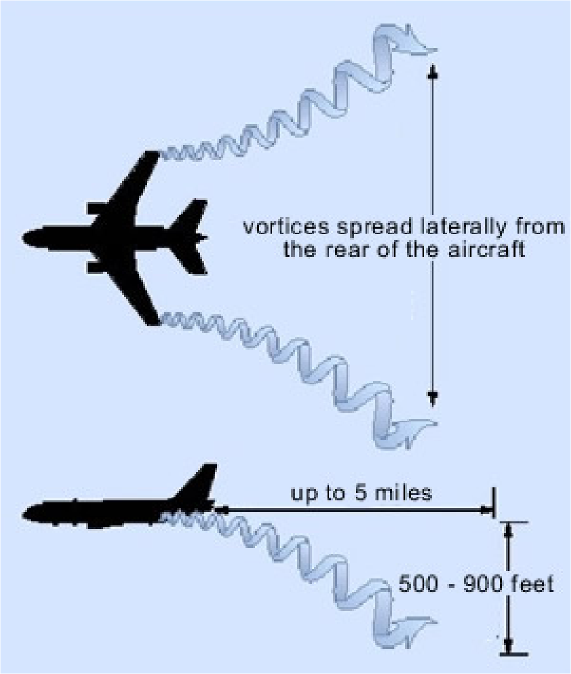
Matthias Steiner

National Center for Atmospheric Research
Boulder, Colorado, USA

Email: msteiner@ucar.edu

Short-Term Weather Forecasting for Probabilistic Wake Vortex Prediction
DLR, Oberpfaffenhofen, Germany, 10 – 11 May 2010

Motivation



Wake vortex impact mitigation builds upon . . .

- **Monitoring & prediction of wake vortices**
  - along aircraft path for approach/landing & takeoff/departure
  - counter-rotating vortices depend on weight, wingspan, & speed of aircraft
  - evolution & lifetime depends on ground proximity, wind & shear, turbulence, & thermal stratification

- **Grappling with uncertainties**
  - **observational limitations** (wind & temperature in boundary layer, including variability in space & time) & aircraft parameters (weight & speed)
  - numerical **prediction limitations** (physical understanding & parameterization, data assimilation, initialization, etc.) for both weather & wake vortices
  - sub-grid scale variability (parameterization or down-scaling)
  - **probabilistic forecasting** (e.g., **ensembles**)
  - critical thresholds (e.g., runway crosswind)
  - post-processing/**calibration** & verification

Ensemble Techniques

Various approaches possible . . .

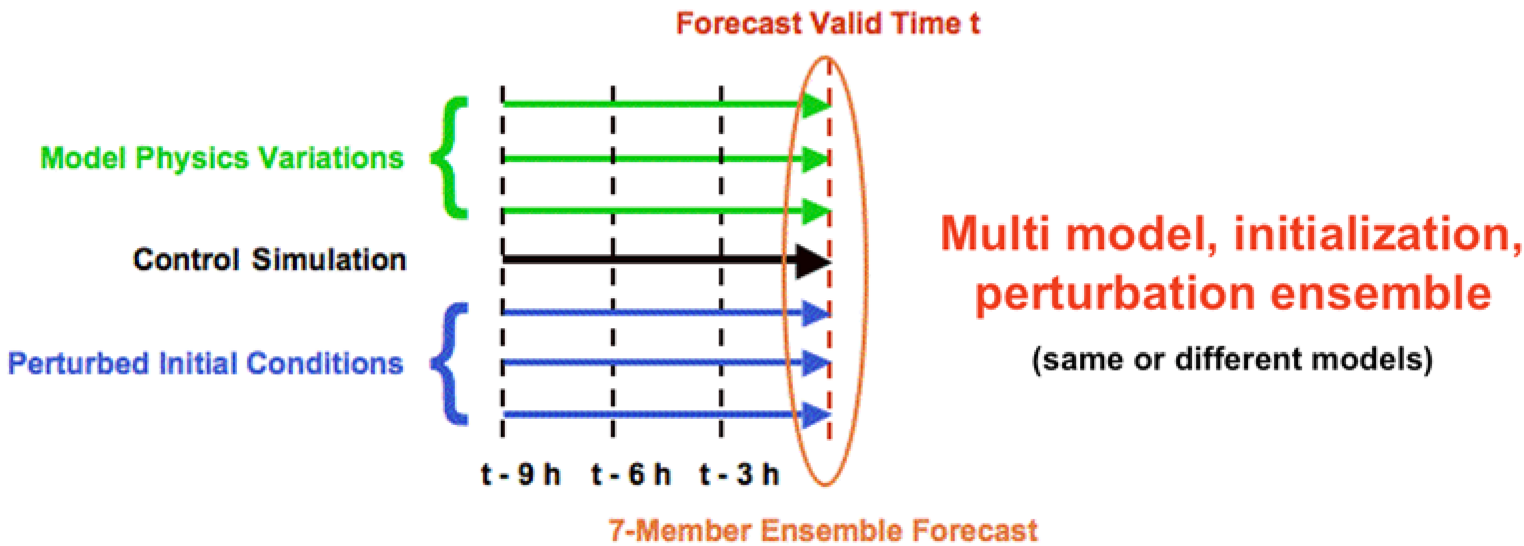
- **Multi model, initialization, perturbation ensemble**
  - Example #1: Hurricane track & intensity prediction
  - Example #2: Overbank flow & flood prediction
  - Example #3: En route air traffic capacity prediction
  - Example #4: Surface wind & energy prediction
  - Example #5: Winter road maintenance prediction

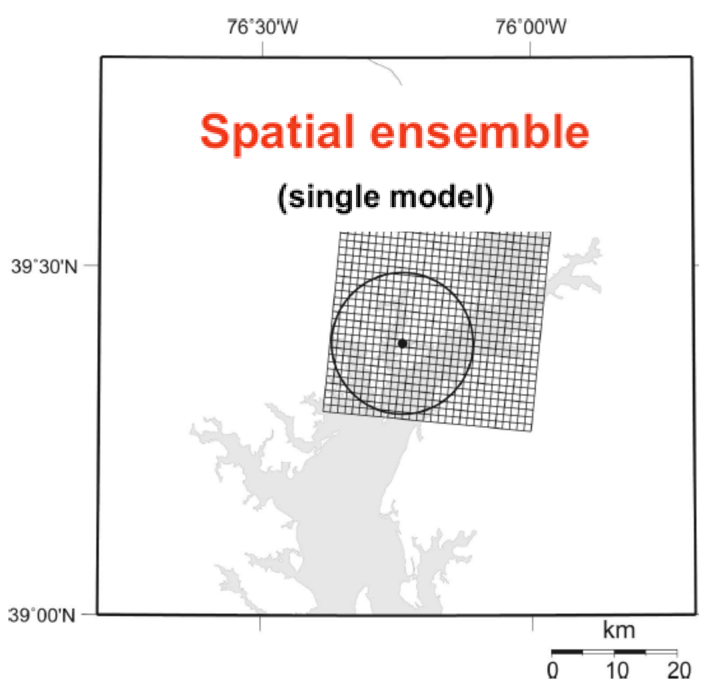
"rich man's ensembles"

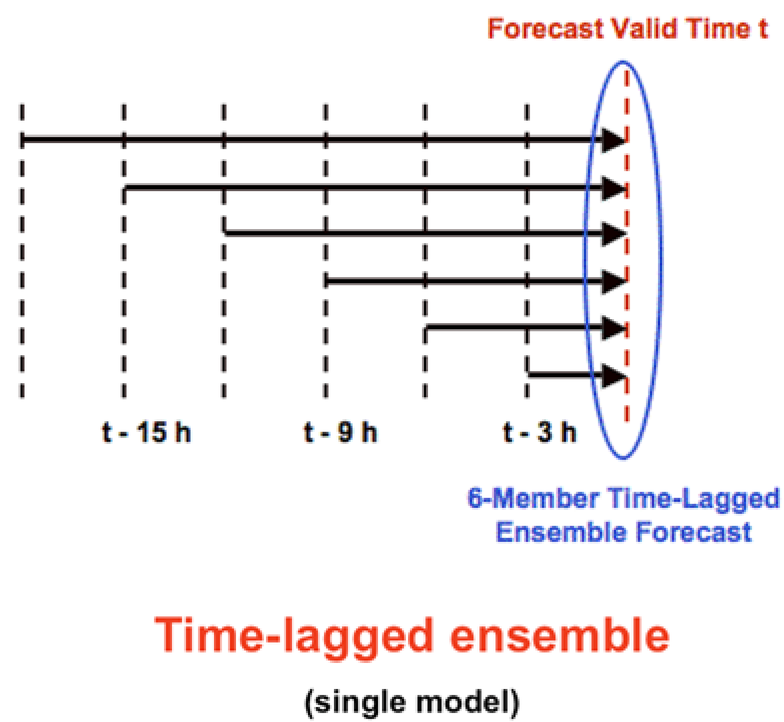
- **Time-lagged ensemble**
  - Example #6: En route air traffic capacity prediction
- **Spatial ensemble**
  - Example #7: Missile trajectory prediction
  - Example #8: Noise propagation prediction
  - Example #9: Pollution dispersion prediction
- **Diagnostic ensemble**
  - Example #10: Graphical turbulence guidance (GTG)

"poor man's ensembles"

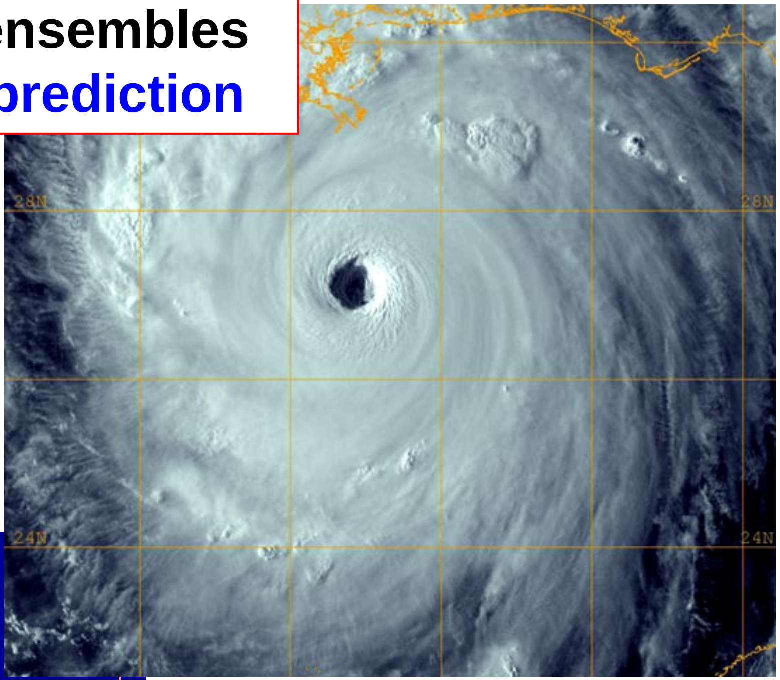
- **Combinations of the above**


Forecast Valid Time t
Model Physics Variations
Control Simulation
Perturbed Initial Conditions
t - 9 h
t - 6 h
t - 3 h
7-Member Ensemble Forecast
Multi model, initialization, perturbation ensemble
(same or different models)

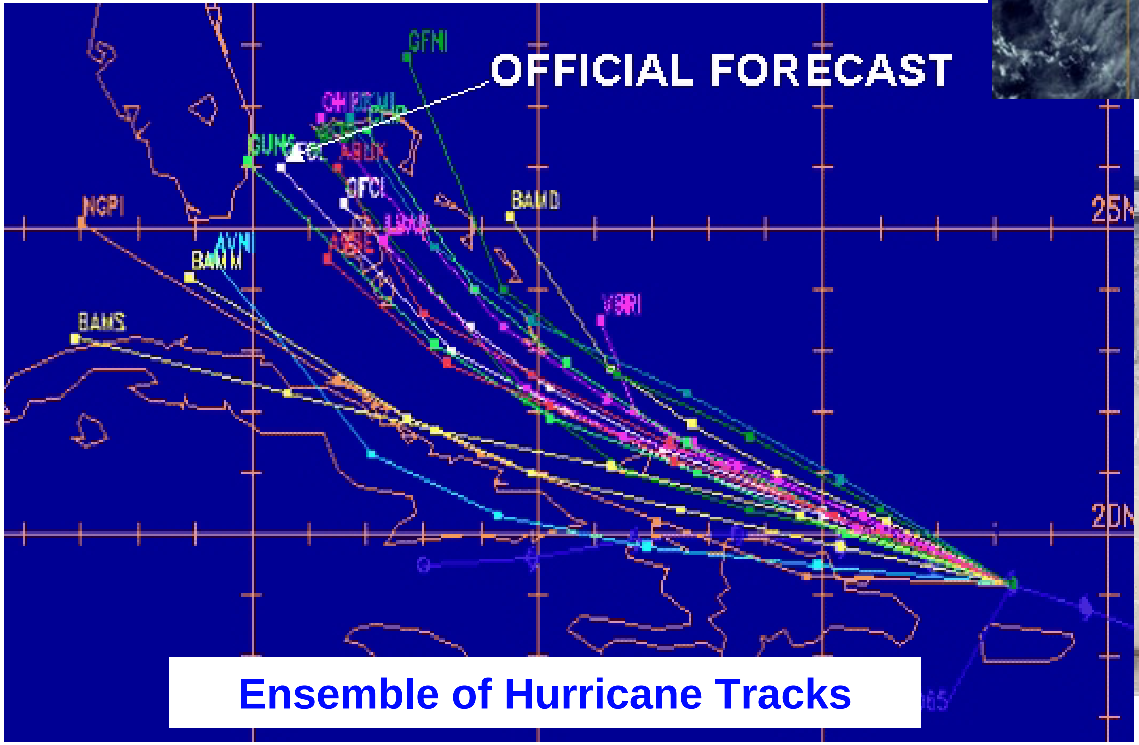

76°30'W
76°00'W
39°30'N
39°00'N
Spatial ensemble
(single model)
km
0
10
20


Forecast Valid Time t
t - 15 h
t - 9 h
t - 3 h
6-Member Time-Lagged Ensemble Forecast
Time-lagged ensemble
(single model)

