STATISTICAL POSTPROCESSING

FROM BIASED MODEL VARS TO CALIBRATED USER VARS

Zoltan Toth
OAR/ESRL/GSD

Mark Antolik
NWS/MDL

Malaquias Pena
NWS/NCEP/EMC & IMSG

Roman Krzysztofowicz
University of Virginia

Brian Etherton
OAR/ESRL/GSD

NGGPS Statistical Postprocessing Workshop
College Park, MD, January 19-22 2016

PREAMBLE

- Not much new – **Recommendations from** series of
  - **NCEP Ensemble User Workshops (2004, '06, '08)**

- **Automated forecasts** based on NWP
  - Consider key choices

- Focus on **methodology** as opposed to methods
  - Set up framework/structure (WHAT?) under which
    - Various methods can be tested and interchanged (HOW?)

- Efficient & robust methods for **central processing**
  - Advanced local applications if needed for specific reqmts

FROM BIASED MODEL VARIABLES TO CALIBRATED USER VARIABLES

- NWP model predicts **prognostic vars on model grid**
- Prog. vars. suffer from **lead-time dependent biases**
  - First, second, & higher moments of ensemble
    - With respect to NWP analysis fields
  - Feedback to model developers
- **Biases reside on model grid** =>
- **Calibrate ensemble against analysis on 3D model grid**
  - **Toolbox #1 – Calibration of model prognostic variables**

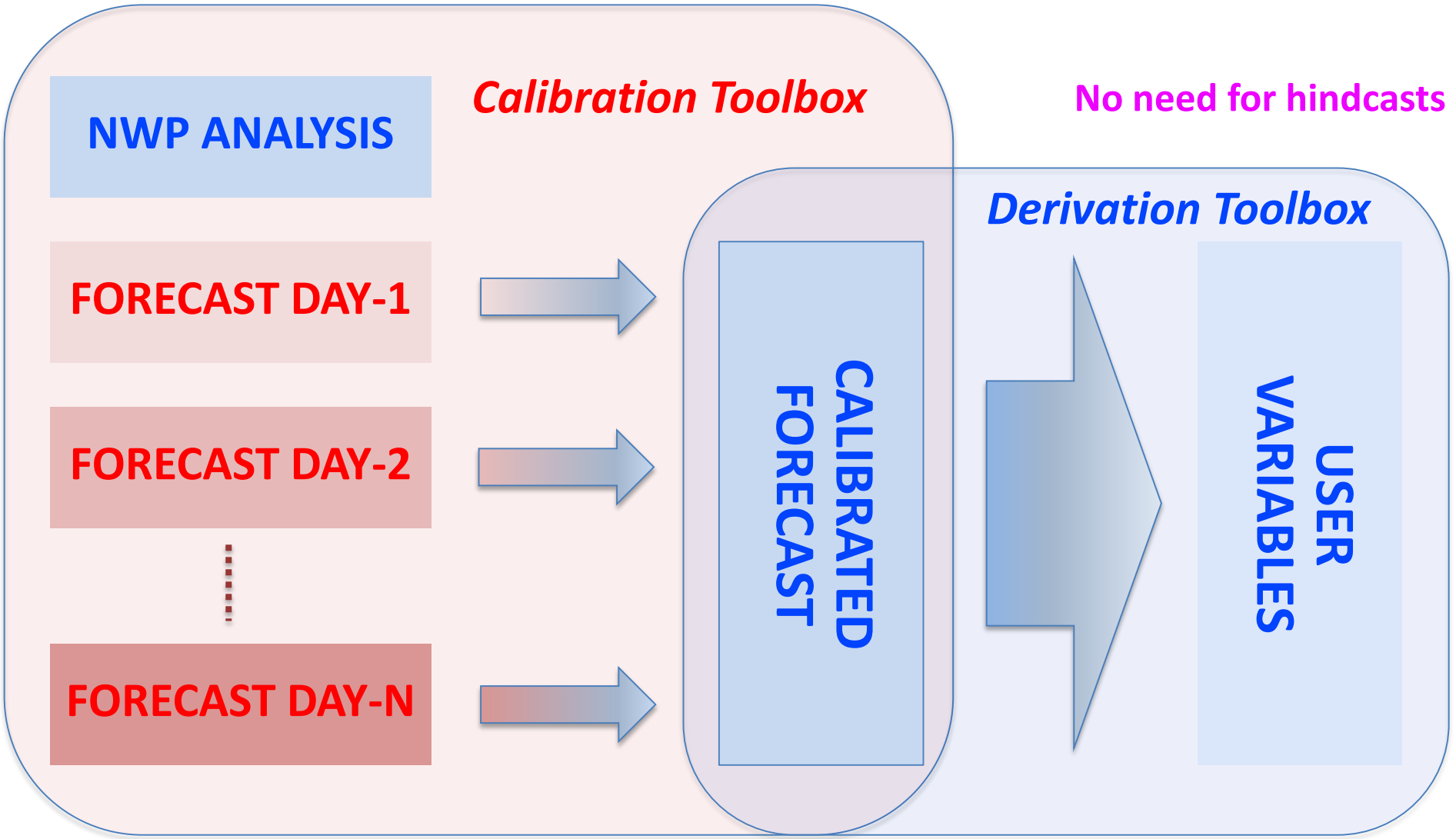
FROM CALIBRATED MODEL VARIABLES TO USER VARIABLES

