Sunspots, Seismicity and Statistics: Recognizing Hidden Patterns in Science Data

Michael Turmon
Data Understanding Group
Explorations Systems Autonomy
(Section 367)

24 September 2002

I hope you come away with either

- An appreciation for the scope and capabilities of machine learning models and algorithms
- A basic understanding of why the methods advocated here work

`turmon@aig.jpl.nasa.gov`
`http://www-aig.jpl.nasa.gov/home/turmon/`

Themes of This Talk

- **Learn within uncertain environments**

  Deal with uncertainty using probability

  Model dependence of the things we care about on
  the things we can measure

  Gererally emphasize *modeling*...
  ...sometimes, we can and do avoid models

- **Today's plan: Visit families of useful models**

  More complex linkages $\Longrightarrow$ more complex models
  Unlinked item-by-item decisions: baseline case
  One-dimensional linkage: state-based models
  Two dimensions and up: Various random fields
  Ad hoc linkages: Stochastic grammars

Ends Justify the Means

- **Algorithms**

  Learn explanations by maximizing posterior probability.

  Fit models to data using general iterative methods.

  Common toolbox:
  ...optimization, sampling, simulated annealing
  ...applied math, statistical physics, signal processing

- **Applications**

  Solar object identification, geophysical time series, volcano classification, cyclone classification

  Where does your data fit in?

Machine Learning in Science

- **Automation**

  Cope with growing data volume

  Generate results faster

  Data center operations often underfunded

- **Repeatability**

  Well-defined algorithm produces results

  Uniformity among distributed investigators

  Crucial for charged subjects like climate change

- **Consensus**

  Ubiquitous algorithms factor out squabbles

  Develop cross-domain solutions

  Exchange models and algorithms as well as data

Machine Learning in Science

- **Falsifiability**

  Quantitative models make checkable assertions

  Popper: Falsifiability characterizes science

- **Quality**
  - Gauss found Ceres' orbit by least squares, 1801
  - Earth's size and oblateness: Laplace, Legendre
  - Kalman filters guided Apollo (onboard!)
  - Viterbi decoder increased Galileo bandwidth

  Optimal inference gives better results

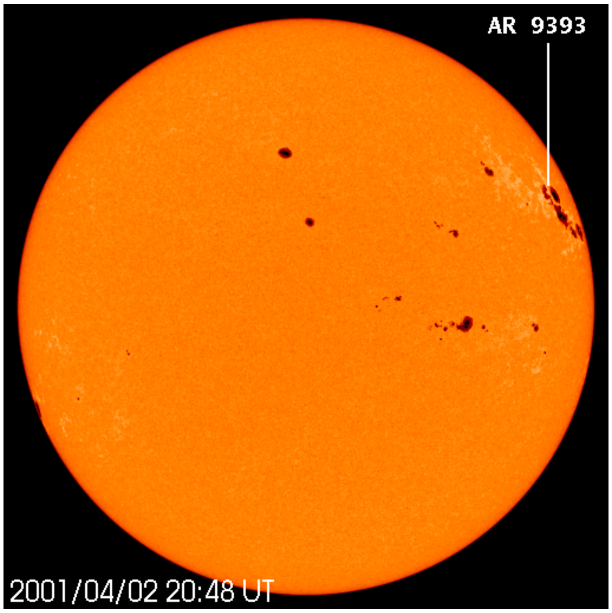
- **Comprehensiveness**

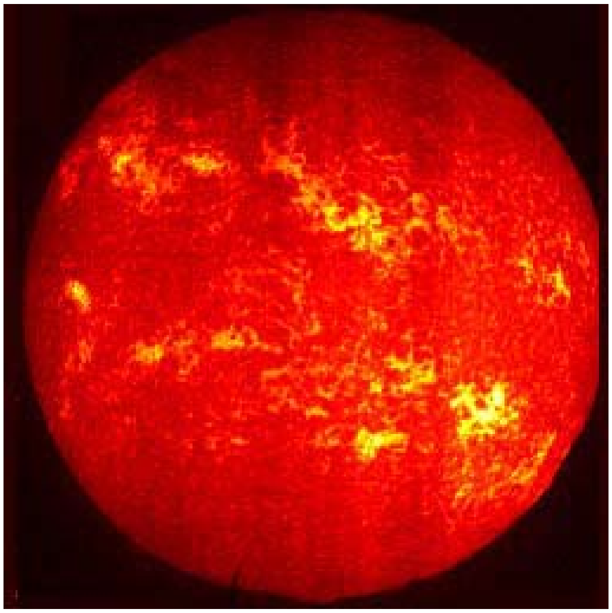
  Data fusion

  Integrate more data into an interpretation

  Achieve total spatial/temporal coverage

Pixel Classification: Applications





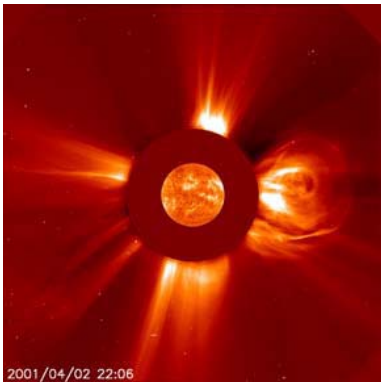
- **Solar Physics**

Reliably identify solar magnetic structures
  photosphere: sunspots, faculae
  chromosphere: plage

Irradiance changes: weather, climate

Tracers for flow measurement

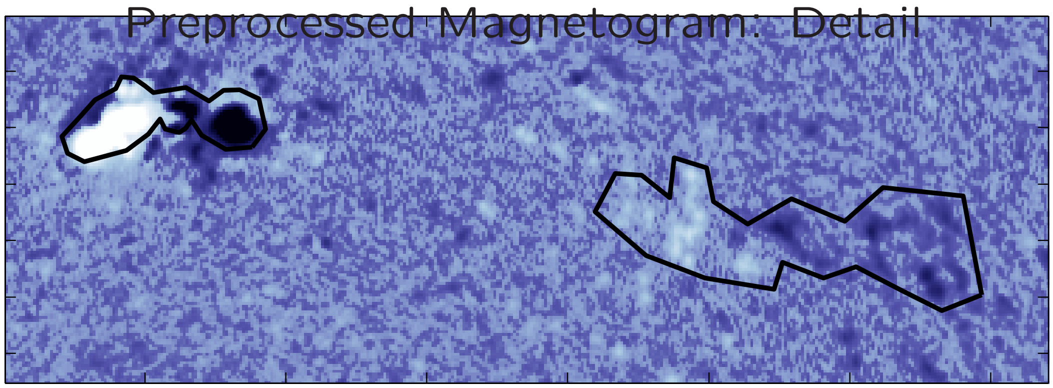
Space weather: δ-spots cause flares

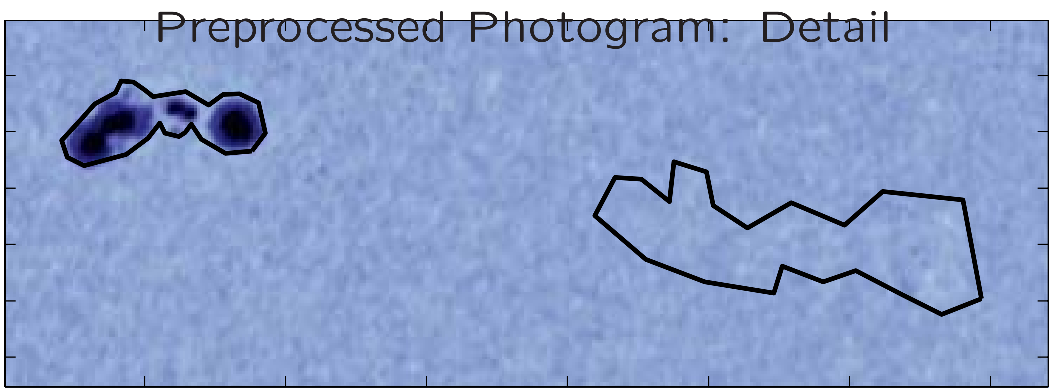


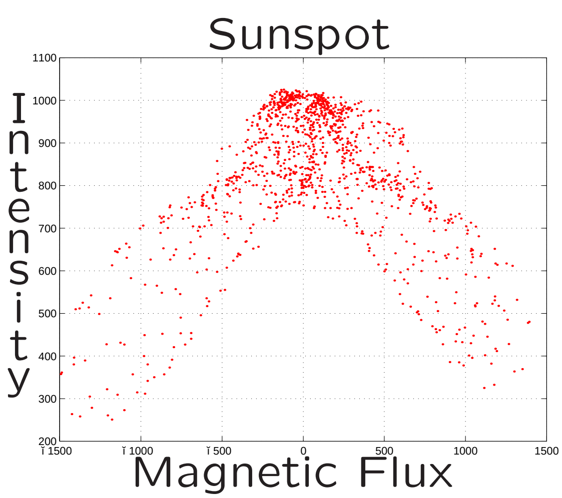
Solar Data

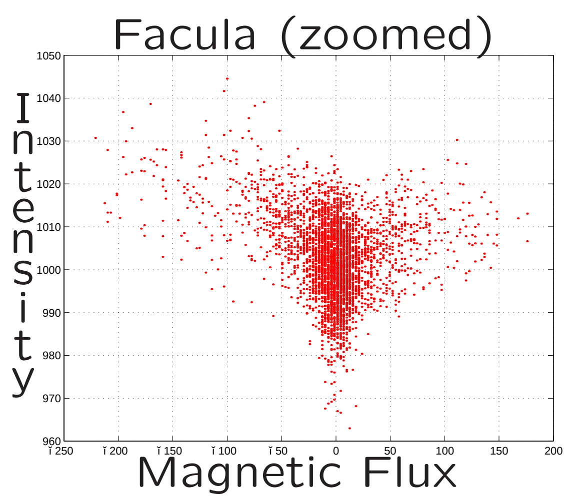
Many observatories, many images.