Multi model, initialization, perturbation ensembles
Example #1: **Hurricane track & intensity prediction**



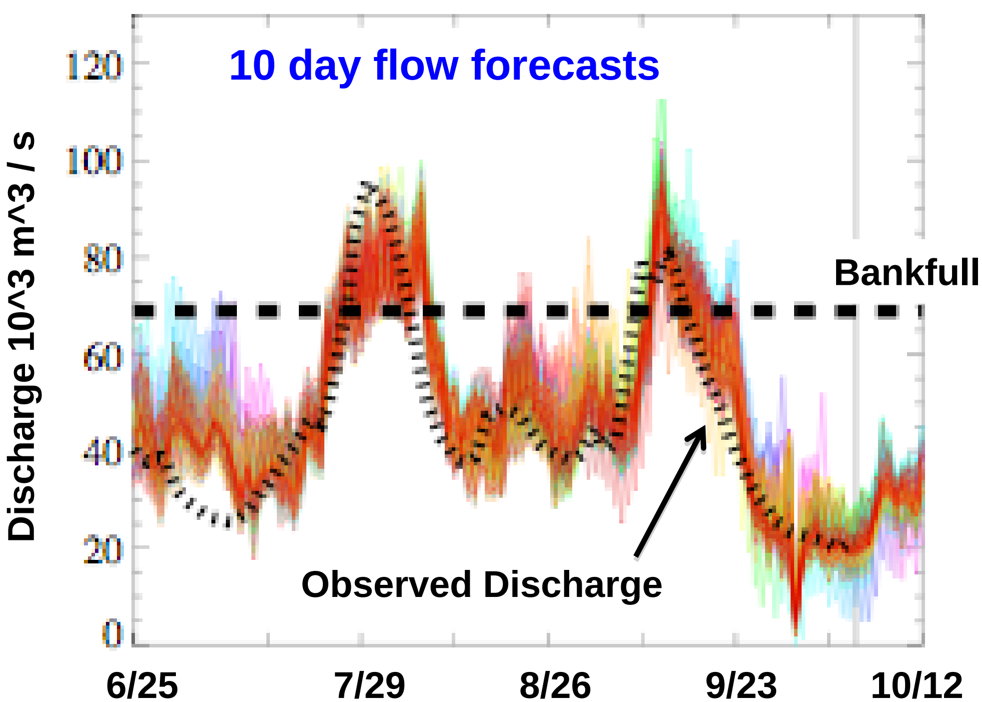
- Hurricanes can cause huge societal impacts
- Focus on storm track, timing, & intensity (both precipitation & wind)
- Translation of hurricane track, size & intensity ensemble into probabilistic evacuation area, storm surge, damage, disruption of services, economic impact, etc.

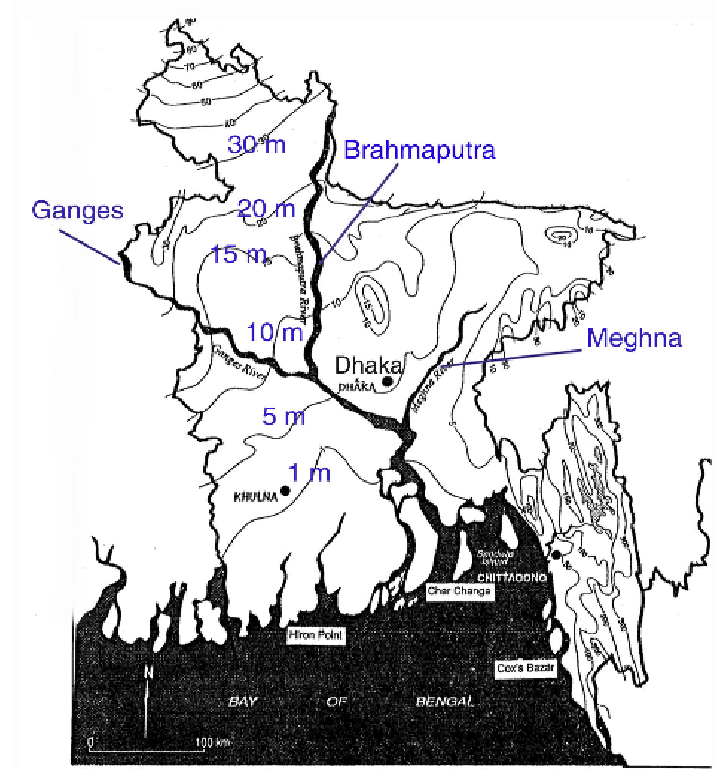


Ensemble of Hurricane Tracks

Multi model, initialization, perturbation ensembles
Example #2: **Overbank flow & flood prediction**

- Bangladesh has little upstream information for Brahmaputra & Ganges rivers
- Focus on amount of precipitation in neighboring countries
- Translation of ECMWF precipitation ensemble into probabilistic discharge, flow depth, inundation & evacuation area, etc.

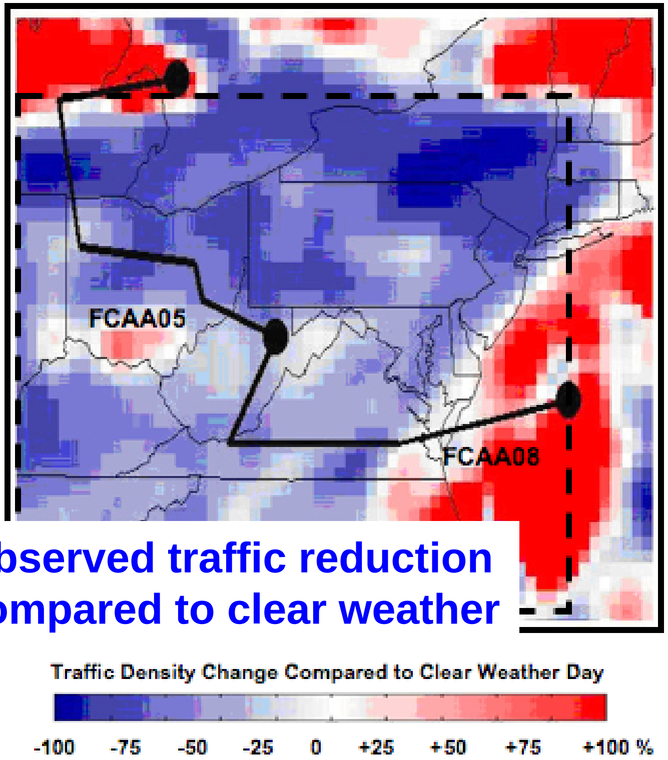




Multi model, initialization, perturbation ensembles

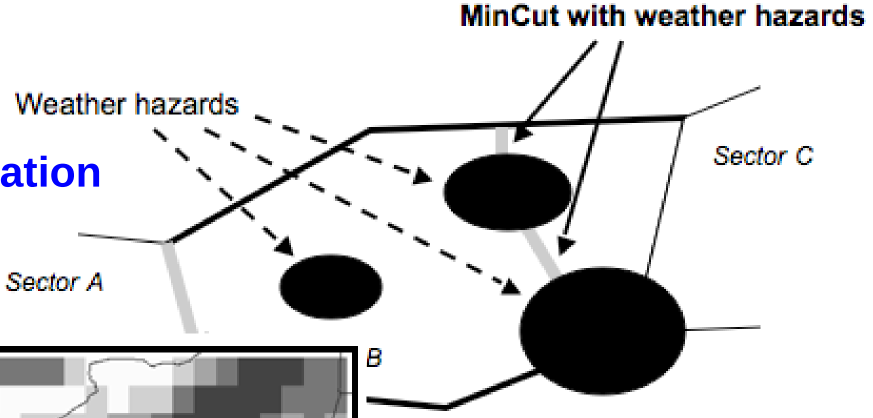
Example #3: **En route air traffic capacity prediction**

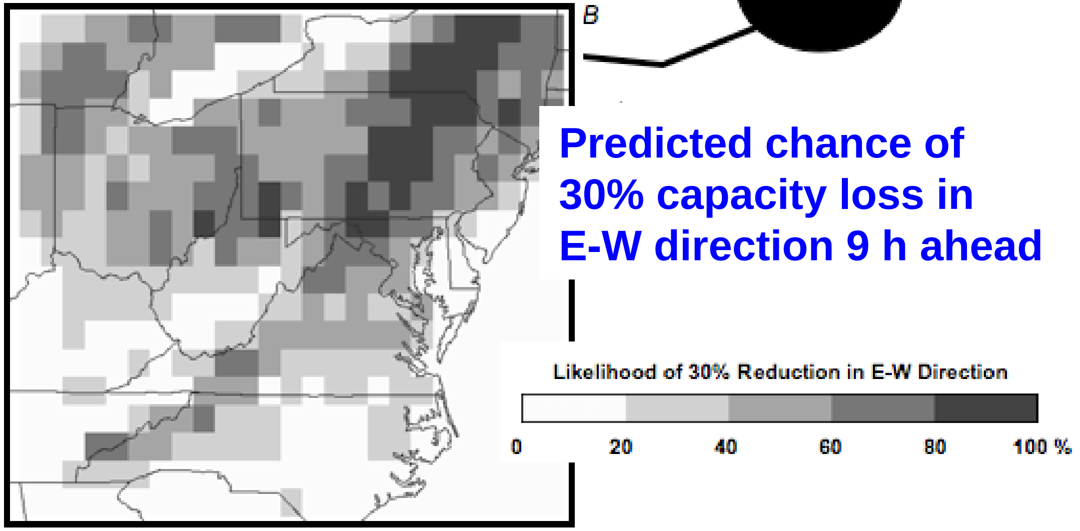
- Weather reduces usable space for air traffic
- Focus on storms & organization
  (e.g., intensity, echo tops, porosity of pattern)
- Translation of precipitation ensemble into
  probabilistic capacity loss



Observed traffic reduction compared to clear weather

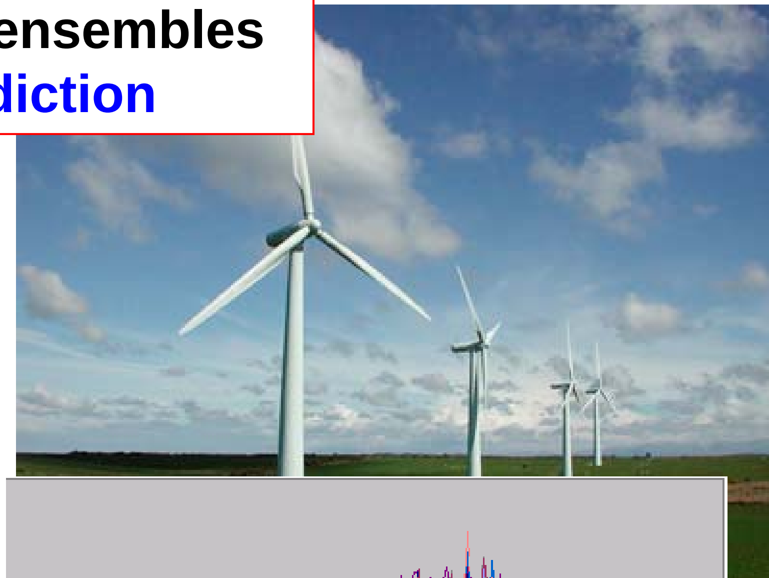
Translation



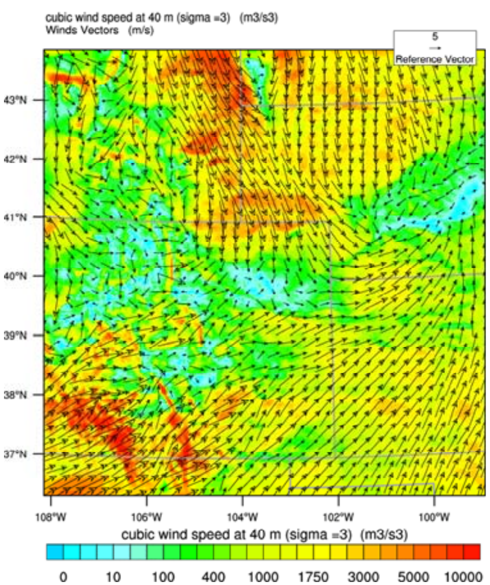


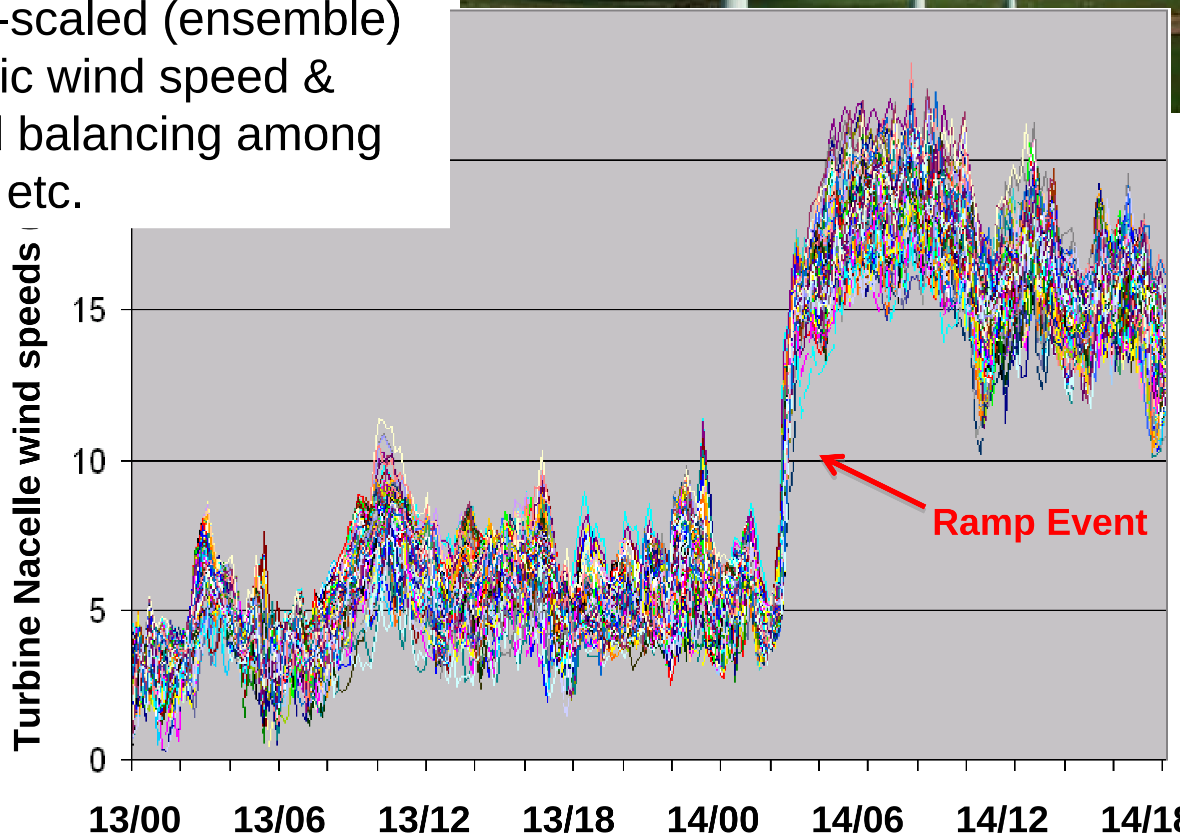
Predicted chance of 30% capacity loss in E-W direction 9 h ahead