- **Model does not represent majority of user variables**
  - Derive user variables (eg, visibility) from calibrated prognostic vars
    - Prognostic variables on finer scales; Additional user variables

- **Physically & statistically based methods** unified across
  - NWP model post-processing (EMC)
  - Satellite product generation (NESDIS)
  - NWP Data Assimilation (DA) forward models JCSDA)
  - Statistical product derivation methods (MDL)
  - **Toolbox #2 – Derivation of user variables**

- Create **fine scale gridded analysis of user vars -** ref. for truth
  - More accurate than single observations
  - **Toolbox #3 – Fine scale observationally based analysis**

2-STAGE POST-PROCESSING



FORM OF CALIBRATED FORECAST DATABASE

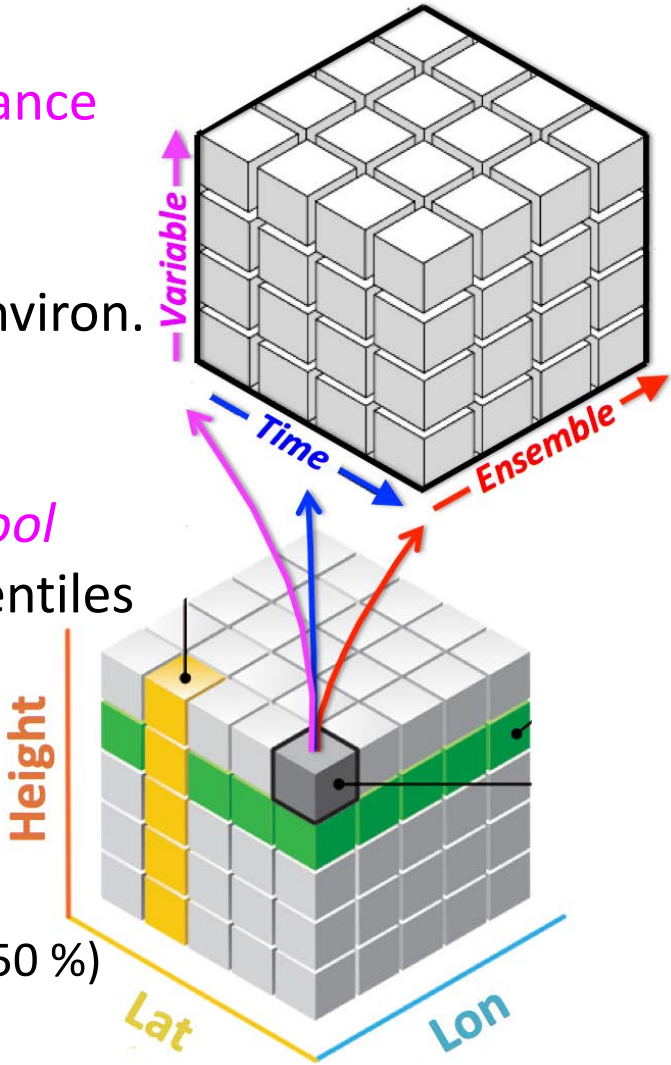
- Large **variety of needs** from wide range of users
  - Univariate single value, exceedance probability, percentile
  - Joint probabilities – precip above, temp below threshold
    - Infinite combinations
    - Impractical to individually calibrate / precompute
  - Calibrated ensemble scenarios
    - For sophisticated downstream applications / decision making