Below: SoHO/MDI, 1997 Sept. 7 at 17:58 UTC







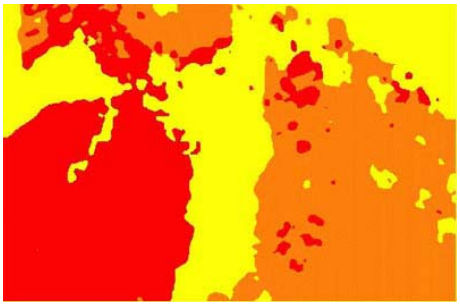


Pixel Classification: Applications

- **Mars Geology**

  Soils: dust, sand, pebbles

  Rocks: sedimentary/igneous, weathering



Pixel Classification: Applications

- **Earth Remote Sensing**

  Cloud/ice; ocean/ice boundaries

  Land usage in multispectral imagery

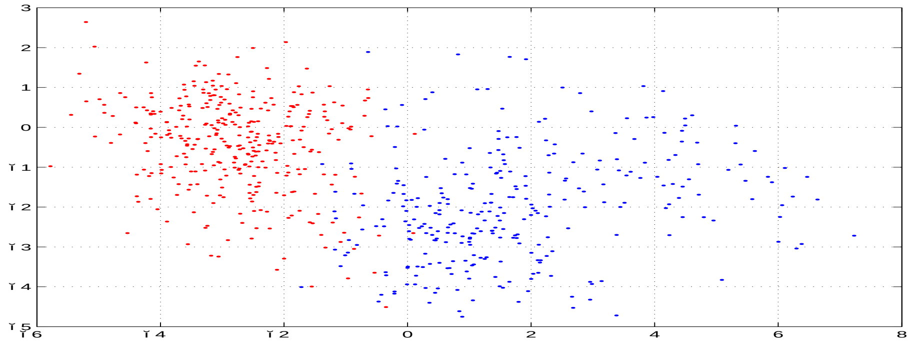
- **Review two applications**
  - Solar: Unsupervised clustering
  - Volcano: Supervised classification

Feature Vector Outlook

Observe vector data $x_1, \ldots, x_N$, $x_i \in R^d$
Goal: tag each x_i with a label $y_i \in \{1, \ldots, K\}$

- **Learning and Clustering**

Supervised: Use *training data* (x, y) supplied by an oracle (e.g., expert)



Unsupervised: *cluster* nearby points: uncover latent structure

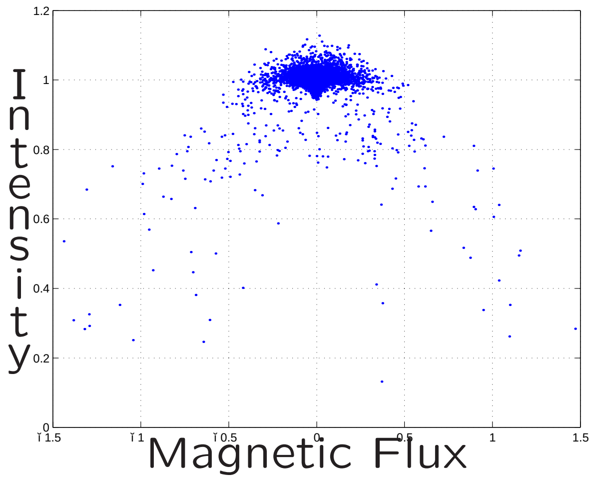
- **Ground truth**

Hard to find trustworthy experts with time on their hands

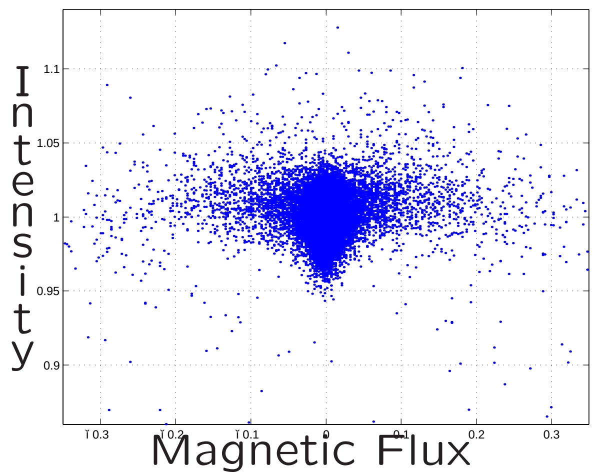
Learning Classes of Pixel Data

Cloud of points $x \in R^d$ (d moderate: 2–50)
E.g., pooled data from SoHO/MDI:

Feature Vectors



Feature Vectors (zoomed)



- **Learning classes**
  Partition data into clusters
  Associate each vector x with a class $k \in \{1 \cdots K\}$

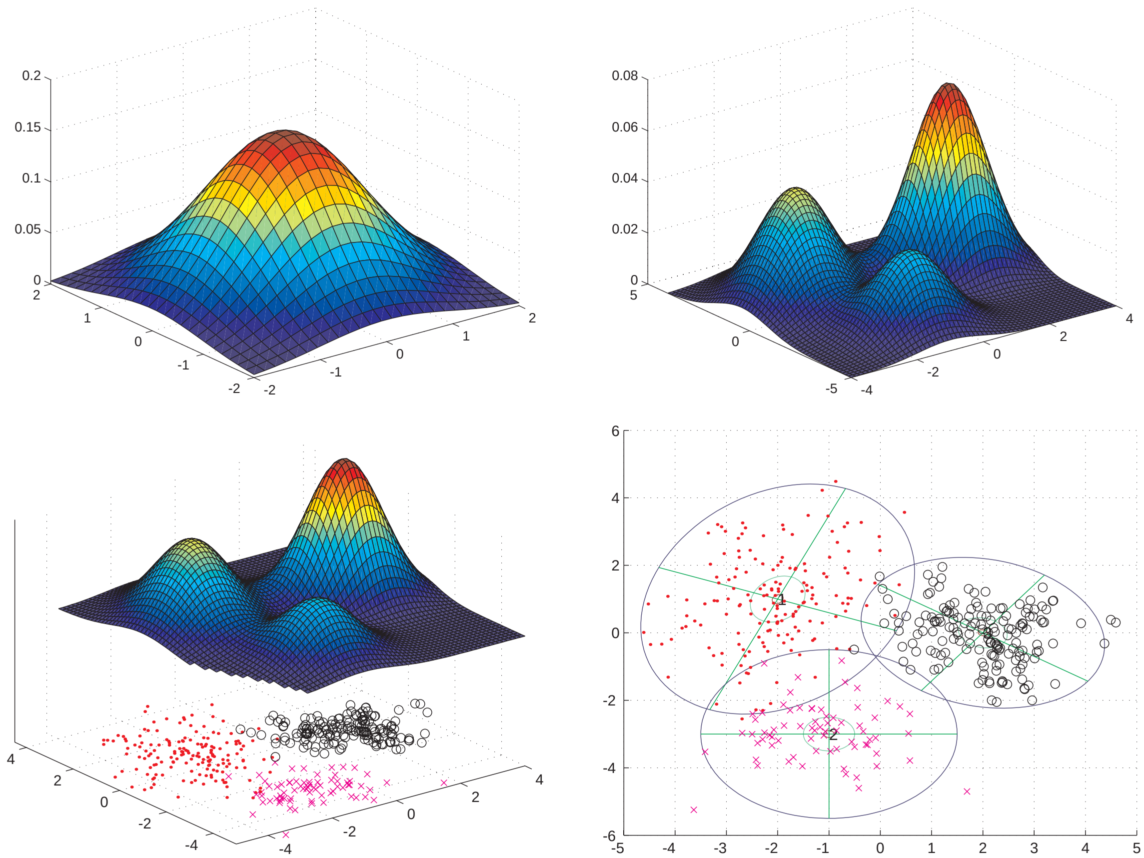
- **Non-pixel examples**
  Sky objects in survey database
  Rock composition and shape
  Spectral signatures