Multi model, initialization, perturbation ensembles
Example #4: **Surface wind & energy prediction**



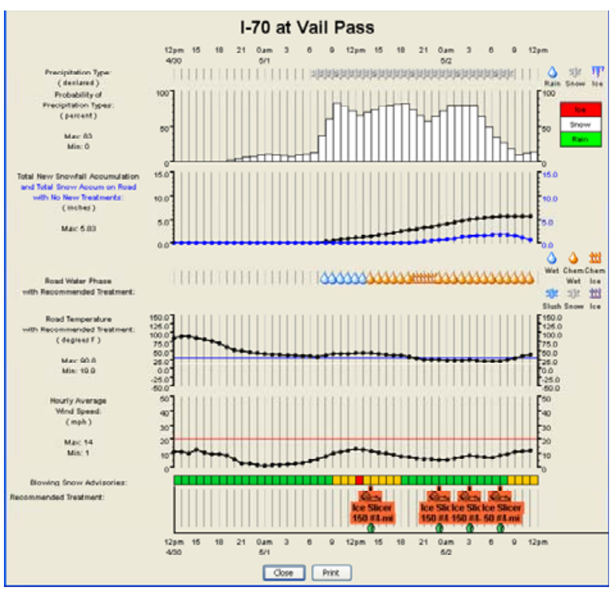
- Managing power grids based on different energy sources is significant challenge, especially for wind & solar energy harvesting
- Wind-generated power is non-linear function of wind speed
- Translation of LES down-scaled (ensemble) forecasts into probabilistic wind speed & ramp events timing, load balancing among different power sources, etc.

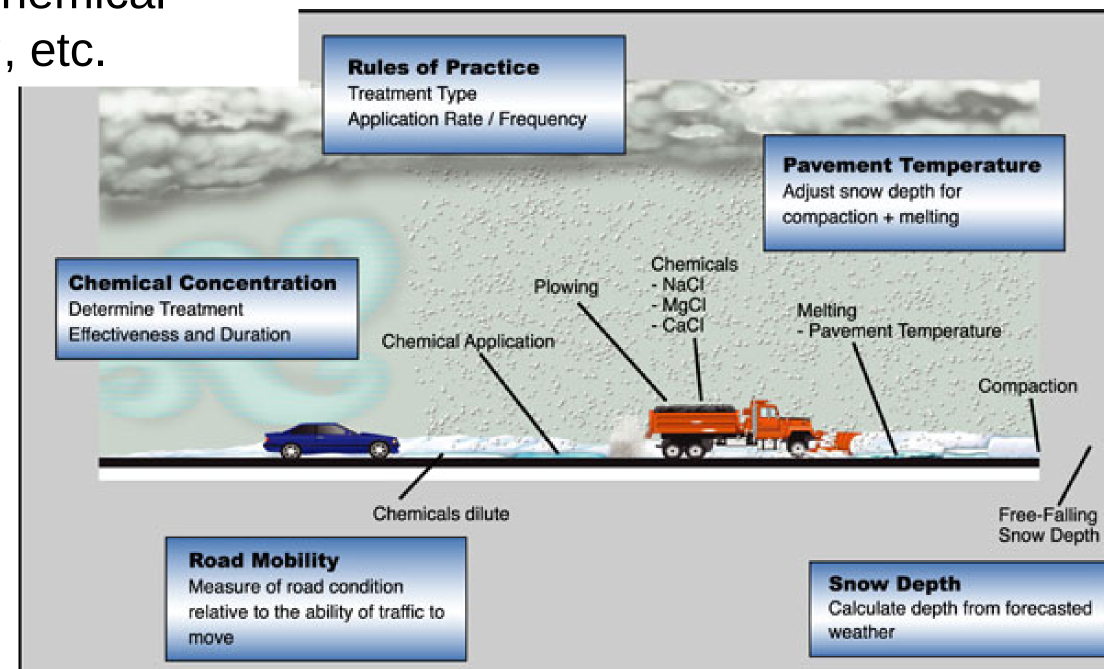




Multi model, initialization, perturbation ensembles
Example #5: **Winter road maintenance prediction**

- Effective winter road maintenance requires knowledge of weather (e.g., temperature, relative humidity, wind & precipitation), road surface information (temperature) & treatment (type, amount & location)
- Translation of (ensemble) weather, road conditions, & treatment instructions into probabilistic road plowing & chemical treatment applications, timing, etc.

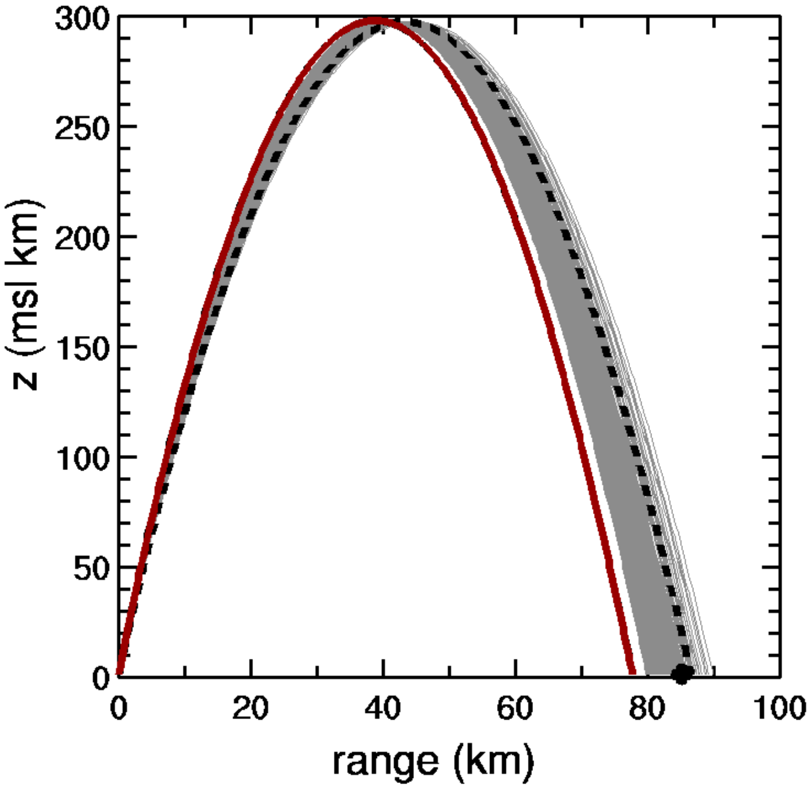


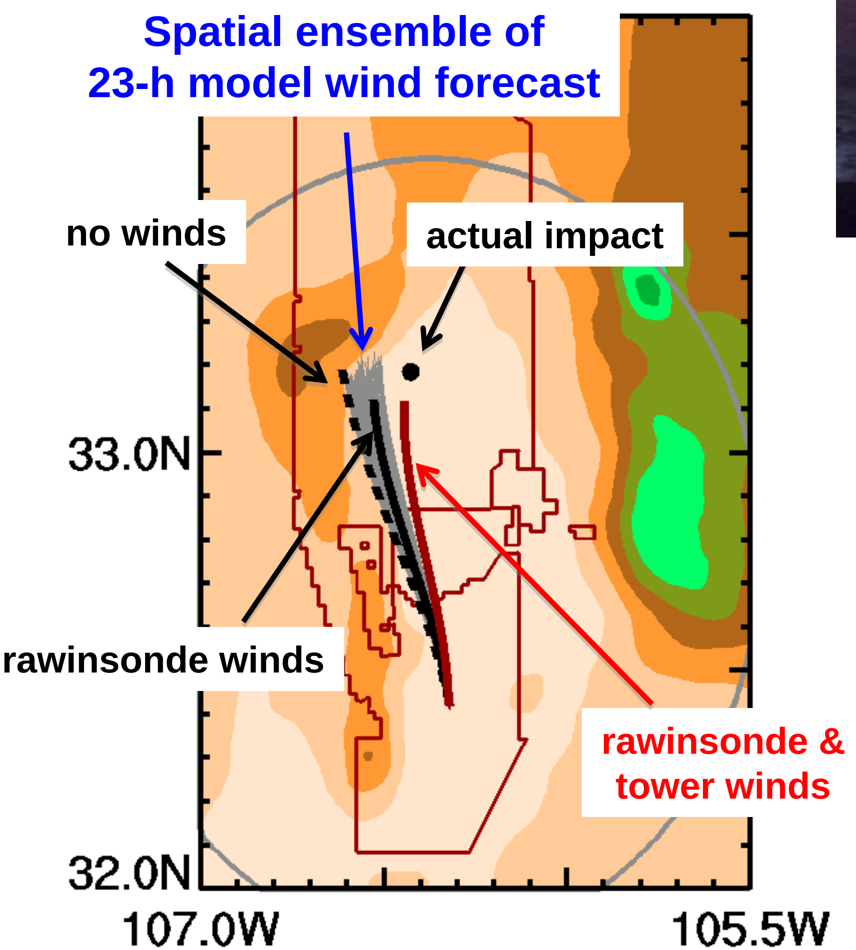


Spatial ensembles

Example #7: **Missile trajectory prediction**

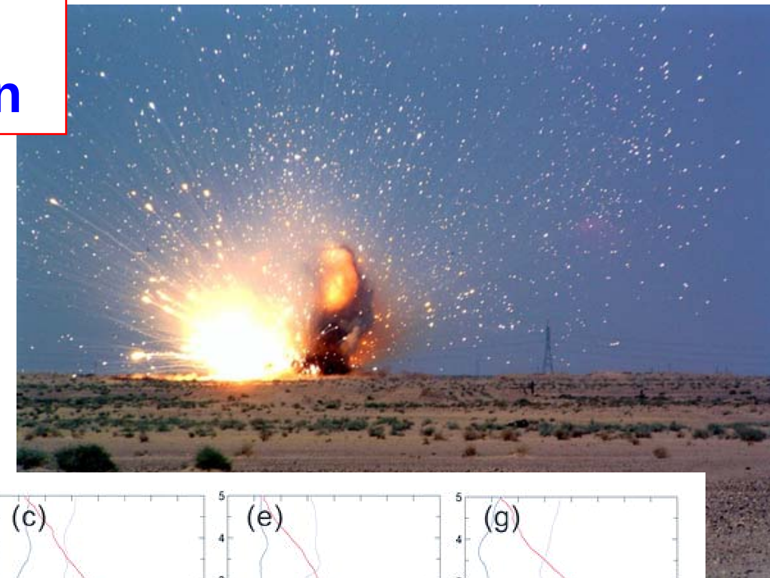
- Missile testing is sensitive to atmospheric conditions (e.g., wind profiles)
- Translation of spatial ensemble of model soundings into probabilistic missile trajectories & impact locations



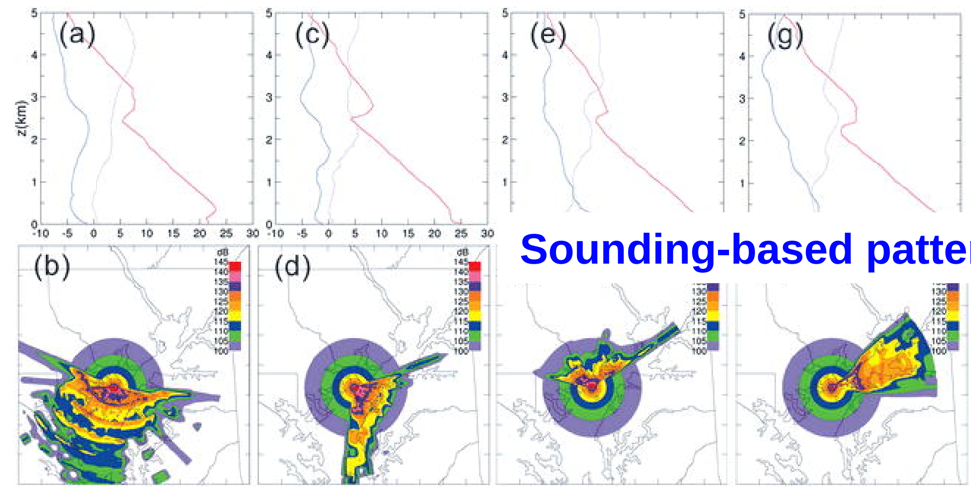


Spatial ensembles

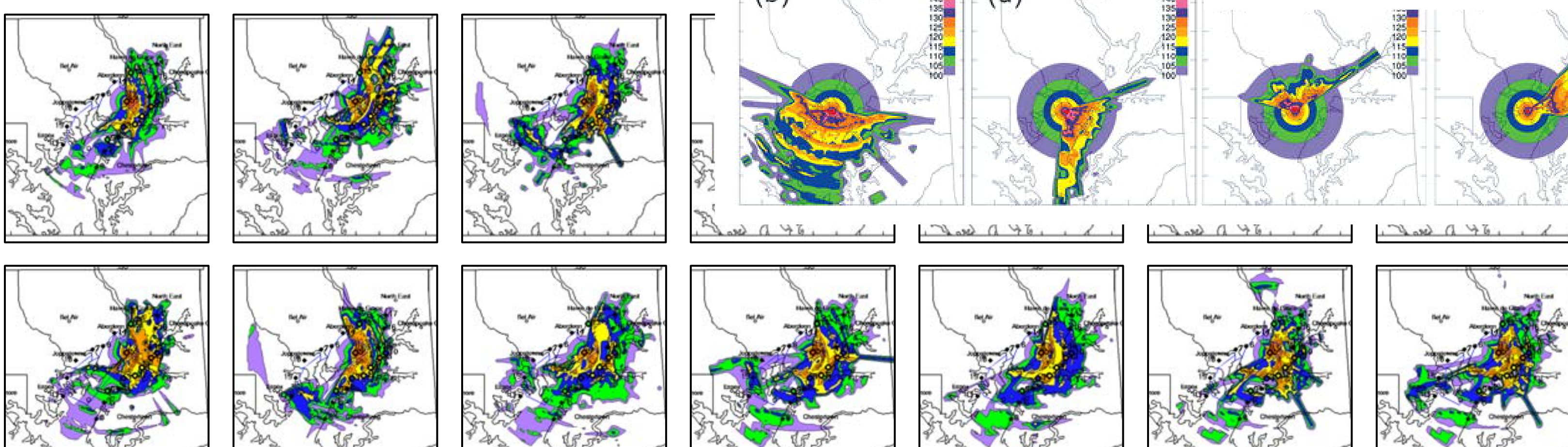
Example #8: **Noise propagation prediction**