- **Calibrated ensemble forecasts as basic database**
  - All other forecast forms derived from this
    - Prognostic variables on model grid
    - Derived user variables on finer scale grid
    - 3D in space, 1D each in lead time, variables, ensemble members
  - Toolbox #4 – "Seamless" 6D-DataCube (6D-DC)

6D-DC – CENTRAL TENET OF FORECAST PROCESS

Connects NWP output w. users via value added steps

- **Role**
  - Repository of authoritative probabilistic guidance
    - Forecaster modification via *editor tool*
- **Function**
  - Holds answer to all questions about future environ.
    - Access info via *interrogator tool*
- **Content**
  - Calibrated **prog. vars.** from *post-processing tool*
  - Calibrated **user variables** – incl. climate percentiles
- **Format**
  - **6D** - Space (3), time, variables, uncertainty
- **Implementation**
  - Phase 1 – 3-fold expansion of NDFD
    - 10 & 90 percentiles added to central tendency (50 %)
  - Phase 2
    - Cdf, pdf analytical distributions added via ~3 parameters for each
    - Ensemble members for joint probabilities & scenarios



MODIFICATION & INTERROGATION TOOLS

- **Forecaster** over the loop **intervention tools**
  - Automatic monitoring & versatile display tools to inspect 6D-DC
  - High level, intuitive, efficient editor
  - Toolbox #5 – Forecast Editor

- Derive variety of consistent **products from 6D-DC**
  - Interrogation by forecasters
  - Product generation & service for end users
  - Uni- & multivariate exceedence probabilities, forecast & climate percentiles, etc
  - Toolbox #6 – 6D-DC Interrogation

