Product Formulas for Positive Functions and Applications to Network Data

Peter Jones (Yale)

Joint Work with Devasis Bassu (Telcordia), Linda Ness (Telcordia), and Vladimir Rokhlin (Yale)

DIMACS Workshop on Statistical Issues in Analyzing Information from Diverse Sources

Outline of the Talk

- Representation of Positive Functions
- The Product Formula
- Coefficients in the product representation and statistical properties
- Some background from Analysis, Geometry, and Mathematical Physics
- Visual Displays
- Several Sources
- Diffusion Geometry
- Preprocessing Tools

SLE Price Chart and "Bursty" Volume from May, 2009

How would we best represent the volume?

Some Haar-like functions

"The Theory of Weights and the Dirichlet Problem for Elliptic Equations" by R. Fefferman, C. Kenig, and J.Pipher (Annals of Math., 1991). We first define the "L^∞ normalized Haar function" h_I for an interval I of form $[j2^{-n},(j+1)2^{-n}]$ to be of form

$$h_I = +1 \text{ on } [j2^{-n},(j+1/2)2^{-n})$$

and

$$h_I = -1 \text{ on } [(j+1/2)2^{-n},(j+1)2^{-n}).$$

The only exception to this rule is if the right hand endpoint of I is 1. Then we define

$$h_I(1) = -1.$$

The Product Formula

- **Theorem (F,K,P)**: A Borel probability measure μ on [0,1] has a unique representation as

$$\prod(1 + a_I h_I) ,$$

where the coefficients a_I are $\in [-1,+1]$. Conversely, if we choose any sequence of coefficients $a_I \in [-1,+1]$, the resulting product is a Borel probability measure on [0,1].

Note: For general positive measures, just multiply by a constant. Similar result on $[0,1]^d$.

Note: See "The Theory of Weights and the Dirichlet Problem for Elliptic Equations" by R. Fefferman, C. Kenig, and J.Pipher (Annals of Math., 1991)

Relative "Volume"

The coefficients a_I are computed simply by computing relative measure ("volume") on the two halves of each interval I. Let L and R = left (resp. right) halves of I. Solve:

$$\mu(L) = \tfrac{1}{2}(1 + a_I)\,\mu(I)$$

$$\mu(R) = \tfrac{1}{2}(1 - a_I)\,\mu(I)$$

Then $-1 \leq a_I \leq +1$ because μ is nonnegative.

Why Use It Instead of Wavelets?

- The coefficients measure relative measure instead of measure. All scales and locations are counted as “equal”.
- Allows another method to represent the signal, where one immediately detects large changes in relative volume. (Anomaly detection)
- Multiple channels are immediately normalized if one looks just at the coefficients instead of absolute volume.
- One cannot easily synthesize using Haar wavelets and the “usual” expansion”. (How to keep the function positive?)

History in Analysis

This method and its “cousins” have been around for a while for analysis of various classes of weight functions on Euclidean space. See e.g.

(P. Jones, J. Garnett) *BMO from dyadic BMO.* Pacific J. Math. (1982)

One idea is to use the dyadic setting, analyze functions or weights, and average over translations of the dyadic grid.

Support of Measures and Information Dimension

An Example of use. Suppose a probability measure is obtained in the previous manner by using the same PDF for every coefficient.

Entropy: $\Sigma\, p_k \log_2(p_k) = -\, nh,$

Here we sum over intervals of length 2^{-n}. The (expectation of the) entropy, h, is derived from the PDF, and is independent of n (so set n = 1).

Theorem: The information dimension h

(= dimension of the support of the measure) is (a.s.) given by the expectation for n = 1. Here $p_1 = 1 + a_{[0,1]}$ and

$p_2 = 1 - a_{[0,1]}$. (Just average over the PDF.)

Remark: Small h means the signal is more “bursty”. Can calculate the full “multifractal spectrum”.

Synthesis

Fix a Probability Density Function on [-1,+1]. Choose the coefficients a_I independently from that PDF. In the following we show a simulated signal, followed by a color chart (using "Jet") of the coefficients.

