Odds Are It’s Wrong Reading

Assignments/Dates

- Chapter 4 notes due Feb 24
- Review Feb 24
- Exam March 1

Model: tool for simulating or predicting the behavior of a dynamical system such as the atmosphere

- heuristic: rule of thumb based on experience or common sense
  - Not strictly accurate or always reliable
  - Example: If the winds get strong, there'll be a lot of damage
- conceptual: framework for understanding physical processes based on physical reasoning
  - Very useful- that's what fills textbooks
  - Example: LIMBS
- empirical: prediction based on past behavior
  - Can tell us what has been likely in the past: record values, typical values, etc.
  - Example: average daily temperature in June vs. January
- analytic: exact solution to "simplified" equations that describe the atmosphere
  - Very useful to understand how things work
  - Example: many of the conceptual models described in the textbook rely on analytic models
- numerical: integration of governing equations by numerical methods subject to specified initial and boundary conditions
  - What is used for day-to-day weather forecasting
  - Example: Global forecast system (GFS) model

PRISM: http://www.prism.oregonstate.edu/

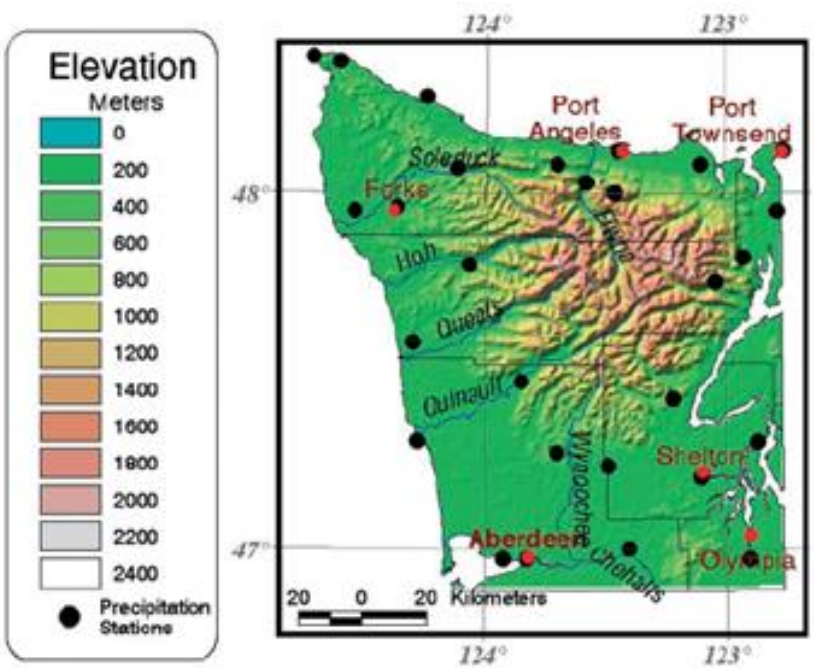
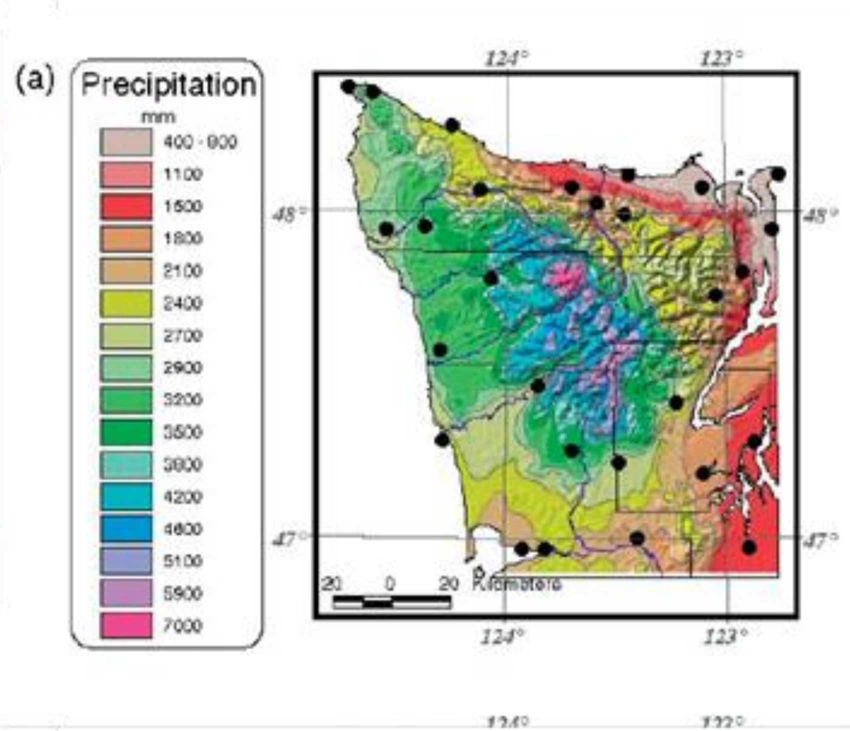


Fig. 2. Terrain map of the Olympic Peninsula, in the northwest corner of Washington State, USA. Terrain resolution is 500 m. Locations of precipitation stations used in mapping are shown as black dots, town locations are red dots



Climatology & Persistence: Empirical Models

- Climatology: what has happened in the past
- Best forecast for specific conditions a week or two in advance
- Really bad forecast if the present conditions are far from what has happened in the past
- Persistence: what is happening is likely to continue to happen
- Best forecast usually for the next few minutes
- Really bad forecast if the weather is changing

Numerical Weather Prediction Model

To make a forecast we need:

(1) Observations of the present state of the atmosphere, ocean, and land surface (snow, soil moisture, etc.)

(2) Description of the behavior of the atmosphere in quantifiable manner (requires equations)

(3) Numerical methods to use information on present conditions and project forward what will happen next

(4) Computer resources sufficient to make a forecast in reasonable amount of time