NWP OUTPUT
Ensembles, high res. controls

PROBABILISTIC FORECAST SERVICES

Consolidate all predictive info

Calibrate model prog. variables

Downscale to user grid

Add user variables

PROB NDFD 6D-DATACUBE

Calibrated

Phase-1: *Quantiles*

Phase-2: *Cdf, pdf, ens. members*

6D-DC forecast editor

Third party web & other applications

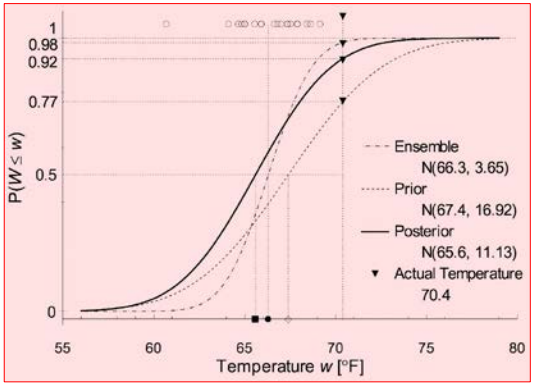
Private sector services

Derive forecast information

User request

Forecaster interrogation

U S E R S

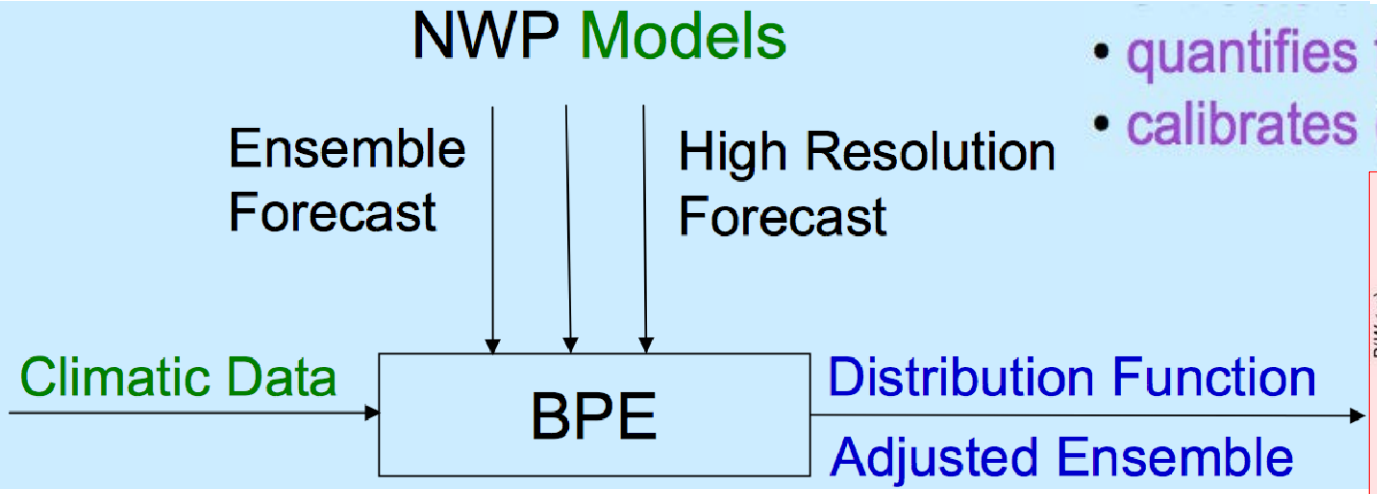


R. Krzysztofowicz

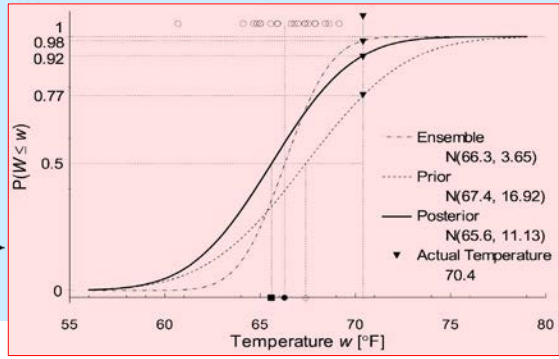
F O R E C A S T E R S

BAYESIAN PROCESSOR OF ENSEMBLE - BPE

To calibrate prog vars – Can be reconfigured to derive user vars



- quantifies total uncertainty
- calibrates (de-biases) ensemble



- Distributional approach – Extremes handled well
- Metagaussian transformation – All continuous variables
- Bayesian combination of
  - Prior – Climatological distribution – Reduced fcst sample need
  - New info – NWP forecasts / ensembles / latest observations
- Multiple Linear Regression at its core – Like MOS
  - Krzysztofowicz & Toth 2008, Krzysztofowicz 2010

OUTLINE / SUMMARY

- **2-Stage post-processing on model / user grids**
  - Calibration of prognostic variables on model grid
  - Derivation of user variables on fine scale grid
- **6D-DataCube - Probabilistic NDFD**
  - Ensembles of prognostic & user variables
- Set of interconnected **toolboxes**
  - Calibration, Derivation, Obs-Analysis, 6D-DC Synthesis, Editor, Interrogator
  - Downstream applications driven by reanalysis-like fcsts
- **Bayesian Processor of Ensemble**
  - Calibrates all continuous variables
  - Extremes handled well
  - Reduces forecast sample need

BACKGROUND

DATA & SAMPLE SIZE REQUIREMENTS

Calibration of prognostic variables on model grid

- **(Re)analysis, (re)forecasts** on model grid
  - Sample size requirements depend on choice of methods
    - Need representative sample
    - Number of cases limited by size of reanalysis
      - Estimation of bias in 1st & 2nd moments

