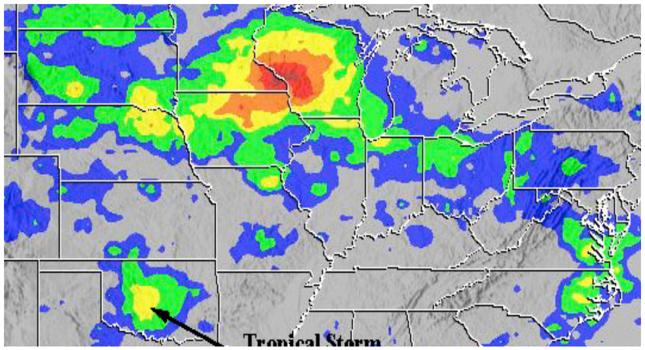
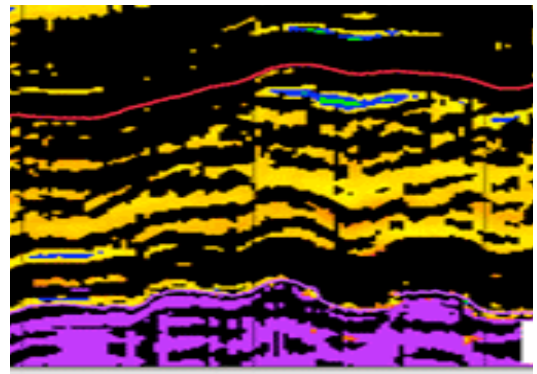
Ensemble-Based Characterization of Uncertain Features

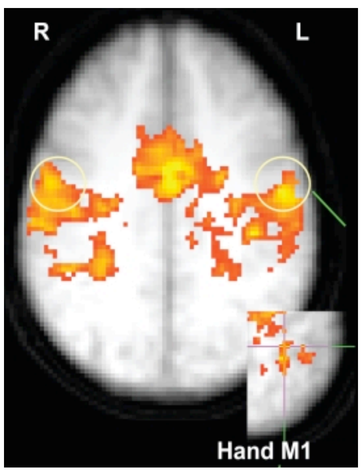
Dennis McLaughlin, Rafal Wojcik



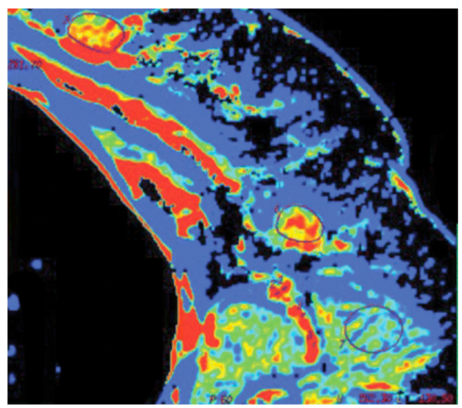
Hydrology
TRMM TMI/PR satellite rainfall