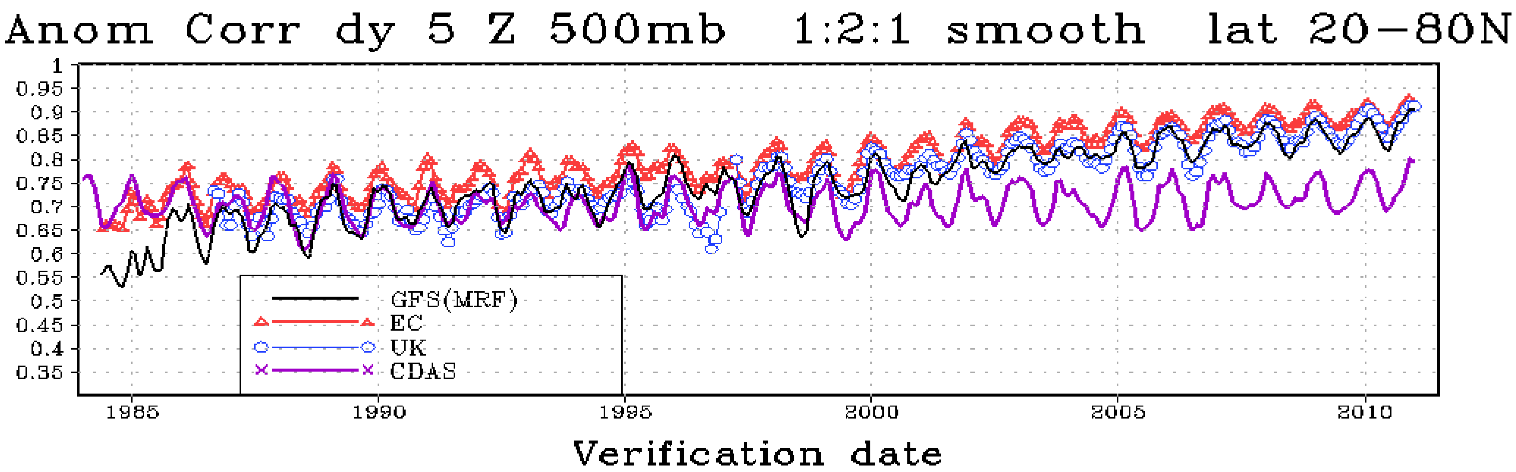
Determining Forecast Error

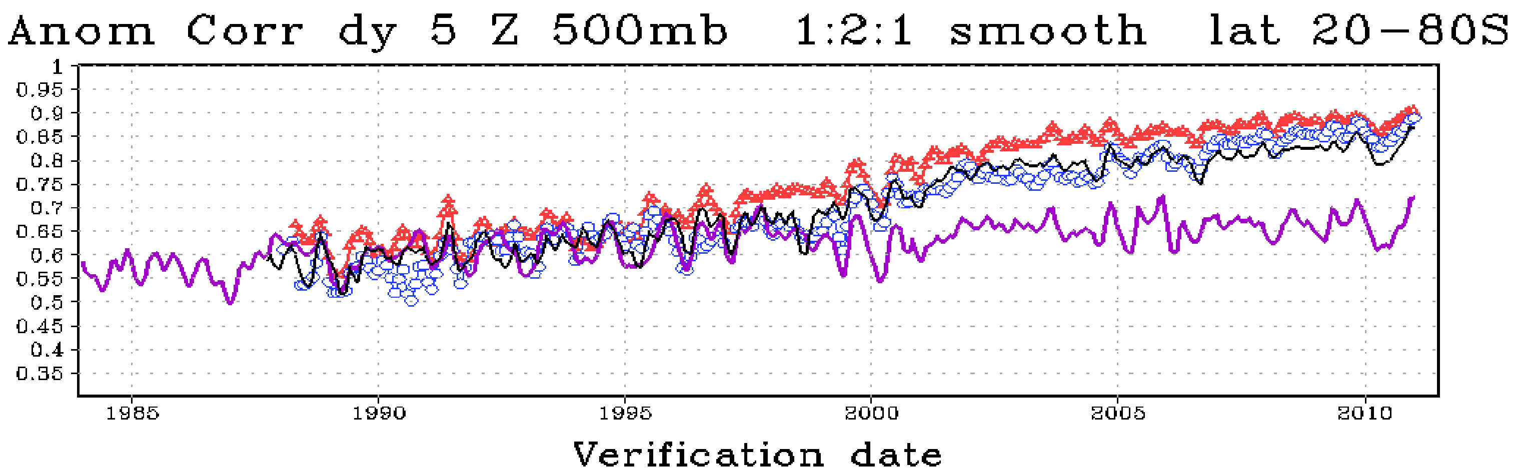
- All observations have errors
- All forecasts have errors
- Some models take into account observational errors, others do not.

Define:

- T_{ti}- unknown truth for the ith value
- T_{oi}- ith observation
- T_{fi}- forecast for the ith value

Comparing Forecast Anomaly Maps to Analyses





Aggregate Forecast Error

- ε_{oi}- error of ith observation = T_{oi}- T_{ti}
- ε_{fi}- error of ith forecast = T_{fi}- T_{ti}
- If making many forecasts, interested in accuracy of sample as a whole
- Expected value is the mean value denoted by an overbar.
- Assume that observations and model forecasts are unbiased, which means that:

- $\overline{T}_t = \overline{T}_o = \overline{T}_f$. Then, the mean errors are: $\bar{\varepsilon}_o = \bar{\varepsilon}_f = 0$ where we have summed over all values.

- common measure of the spread of the errors (root-mean squared error, E, or rms error) is equivalent to the sample standard deviation of the forecasts or observations relative to the unknown truth.

$$E_f = \sqrt{\overline{\varepsilon_{fi}^2}} = \sigma_{ef} \qquad E_o = \sqrt{\overline{\varepsilon_{oi}^2}} = \sigma_{eo}$$

Error of Climatological Forecast

Assume forecast SLC temperature at 00 UTC on March 1, 2011 based on average of 30 values of temperature for that time and date during the period 1981-2010.

$T_{fi} = T_{ci} = \overline{T}_o = \overline{T}_t$ where i day of the year

Assuming unbiased

What is the error of a climatological forecast evaluated over many such forecasts?

$$E_c = \sigma_{ec} = \sqrt{\overline{\varepsilon_{ci}^2}} = \sqrt{\overline{(T_{ci} - T_{ti})^2}} = \sqrt{\overline{(\overline{T}_t - T_{ti})^2}} = \sigma_t$$