- **Generation of re-analysis / forecast**
  - Retrospective – Logistic challenges, frozen systems
  - Real time – NWP system updates; Testing opportunity

DATA & SAMPLE SIZE REQUIREMENTS – 2

Derivation of fine scale user variables from prog vars

- **Instantaneous relationships**
  - **No need for hindcasts!**
- **(Re)analysis on model & finer grid**
  - To establish statistical relationships
- **Downstream applications use reanalysis**
  - No need for reforecasts – eg, hydromet ensembles

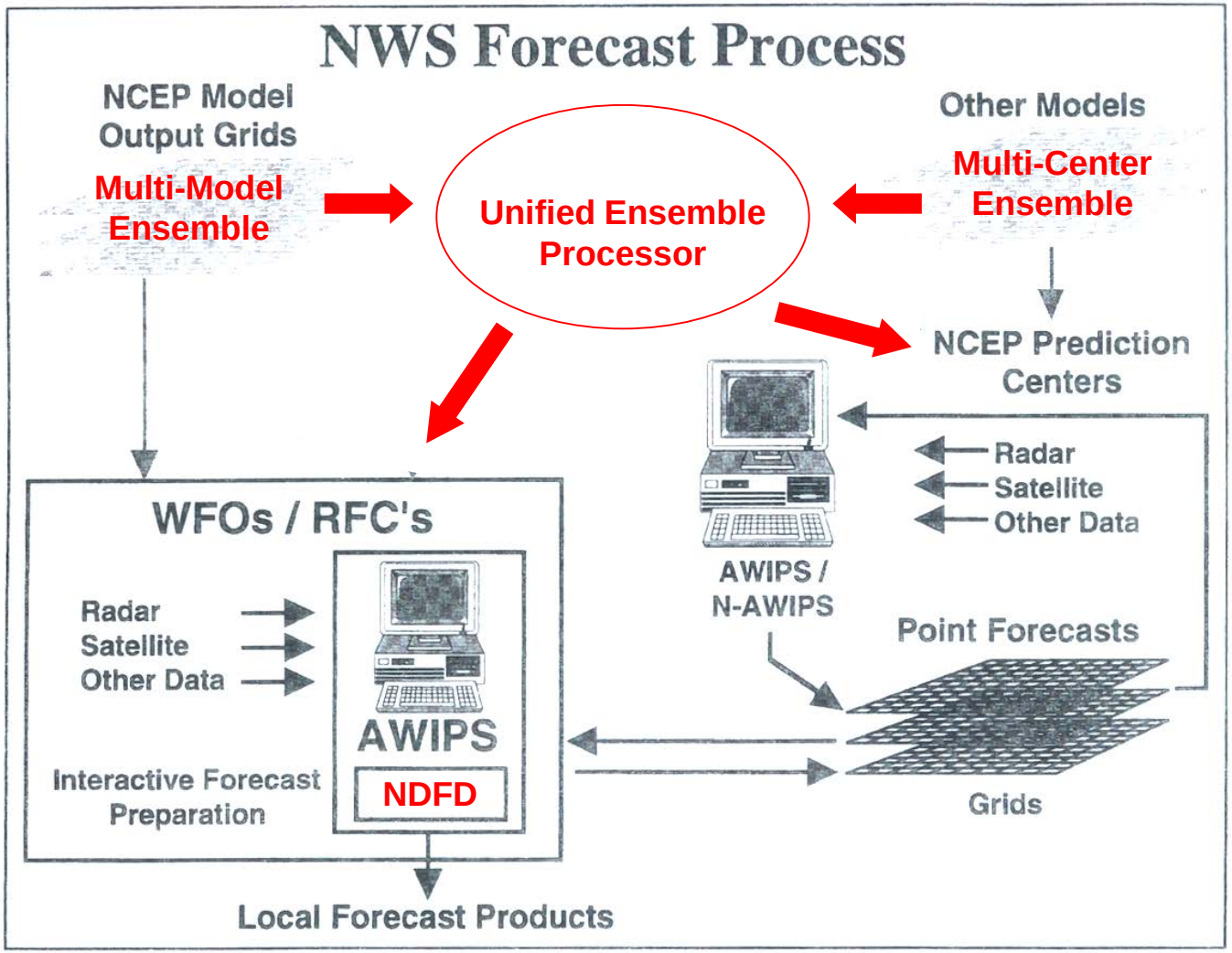
INVEST IN SAMPLE SIZE VS REAL TIME FORECAST