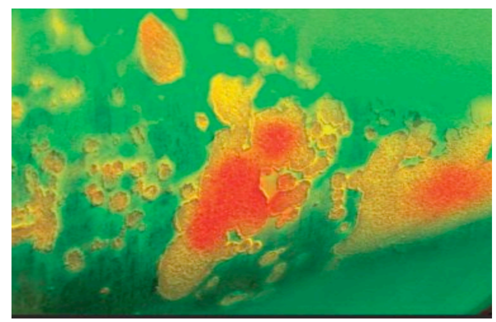
Geophysics – Seismic



Neuroscience -- MRI



Medicine -- CAT

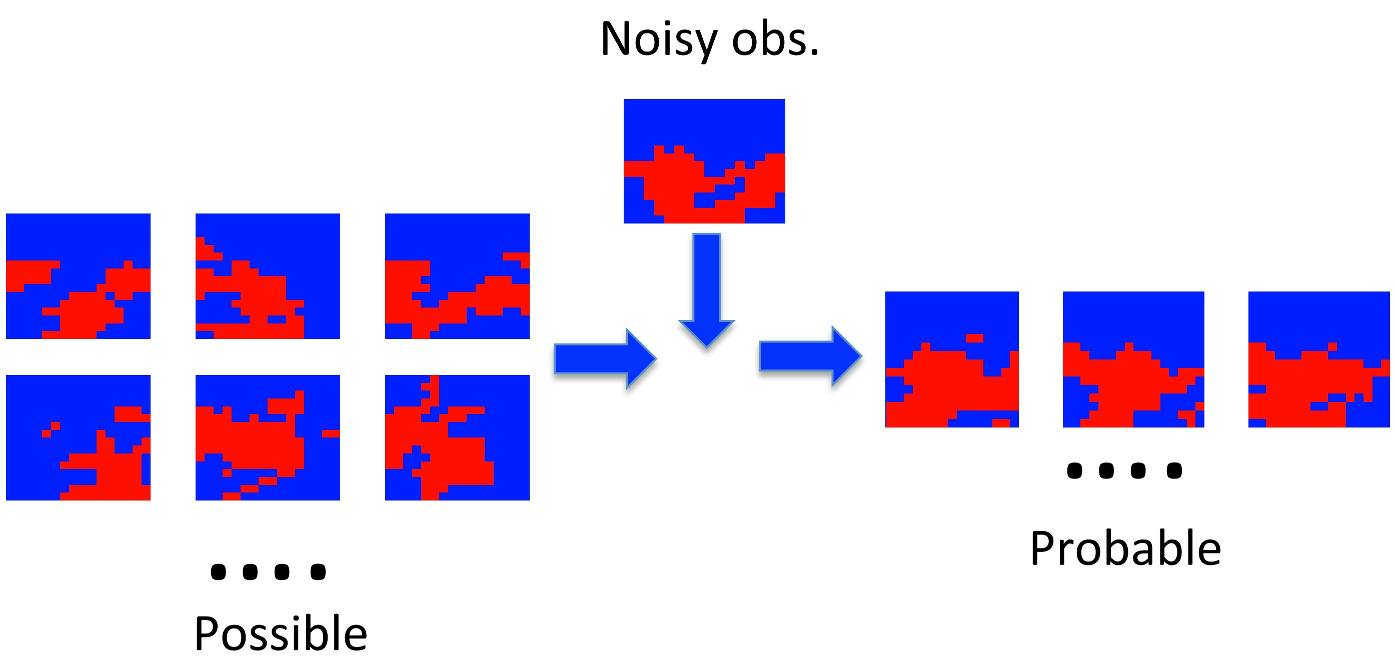


Material testing
Laser scanner

Problem Statement

Problem: Characterize complex geometric features that are difficult to observe.

Solution: Identify a set of **probable** features selected from a large ensemble of **possible** candidates :



Use the set of probable images for **screening** purposes, **risk analysis**

Bayesian Approach -- Importance Sampling

Bayesian estimation theory provides a general framework for solving the feature characterization problem:

- **Possible features: Prior probability** conveys distinctive structure and uncertainty
- **Likelihood function:** Conveys effect of measurement errors
- **Probable features: I**dentified by **posterior probability** derived from prior and likelihood

$$P_{\vec{X}|\vec{Z}}(\vec{X} \mid \vec{Z}) = cP_{\vec{Z}|\vec{X}}(\vec{X} \mid \vec{Z})P_{\vec{X}}(\vec{X})$$

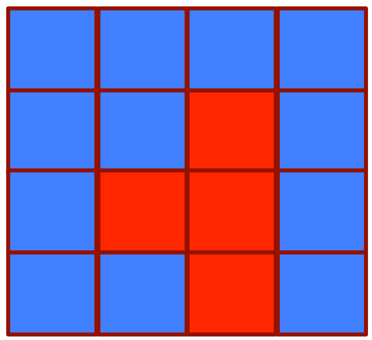
$\vec{X}, \vec{Z} =$ Actual and measured images

Non-parametric approach:

Use discrete prior and posterior probabilities described by finite **ensembles** of possible features.

Quantify Images for Discrete Computation

"Natural" description:
Vector of individual **pixel values** (e.g. 0 or 1)



- General
- Inefficient (high redundancy)
- High-dimensional

Alternative description:
Vector of a few distinctive **attributes** (which ones?)

Our approach:

- Start with **pixel-based** prior and measurement images/vectors
- Transform all images into **attribute** vectors
- Perform Bayesian computations in **attribute** space
- Assign updated probabilities to **pixel-based** prior images

Types of Measurements Required

Historical measurements (archived):

- **High-quality** measurements (ground truth)
  May be expensive or only available at limited times/locations.
  Use some of these to construct prior replicate **training image**
- **Low-quality** measurements
  More readily available but are indirect, noisy, coarser resolution, etc.

Current measurements (real time):

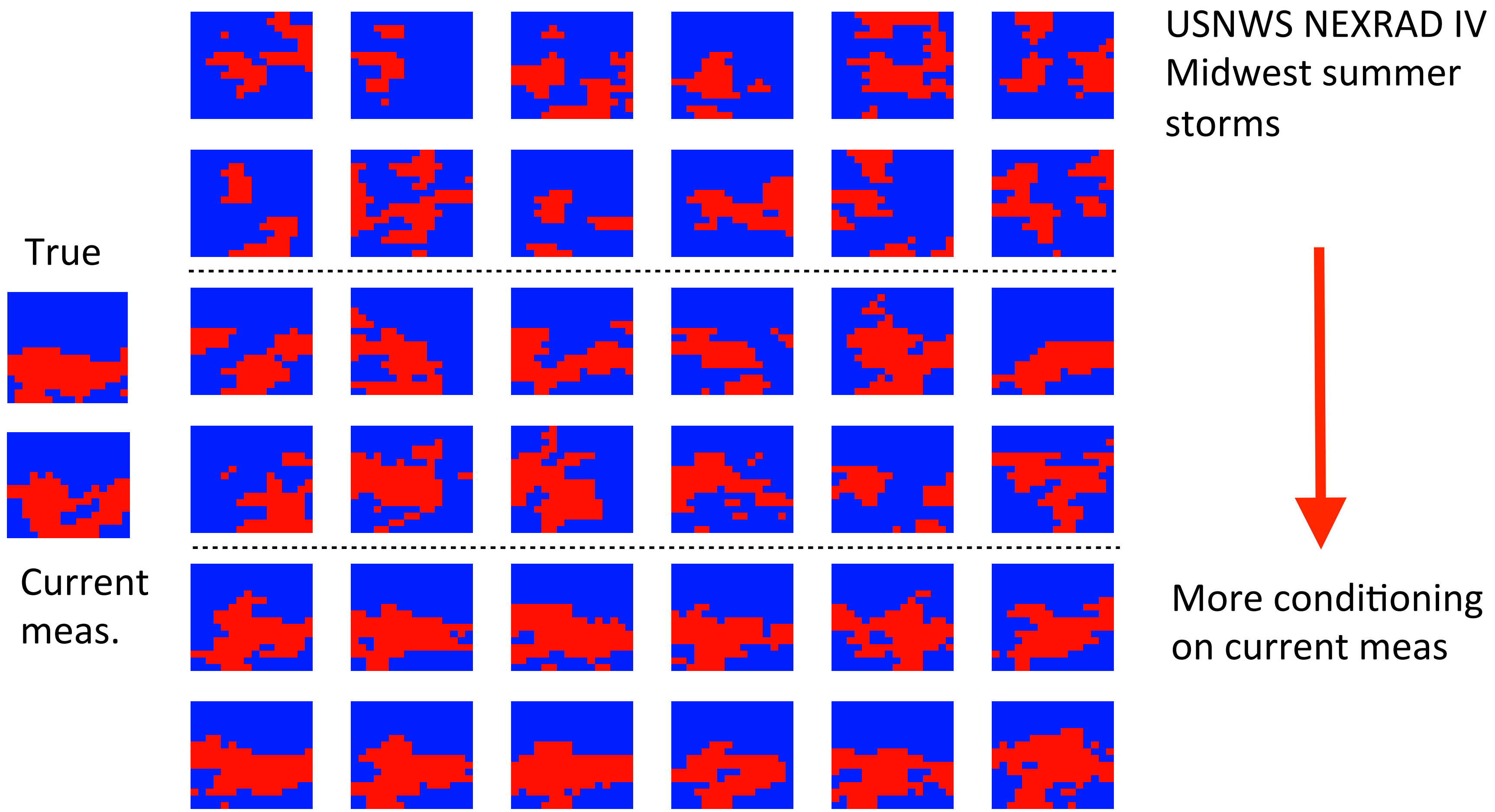
Observations of image of current interest, possibly with different instruments