Clustering with Normal Mixtures

Gaussian bump = a cluster with a given shape

Gaussian mixture = weighted sum of K bumps

Cluster membership is the identity of the generating bump



- **Clustering** = find cluster centers, shapes, and weights to fit data

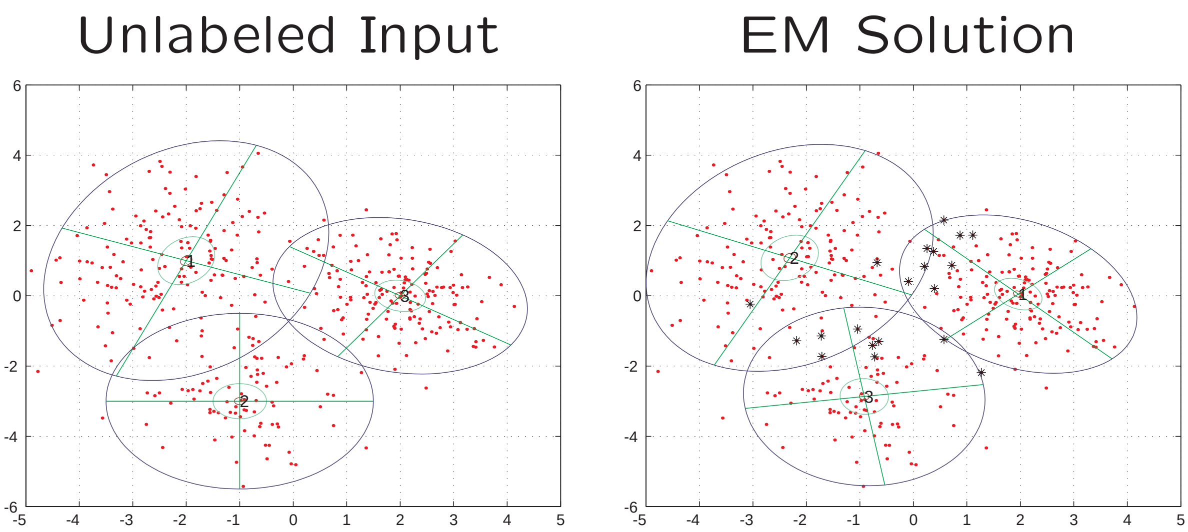
Clustering Algorithm

From data $X = [\vec{x}_1 \cdots \vec{x}_n]$, find a mixture $p(\vec{x}; \widehat{\theta})$

Find parameters by maximum-likelihood:

$$\widehat{\theta} = \arg\max_{\theta} \log P(X;\, \theta)$$

EM algorithm updates parameter estimates and cluster assignments until fixed point

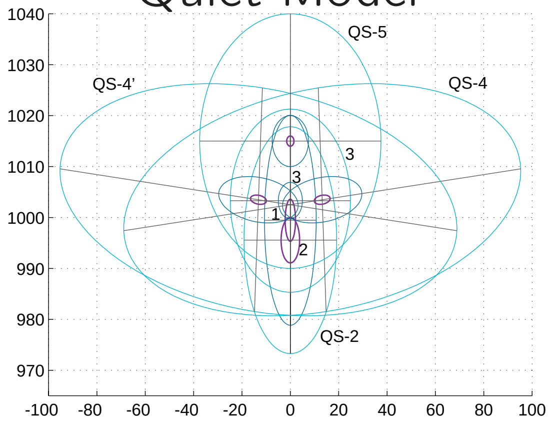


Unsupervised mode: The mixture partitions X into clusters on its own

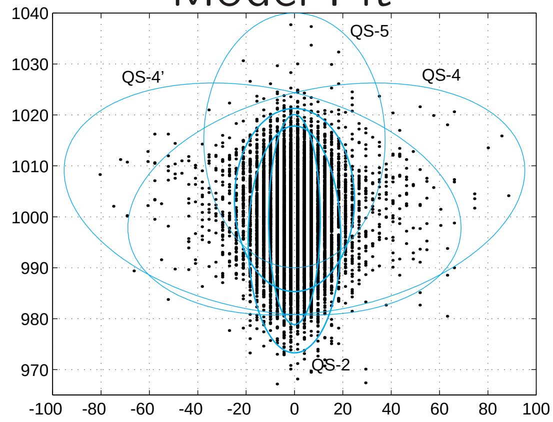
Partial information: Use partial cluster assignments of sample x's to guide solution

Learned Clusters: SoHO/MDI

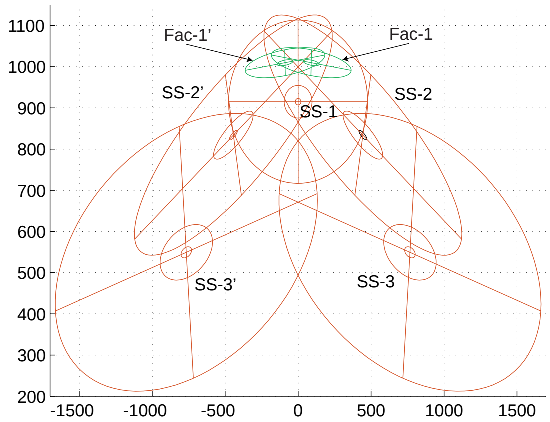
Quiet Model



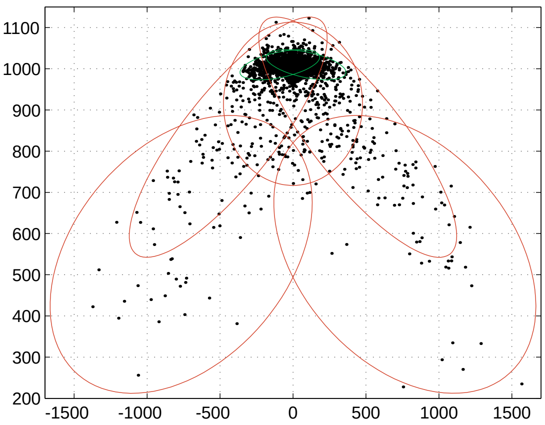
Model Fit



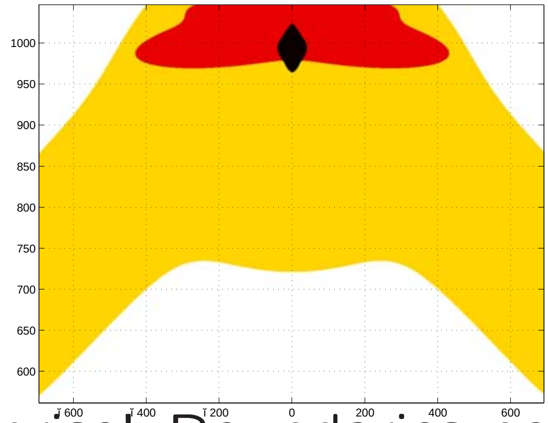
Active Model



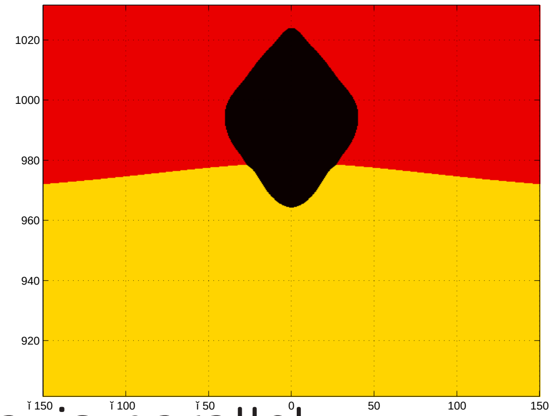
Model Fit



Most Likely Class



Class Map (detail)