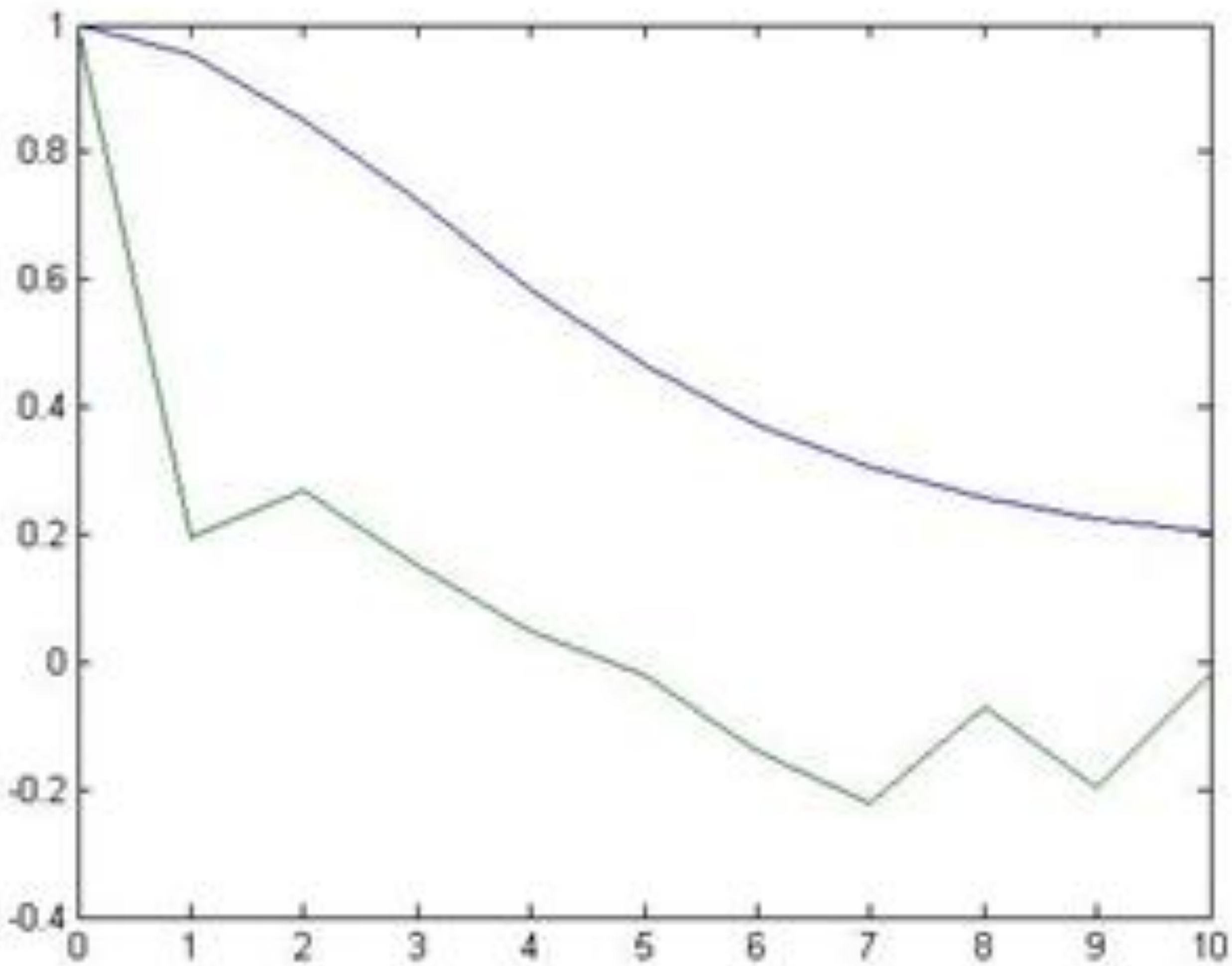
standard deviation of the true values

Climatological forecast has same variability as truth

Persistence: Serial Correlation

- linear autocorrelation: measure of persistence
- relates pairs of data from the same sample separated by lag τ.
- As long as we have a very long time series of data such that
- $\tau \ll n$, then $$r(\tau) = (\overline{x'(t)x'(t+\tau)}) / s_x^2$$
- and $r(0)=1$ and $-1 \le r(\tau) \le 1$
- As τ increases, $r(\tau)$ to decrease. T
- rapidity at which it does is a measure of the "memory" in that component of the environment.
- an autocorrelation of 0 at lag τ may reflect the frequent occurrence of wavelike propagating features
- the temporal period of the wave phenomenon may be crudely estimated as 4τ

Persistence of Great Salt Lake Level vs. Utah annual rainfall



Error of Persistence Forecast

autocorrelation (r(τ)) can be used to predict a future value τ time steps later than the current value:

$$T_p(t+\tau) - \overline{T}_o = r(\tau)(T_o - \overline{T}_o) + (1 - r^2(\tau))\sigma_o^2 e$$

where e is a random number defined from Gaussian distribution mean 0 and sd = 1, Assume observations unbiased and no observational error (perfect obs)
Then, persistence forecast error is:

$$E_p = \sqrt{(r-1)^2 \overline{(T_t - \overline{T}_f)^2} + (1-r^2)\sigma_t^2 \overline{e^2}} = \sigma_t \sqrt{r^2 - 2r + 1 + 1 - r^2} = \sigma_t \sqrt{2(1-r)} = E_c \sqrt{2(1-r(\tau))}$$

E_p = 0 at the initial and as lag gets very large then $E_p = \sqrt{2} E_c$

The first order autoregressive model will have an error equal to that of a climatological forecast for the lead time when r=0.5

Perfect Model with Imperfect Initial Conditions

(1) model unbiased and reproduces the variability of the true state of the environment $\sigma_f = \sigma_t$

(2) as forecast duration increases, the departures of the forecast from their mean and the departures of the true state from their mean become unrelated when averaged over many forecasts,

$$E_f = \sqrt{\overline{(T_{fi} - T_{ti})^2}} = \sqrt{\overline{((T_{fi} - \overline{T_f}) - (T_{ti} - \overline{T_t}))^2}} = \sqrt{\overline{(T_{fi} - \overline{T_f})^2} - 2\overline{(T_{fi} - \overline{T_f})(T_{ti} - \overline{T_t})} + \overline{(T_{ti} - \overline{T_t})^2}}$$

forecasts eventually have no correspondence to what is actually happening- they are random

$$E_{f\infty} = \sqrt{2}\sigma_t = \sqrt{2}E_c$$

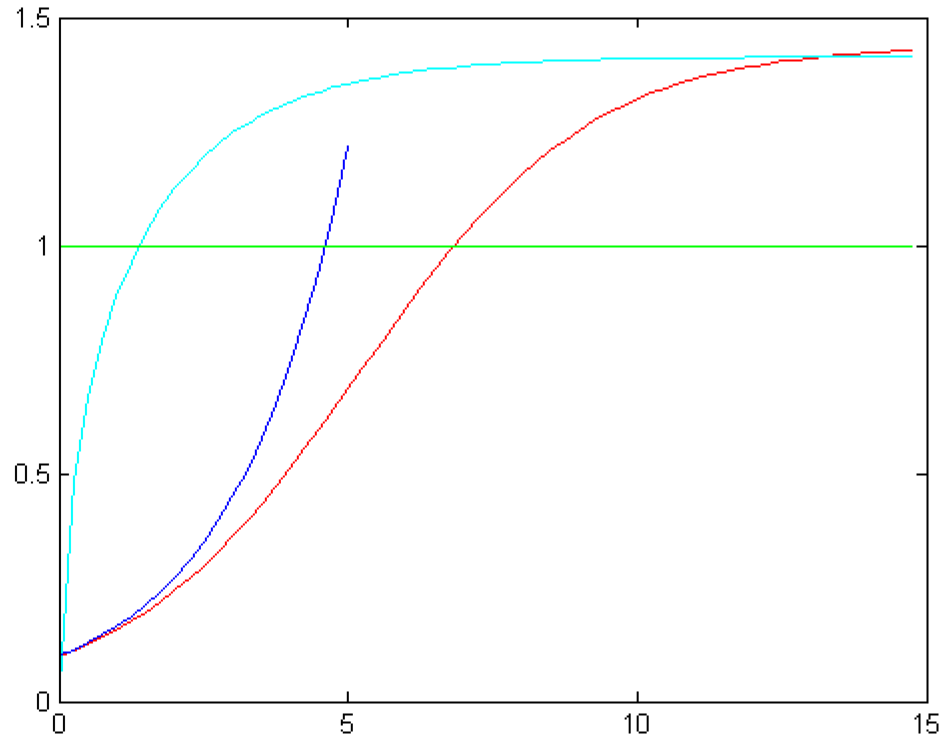
Comparing Error Growth of Perfect Model to Climo and Persistence

Green- climo forecast

cyan- persistence

Blue- exponentional error growth

Red- slower error growth



Look at forecast_error.m

Summary

persistence forecast is better empirical forecast than a climatological forecast at short lead times

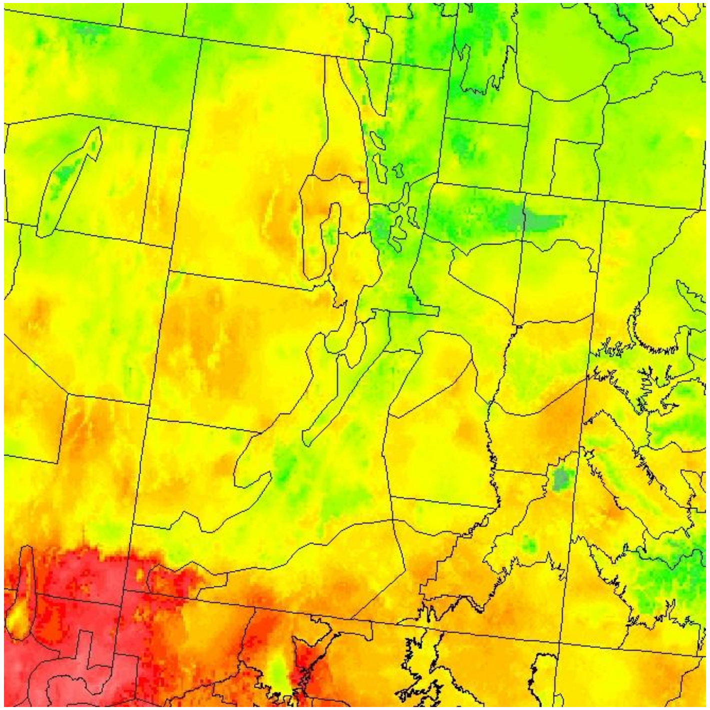
numerical weather prediction models should out perform persistence and climatological forecasts at lead times out to some lead time

for all models: model accuracy is often evaluated in a least squared sense relative to the unknown truth.