Quality is similar to low quality measurements in historical archive

Errors in current measurements are revealed by comparisons between high and low-quality historical observations.

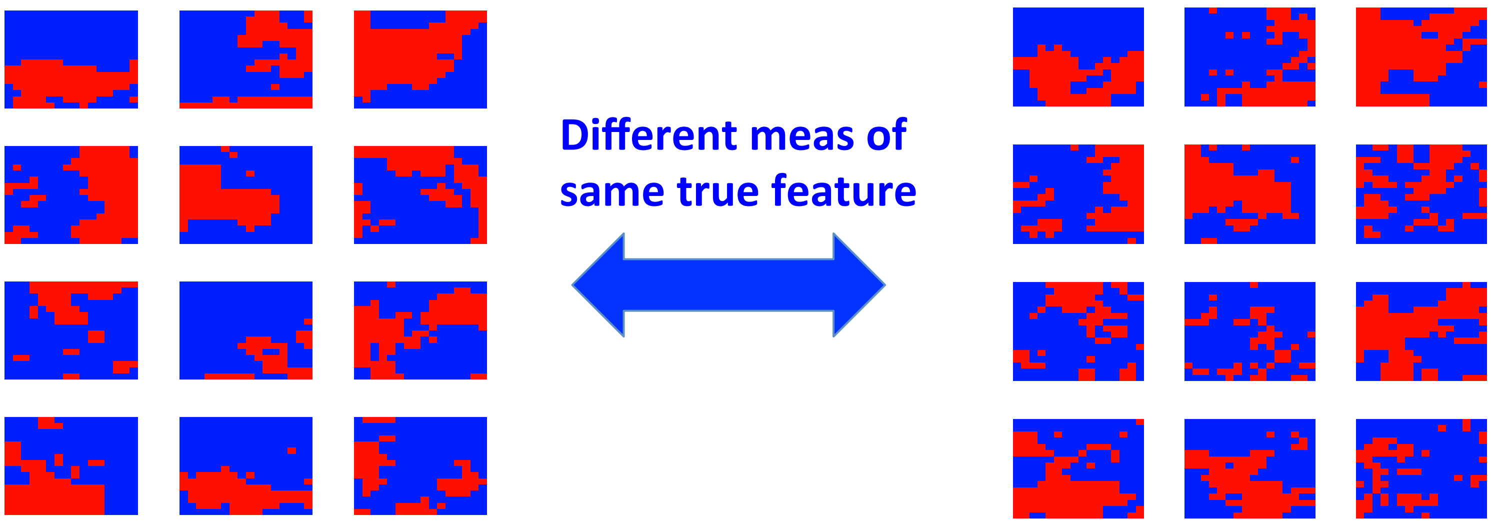
Generate Prior Ensemble of Images – Rainfall Example

Use **multi-point** methods to generate realistic prior replicates from **training image**



Compile Measurement Error Information for Likelihood

Errors in current meas. are **feature-dependent**:
Identify error properties by comparing high & low quality historical meas:



High-quality historic meas
Treat as "ground truth"
NEXRAD IV weather radar

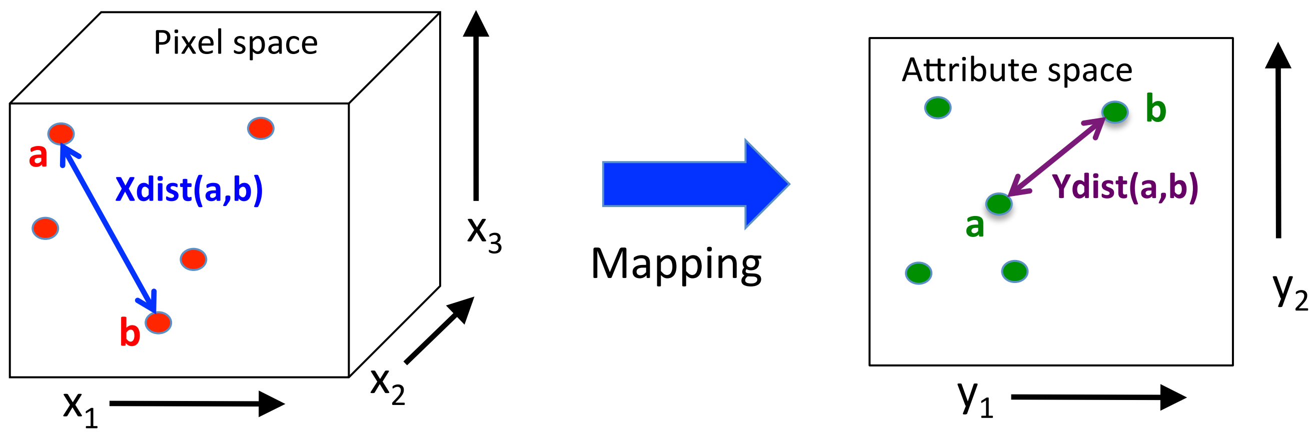
Low-quality historic meas
Similar to current meas
Surrogates for NOAA AMSU