- Sound propagation models for test range neighborhood annoyance mitigation & avoiding window shattering are sensitive to atmospheric conditions (e.g., temperature inversion and wind shear)
- Translation of spatial ensemble of model soundings into probabilistic sound propagation



Sounding-based patterns

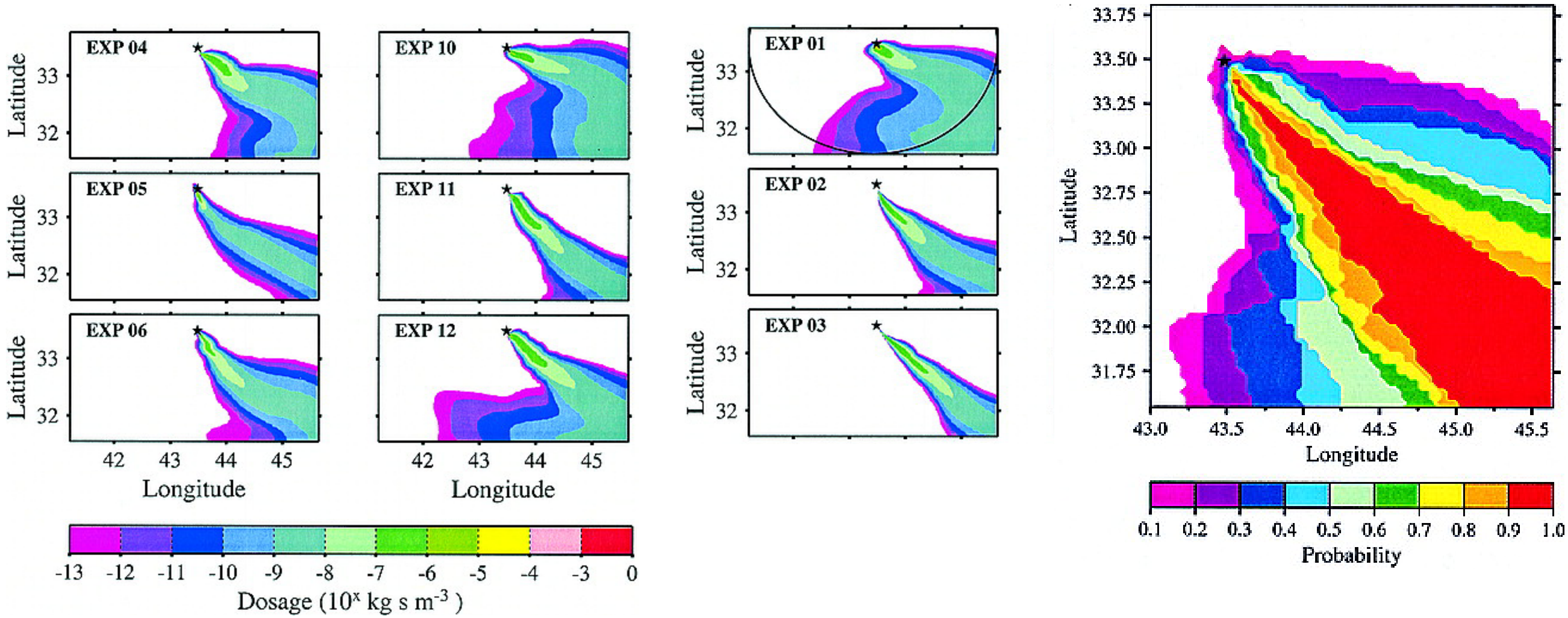


Ensemble of modeled noise propagation from a detonation

Spatial ensembles

Example #9: **Pollution dispersion prediction**

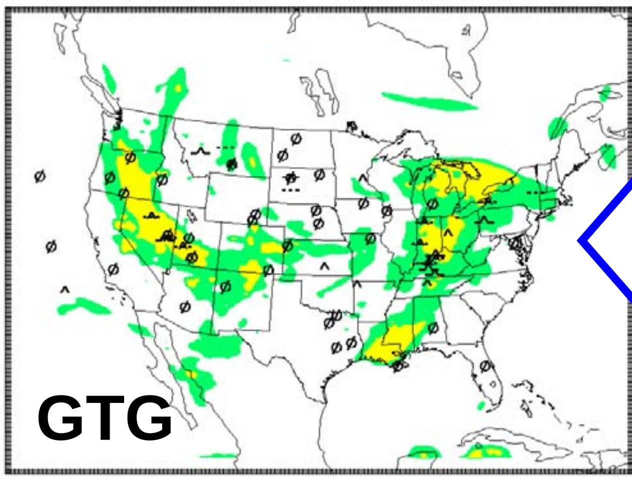
- Atmospheric diffusion models to predict concentration of smoke or chemical/biological agents require spatially & temporally varying wind, temperature, surface heat flux, & PBL depth forecasts
- Translation of ensemble weather forecasts into probabilistic dispersion of pollutants & concentration



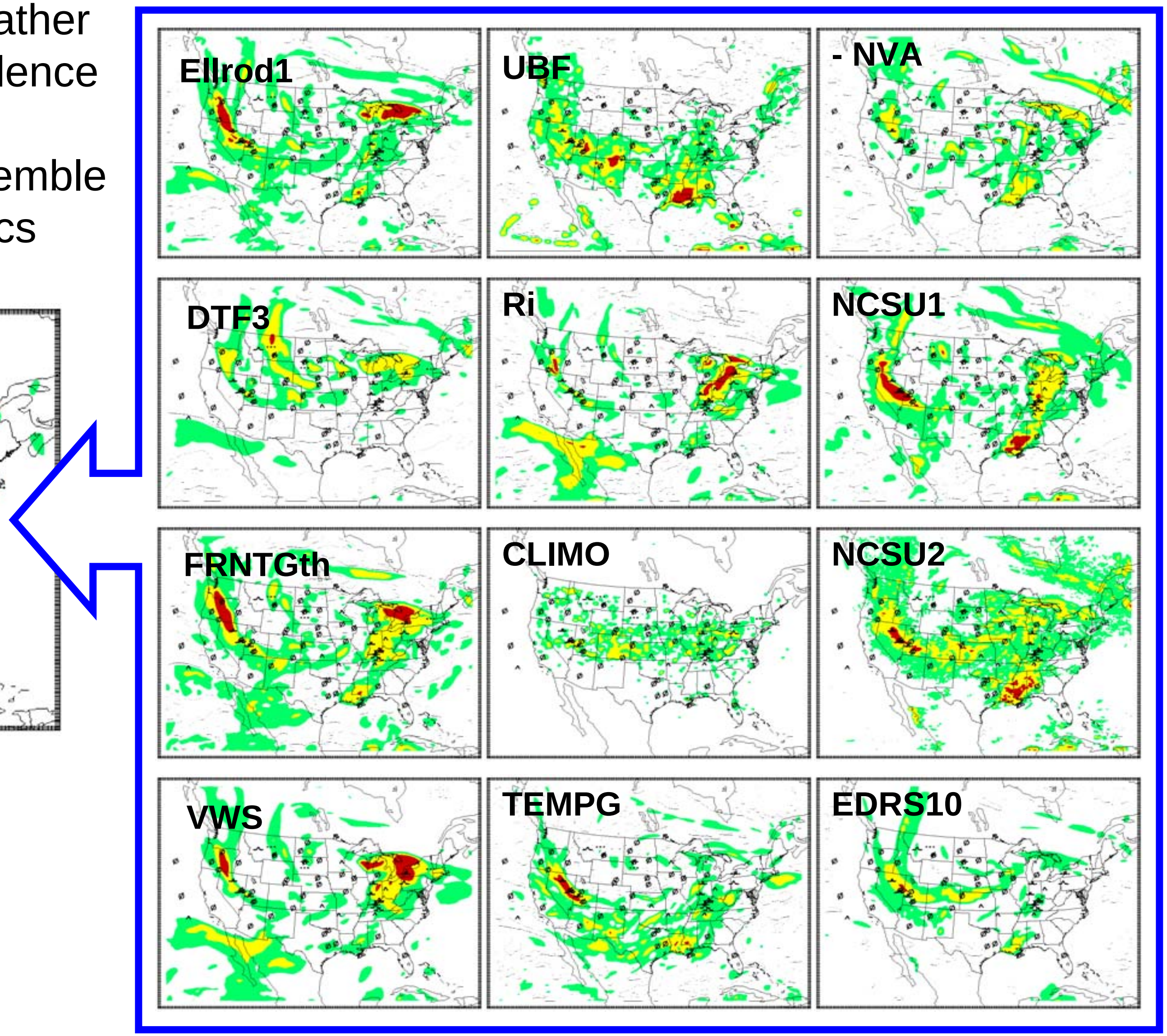
Diagnostic ensembles

Example #10: **Graphical turbulence guidance (GTG)**

- Translation of NWP weather data into multiple turbulence diagnostics
- GTG is a weighted ensemble of turbulence diagnostics



0 h forecast valid at 22 Sep 2006 15Z



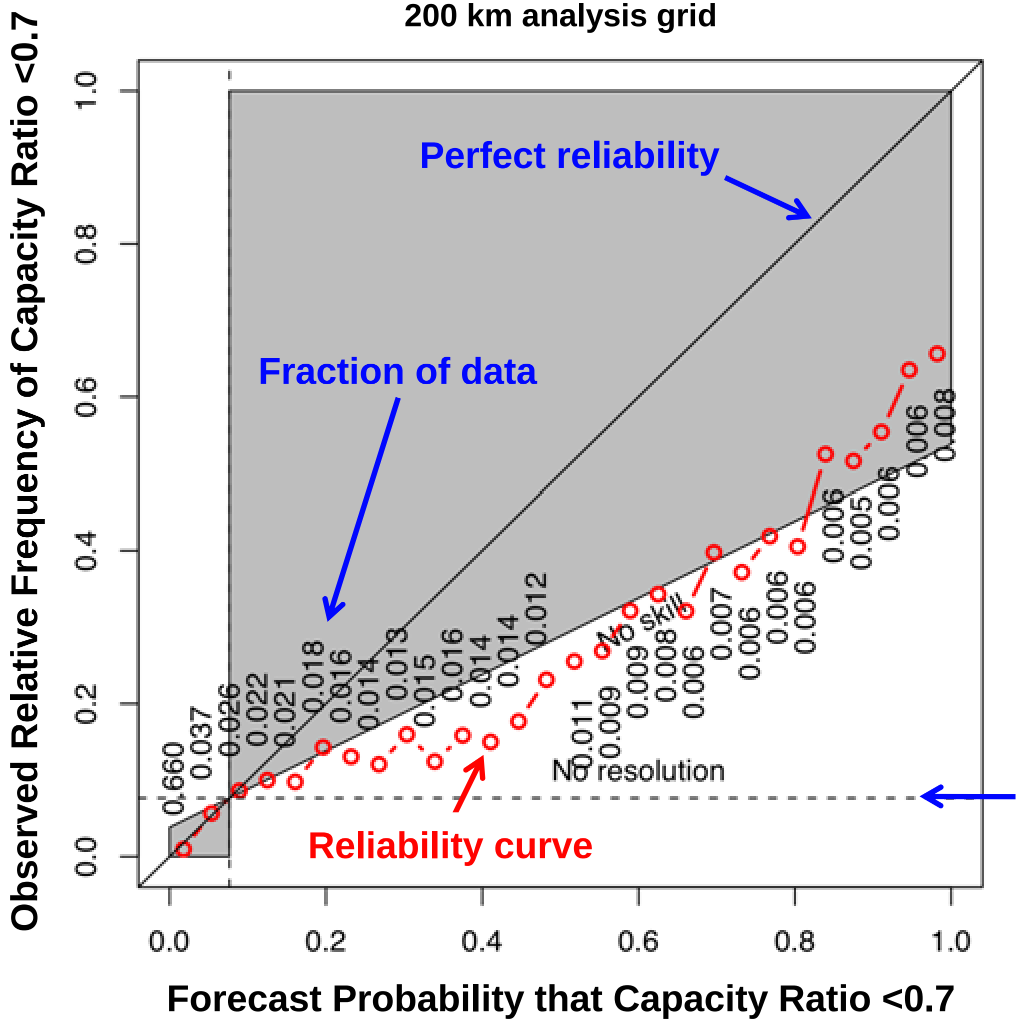
Performance Assessment

Focusing on . . .