This allows one to simulate for purposes of training. Similar (but also different!) from other methods (e.g. Barral and Mandelbrot). They need to normalize measure "at every step".

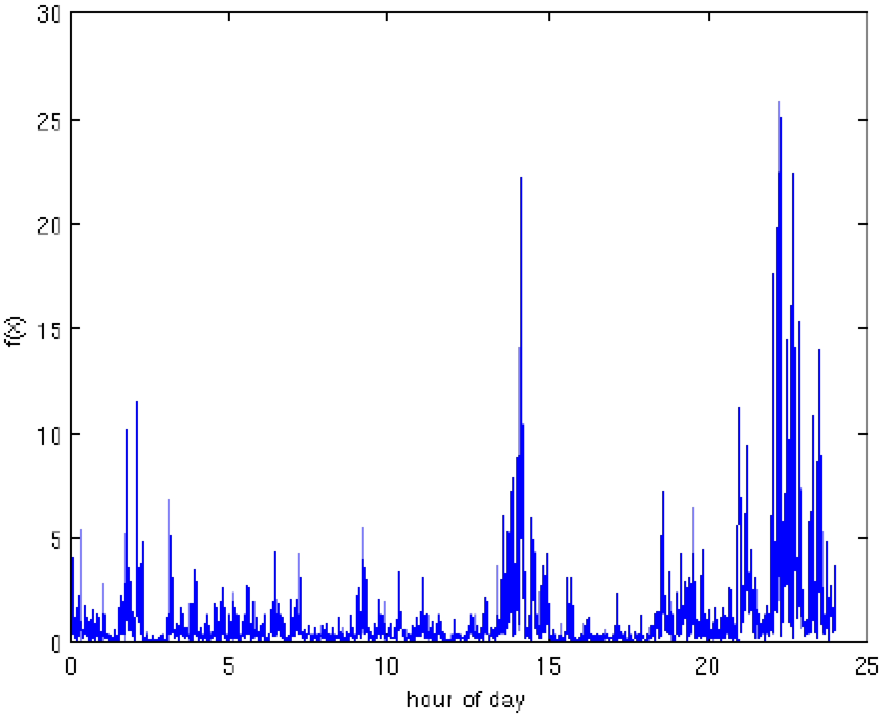


Figure 1 Volume for PDF 1/2(1 + cos(πx))