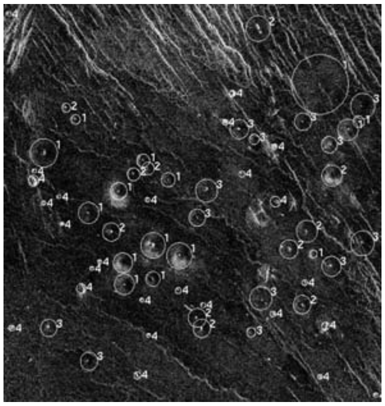


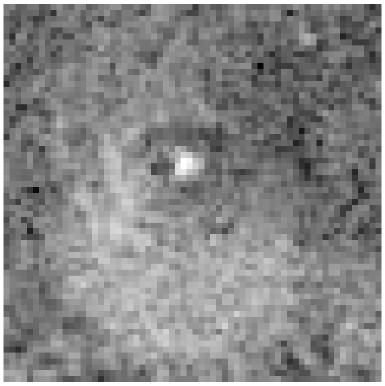
Surprise! Boundaries not axis-parallel

Object Detection

- **Known object**

  Volcanoes on Venus in Magellan SAR



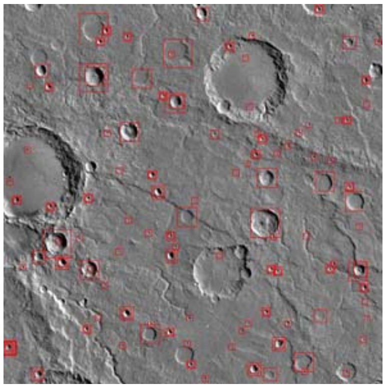


- **Known object family**

  E.g., craters

  — scale variation

  — overlap

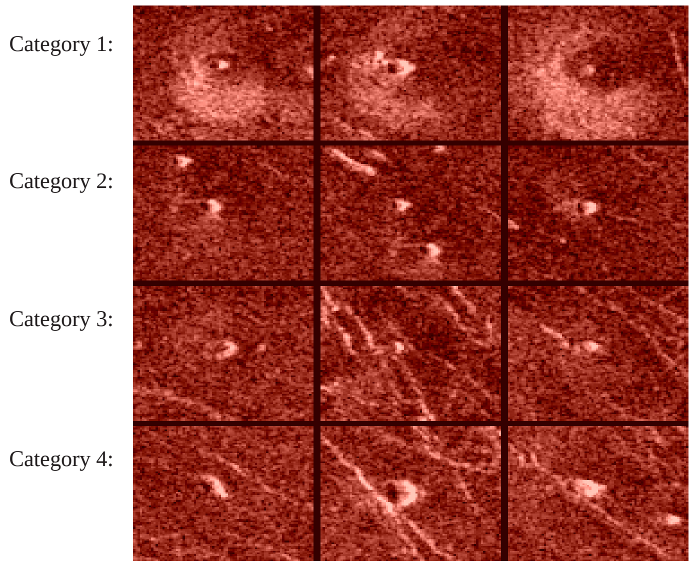


- **Unknown objects**

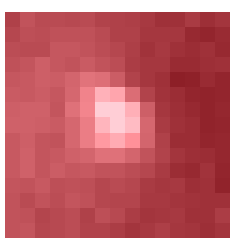
  Detect local variations in a background

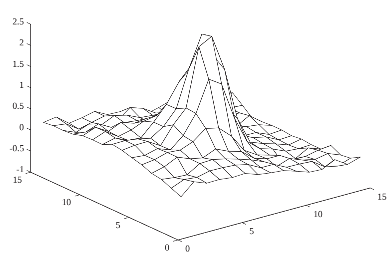
Scheme for Finding Objects

Due to M. Burl (formerly JPL) and collaborators
Train with scientist-supplied ‘chips’:



First: Focus of attention via “matched filter,” the average of all volcanoes:

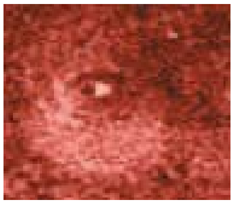


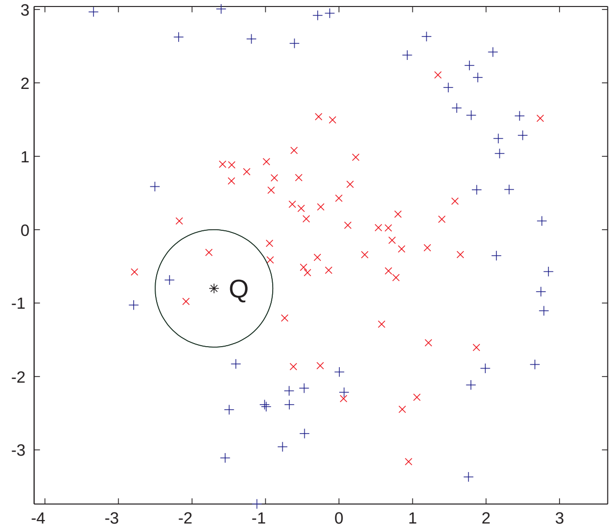


This filter sweeps whole image, identifying possible volcano sites which are extracted as square ‘chips’

Phase Two: Classify Candidate Chips

Label query chip (15^2 pixels) as V or NV

$Q =$ 



- ***K*-Nearest Neighbors**

Find closest K training chips to the query chip

Majority class (V/NV) among them wins

Neighbors via weighted Euclidean $(x-y)^T R(x-y)$ where R emphasizes pixels near chip center

Accuracy $\approx$ human experts in homogeneous data; degrades markedly in heterogeneous regions

K-NN alternatives: QDA, SVM, etc.

When Are Decisions Linked? (halfway point)

One of the most important problems in the philosophy of the natural sciences is...to make precise the premises which would make it possible to regard any given real events as independent. This question, however, is beyond the scope of this book.

– Kolmogorov, *Grundbegriffe*, 1933

- **Unlinked decisions** not always justified
  Pixels, or pixel-groups, in a bag
  Unsupervised: cluster pixels (solar feature vectors)
  Supervised: classify pixel-groups (volcano chips)

  Simplicity is powerful but limits applications

- **Science data** typically linked by time or space:

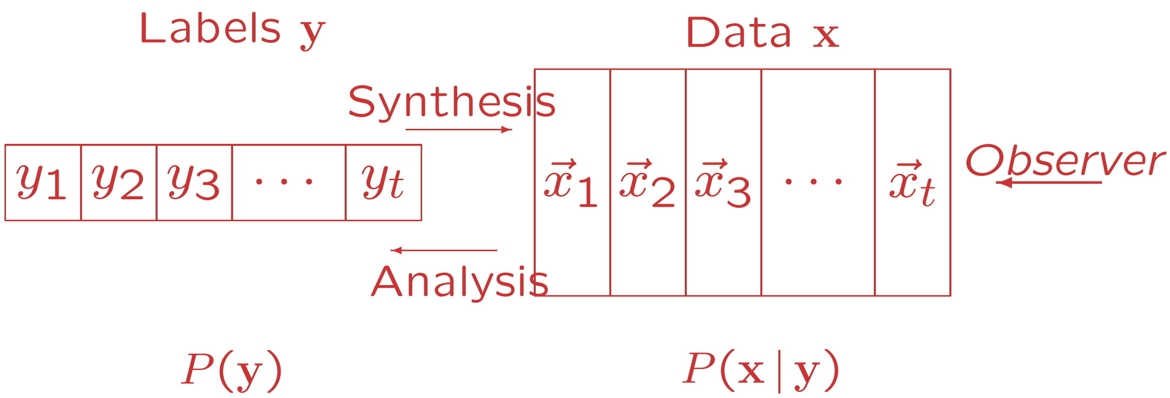
  Deduced labels are not interchangeagble; they have continuity or smoothness

  Use these links to our advantage to learn better

Probabilistic Sequence Models

At t, one of K physical processes y_t is dominant.

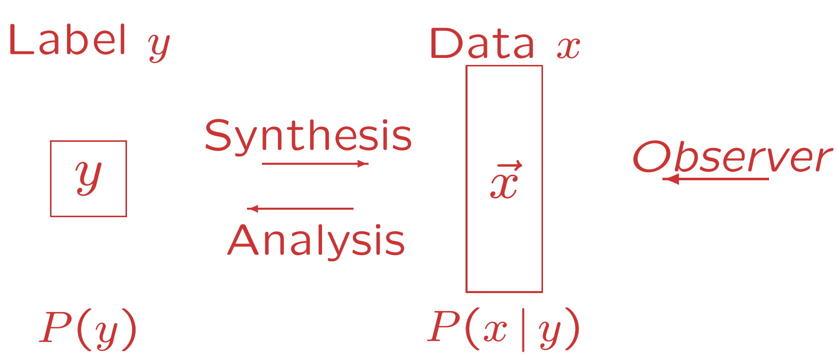
Observables $\vec{x}_t$ arise depending on this dominant process.



Generation of data adds uncertainty (noise) to the underlying dominant process.

- **Statistical model**: distributions $P(\mathbf{y})$ and $P(\mathbf{x} \mid \mathbf{y})$

Incidentally, unlinked decisions model is:



Linking the Labels in Time

Let y_{t+1} depend on y_t. Locality is key: next state depends only on current state.