- **Air traffic capacity prediction**
  - multi model, initialization, perturbation ensemble (Example #3)
  - time-lagged ensemble (Example #6)
  - use of reliability diagrams & ROC curves
- **Wind & temperature prediction**
  - spatial ensemble (Example #7)
  - multi model, initialization, perturbation ensemble (Example #6)
  - use of spaghetti plots, Taylor diagrams, & wind roses, etc.
- **Ensemble spread – skill relationship**
  - multi model, initialization, perturbation ensemble
  - focus on air traffic capacity (Example #3), &
    surface wind & temperature prediction (Example #4)

Multi model, initialization, perturbation ensembles
Example #3: **En route air traffic capacity prediction**

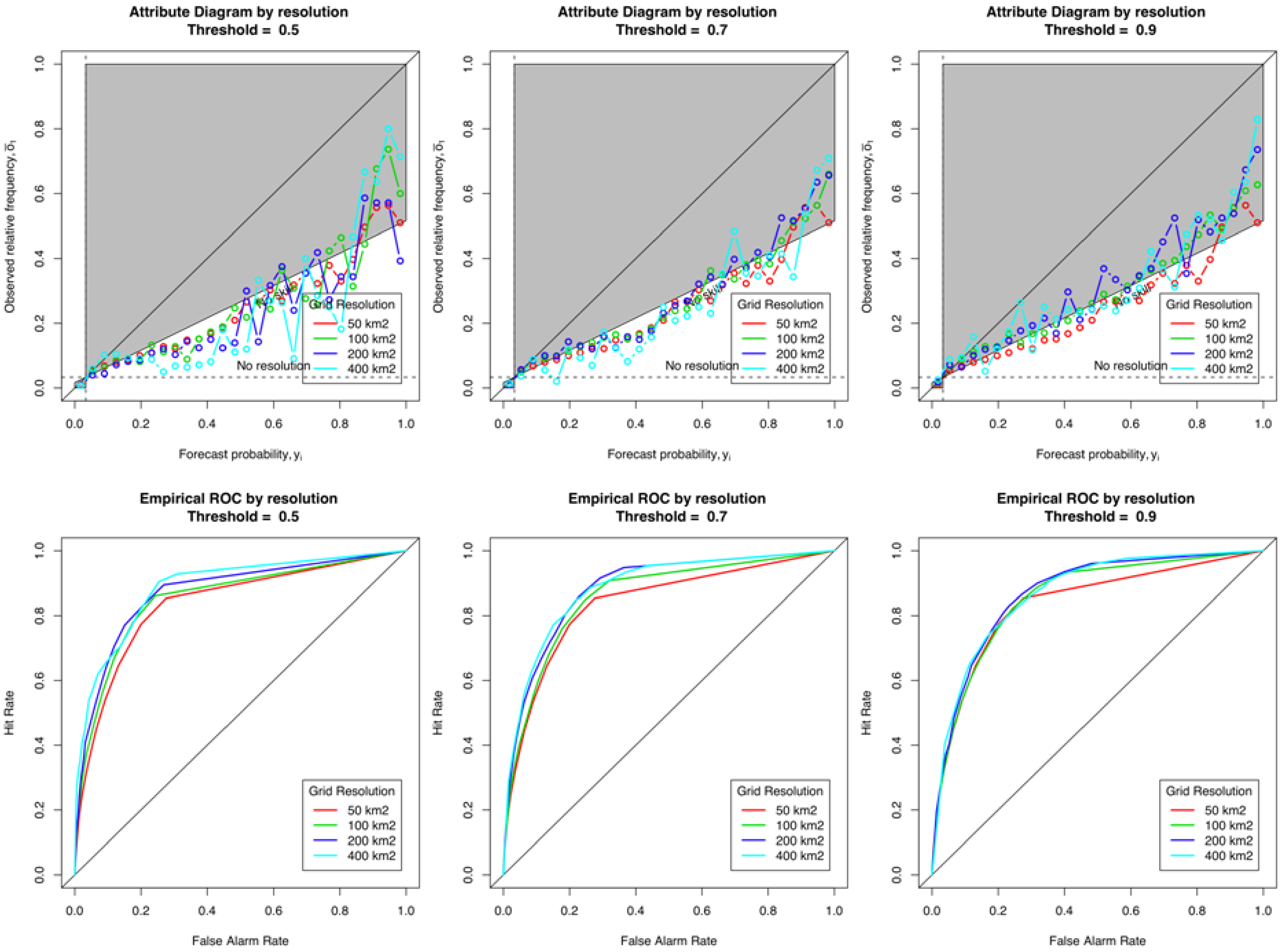
Reliability Diagram based on Brier Skill Scores
200 km analysis grid

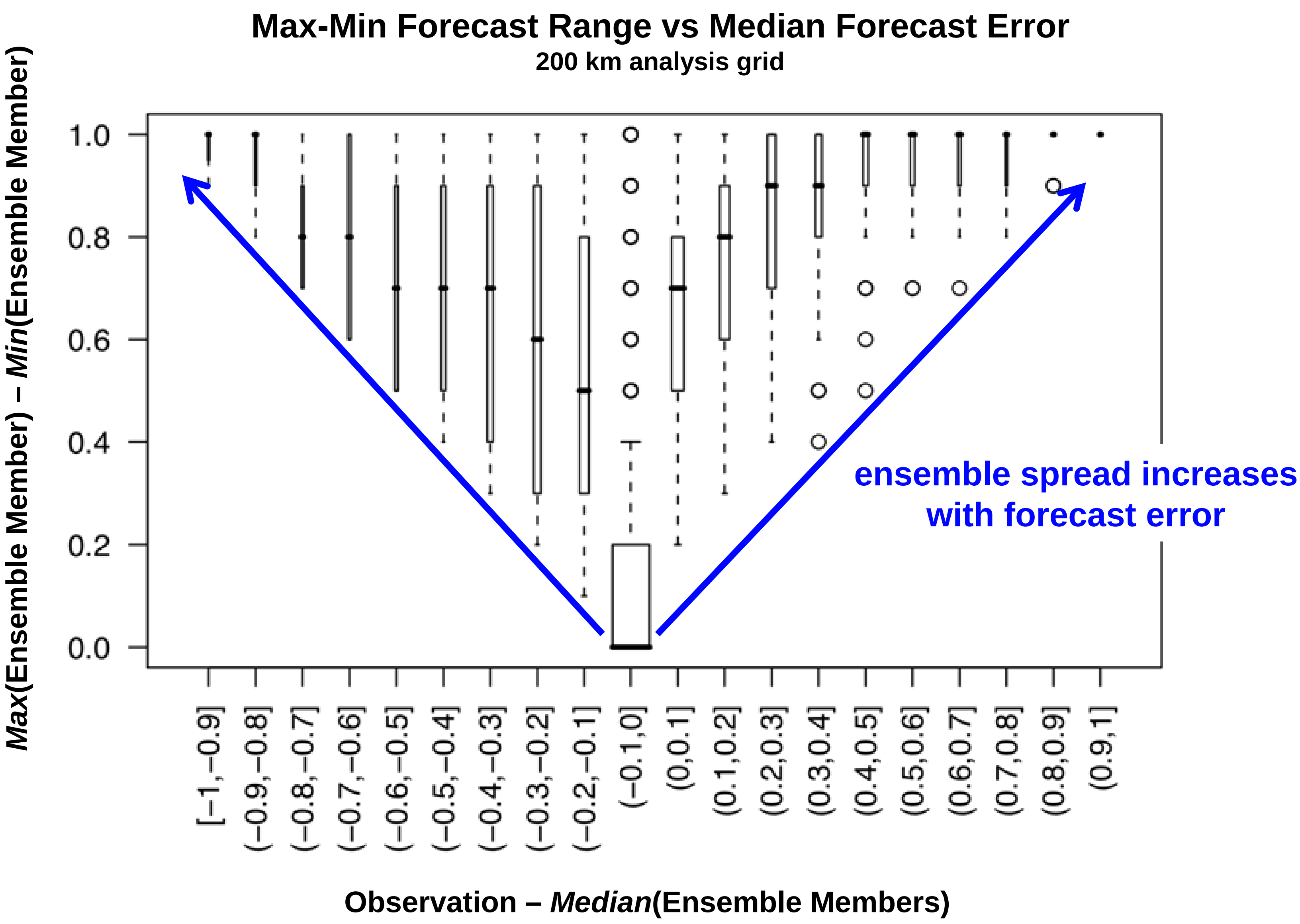


Climatological forecasts (reference) have perfect reliability but no resolution

- calibration needed to create reliable ensemble performance
- on average, similar performance across many spatial scales, forecast horizon, & user threshold
- but variability of performance from day to day

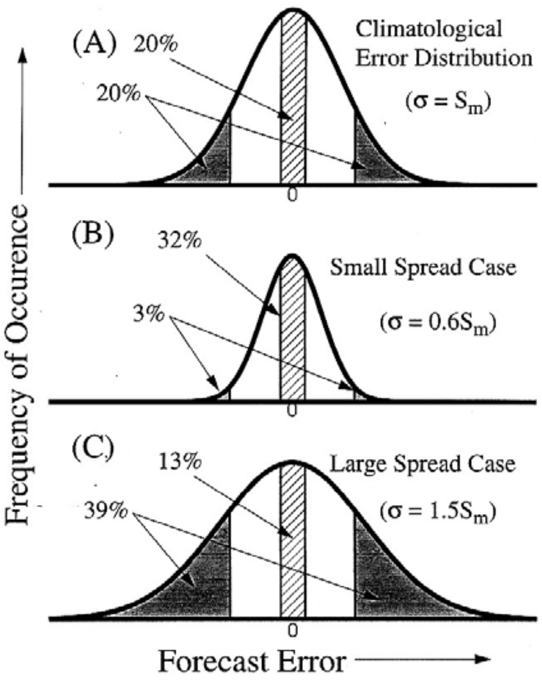
Steiner et al. (ATCQ 2010)


Attribute Diagram by resolution
Threshold = 0.5
Attribute Diagram by resolution
Threshold = 0.7
Attribute Diagram by resolution
Threshold = 0.9
Observed relative frequency, $\bar{o}_i$
Forecast probability, y_i
No resolution
No skill
Grid Resolution
50 km2
100 km2
200 km2
400 km2
Empirical ROC by resolution
Threshold = 0.5
Empirical ROC by resolution
Threshold = 0.7
Empirical ROC by resolution
Threshold = 0.9
Hit Rate
False Alarm Rate
0.0
0.2
0.4
0.6
0.8
1.0

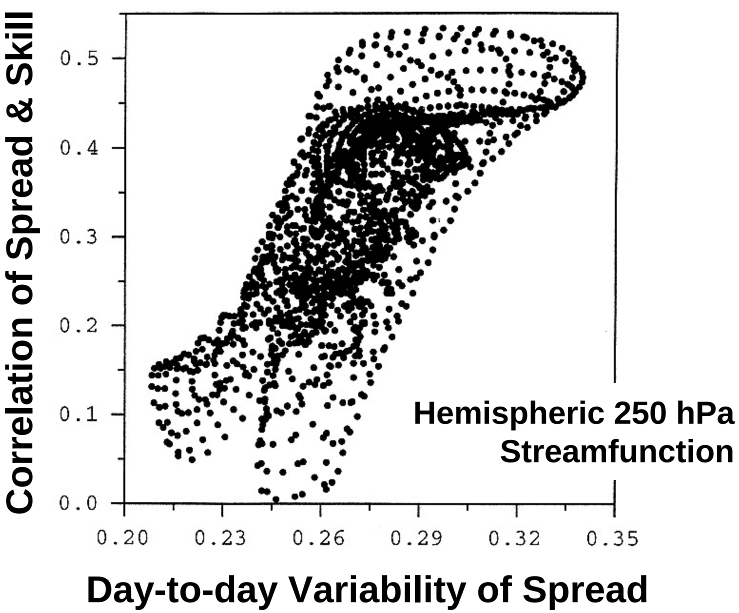

Max-Min Forecast Range vs Median Forecast Error
200 km analysis grid
Max(Ensemble Member) – Min(Ensemble Member)
1.0
0.8
0.6
0.4
0.2
0.0
[−1,−0.9]
(−0.9,−0.8]
(−0.8,−0.7]
(−0.7,−0.6]
(−0.6,−0.5]
(−0.5,−0.4]
(−0.4,−0.3]
(−0.3,−0.2]
(−0.2,−0.1]
(−0.1,0]
(0,0.1]
(0.1,0.2]
(0.2,0.3]
(0.3,0.4]
(0.4,0.5]
(0.5,0.6]
(0.6,0.7]
(0.7,0.8]
(0.8,0.9]
(0.9,1]
ensemble spread increases
with forecast error
Observation – Median(Ensemble Members)

Ensemble Spread – Skill Relation

Some relevant thoughts on the side



- even for a perfect ensemble (one in which all sources of forecast error are sampled correctly) there need not be a high correlation between spread & skill
- correlation between spread & skill should be larger where the day-to-day variability of spread is large
- spread is likely to be most useful as predictor of skill when it is "extreme"—i.e., when it is either very large or very small compared to its climatological mean value
- existence of analytical limit to the strength of spread – error correlation even for ideal ensemble prediction system