Differences depend on true feature, do not have simple stationary structure

Data-driven Method for Defining Attributes

- Compute "distances" between all **pairs** of points in pixel space.
- Randomly distribute points in **attribute space** (one for each image)
- Compute "distances" between all **pairs** of points in attribute space.
- Iteratively **adjust** attribute point locations so pair distances in attribute space **match** pair distances in pixel space (as much as possible)

5 Images plotted as points in each space



Multi-dimensional scaling: Defines **mapping** from pixel to attribute values

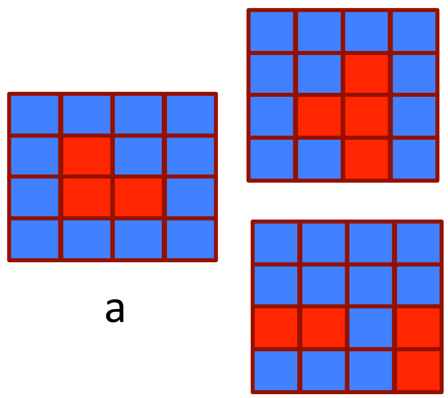
Select Image Similarity Measures – "Distances"

How do we measure "distance" between 2 images ?

Pixel space: Use "Simple Matching Coefficient" (SMC):

$$Xdist_{ij} = \left[\frac{n_{01} + n_{10}}{n_{01} + n_{10} + n_{11} + n_{00}}\right]_{ij}$$ Range: 0 - 1

b is "closer" to a



a

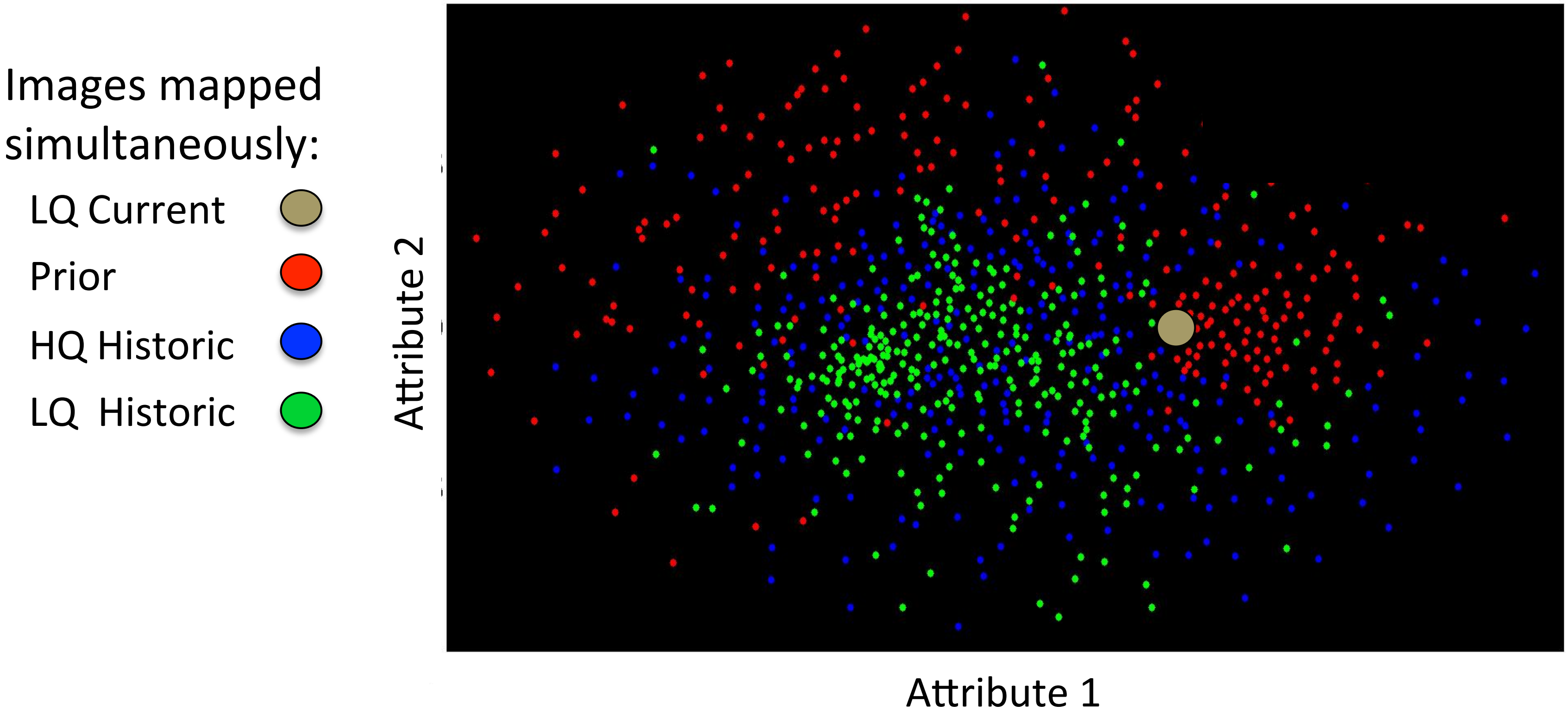
b $$d_{ab} = \left[\frac{2+1}{2+1+2+11}\right]_{ab} = 0.1875$$

