

Underpinnings?

Sheela Nair and Mark Hansen
Department of Statistics



Thermal mapper

- In the statistical literature, these kinds of structures often fall under the title of “functional” data
- Let's consider a recent NIMS-RD deployment in the White Mountains
- The NIMS node was pared down to just a single sensor which collected spot temperature measurements on the ground over a 3m transect



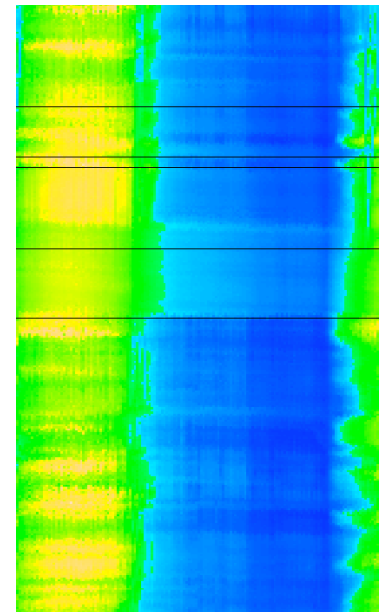
NIMS and complex data

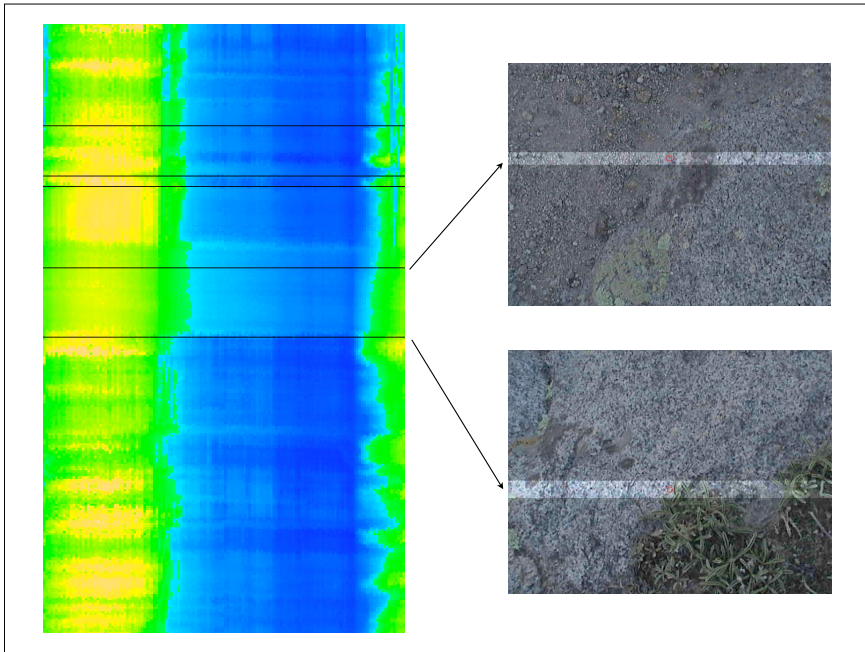
- For the most part, NIMS-RD deployments have involved regular sweeps over a *transect*
- This intuitively simple structure yields remarkably complex data; the regular patterns fill out a space-time matrix that asks for somewhat more refined analyses



Thermal mapper

- Here is one display of the data; pixel intensities represent temperature
- Along the x-axis we have time, the y-axis location along the transect
- What do we see?