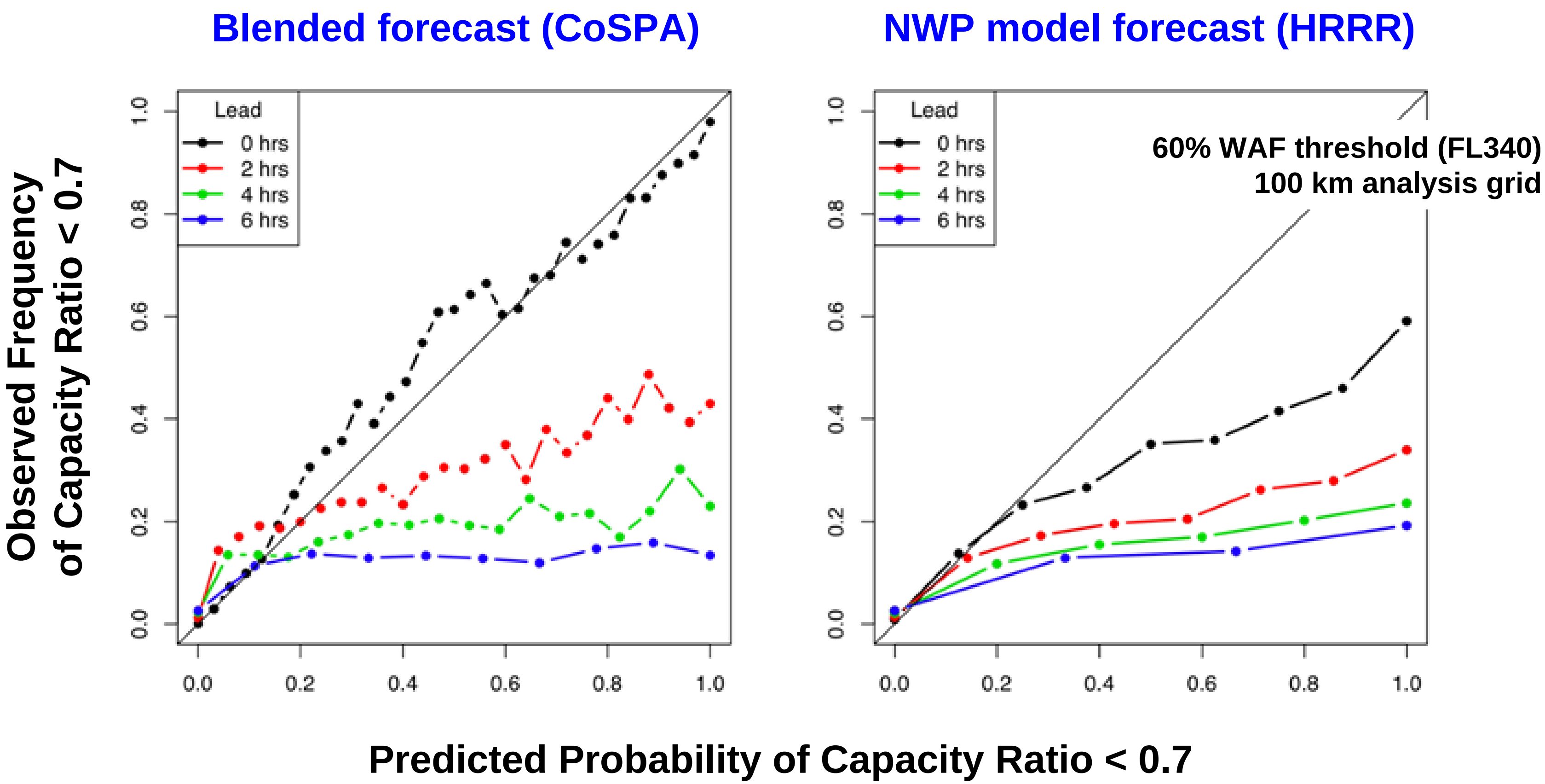
Whitaker & Loughe (MWR 1998)

Grimit & Mass (MWR 2007)

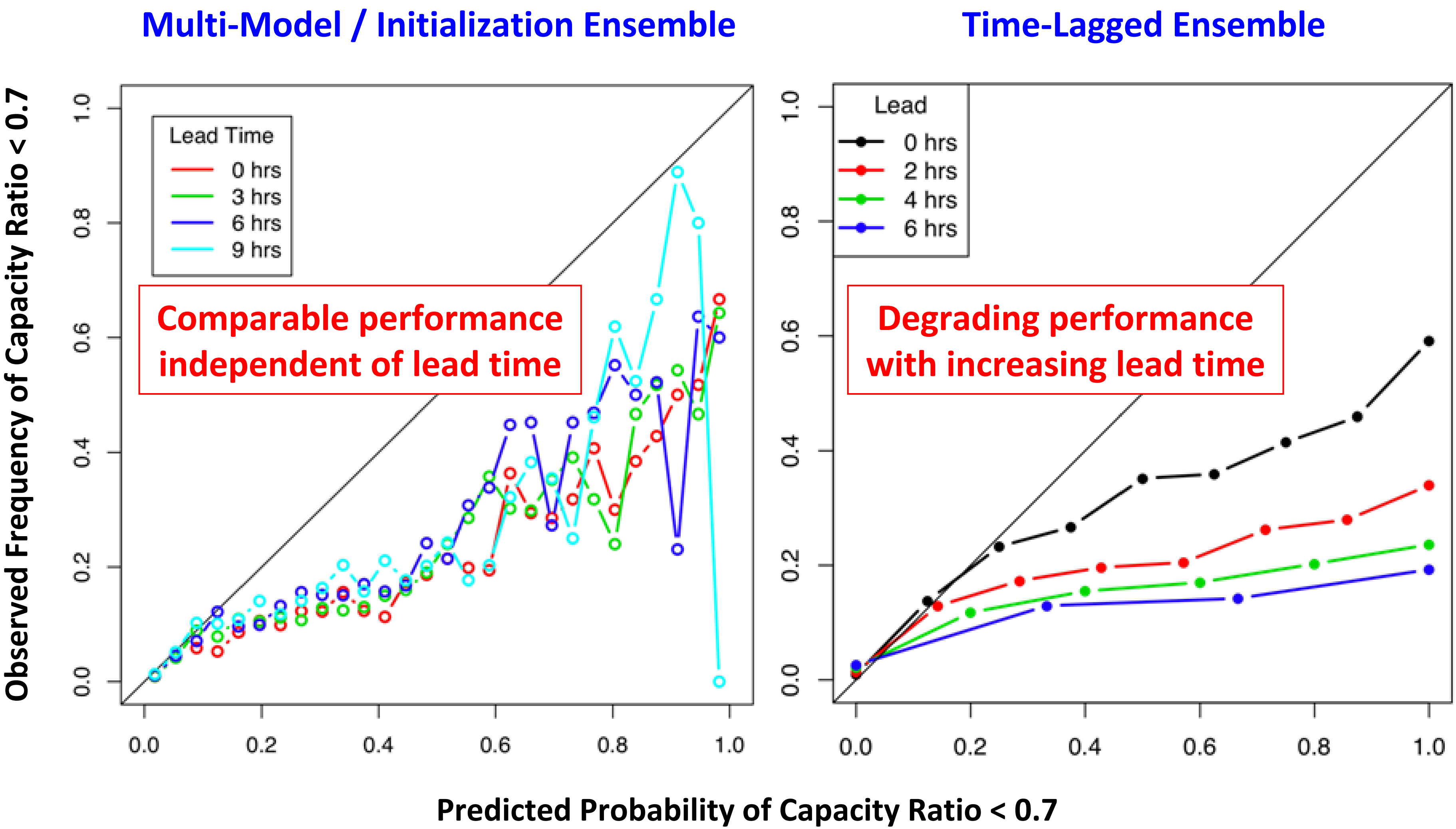
Time-lagged ensembles

Example #6: **En route air traffic capacity prediction**

- size of membership for time-lagged ensembles depends on forecast horizon
- forecast skill rapidly decreases with increasing lead time
- blending of NWP model with extrapolation forecasts adds skill, particularly for shorter forecast lead times



Multi model, initialization, perturbation vs time-lagged ensembles
Examples #3 & #6: **En route air traffic capacity prediction**



- NOTE: previous comparison is not clean, because . . .

	Multi-model/initialization	Time-lagging
NWP Model	* RT-FDDA * MM5 & WRF cores * 10 km resolution * 6 h cycling * 4D data assimilation * parameterized microphysics	* HRRR * WRF core * 3 km resolution * 1 h cycling * GSI & DFI assimilation * explicit microphysics
Ensemble	* 30 members * same for all lead times	* max. 25 members * decreasing with lead time
Analysis	* hourly precipitation * 2 mm threshold	* WAF based on VIL & ETOP * 60% threshold
Data	* 30 days from 2007 & 2008	* 20 days from 2009

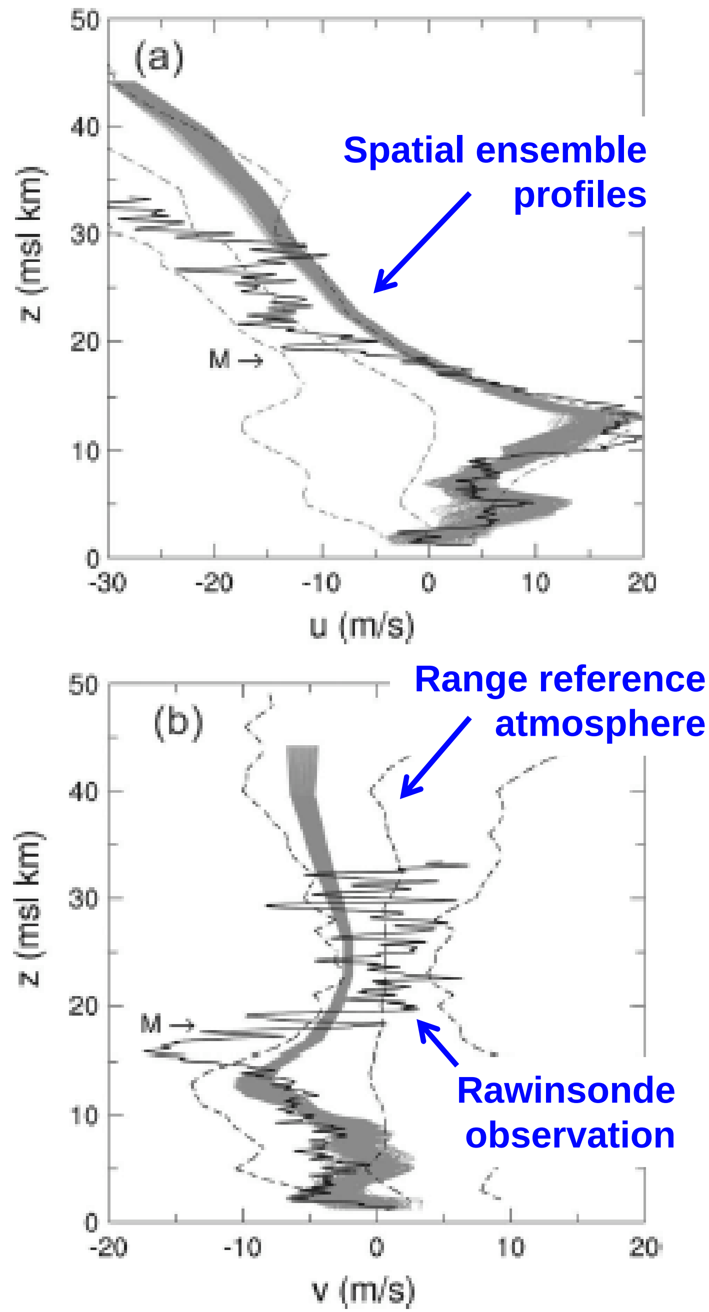
- requires creation of suitable dataset & in-depth analysis

Spatial ensembles
Example #7: **Missile trajectory prediction**

- ensemble wind predictions evaluated against soundings show significant bias & RMSE

p level (hPa)	Speed ($m\ s^{-1}$) Bias	Speed ($m\ s^{-1}$) RMSE	Direction (°) Bias	Direction (°) RMSE	RMSVE ($m\ s^{-1}$)	N
1–12-h forecast wind statistics						
850	1.4	3.6	−18.5	69.4	3.9	170
700	1.8	4.4	−0.5	36.8	4.9	153
500	1.1	5.0	9.2	29.7	5.0	106
400	0.6	4.1	12.1	40.1	4.8	94
300	0.8	4.3	8.2	19.9	4.7	55
250	2.8	4.8	11.2	16.0	5.7	42
200	5.4	6.9	17.7	24.6	7.5	23
150	3.8	7.3	11.4	19.2	7.4	24
100	6.7	8.0	4.4	23.6	7.7	43
13–24-h forecast wind statistics						
850	1.9	3.7	−18.6	90.6	4.6	179
700	2.7	4.8	−16.8	61.3	7.2	166
500	−0.3	4.8	11.1	43.2	6.1	110
400	0.0	5.2	12.6	51.2	6.9	101
300	−0.5	5.6	4.6	18.0	6.0	57
250	2.8	5.3	18.7	23.6	7.1	46
200	3.0	4.5	29.1	36.7	6.7	22
150	4.1	7.9	14.7	25.3	8.8	25
100	7.0	8.6	2.6	24.1	8.5	38