Statistical Interpolation

- A common goal in environmental fields is to take observations of environmental conditions scattered over a spatial domain and interpolate/extrapolate those values to a regular grid.
- Simple schemes (Cressman) were developed in the atmospheric sciences over 50 years ago to give greater weight to observations close to the location at which the analysis value is desired compared to more distant observations
- Rather than attempting to interpolate fields without any other information, early researchers recognized that defining a “first guess” or “background” from a source such as a model forecast and weighting corrections between the observations and first guess fields was a superior approach.

Objective Analysis



- A map of a meteorological field
- Relies on:
  - observations
  - background field
- Used for:
  - Initialization for a model forecast
  - Situational awareness
  - Verification grid

Discussion Points

- Why are analyses needed?
  - Application driven: data assimilation for NWP (forecasting) vs. objective analysis (specifying the present or past)
- What are the goals of the analysis?
  - Define microclimates?
    - Requires attention to details of geospatial information (e.g., limit terrain smoothing)
  - Resolve mesoscale/synoptic-scale weather features?
    - Requires good prediction from previous analysis
- How is analysis quality determined? What is truth?
  - Evaluating analysis by withholding observations

Discussion Points (cont.)

- What causes large variations in surface temperature, wind, moisture, precipitation over short distances?
  - Terrain, convection, etc.

- How well can we observe, analyze, and forecast conditions near the surface?
  - What errors should we tolerate?

- To what extent can you rely on surface observations to define conditions within 2.5 x 2.5 or 5 x 5 km^2 grid box?
  - Do we have enough observations to do so?

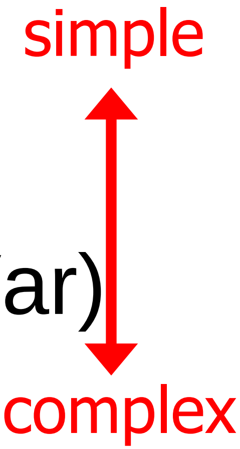
ABC's

Analysis value = Background value + observation Correction

- An analysis is more than spatial interpolation
- A good analysis requires:
  - a good background field supplied by a model forecast
  - observations with sufficient density to resolve critical weather and climate features
  - information on the error characteristics of the observations and background field
  - appropriate techniques to translate background values to observations (termed "forward operators")

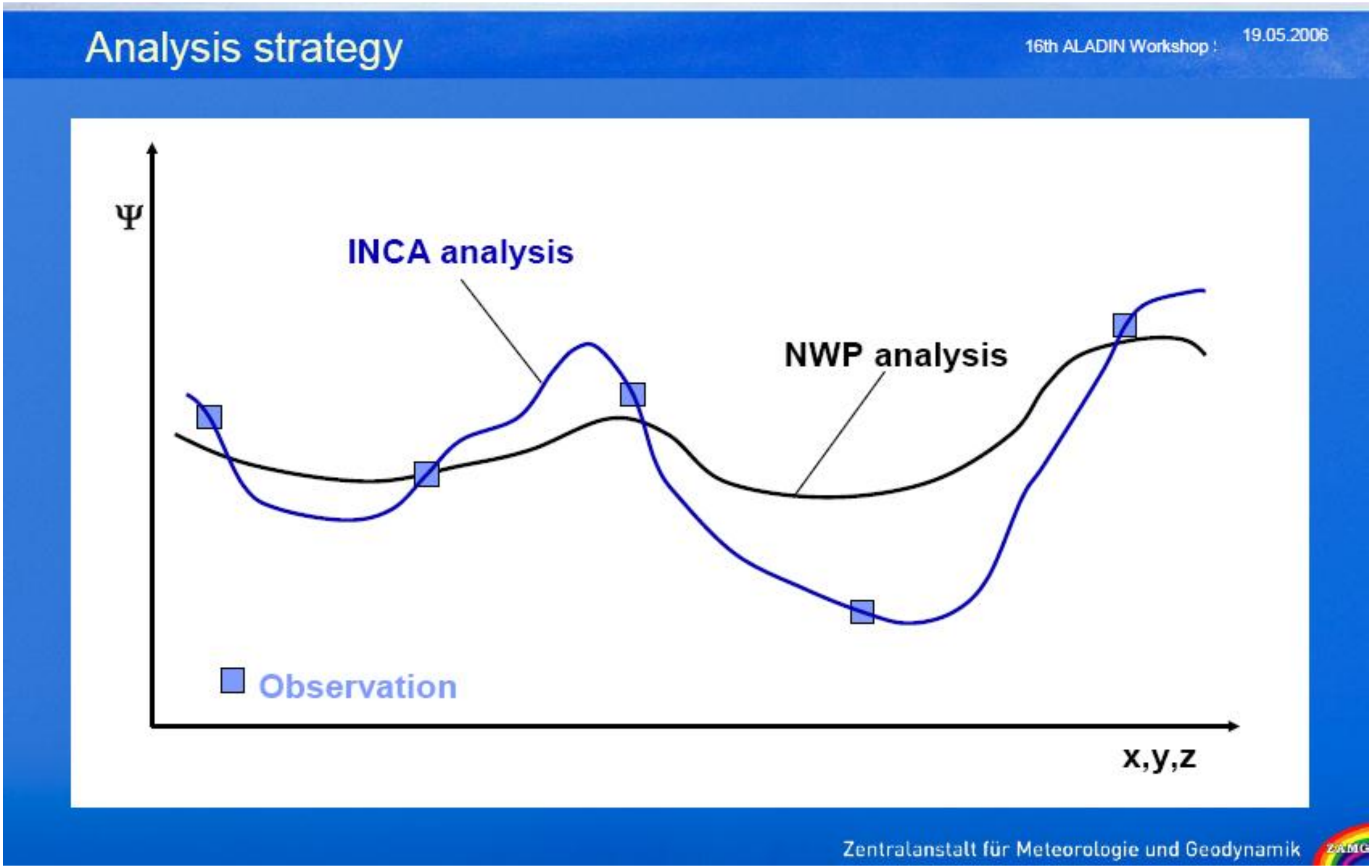
Objective Analysis Approaches

- Successive Corrections
- Optimal Interpolation
- Variational (2DVar,3DVar, 4DVar)
- Kalman or Ensemble Filters



- Kalnay (2003) Chapter 5 – good overview of different schemes

One Approach: Adjust Model Guidance to Match Observations (INCA and MatchObsAll)