- **Zero-sum cpu game** – Need cost-benefit analysis
  - More hindcasts vs more sophisticated real time forecast system?

- **What info can be calibrated?**
  - Basic parameters (1st & 2nd moments) with achievable samples
  - Most other info - higher moments, covariances
    - Needs orders of magnitude larger sample
    - Derive such info from calibrated NWP prog vars

- **How to potentially lower hindcast requirements?**
  - Use Bayesian methods
    - Climate cdf for reliability, extremes
  - Single (not an ensemble) hindcast
    - Is spread / error relationship stationary?
    - Tune ensemble spread to error in unperturbed forecast
      - Ensemble spread well controlled by a single parameter
  - Downstream applications should be tuned using reanalysis
    - Calibrated prog var forecasts behave like analyses

PRECURSERS TO PROBABILSITIC FORECASTING



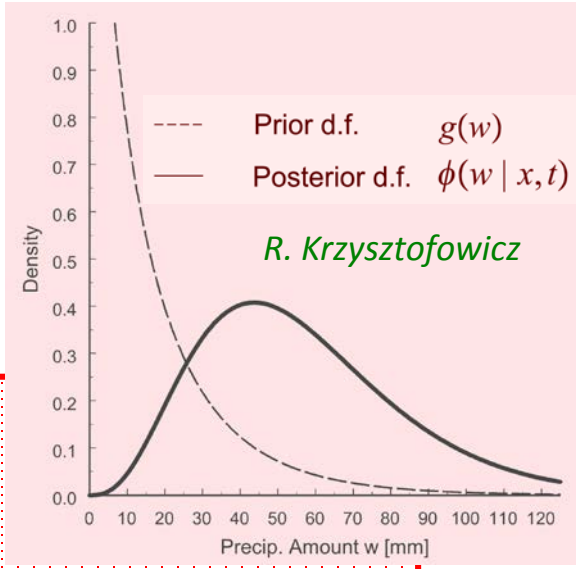
Critical elements of probabilistic forecasting (in red, Toth et al 2009) illustrated in general NWS forecast process design of Wernly & Uccellini (1999, in grey)

Wernly & Uccellini, 1999: Storm Forecasting for Emergency Response. Storms Vol. I, Ed. Pielke & Pielke, Routledge, pp. 70-97.

Toth, Z. et al., 2009: NWS Unified Ensemble Post-processing System (NUEPS). White Paper, pp. 8, available from Editor.

FORECAST STEPS

- **NWP**
  - Predict prognostic variables
    - Based on temporal relationships across variables