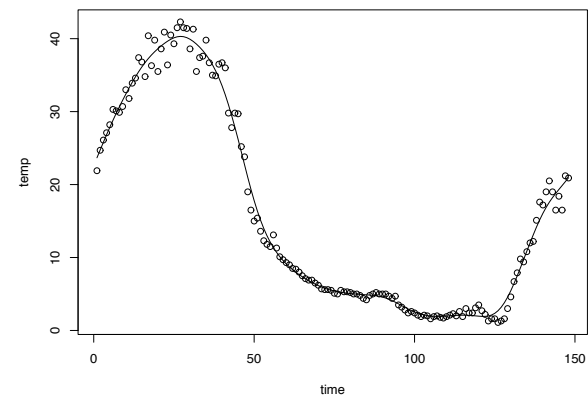


Functional data analysis

- In these settings, we consider “functional” properties of the object we’re looking at
- Slopes, curvatures, the location of extremes, inflections points; all of these might prove useful in describing the major characteristics of the data

Accounting for uncertainties

- The conditions were challenging for a deployment, wind kept the node from focusing precisely on a single spot
- The wind itself would cause fluctuations in the measurements; should we consider this noise?
- In understanding these data, it makes sense to think about a classification problem (what features of the data allow us to discriminate between the classes?) and there are uncertainties in our tagged data

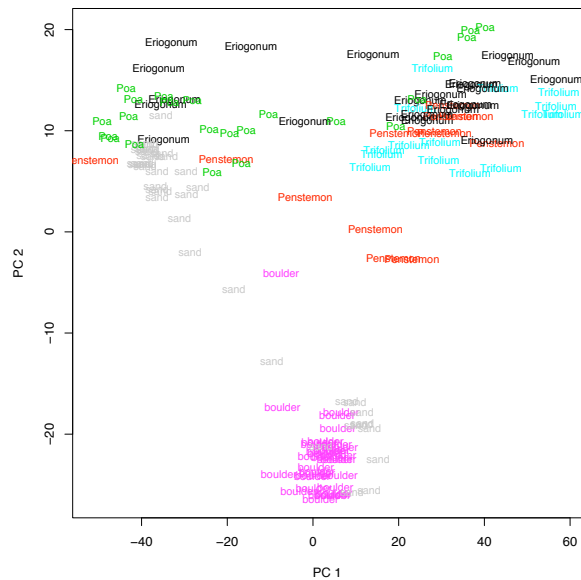
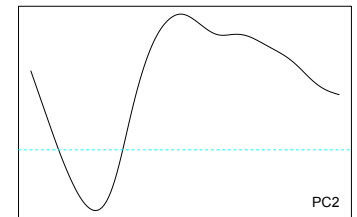
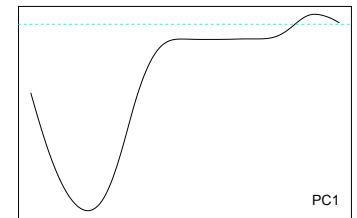


Functional data analysis

- Compare this to the usual “spot measurements” that might ride on a paired t-test or some other simple statistic
- A functional approach might give us greater interpretability and richer descriptions
- Let's see how this would work in the case of clustering; here's a *smoothed principal component decomposition*