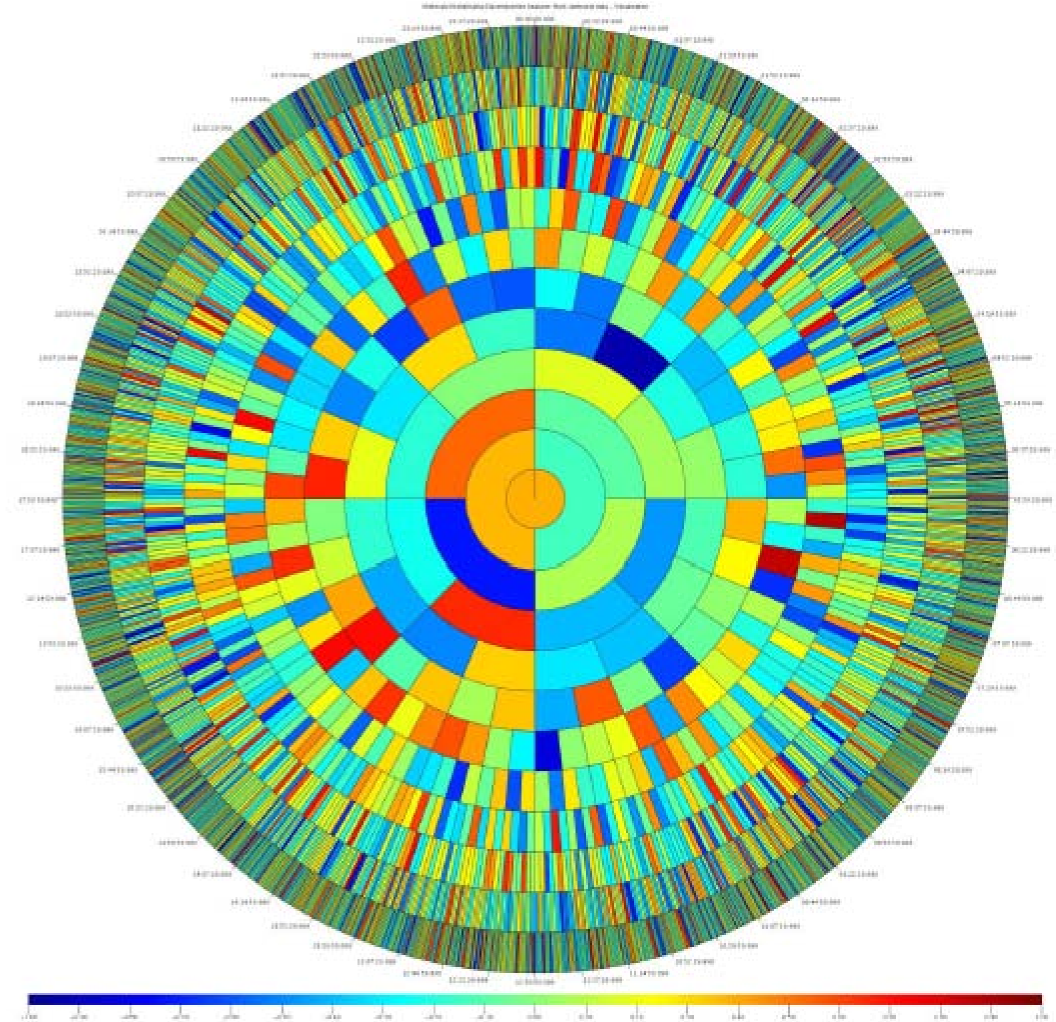