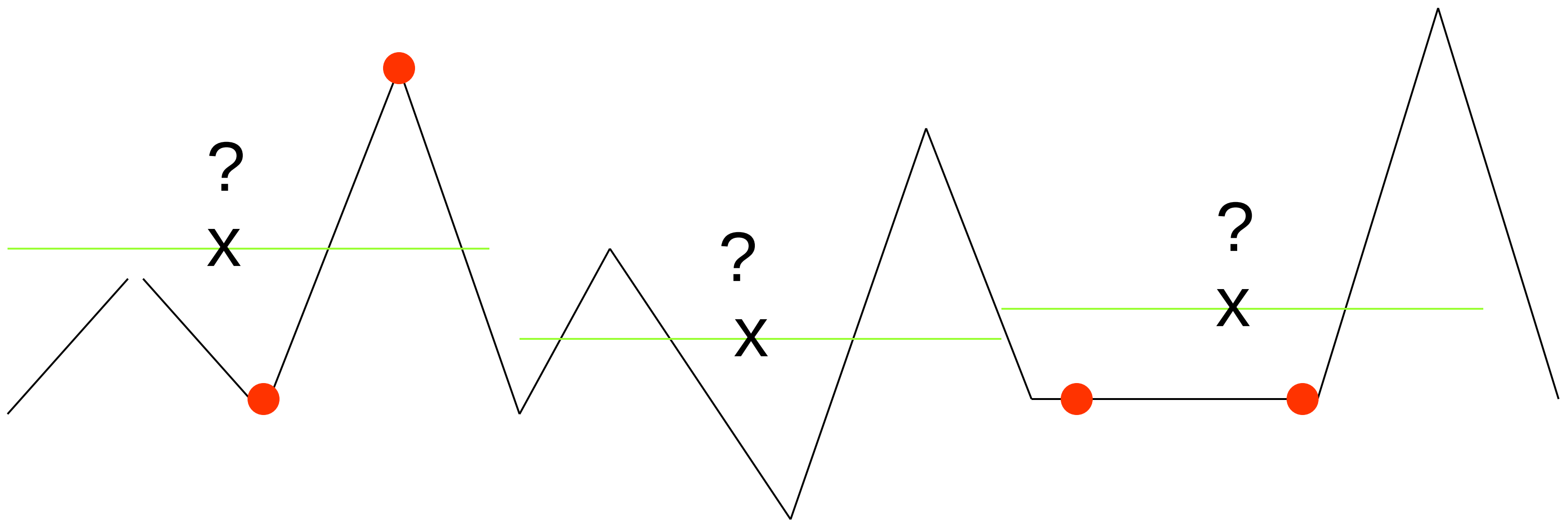


Potential for Confusion

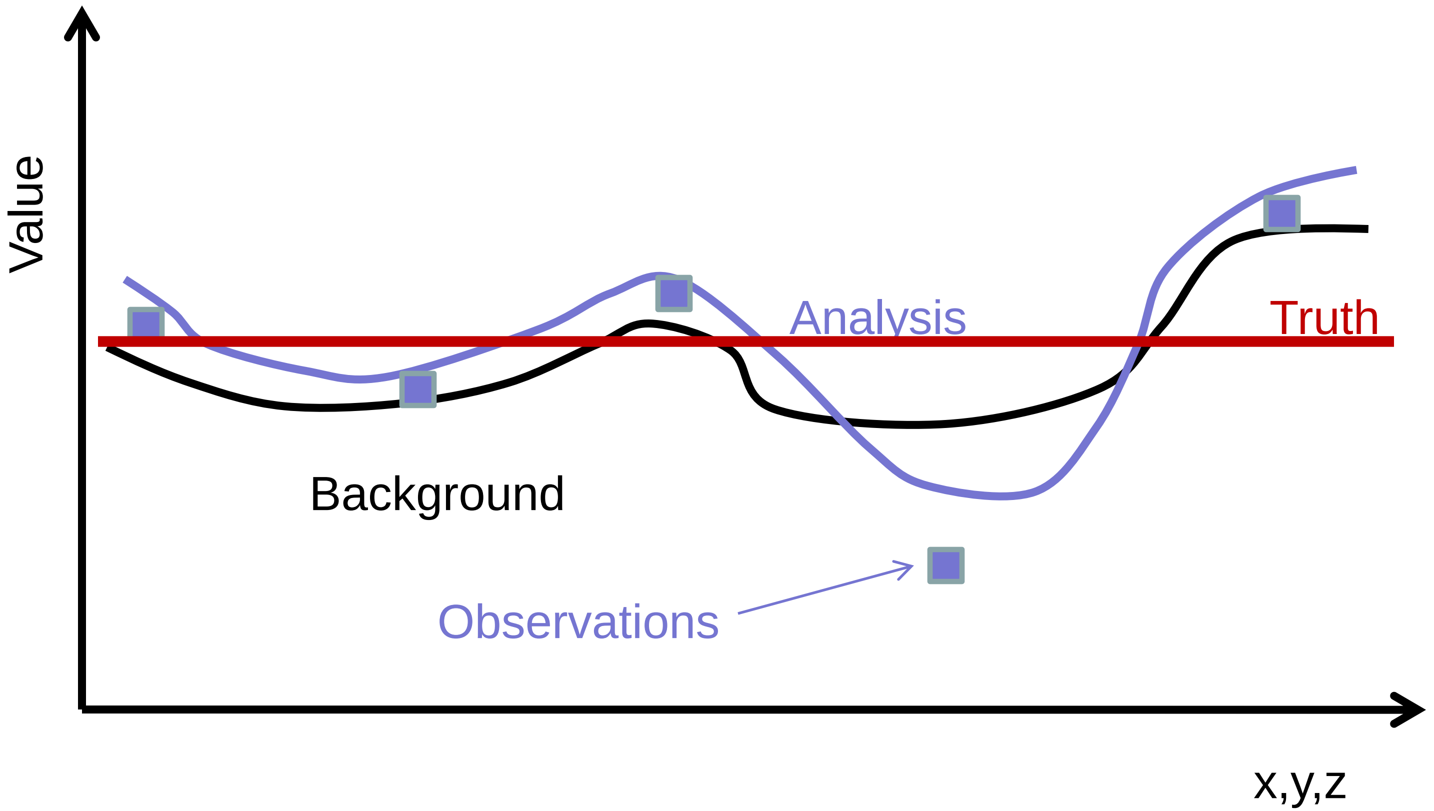
- Analysis systems like INCA suggest that the analysis should exactly match every observation
- Variational or other analysis values usually don’t match surface observations
  - Analysis schemes are intended to develop the “best fit” to the *differences* between the observations and the background taking into account observational and background errors when evaluated over *a large sample* of cases

What are appropriate analysis gridpoint values?

- Inequitable distribution of observations
- Differences between the elevations of the analysis gridpoints and the observations