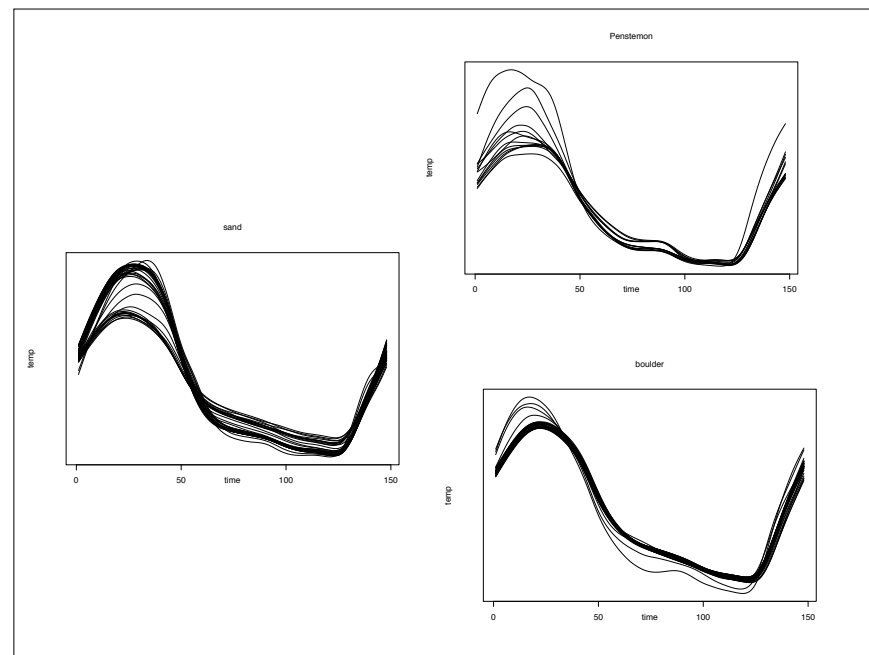
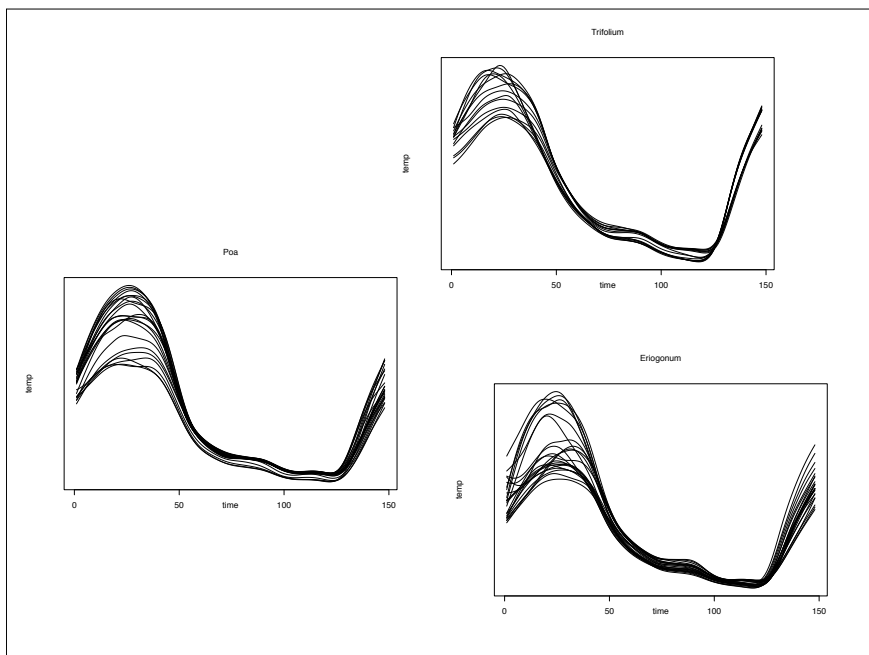