c $$d_{ac} = \left[\frac{3+2}{3+2+1+10}\right]_{ab} = 0.3125$$

Attribute space: *YDist* is Euclidean distance between attribute vectors

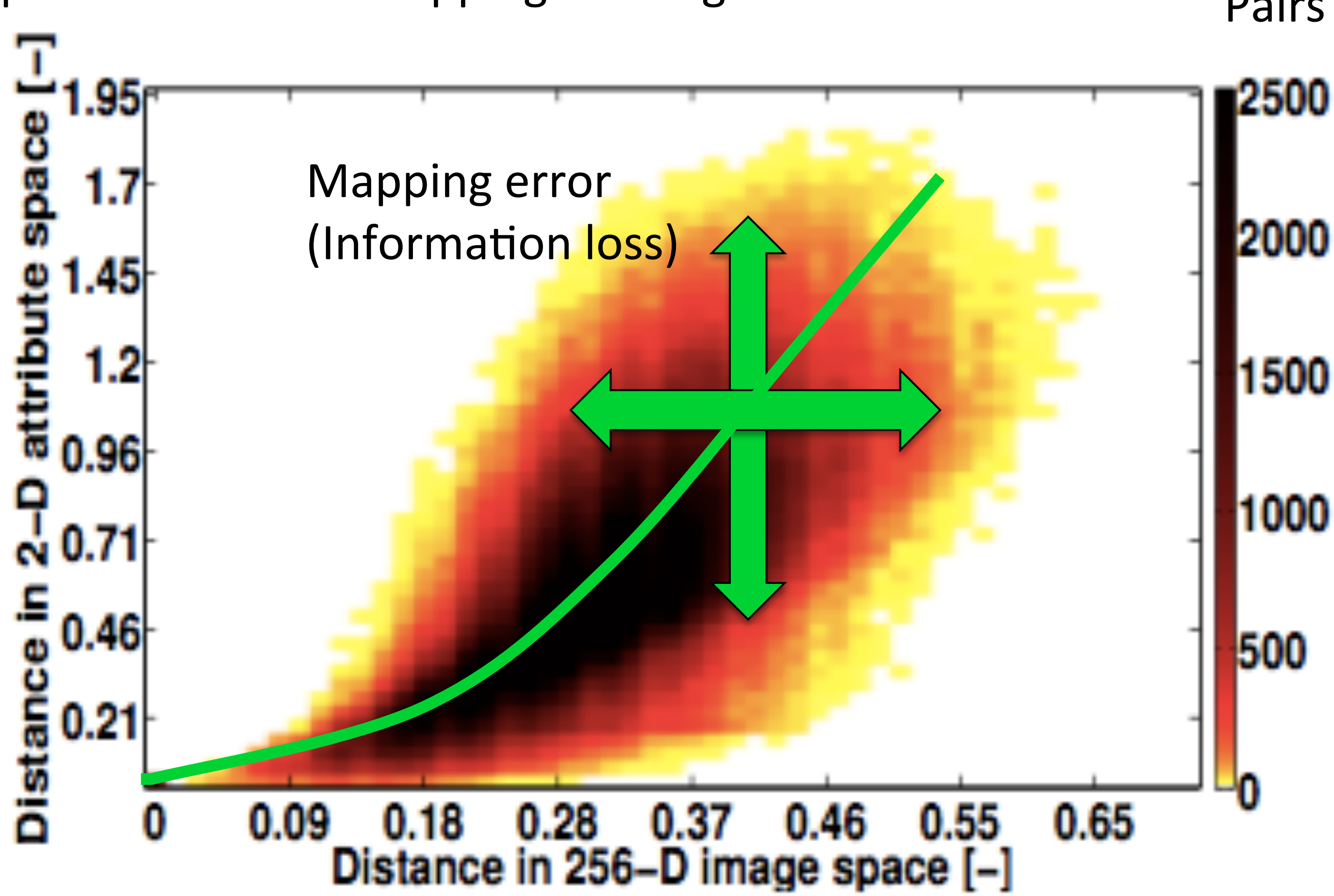
Map All Images to Attribute Space

Use distance-preserving (MDS) technique to map 16 X 16 (256) pixel images to 2 attributes