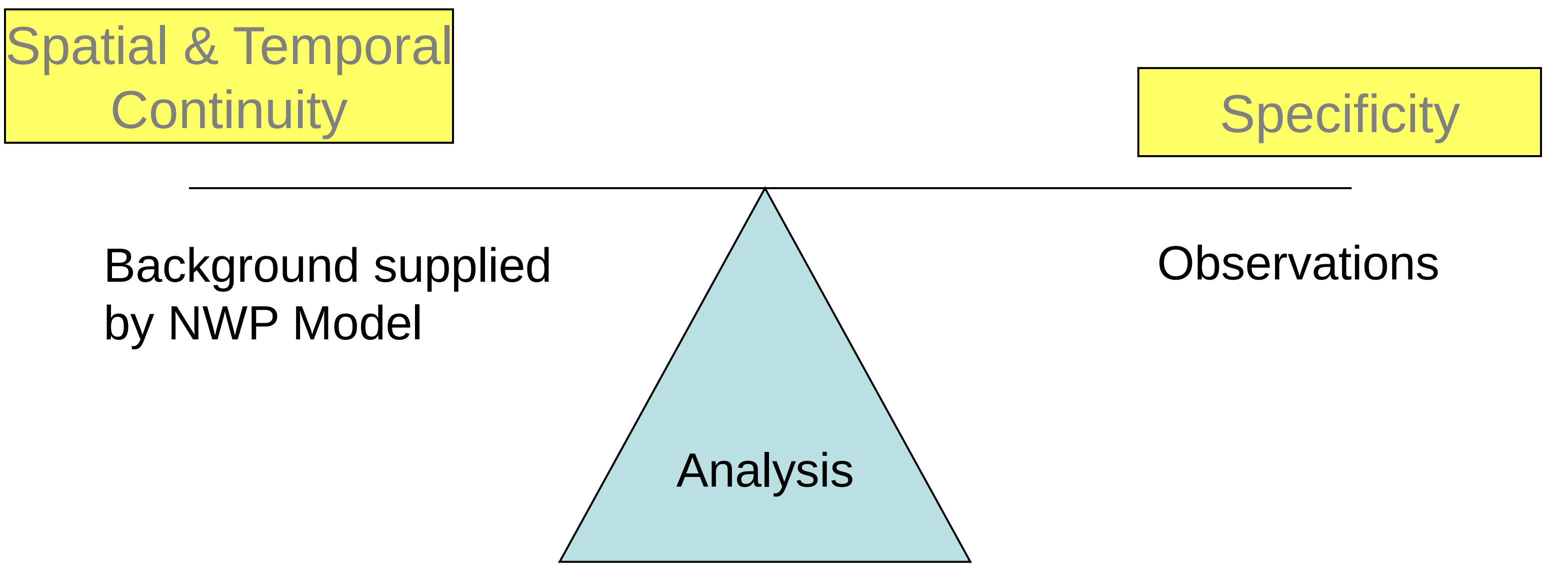


Predominant Approach: Constrain Imperfect Model Guidance by Imperfect Observations

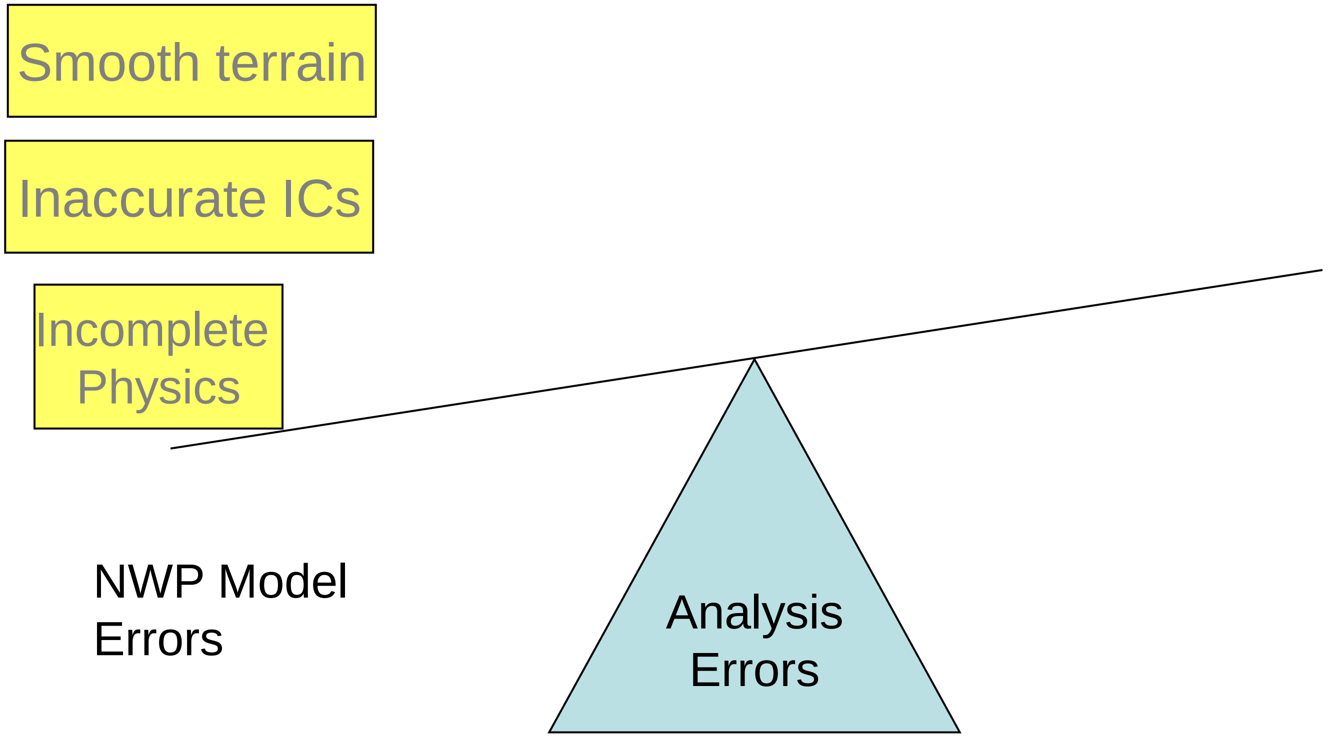


Need for balance…

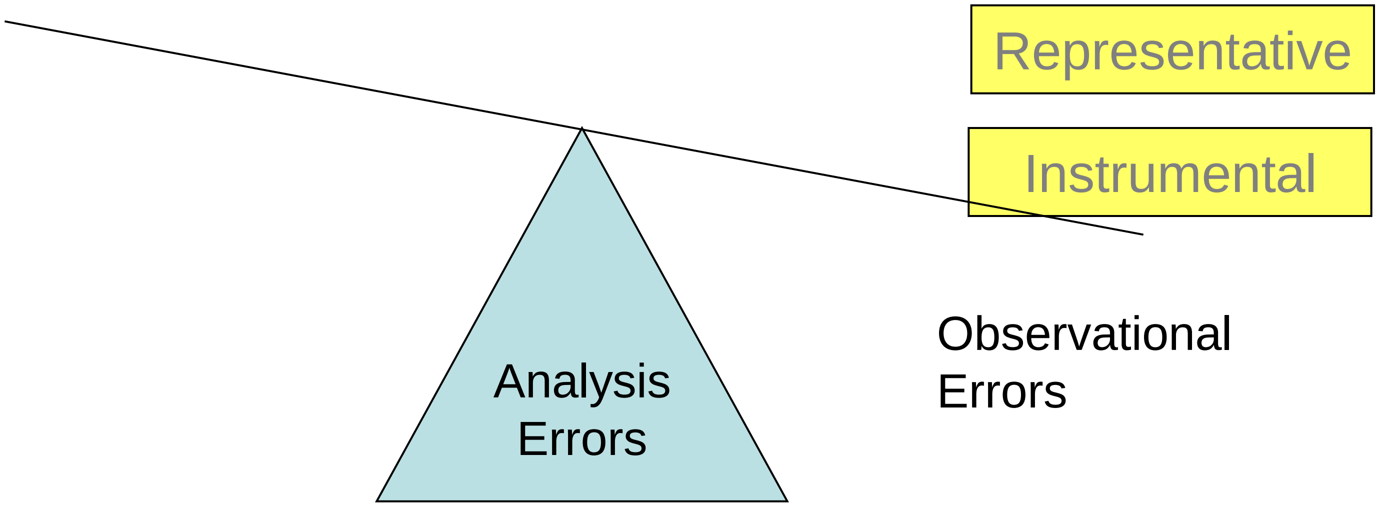
Models or observations cannot independently define weather and weather processes effectively



Recognition of Sources of Errors



Recognition of Sources of Errors

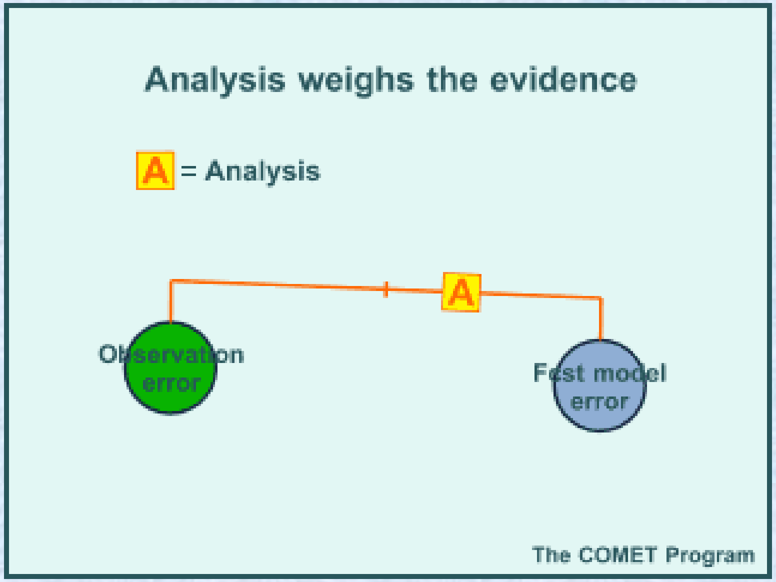


Which is weighted more – observation or model value?