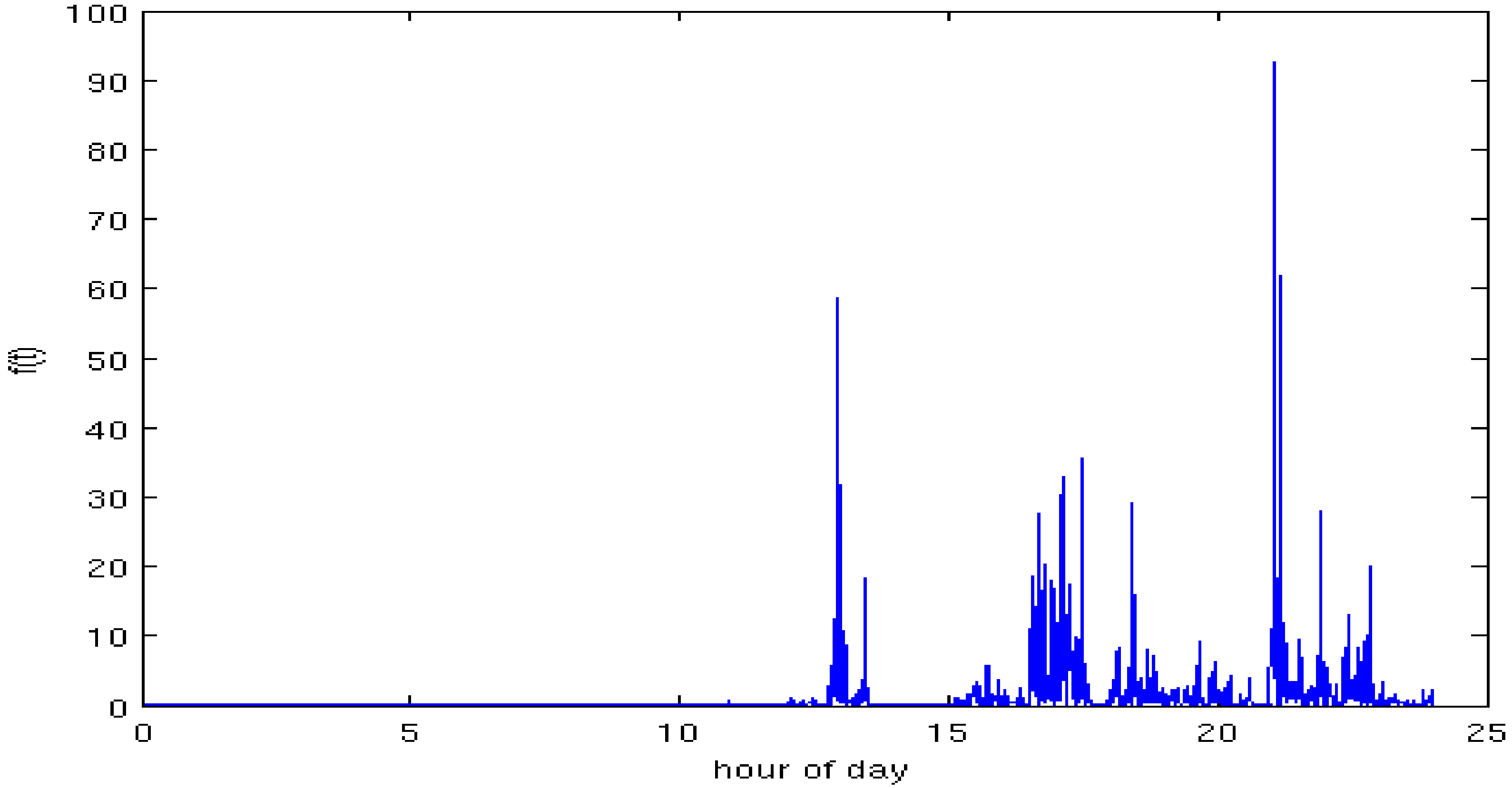