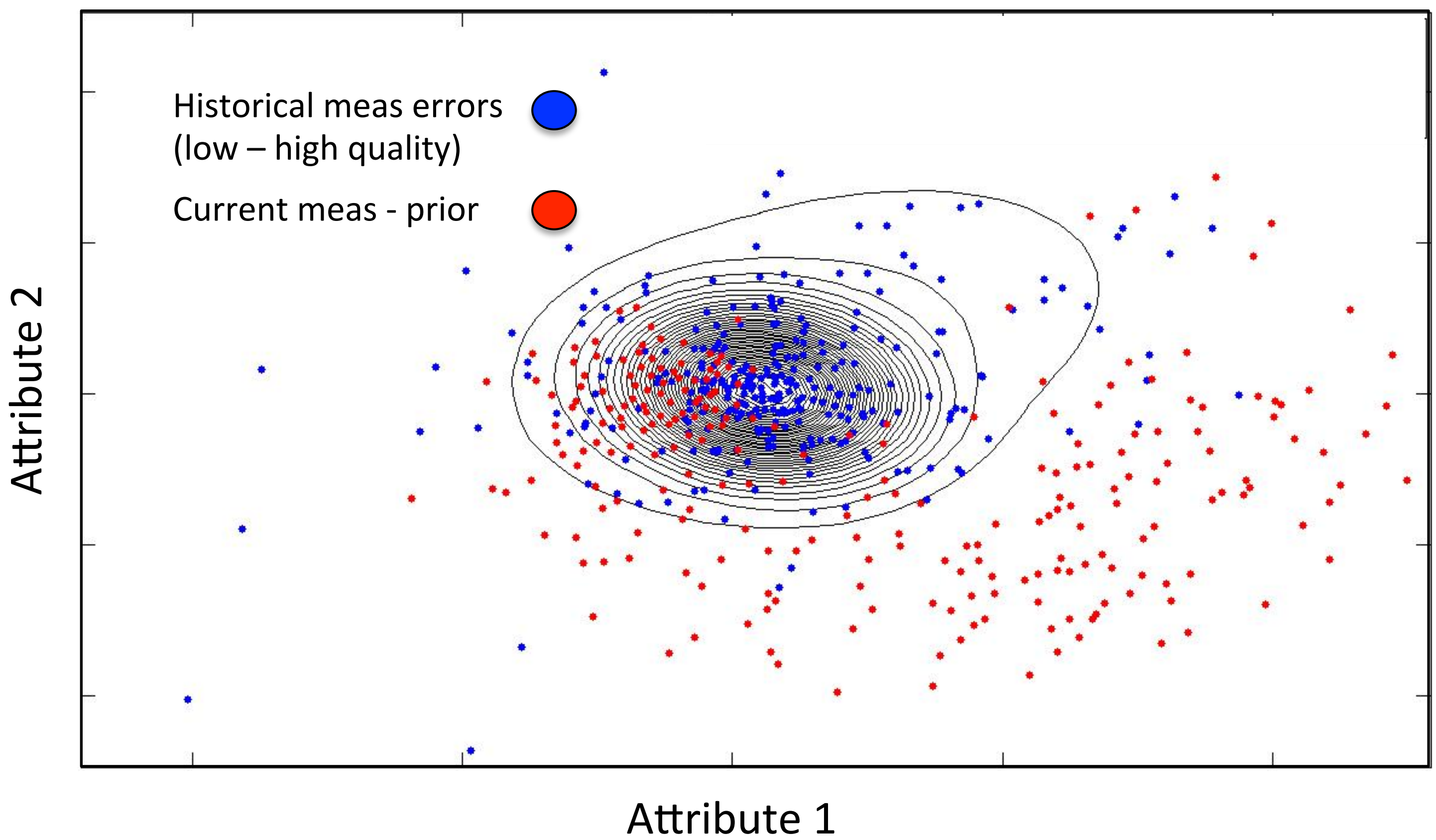
Quality of Mapping from Pixels to Attributes

An ideal pixel to attribute mapping would give a narrow one-to-one curve:



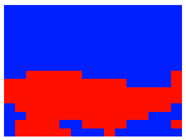
Construct Likelihood Function in Attribute Space

Assume errors in meas. attributes are additive.
Fit a continuous bivariate distribution (grey scale
Evaluate likelihood at current meas –each prior replicate



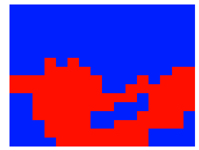
Conditioning Results: 1 Current Measurement

500 diverse images in prior ensemble ….