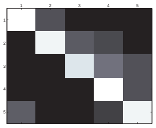
H.avivoca courtesy M. Pingleton, UIUC

- **Frog on lily pad**

  Destination depends only on current position

  Need $K \times K$ matrix Φ of *transition probabilities*

  $$\Phi_{k,l} = P(y_{t+1} = l \mid y_t = k)$$

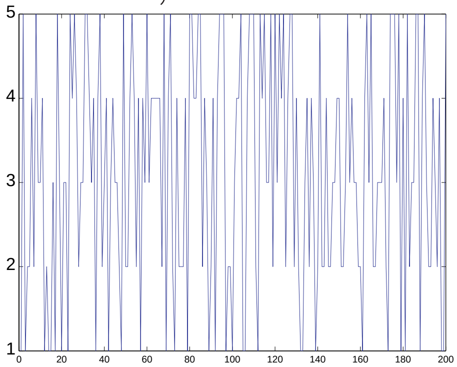
  

  — Expected staying time in a state

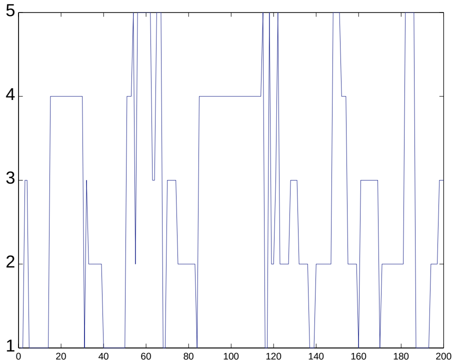
  — Which states are likely to follow a state

- **Example State Sequences**

Independent y_t
$(\Phi_{k,l} = 1/K)$



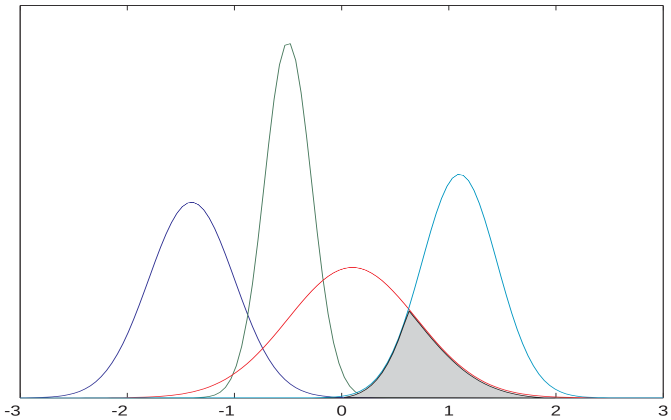
Dependent y_t



Observations Come from Labels

- **Generating Observables with $P(x \mid y)$**

  Need K distributions, one for each event class

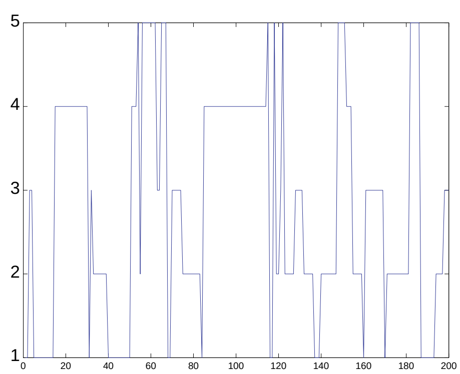


  Can learn these from data, or use scientist-labeled series and prior knowledge to constrain them

  Linked labels will inform about "gray areas" above

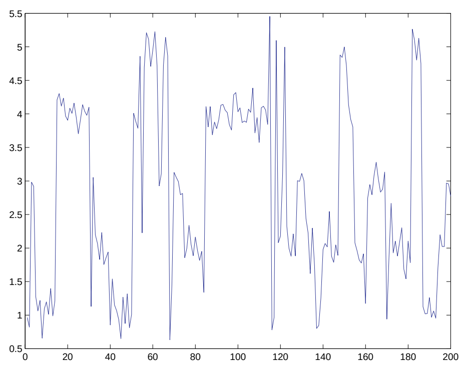
- **Example Output Sequence**

Dependent y_t



Observed Data x_t

$(\mu_k = k, \Sigma_k = 0.2^2)$



Learning Labels and Models

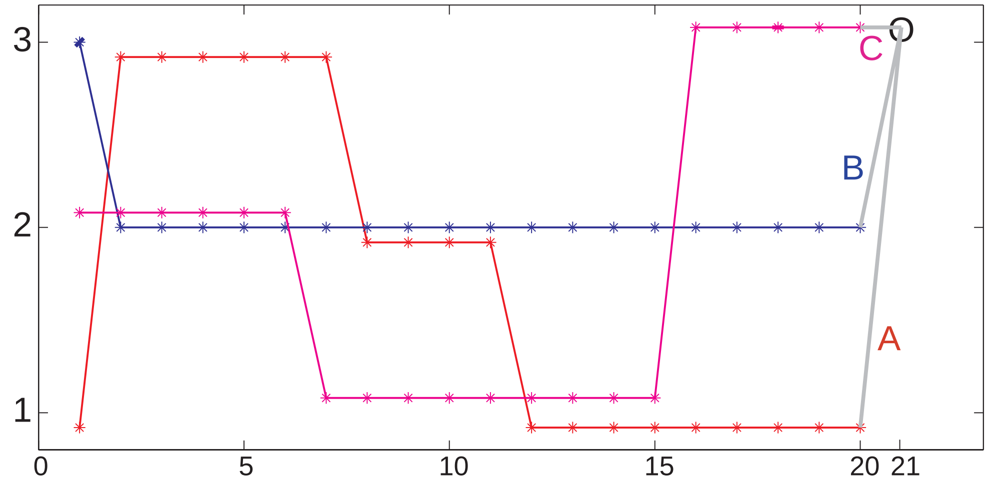
- **Learning Labels**

  Choose the most likely "interpretation"

  $$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}} P(\mathbf{y} \mid \mathbf{x})$$

  Price paid for linking the labels:
  there are now K^T interpretations to consider

  Viterbi algorithm:
  Recursively update the likeliest path to each state



- **Learning model parameters**

  Done using a variant of the iterative EM algorithm described for independent data

  Find likeliest $\mathbf{y}$, use it to estimate Φ, re-compute $\mathbf{y}$
  Iterate to convergence: maximizes likelihood

Classifications: Seismicity

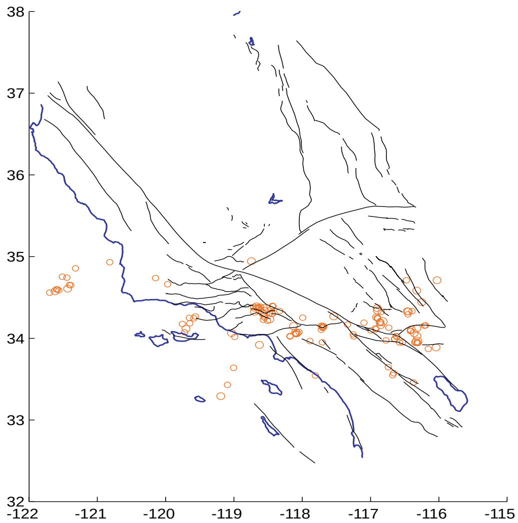
SCEC catalog, 1960–1999, $M > 4$

$K = 17$ labels. Inputs: position, M, time to next, time since prior. (R. Granat, A. Donnellan)

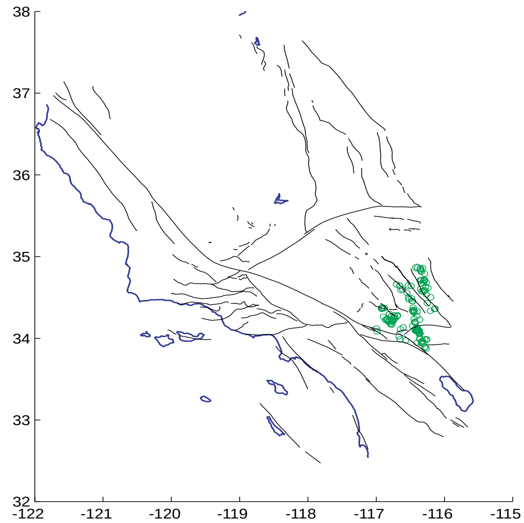
Blue lines: coastlines; black: major faults

Circles for earthquakes; circle size for magnitude.