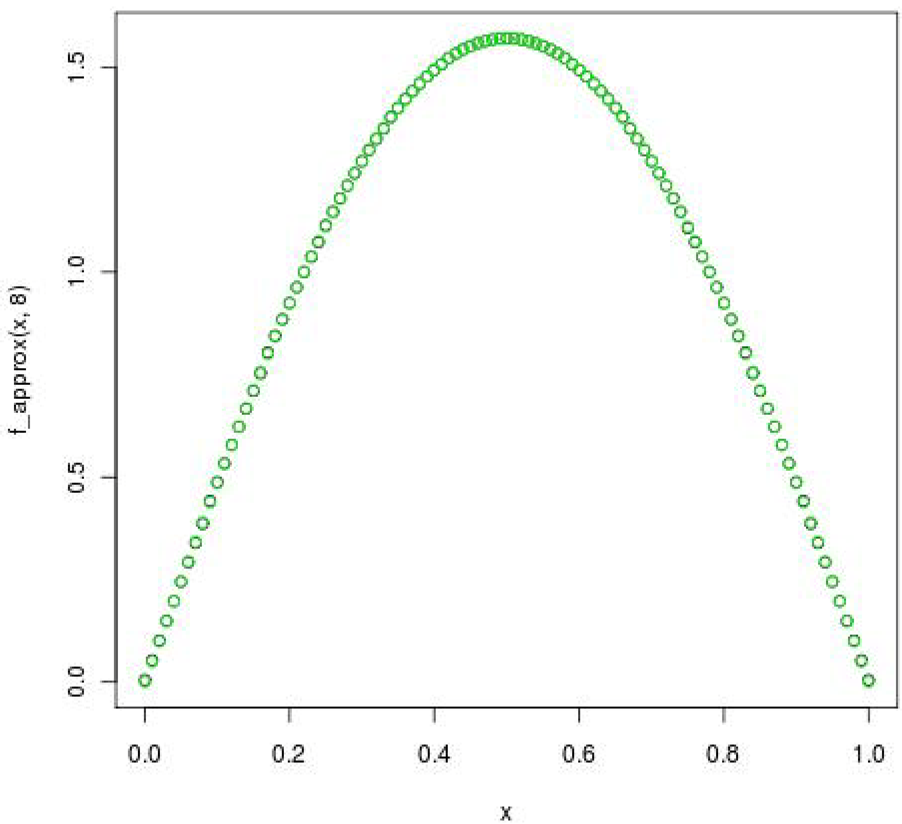
Figure 3 Coefficients for PDF 1/2(1 + cos(πx))



Synthesized signal

Here all the coefficients have been chosen from one PDF for the intervals in the "second half of the day".