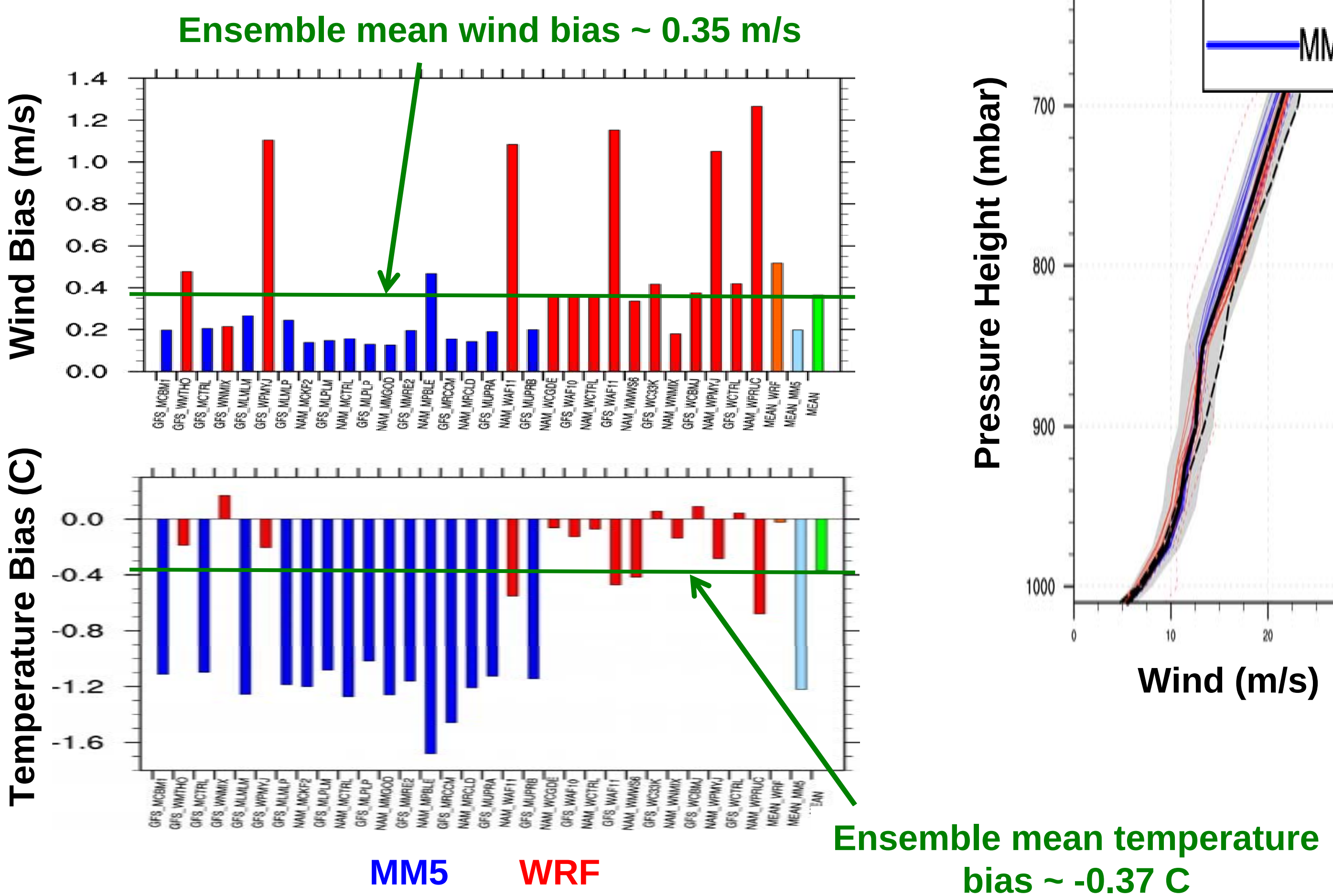
Sharman et al. (JAMC 2008)

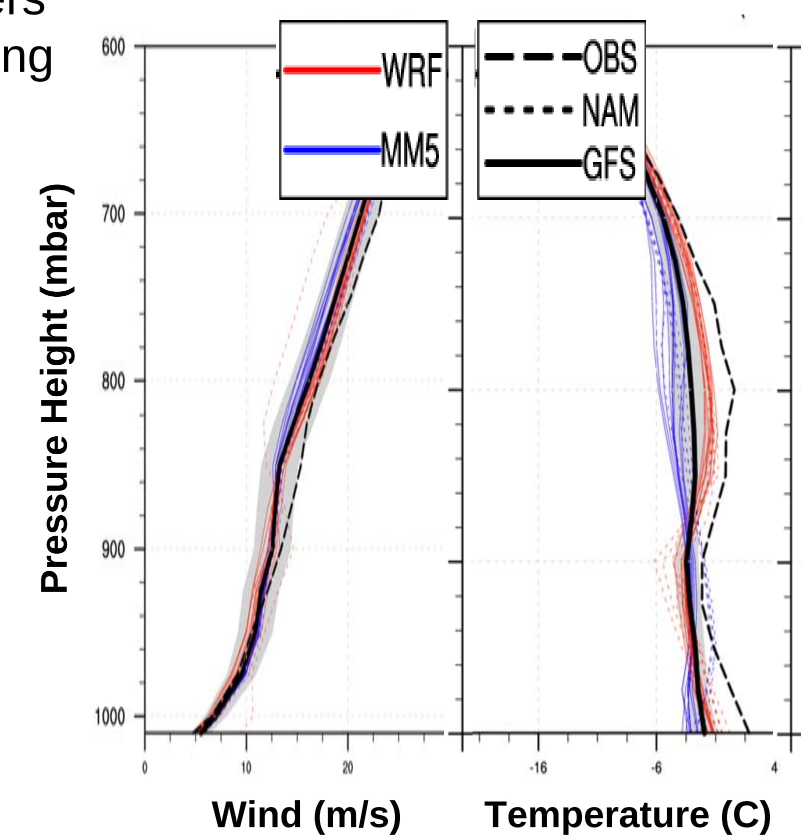


Multi model, initialization, perturbation ensembles

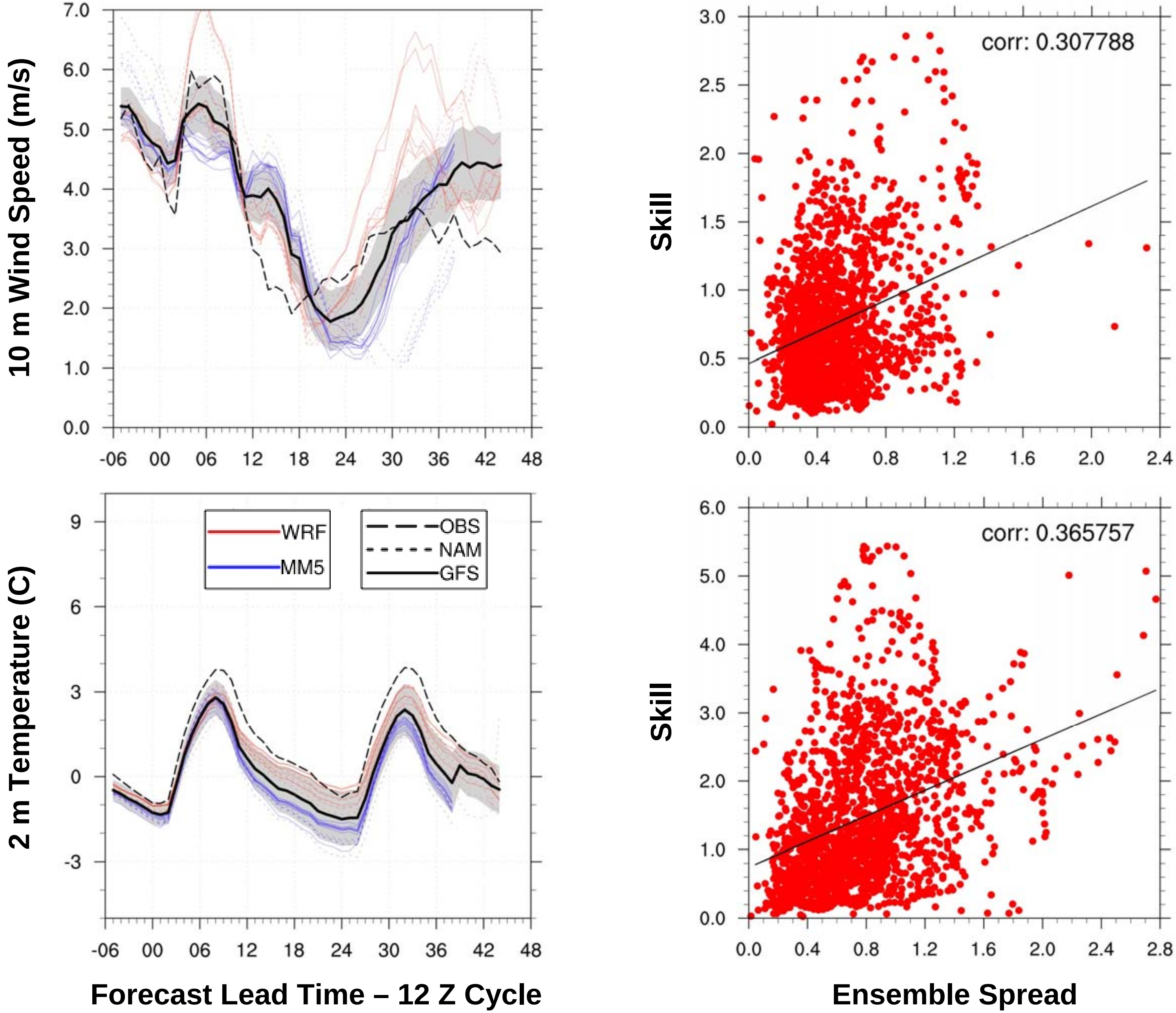
Example #4: **Surface wind & energy prediction**

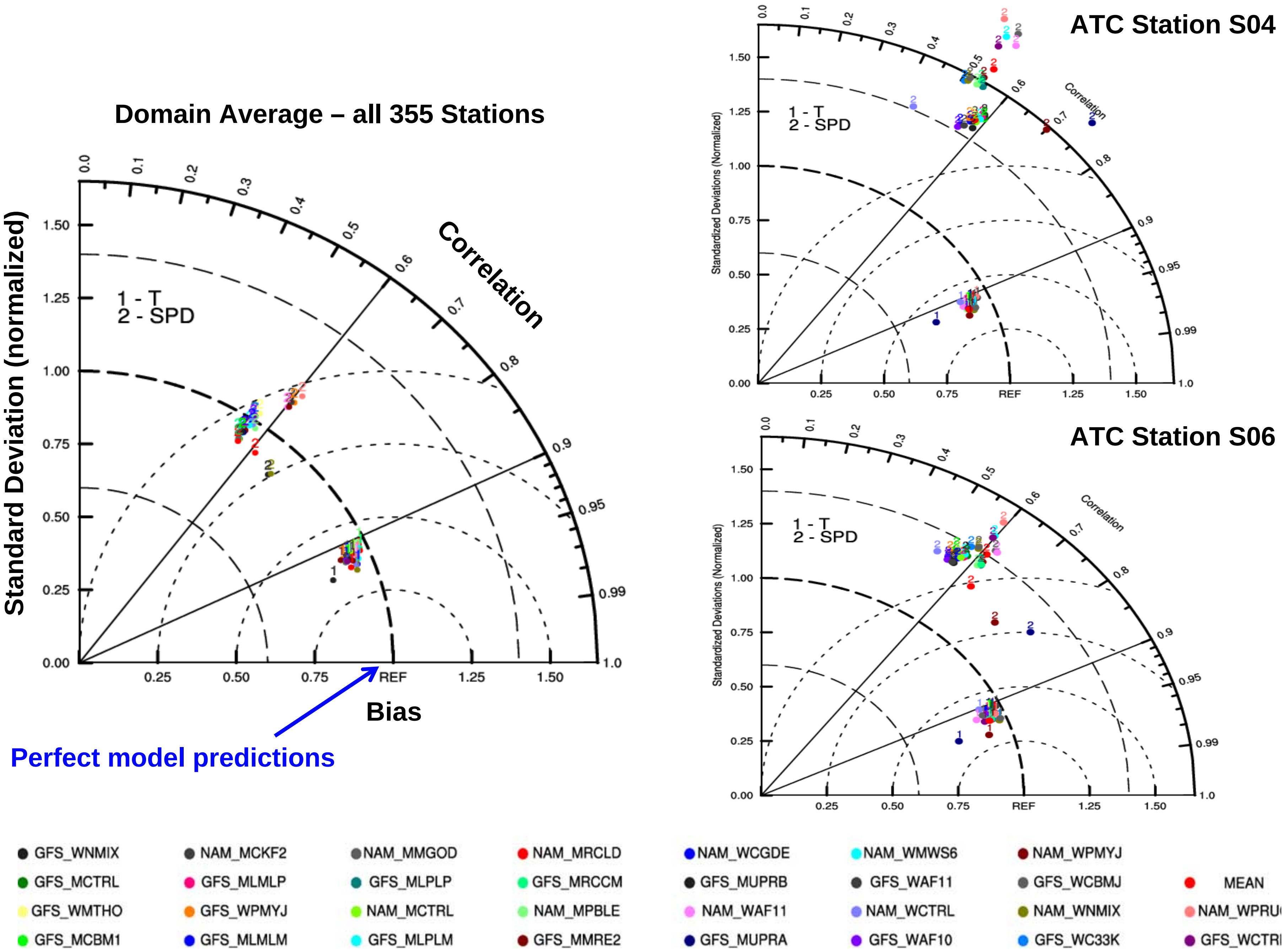
- ensemble performance is model specific
- significant variability among different members
- MM5 & WRF members show distinct clustering





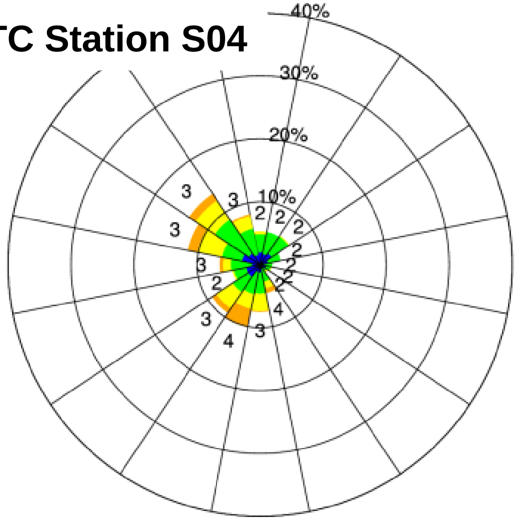
Domain Average Values for Dec 2008 – Jan 2009 based on 355 Stations

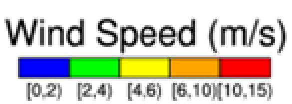



Domain Average – all 355 Stations
Standard Deviation (normalized)
Correlation
1 - T
2 - SPD
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 0.95 0.99 1.0
0.00 0.25 0.50 0.75 1.00 1.25 1.50
0.25 0.50 0.75 REF 1.25 1.50
Bias
Perfect model predictions
ATC Station S04
ATC Station S06
Standardized Deviations (Normalized)
GFS_WNMIX
NAM_MCKF2
NAM_MMGOD
NAM_MRCLD
NAM_WCGDE
NAM_WMWS6
NAM_WPMYJ
GFS_MCTRL
GFS_MLMLP
GFS_MLPLP
GFS_MRCCM
GFS_MUPRB
GFS_WAF11
GFS_WCBMJ
MEAN
GFS_WMTHO
GFS_WPMYJ
NAM_MCTRL
NAM_MPBLE
NAM_WAF11
NAM_WCTRL
NAM_WNMIX
NAM_WPRU
GFS_MCBM1
GFS_MLMLM
GFS_MLPLM
GFS_MMRE2
GFS_MUPRA
GFS_WAF10
GFS_WC33K
GFS_WCTR