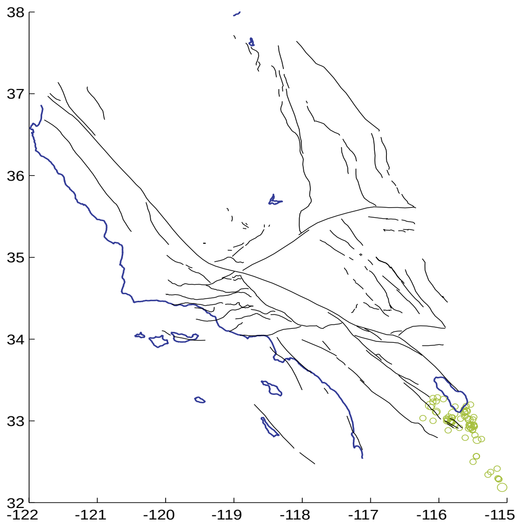
Transverse Range events



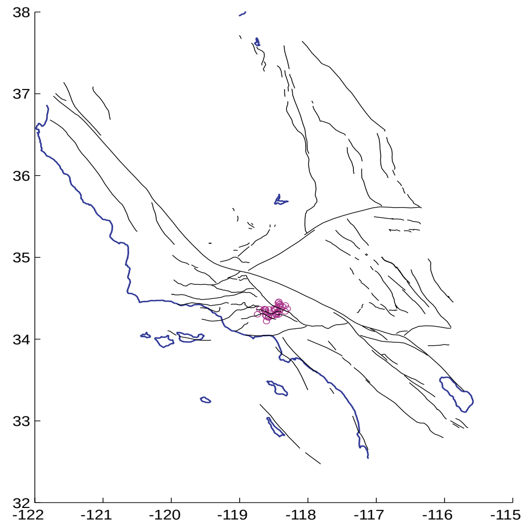
HMQ & Landers



Salton Sea swarm events



Northridge aftershocks



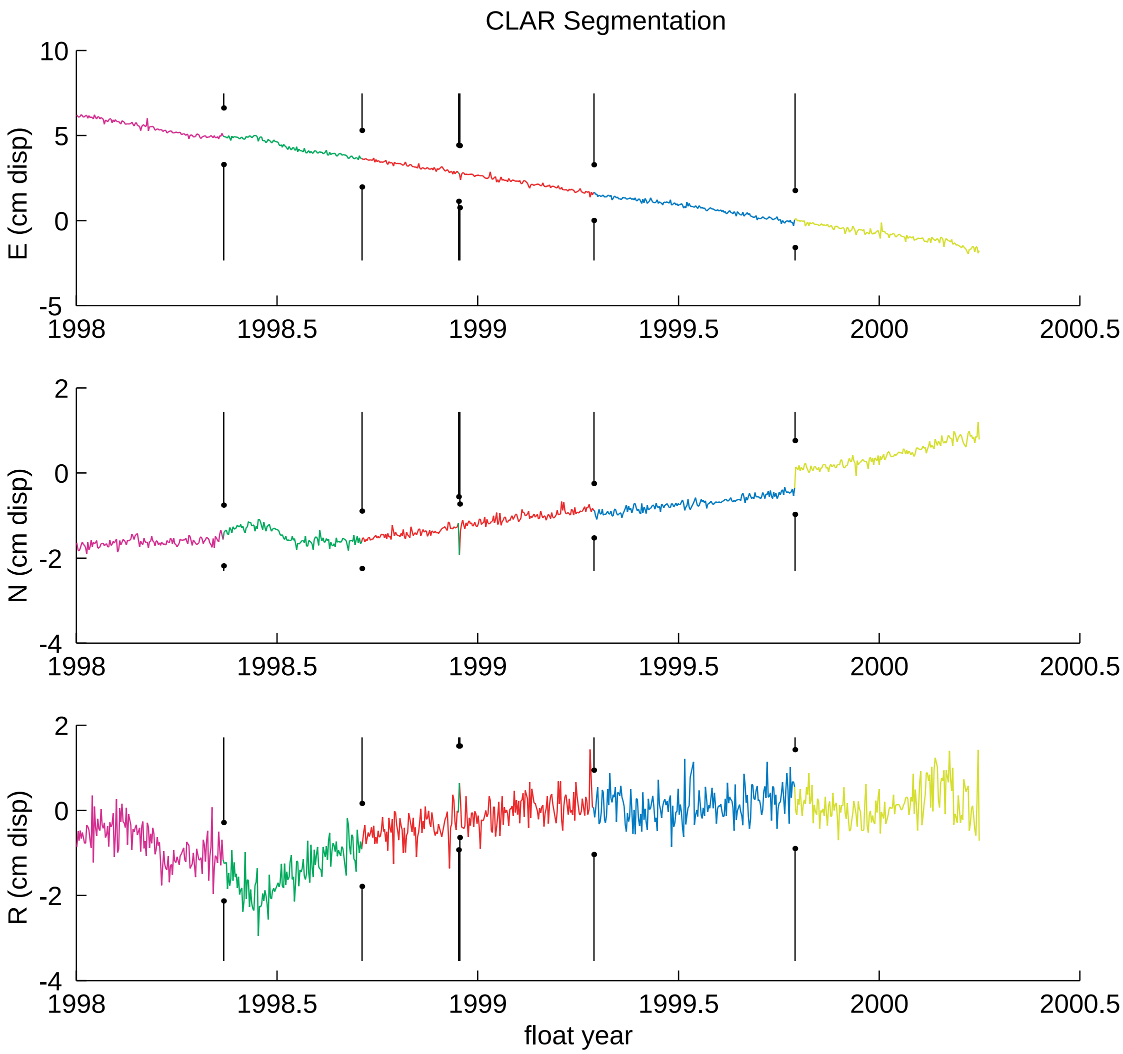
Classifications: Geodesy

SCIGN GPS signal, Claremont, CA: R. Granat

Inputs: Daily station displacement (mm accuracy)

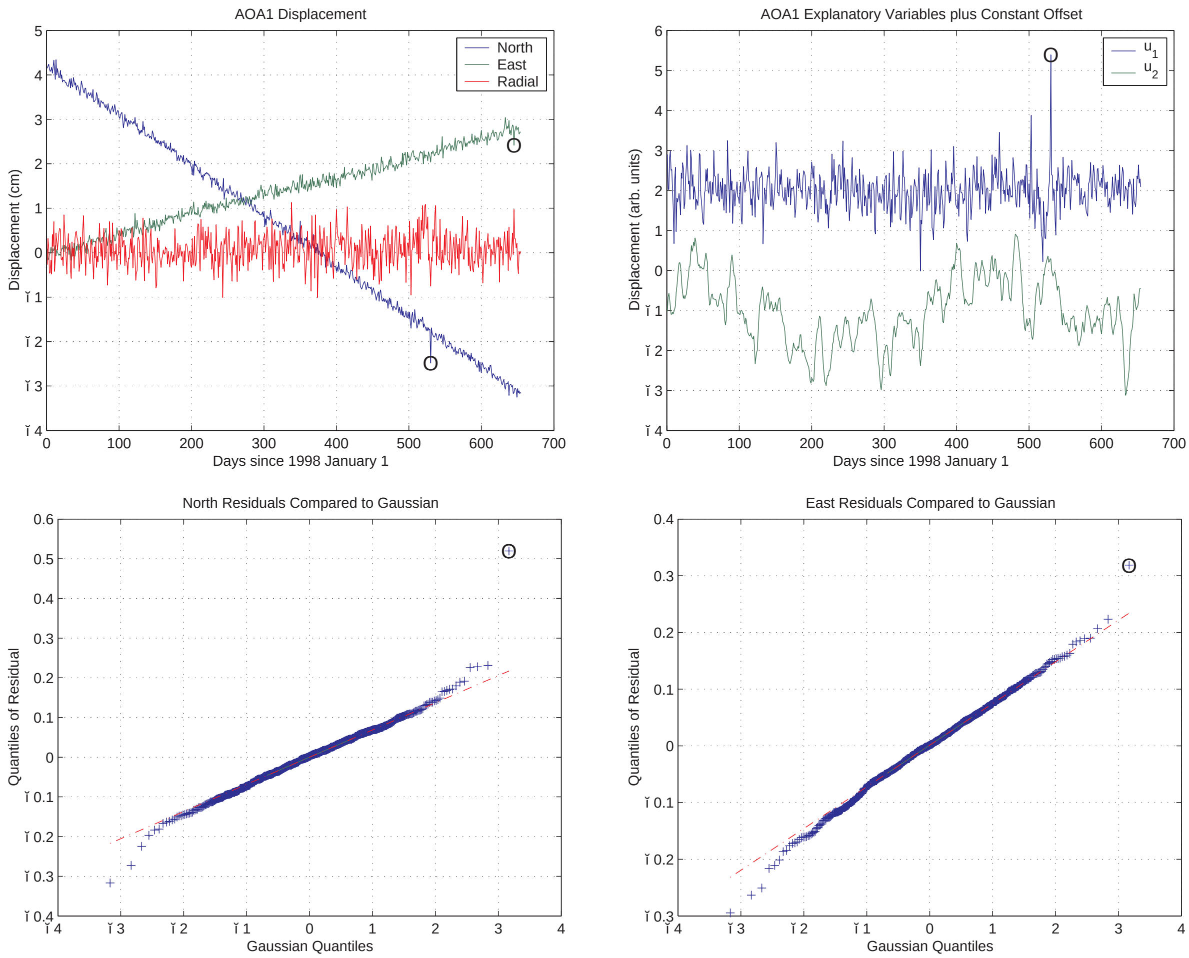
The HMM finds different modes of the signal:

- Dip 1998 from local ground water pumping
- The 1999.8 Hector Mine earthquake



Continuous Models

Kalman filter models for SCIGN station AOA1



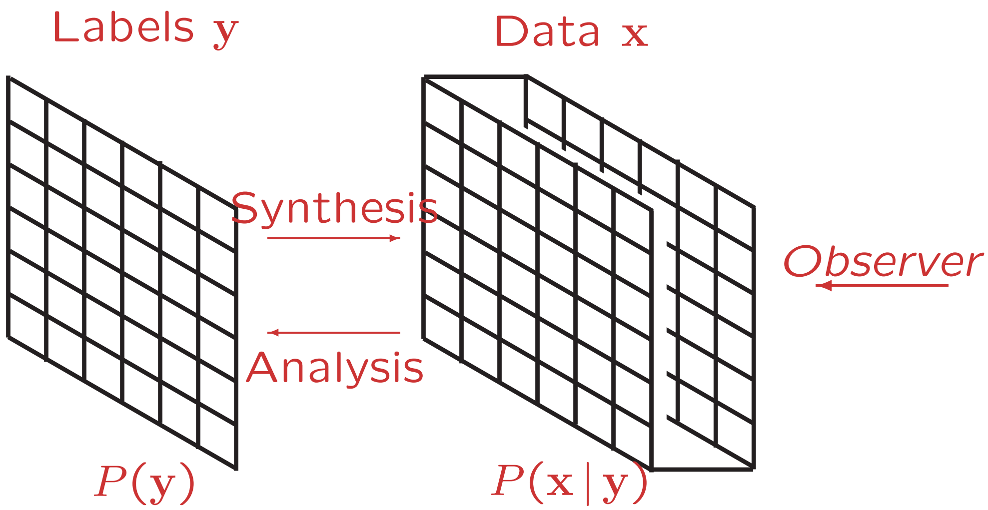
Two overlaid motion patterns:

- High-frequency, Brownian-type process
- Slower months–years trend

Residuals (N, bottom left; E, bottom right) show two deviations from Gaussianity at times 530 and 645 (circled).

Probabilistic Image Models

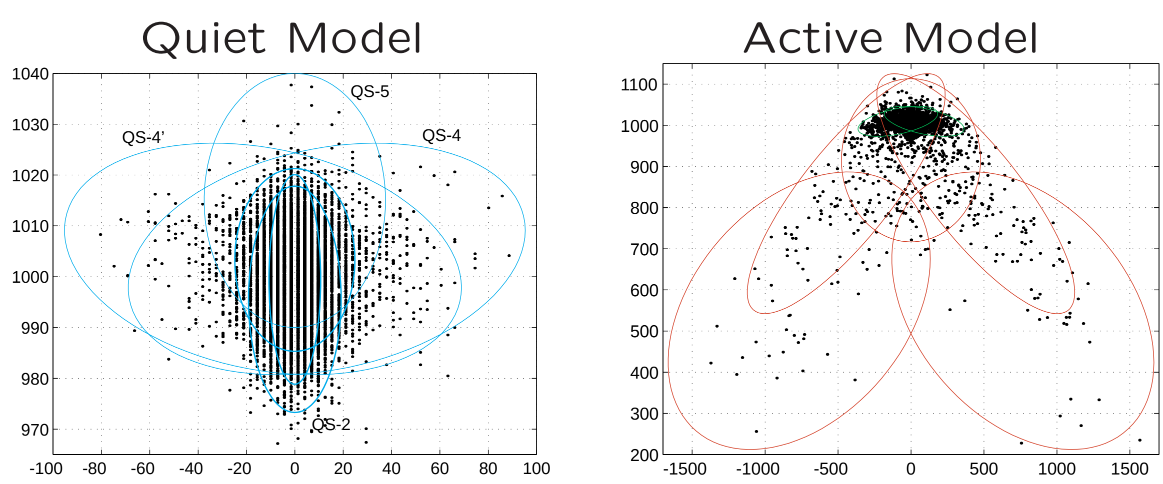
Two-dimensional version of time-series models



Need two distributions: $P(\mathbf{y})$ and $P(\mathbf{x} \mid \mathbf{y})$

- **Solar data: Already know $P(\mathbf{x} \mid \mathbf{y})$**

Mixture models from first part show how to translate labels to data



Drawing Random Images

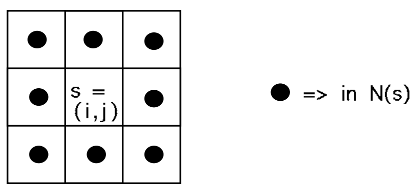
- **Quantifying Spatial Smoothness with $P(\mathbf{y})$**

Charge β for each disagreement of nearby pixels to enforce spatial coherence of labelings

Typically $\beta \geq 0$ controls smoothness in the prior

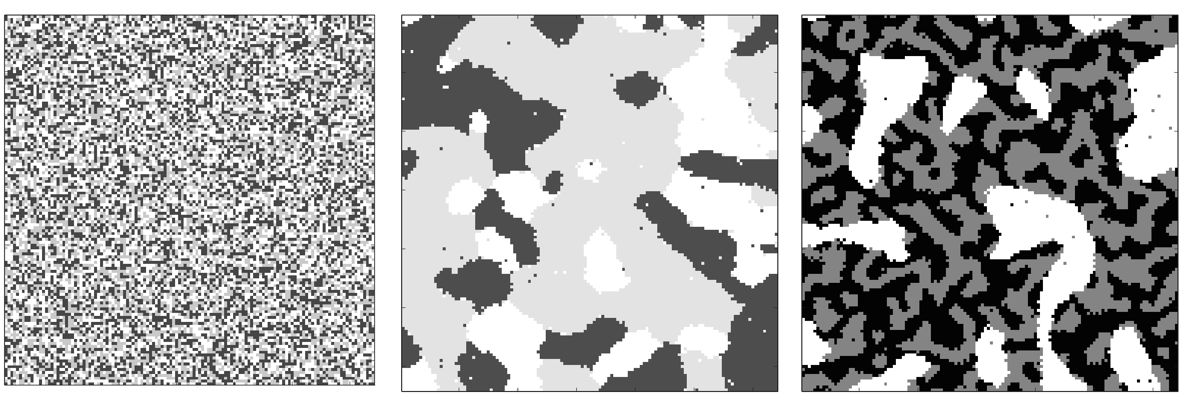
$$P(\mathbf{y}) = \frac{1}{Z} \exp\left(-\beta \sum_{s \sim s'} 1(y_s \neq y_{s'}) \right)$$

where $s \sim s'$ when site s within one pixel of site s'



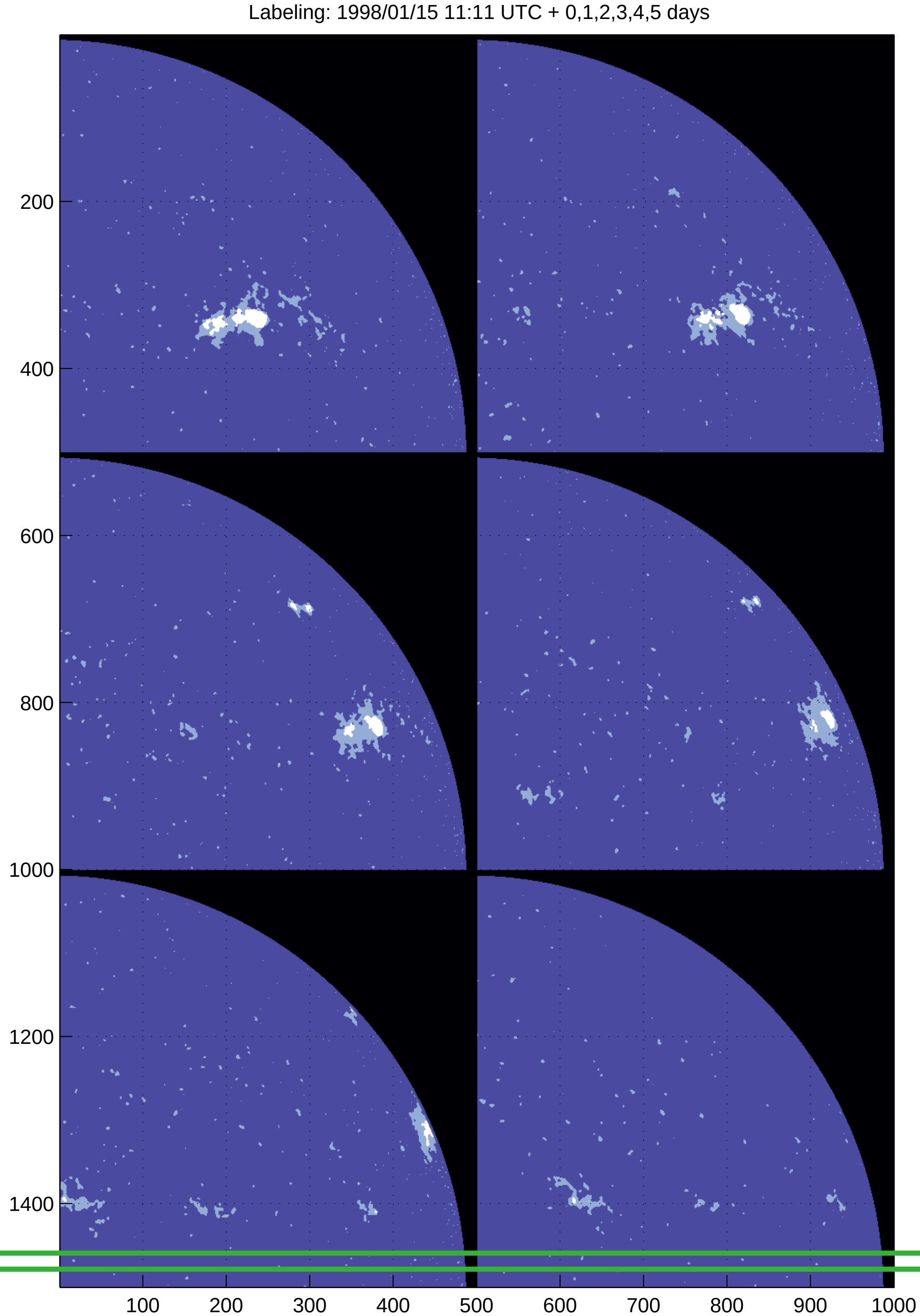
At $\beta = 0$, penalty and spatial constraint vanish