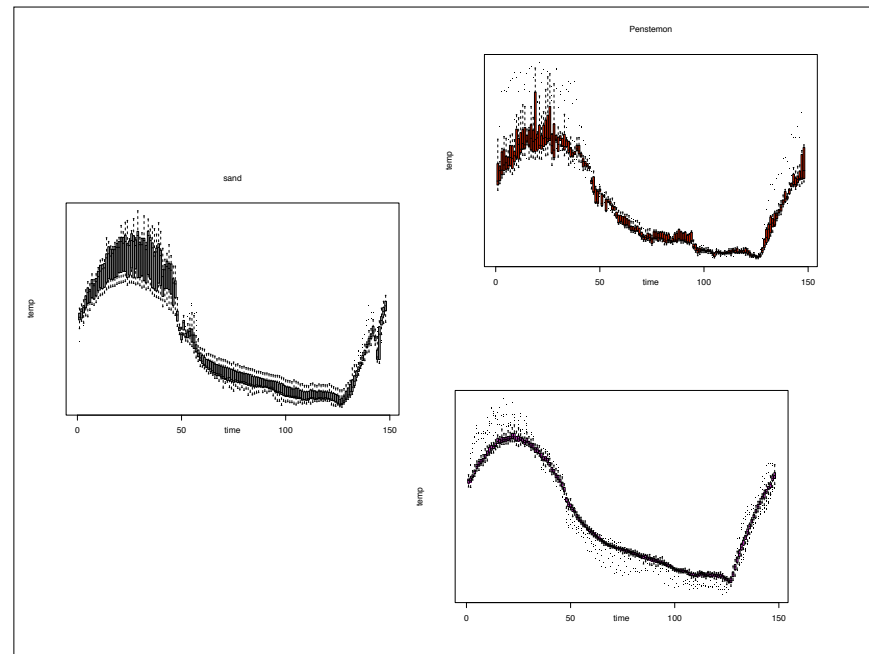
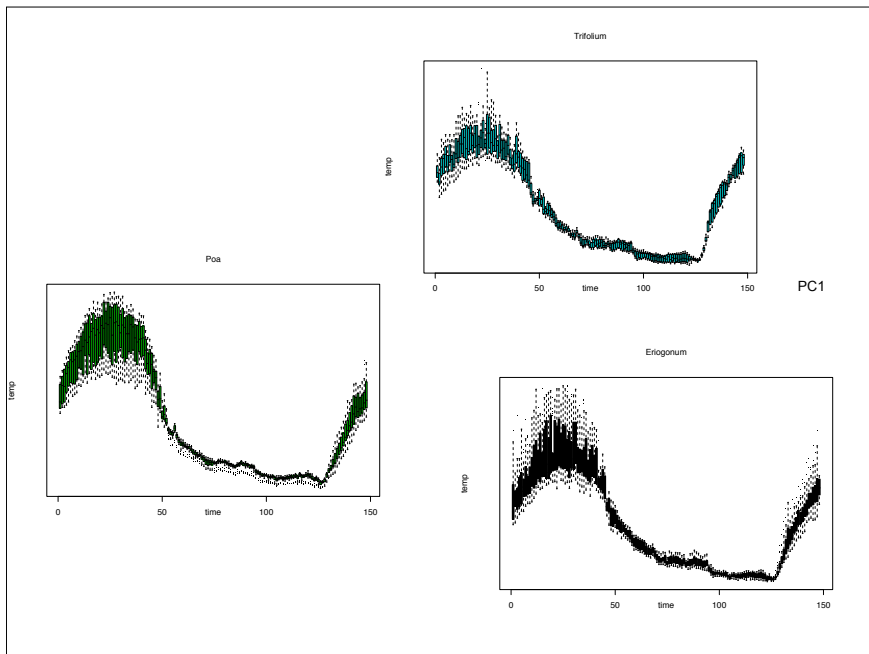
The smoothed PCs