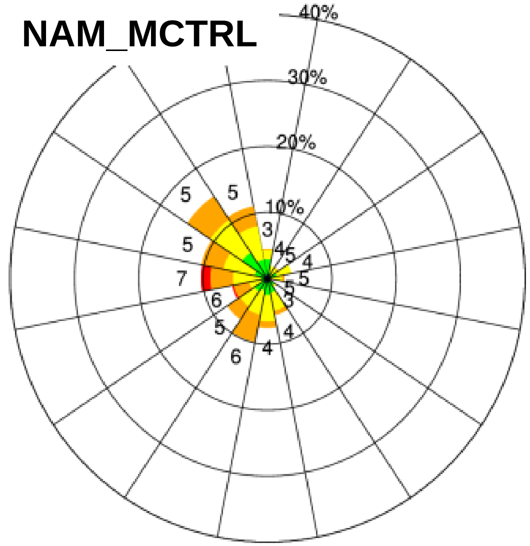
Wind roses comparison

ATC Station S04

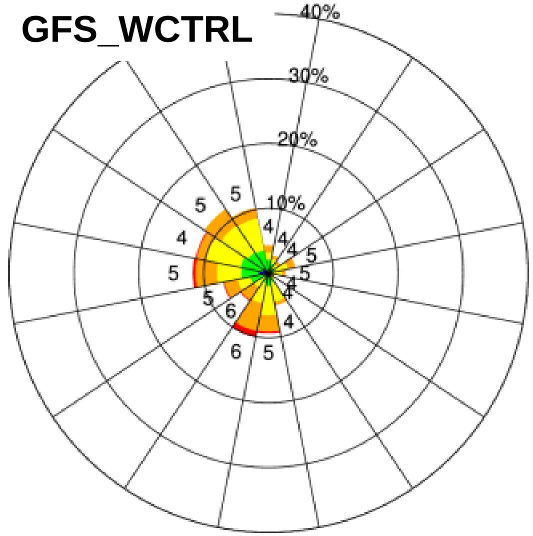




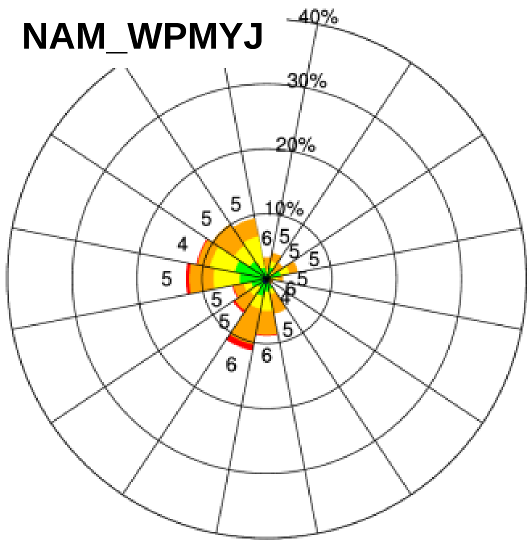
NAM_MCTRL



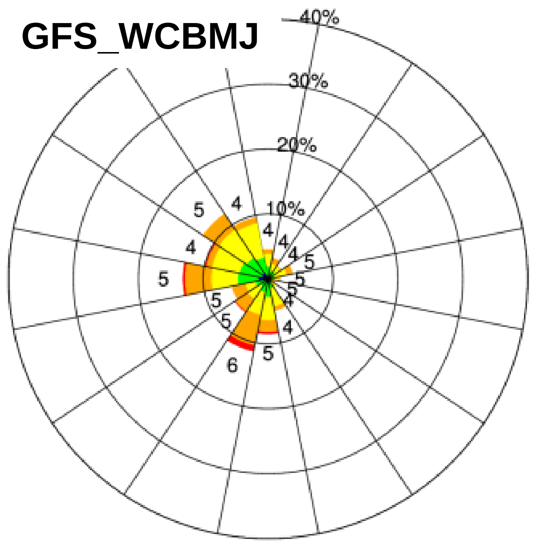
GFS_WCTRL



NAM_WPMYJ

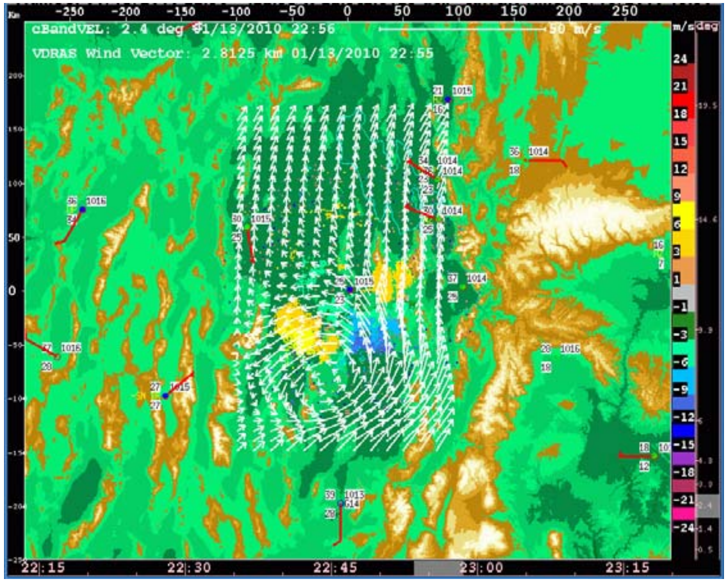


GFS_WCBMJ



4D Wind & Temperature Analysis – VDRAS/VLAS

A note on the side



Synthesis of observations with model data

- **variational Doppler radar/lidar analysis**
  - assimilation of available observations, including surface observations, soundings, profilers, radar or lidar data
  - synthesis of observations & NWP model data based 4D variational analysis

- **thermodynamic retrieval**
  - synthesis provides not only 3D wind fields but also 3D thermodynamic fields
- **nowcasting**
  - synthesized analysis provides initialization for model run
  - nowcasts of 3D wind & thermodynamic fields can be generated a few hours into future
  - boundary conditions problem for small domain

Summary

Potential value of ensemble approaches for wake vortex prediction:

- **Multi model, initialization, perturbation ensemble**
  - good performance, if properly configured (not trivial)
  - computationally expensive
- **Time-lagged ensemble**
  - easy to create based on frequently updating model runs (hourly or better)
  - rapidly degrading performance with increasing lead time
- **Spatial ensemble**
  - may be a good solution for short-term prediction, if spatial lidar data available
  - potentially meaningful approach also for model predictions
- **Diagnostic ensemble**
  - maybe less applicable for wake vortex problem (?)
- **Combinations of the above**
  - a time-lagged & spatial ensemble hybrid might be of potential interest

Summary (continued)

Additional considerations:

- **Translation of weather into aviation impact**
  - analysis of each ensemble member from an impact perspective
  - ensemble user-relevant information rather than weather data
- **Post-processing**
  - calibration required to achieve reliable & sharp probabilistic predictions
- **Ensemble spread – skill relationship**
  - spread not necessarily a good measure for skill
  - small spread can provide false sense of accuracy (all members may be wrong)
- **Verification & operational evaluation**
  - requires long-term data
  - focus on weather & user impact quantities
  - look for dependencies with forecast lead time
- **Numerical weather prediction models**
  - wind speed & direction prediction may need improvement
  - requires focus on boundary layer physics & parameterizations

References & Contacts

- **Papers on probabilistic wind forecasting**

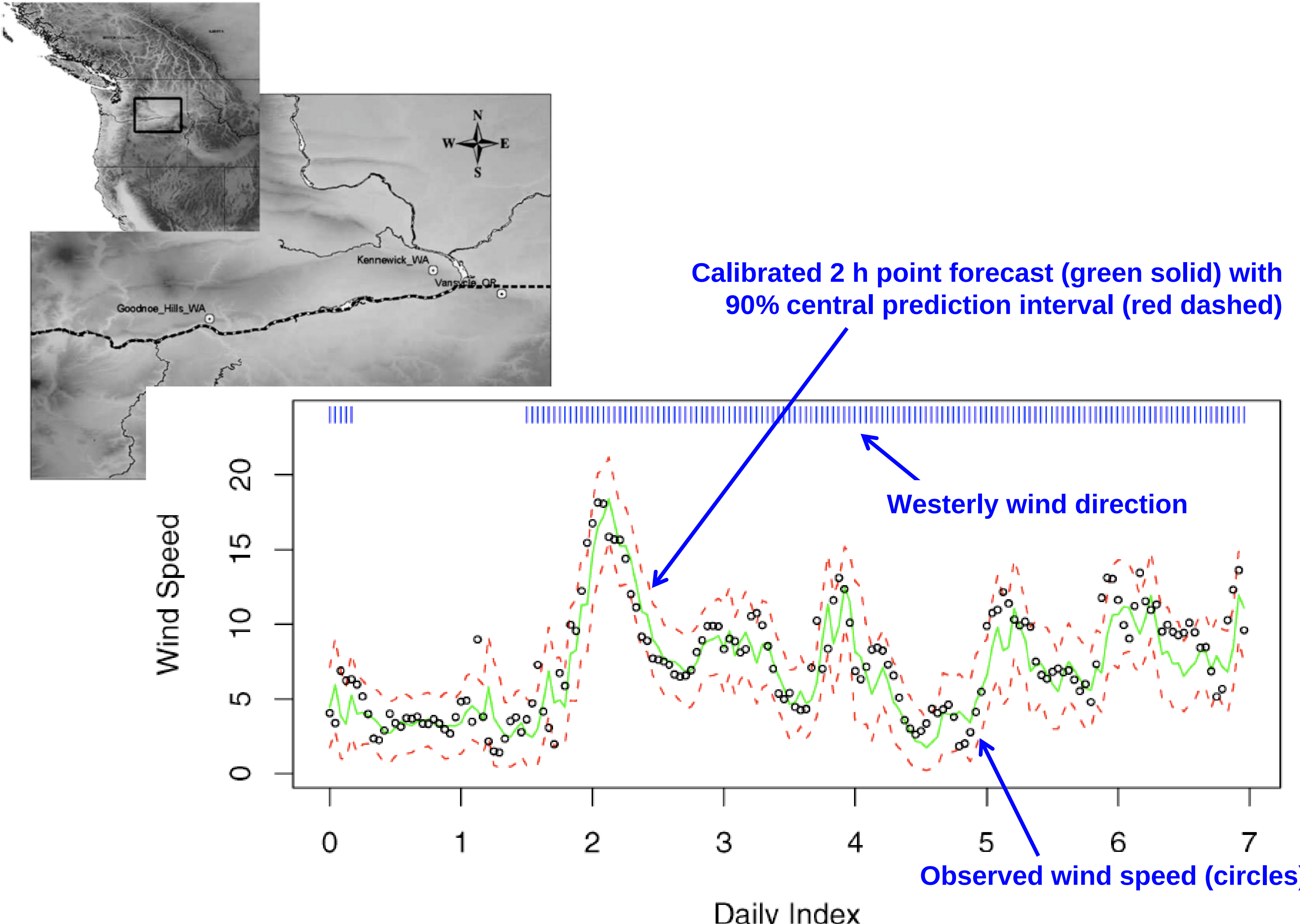
- Gneiting, T., K. Larson, K. Westrick, M. G. Genton, and E. Aldrich, 2006: Calibrated probabilistic forecasting at the Stateline Wind Energy Center: The regime-switching space-time method. *Journal of the American Statistical Association*, **101**, 968 – 979.
- Grimit, E. P., and C. F. Mass, 2007: Measuring the ensemble spread – error relationship with a probabilistic approach: Stochastic ensemble results. *Monthly Weather Review*, **135**, 203 – 221.
- Sloughter, J. M., T. Gneiting, and A. E. Rafter, 2010: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association*, **105**, 25 – 35.
- Whitaker, J. S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Monthly Weather Review*, **126**, 3292 – 3302.

References & Contacts (continued)

- **Papers on weather – impact translation**
  - Sharman, R. D., Y. Liu, R.-S. Sheu, T. T. Warner, D. L. Rife, J. F. Bowers, C. A. Clough, and E. E. Ellison, 2008: The operational mesogamma-scale analysis and forecast system of the U.S. Army Test and Evaluation Command. Part III: Forecasting with secondary-applications models. *Journal of Applied Meteorology and Climatology*, **47**, 1105 – 1122.
  - Steiner, M., R. Bateman, D. Megenhardt, Y. Liu, M. Xu, M. Pocernich, and J. Krozel, 2010: Translation of ensemble weather forecasts into probabilistic air traffic capacity impact. *Air Traffic Control Quarterly*, (to appear in Sep issue).

- **People**

  - **Prof. Tilmann Gneiting, University of Heidelberg, Applied Mathematics**
    Email: t.gneiting@uni-heidelberg.de
  - Dr. Thordis Thorarinsdottir, University of Heidelberg
    Email: thordis@uni-heidelberg.de



Gneiting et al. (JASA 2006)

Scientific Questions & Future Activities

- **Ensemble configuration**
  - How to create an effective ensemble configuration for a specific application?
  - How many ensemble members may be needed?
- **Post-processing & calibration**
  - What are efficient & effective ways to calibrate probabilistic forecasts?
  - Should every ensemble member be calibrated individually or the entire ensemble combined?
  - How much data is needed for a proper calibration?
- **Numerical weather prediction**
  - How to improve boundary layer & sub-grid scale processes?
  - What data assimilation procedures are efficient & effective for wake vortex prediction?
- **Observations**
  - How to design a good network for capturing the relevant atmospheric quantities & their spatial/temporal variability?
- **Verification & operational evaluation**
  - What are effective metrics for diagnostic verification & benefits assessment?