- The analysis procedure (2D-VAR) “knows” the value and limitations of observations using expected observation errors for each data type
- It “knows” model’s behavior by using model forecast error statistics at each grid point and spatial relationships of error patterns
- The analysis assesses penalties-
  - Penalty for deviations from observations
    - Larger penalty if observation type is known to have smaller error
  - Penalty for deviations from background
    - Larger penalty if model forecast is usually good
  - Scheme chooses analysis that pays the smallest total penalty for observations and model combined

- We want the analysis to:
  1. Draw closer to better quality data
  2. Retain more details in the background from a better quality model

But the weighting may be incorrect if error statistics are not appropriate for today’s weather



Background Values

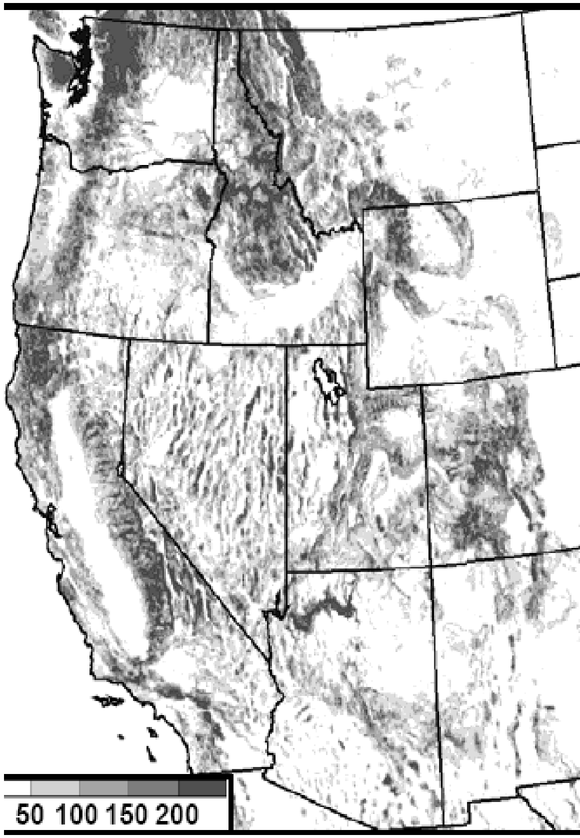
- Obtained from an analysis:
  - Climatology or analysis from prior hour
  - An objective analysis at a coarser resolution
  - Short term forecast
- Most objective analysis systems account for background errors but approaches vary

Observations

- Observations are not perfect…
  - Gross errors
  - Local siting errors
  - Instrument errors
  - Representativeness errors
- Most objective analysis schemes take into account that observations contain errors but approaches vary

Representativeness Errors

- Observations may be accurate…
- But the phenomena they are measuring may not be resolvable on the scale of the analysis
  - This is interpreted as an error of the *observation* not the analysis
- Common problem over complex terrain
- Also common when strong inversions
- Can happen anywhere



Sub-5km terrain variability (m)
(Myrick and Horel, WAF 2006)

Incorporating Errors

- Basic example:

$$T_a = T_b + W(T_o - T_b) \qquad W = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

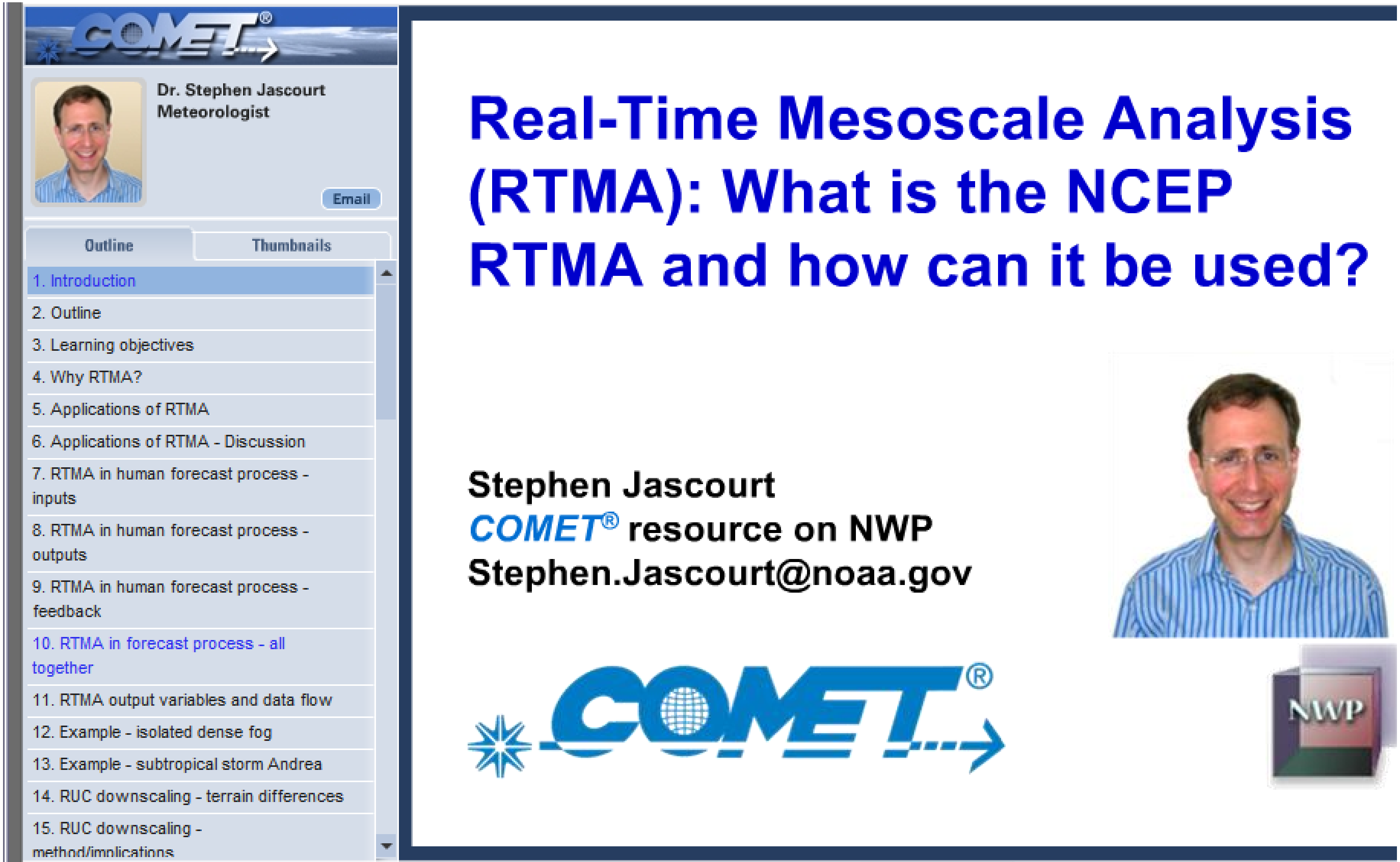
σ_b = background error variance

σ_o = observation error variance

W = 0, distrust observation

W = 1, trust observation

More Info… www.meted.ucar.edu



The actual ABCs…

- The RTMA analysis equation looks like:

$$\left(\vec{P}_b^{\,T} + \vec{P}_b^{\,T}\vec{H}^T\vec{P}_o^{-1}\vec{H}\vec{P}_b\right)\vec{v} = \vec{P}_b^{\,T}\vec{H}^T\vec{P}_o^{-1}\left[\vec{y}_o - \vec{H}\left(\vec{x}_b\right)\right]$$