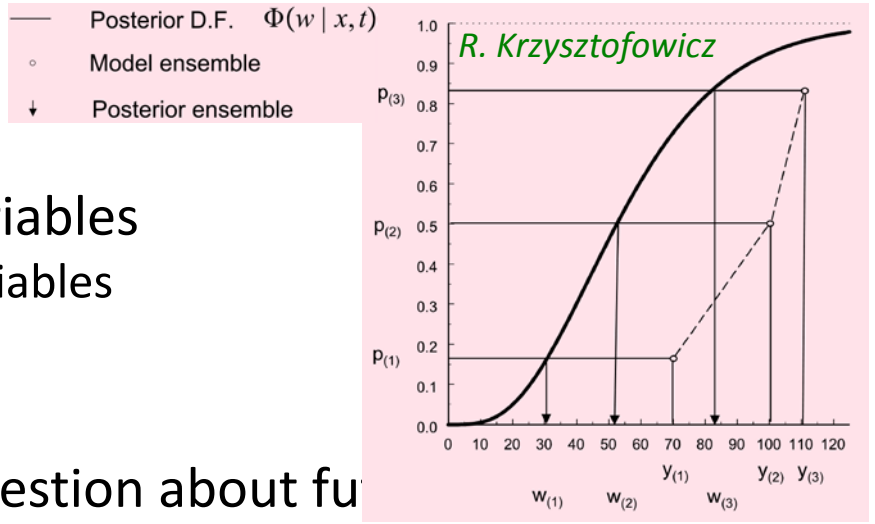


- **Post-process model prog variables**
  - Consolidate various NWP guidance + climate
  - Calibrate model prognostic variables
    - ***Lead-time dependent – forecast sample needed***
  - Distributional approach - Bayesian Processor of Ensemble (BPE)

- **Derive user variables**
  - Based on simultaneous physical / statistical relationships across vars
    - ***Independent of lead time – reanalyses (but no hindcasts) needed***
  - Derive additional user variables from model prognostic variables
    - NWP model - 2m temperature
    - UPP – Precipitation type
    - DA forward models – Reflectivity based on prognostic variables
    - Satellite product generation algorithms – Look-alike images
    - BPE – Statistical relationships btw coarse prognostic & fine scale user variables

FORECAST STEPS – Cont.

- **6D DataCube** (6D-DC)
  - Repository of calibrated prognostic and user variable forecasts
    - Stage 1 – Quantiles (forecast values for 10, 50, 90, or more percentiles)
    - Stage 2 – Continuous & density distribution functions & ensemble members



- 6D-DC **Forecast Editor**
  - Interactive tool to modify calibrated prognostic (& user) variables
    - Changes across space, time, & variables

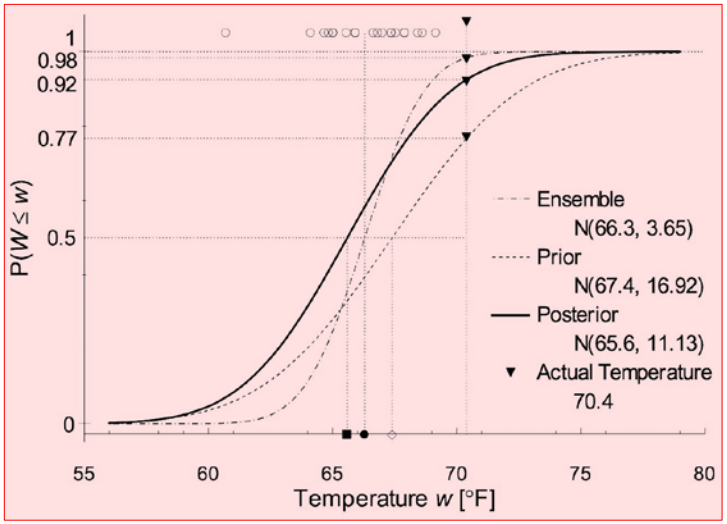
- 6D-DC **Interrogator**
  - Interactive tool to answer any question about fu
    - Eg, joint probability of high winds & heavy precip at point A during period B
    - For forecasters and external users
    - Graphical displays

- Web-based & other **3rd party applications**
  - Linked with 6D-DC (possibly via Interrogator)

BAYESIAN PROCESSOR OF ENSEMBLE - BPE

To calibrate prog vars – Can be reconfigured to derive user vars

- **Distributional approach**
  - Extremes handled graciously
- **Metagaussian transformation**
  - Works with all continuous variables
    - Set of ~30 distribution functions
    - All predictive info extracted / fused



- **Combines info from hires controls, ensembles, mult. vars, latest obs**
  - Multiple Linear Regression at its core – like MOS
- **Bayesian method**
  - **Climatological distribution** used as prior
    - Joint analysis – forecast sample - Can be smaller if representative, extremes captured
- **Outputs**
  - Calibrated cdf, pdf, quantiles, ensemble – all consistent