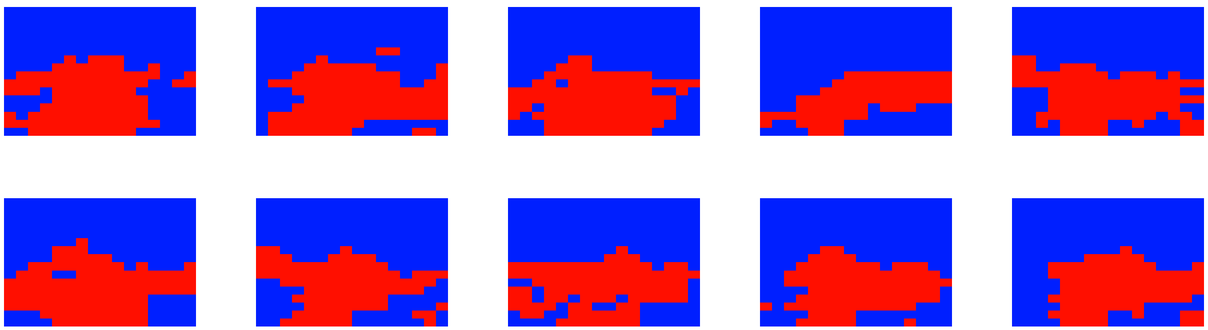


True

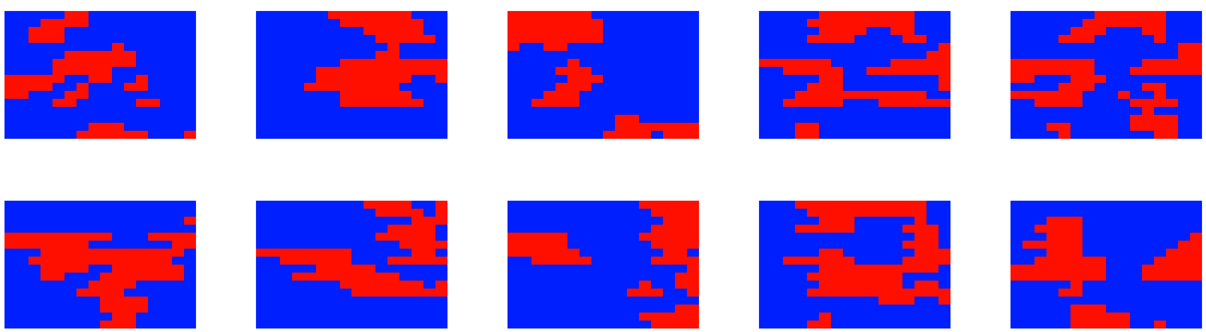
Current meas



10 most probable prior images

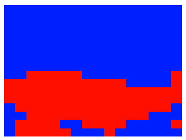


10 least probable prior images



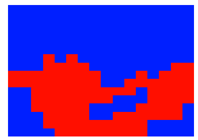
Conditioning Results: 2 Current Measurements (Data Fusion)

500 diverse images in prior ensemble ....

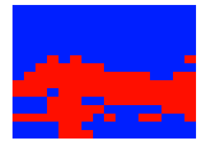


True

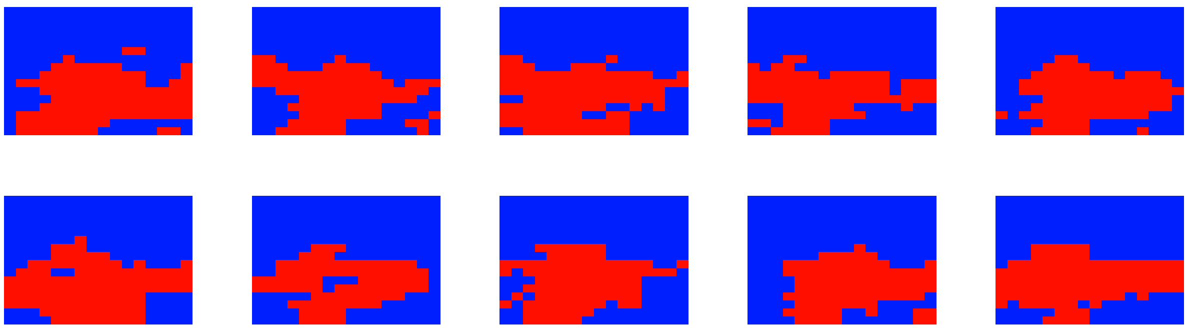
Current Meas 1



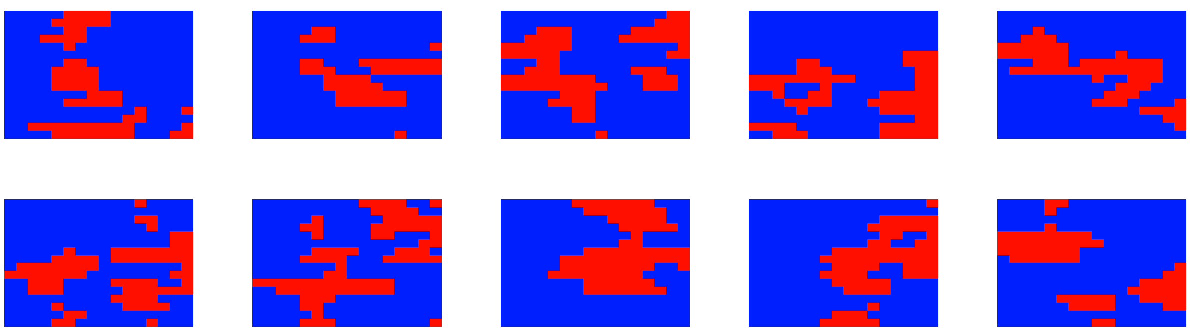
Current Meas 2



10 most probable prior images



10 least probable prior images



Some Issues for Further Consideration

- How many attributes?
  Tradeoffs between number of attributes vs. ensemble size vs. computational effort ?
- Go beyond binary images – include intensity variations within features. Consider implications for measurement error characterization.
- Try other distance metrics for both pixel and attribute spaces
- Improve replicate generation – account for dynamics & temporal correlation (i.e. prior comes from a forecast)
- Test measurement error analysis and likelihood generation with real low quality historical meas.
- Measurement error additivity in attribute space?
- Test approach with Markov Chain Monte Carlo (MCMC) version of Bayesian conditioning

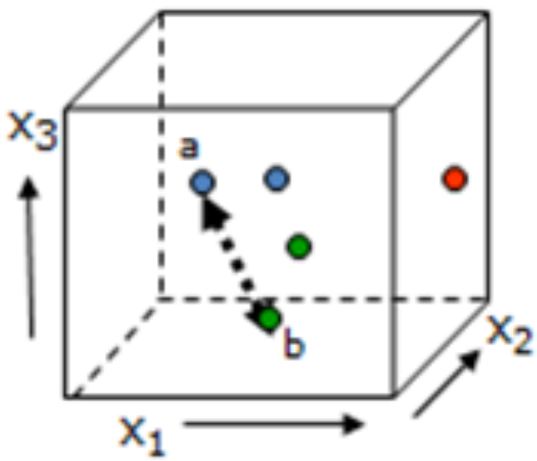
Feature Characterization

Common aspects over many applications:

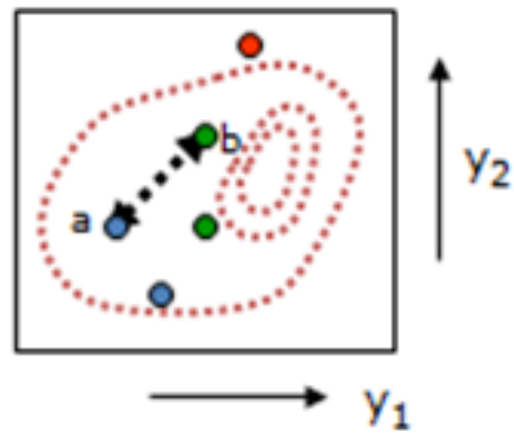
- Features are **heterogeneous, localized, often disconnected**, with complex but **distinctive structure**.
- Features are **uncertain**
- **Higher-quality** (ground truth) observations are only available at **limited times/locations** (or may be more expensive)
- **Lower-quality** measurements are more readily available but are **indirect, noisy, coarser resolution**, etc.
- **Measurement errors** are revealed by comparisons between high and low-quality observations, when and where both are available.

Visual Representation of Alternative Image Spaces

4 images in pixel space -
3 pixel values (x_1 , x_2 , x_3)



Same 4 images in attribute space (y_1, y_2)



In both spaces, define similarity (“distance”) between any 2 images (e.g. a and b)

Efficient Bayesian Characterization

A hybrid approach :

In the high-dimensional **pixel-space**:

- Generate an ensemble of equally likely **prior images**
- Compile sets of high and low-quality **historical** measurements that characterize measurement error.
- Obtain a low-quality **current measurement** of the image of interest

Associate each prior replicate or measurement in the pixel space with a corresponding point in the attribute space.

In the low-dimensional **attribute space**:

- Construct a **likelihood function** from the historical measurement attributes
- Evaluate the likelihood of each prior replicate, given the current measurement (using attribute values)
- Use Bayes Theorem to assign updated **posterior probabilities** to all the prior replicates

The updated probabilities identify the “most likely” prior replicates in either space.

Generate Realistic Prior Replicates

Prior replicates should be:

- **Realistic** (i.e. they should "look like" true features, as revealed by scattered high-quality measurements)
- **Localized** in space (non-stationary).
- Sufficiently variable to adequately reflect **uncertainty** about image of interest.

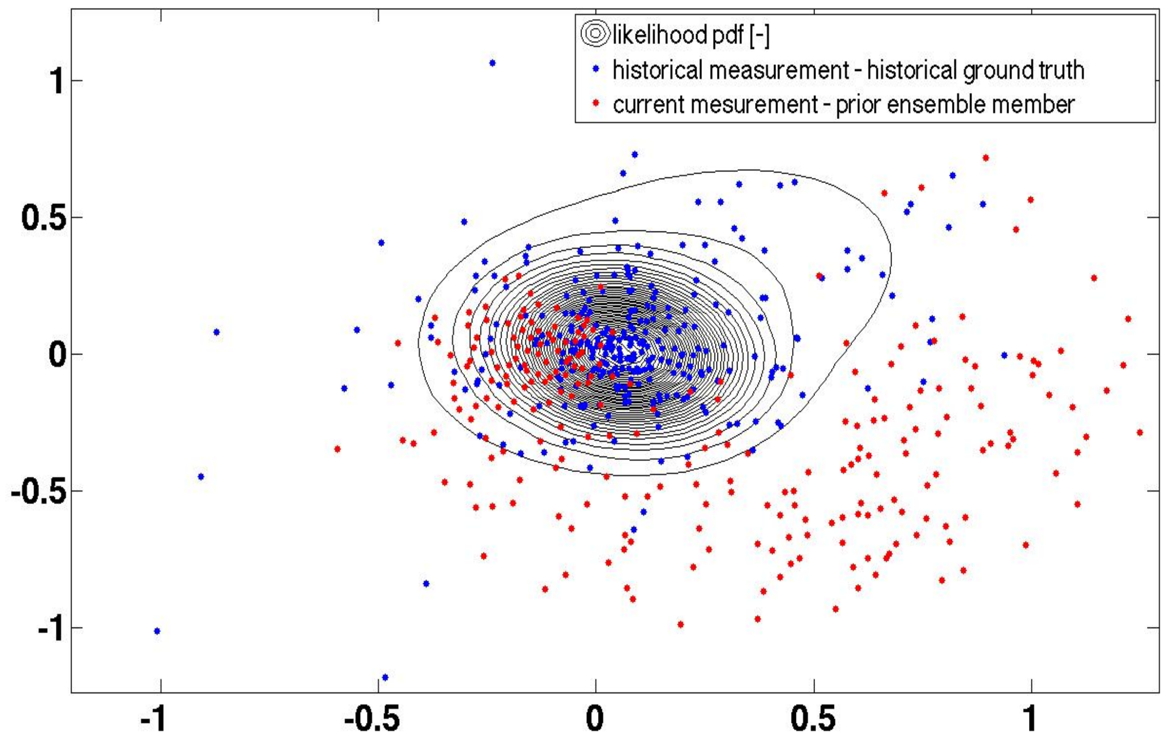
Generating prior replicates:

- Derive **stationary unconditional** replicates from training image(s) obtained from high quality historical observations (use multi-point geostatistics).
- Condition these replicates on scattered low quality current measurement values – the number of conditioning values should be varied.

This yields a set of **nonstationary conditional** replicates.

- Construct the **prior ensemble** from the conditional replicates

Likelihood Function – Conditioning



Likelihood values are used to determine importance sampling weights.
Most likely samples are those closest to the likelihood function peak.