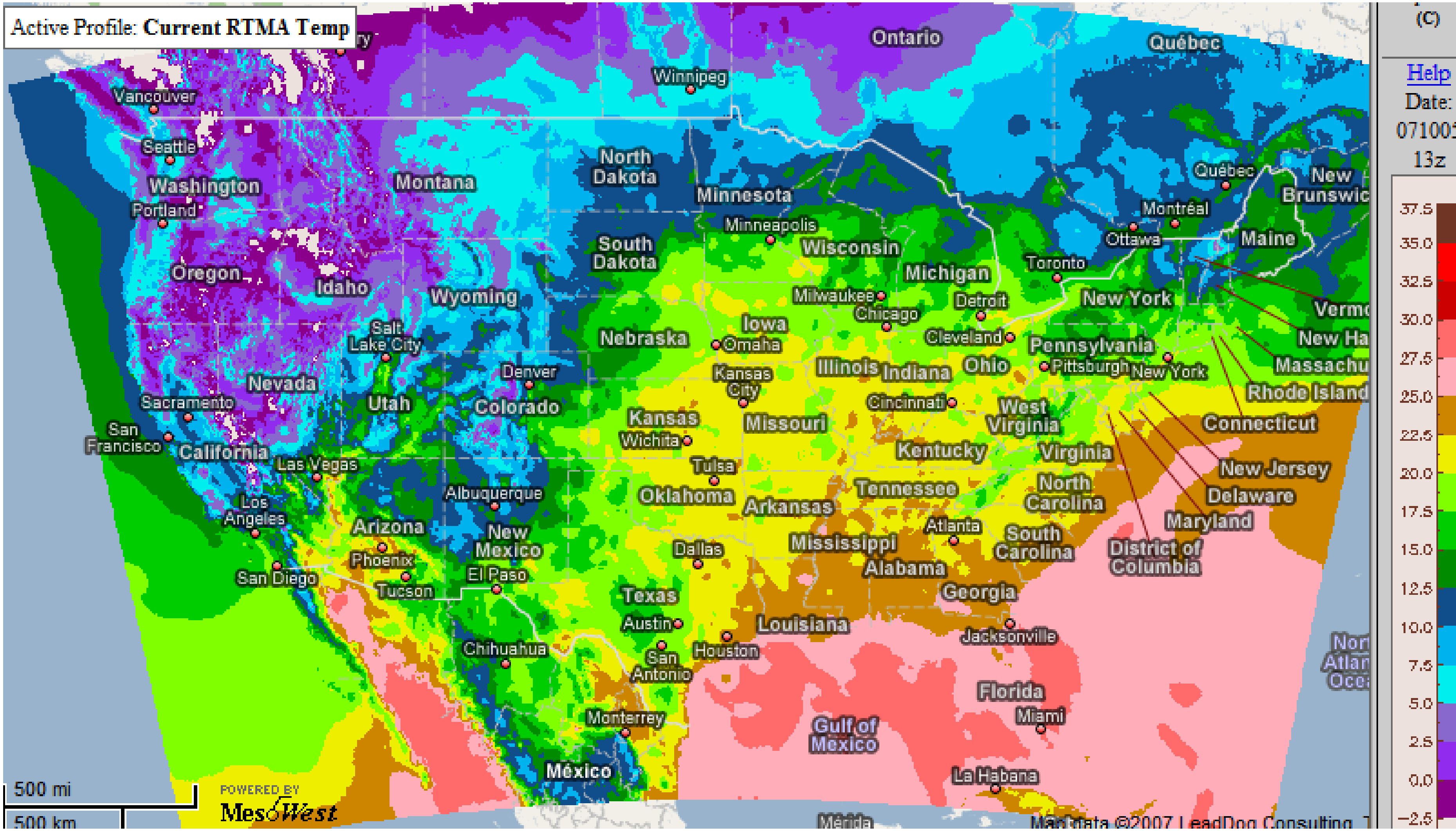
$$\vec{x}_a = \vec{x}_b + \vec{P}_b\vec{v}$$

- Covariances are error correlation measures between all pairs of gridpoints
- Background error covariance matrix can be extremely large
  - 2,900 GB memory requirement for continental scale
  - Recursive filters significantly reduce this demand

Estimation of Observation and Background Error Covariances

- Temperature errors at two gridpoints may be correlated with each other
- Error covariances specify the influence of observation innovations upon surrounding gridpoints
- RTMA used decorrelation lengths of:
  - Horizontal (*R*): 40 km
  - Vertical (*Z*): 100 m
  - Now increased to ~80 km and 200 m respectively
- *Significant limitation to specify error covariances rather than determine them through ensemble methods*

RTMA CONUS Temperature Analysis



Local Surface Analysis

- Solving linear system of form Ax=b using GMRES- generalized minimal residual method

$$\left(\overrightarrow{P_b}' + \overrightarrow{P_b}'\,\overrightarrow{H}'\,\overrightarrow{P_o}^{-1}\,\overrightarrow{H}\,\overrightarrow{P_b}\right)\vec{v} = \overrightarrow{P_b}'\,\overrightarrow{H}'\,\overrightarrow{P_o}^{-1}\left(\overrightarrow{y_o} - \overrightarrow{H}\left(\overrightarrow{x_b}\right)\right)$$

$$\overrightarrow{x_a} = \overrightarrow{x_b} + \overrightarrow{P_b}\,\vec{v}$$

- In matlab x= gmres(A,b)

Assumptions affecting the Analysis

1. **Statistical Assumptions**
   - Observation and model errors are assumed to follow normal distributions
     - Works well for common cases, but not extreme
     - Can’t distinguish different model performances in different regimes
     - Can’t distinguish different local conditions
2. **Assumed Observation Error**
   - Instrument (well known – an engineering matter)
   - Representativeness (not well known)
     - Accurate observation doesn’t represent average value over entire grid box
   - Observation error should vary by weather scenario – but no one knows how to do this
3. **Assumed Background Error**
   - Based on model performance statistics
   - If model performs differently than it usually does for this type of situation, model errors may be inappropriate
4. **Assumed Balance Constraint: Generally Not Known On Mesoscale**
   - Mass –Wind linkage is loosely enforced
   - An “initialization” or “spin-up” step is no longer necessary – balance is achieved within the analysis itself

Summary

- Improving current analyses such as RTMA requires improving observations, background fields, and analysis techniques
  - Increase number of high-quality observations available to the analysis
  - Improve background forecast/analysis from which the analyses begin
  - Adjust assumptions regarding how background errors are related from one location to another
- Future approaches
  - Treat analyses like forecasts: best solutions are ensemble ones rather than deterministic ones
  - Depend on assimilation system to define error characteristics of modeling system including errors of the background fields
  - Improve forward operators that translate how background values correspond to observations

Trauth Text: What Did/Didn't Get Covered

Chapter 1: all

Chapter 2: all

Chapter 3: all but discrete theoretical distributions, chi squared test

Chapter 4: 4.1-4.4

Grad students second half: Chapters 5, 6, 9 and…(?)