- At the right we have our first two principal components, presented as curves (functions of time)
- How do we interpret these? How do they compare to the plot on the previous slide?



Back to the data

- Can we see some of this in the data? Let's consider both the “raw” curves and the smoothed data
- Where is there greater *within-class* variability? And is that noise or is it structured?



Classification

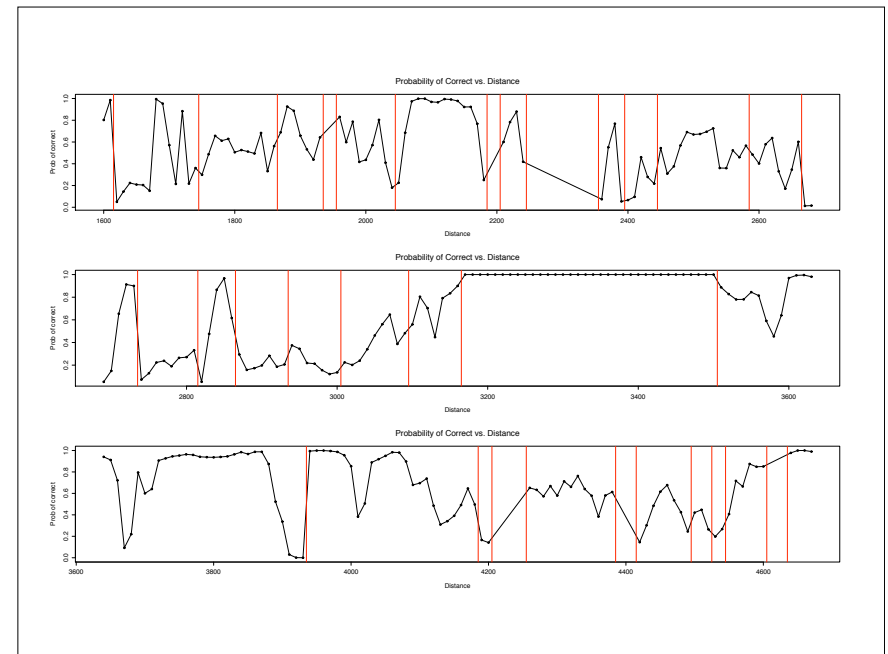
- There are several ways to incorporate functional components into a classification problem
- Dimension reduction via a simple projection

$$f(x) = \beta_1 B_1(x) + \dots + \beta_J B_J(x)$$

and then apply standard multinomial regression to the coefficients $\beta = (\beta_1, \dots, \beta_J)$; if we use something like B-splines, then the coefficients roughly describe the shape of the curve

- Alternatively, we can learn the important features as well as the form of the dependence; $f_1 = 1, f_2, \dots, f_K$ and

$$P(j|y) = \frac{\exp f_j(y)}{1 + \sum_k \exp f_k(y)}$$



Complications

- As I alluded to previously, the classes we're attempting to learn are "noisy"; that is, we might not know precisely what kind of ground cover is being sensed
- Wind blows the node from side to side, creating a kind of average of different classes
- We are now examining how to detangle these effects, again leveraging smoothness in space

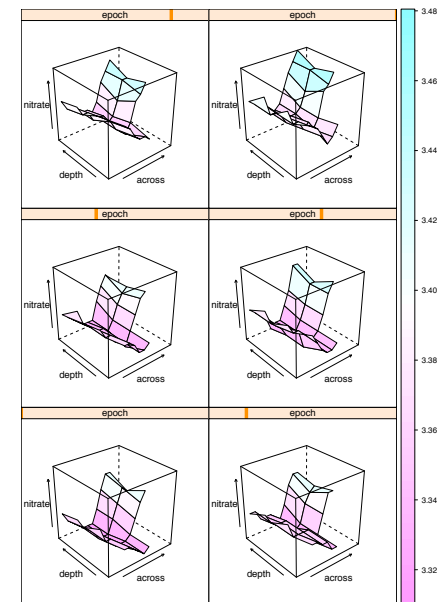
October 24, 2005

Table 1: default

predicted/truth	Boulder	Eriog.	Penst.	Poa	Sand	Trif.
Boulder	34	0	0	0	0	0
Eriog.	0	34	1	0	7	5
Penst.	0	6	25	1	8	3
Poa	0	1	1	37	4	1
Sand	0	6	5	4	73	3
Trif.	0	6	5	4	4	4

Medea Creek

- NIMS-RD has also spawned slightly more complicated structures
- In the Medea Creek deployment, an aquatic sensing node was dipped in a regular pattern giving us surfaces over time
- We've got a similar kind of structure as the White Mountains, but we've added a dimension in the response



Other examples?

- Cyclops style imagers will also provide similar structures as Medea Creek
- With the White Mountains, changing weather conditions let us cluster or classify regions; in the case of imagers, we might depend on other events like rain (shameless plug for Eric Graham's talk later today)

Overlap with statistical computing

- How do we equip biologists and others to work with these data *in situ*?
- Emissary is meant to provide that functionality; but it clearly cannot be a series of canned analyses
- 30 years of statistical computing has taught us that underneath every good GUI there's an expressive, extensible language
- What would that look like for NIMS? for fixed deployments?