Sample realizations from $P(\mathbf{y})$



Labelings

From SoHO/MDI, 1998 January 15–20



Complex Objects

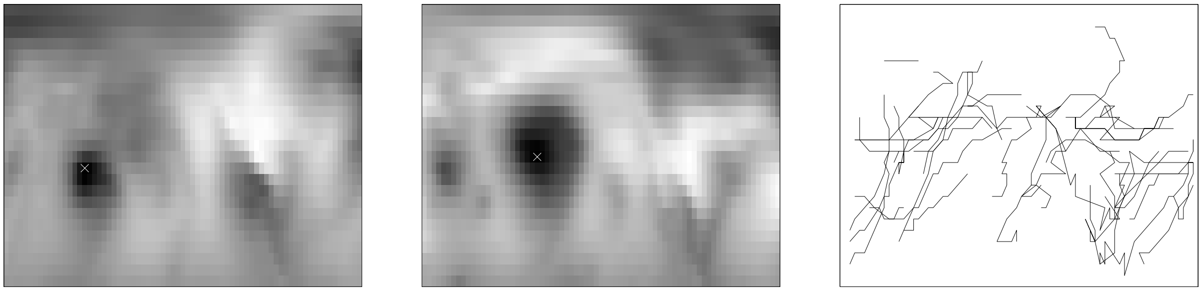
As pattern theory develops in the future it will be imperative to include more and more detailed subject matter knowledge into the regular structures — this is a form of mathematical knowledge engineering.

– U. Grenander, 1993

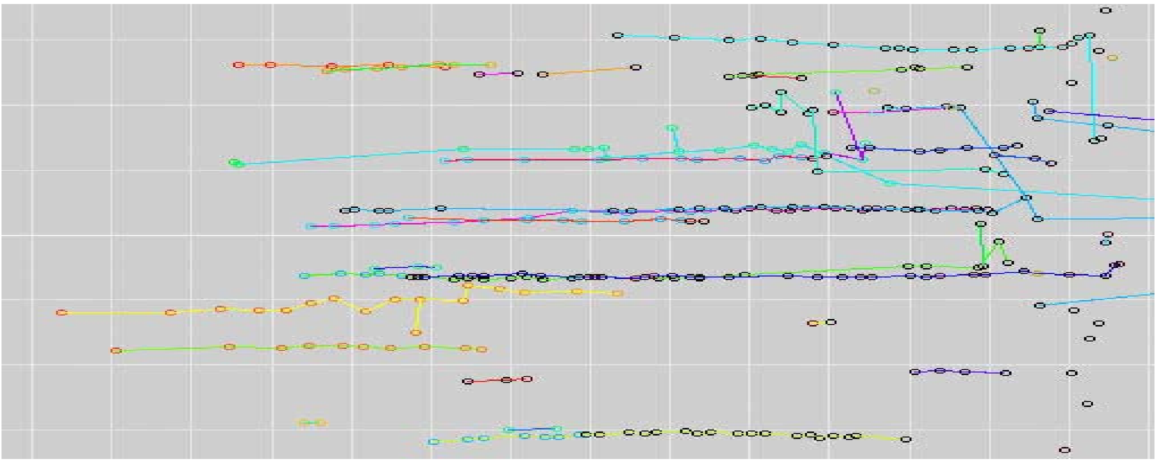
- **Structured index sets** are natural arenas for general-purpose models
- **Objects with complex links** require detailed domain-specific modeling

Full-Sequence Classification

- **Cyclone trajectories**
  Left: Pacific sea-level pressure ($\delta t = 48\text{H}$)
  Right: trajectories from (quantized) observations
  Data from P. Smyth, UC Irvine



- **Sunspot trajectories** from 1996 Aug.–Nov.



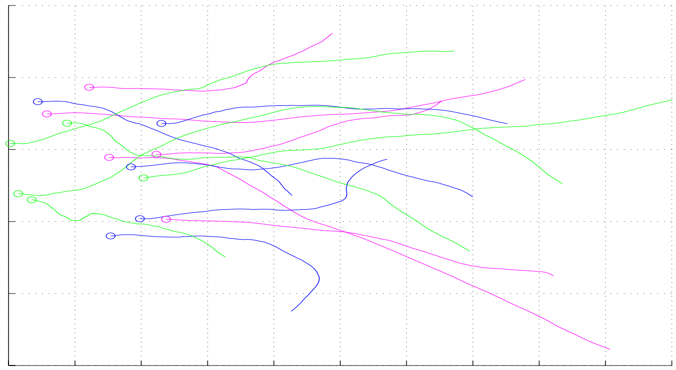
Varying lengths $\Longrightarrow$ no common feature vector

Expressive Temporal Models

- **Sequence Classification**

  Overall mode variable controls whole model

  Approach...
  estimate model parameters automatically
  compare learned models statistically



- **Related variants**

  Temporally evolving mode variable
  Account for mode switches mid-sequence
  Geodetic applications

  Non-gaussian state representations
  Coarse-scale ambiguity in data
  Bifurcations in phase space

More on Complex Models

- **Computational Geology**

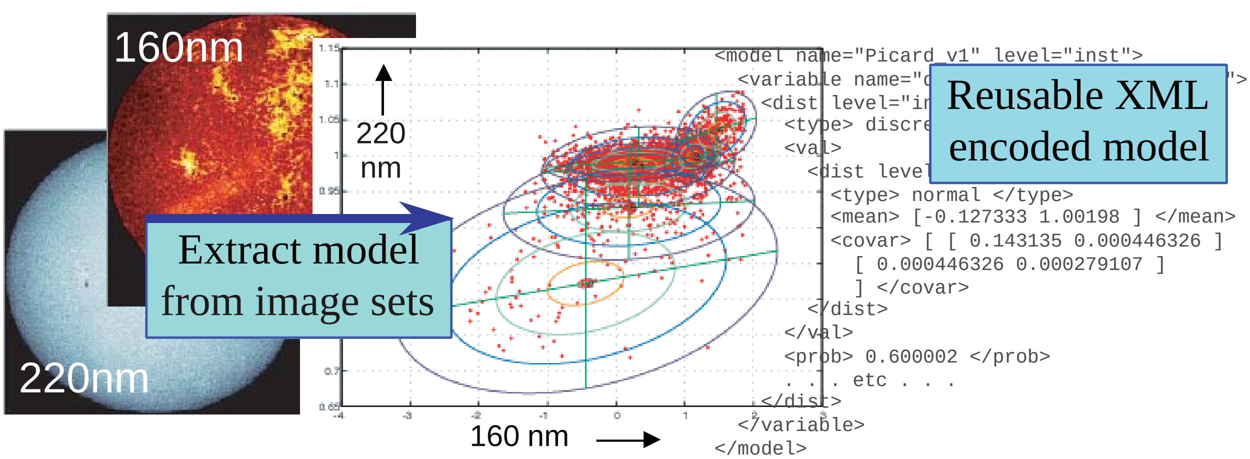
  Forward stochastic model for rock deposits using impact ejection physics

  Run in reverse to learn geological process history from rock observations

- **Languages for diverse models**

  Based on Bayes networks or stochastic grammars

  Encode model in neutral language



  Hand model and data to computational engine to learn hidden variables

Context

Contributors:

- Robert Bao (Stanford)
- Andrea Donnellan
- Vlad Gluzman
- Robert Granat
- Ken Hurst
- Saleem Mukhtar (Caltech)
- Judit Pap (UMBC)
- Kacie Shelton

Colleagues:

- Rick Bogart (Stanford)
- Michael Burl (U. Colorado)
- Igor Cadez (UC Irvine)
- Becky Castaño
- Dennis DeCoste
- Eric Mjolsness
- Padhraic Smyth (UC Irvine)

Support: Autonomy Technology Program, AISRP, SoHO Guest Investigator, REE

http://www-aig.jpl.nasa.gov/home/turmon/