PDPM – An Example (scale 8)

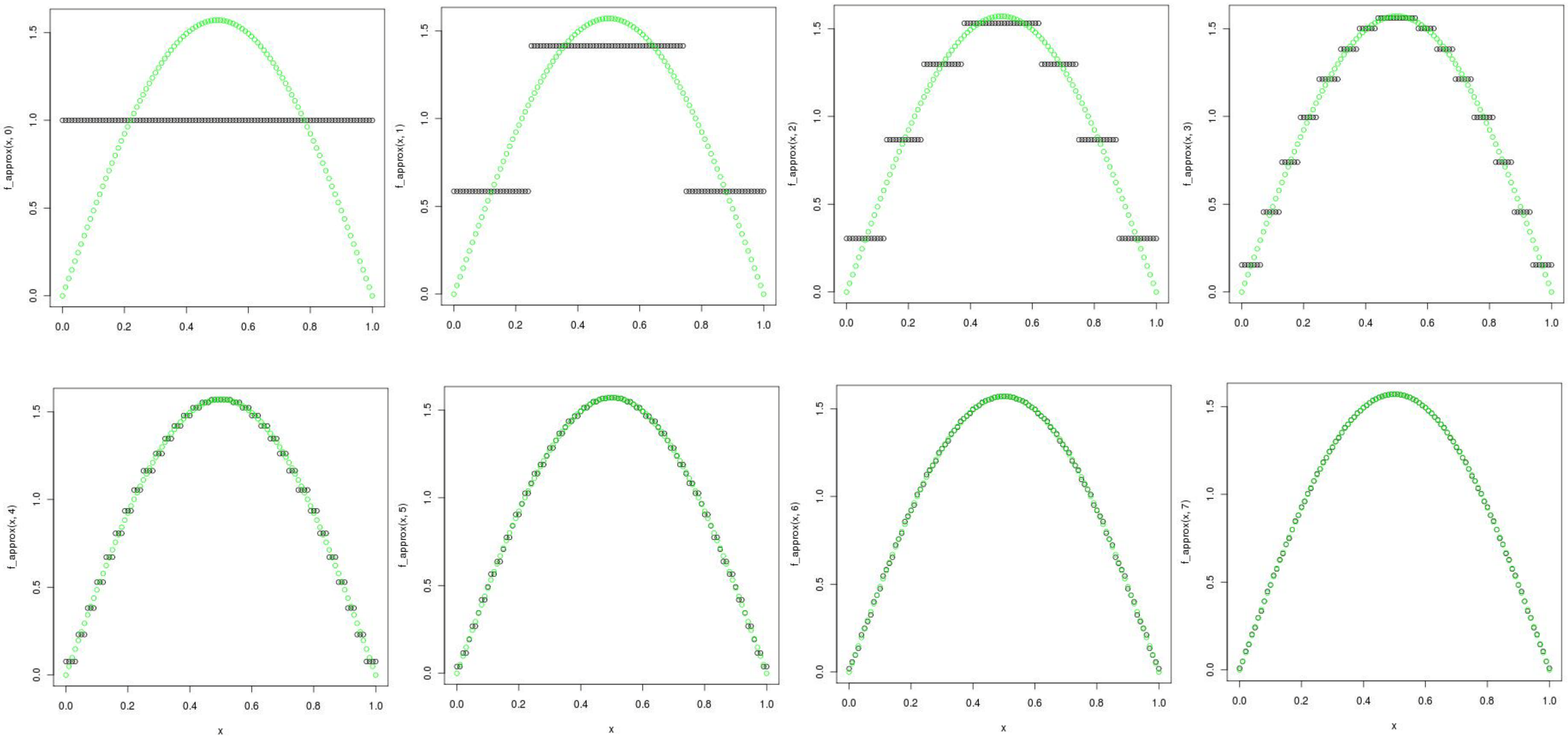


- Function

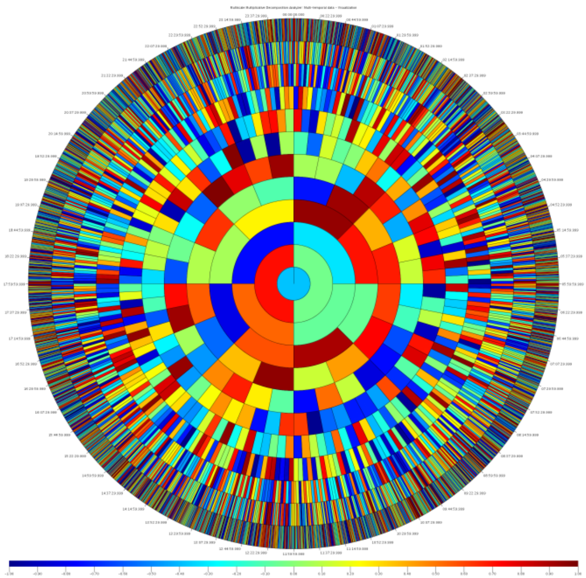
  $f(x) = (\pi/2).\sin(\pi.x)$ defined on [0,1]

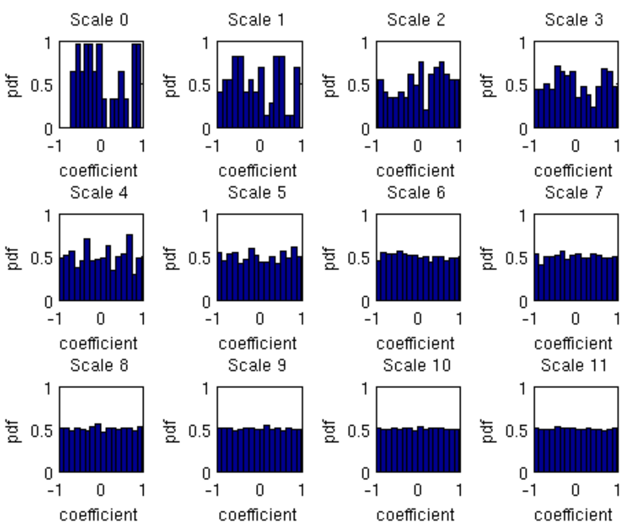
- Legend
  - True function (green)
  - Approximation using PDPM (black)

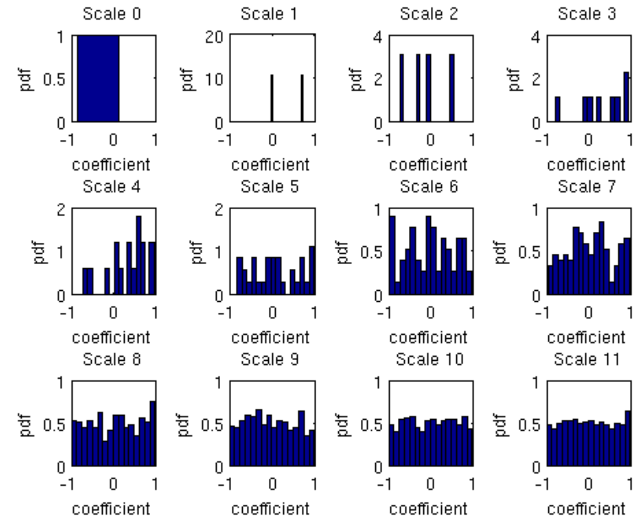
PDPM – An Example (scales 0 to 7)

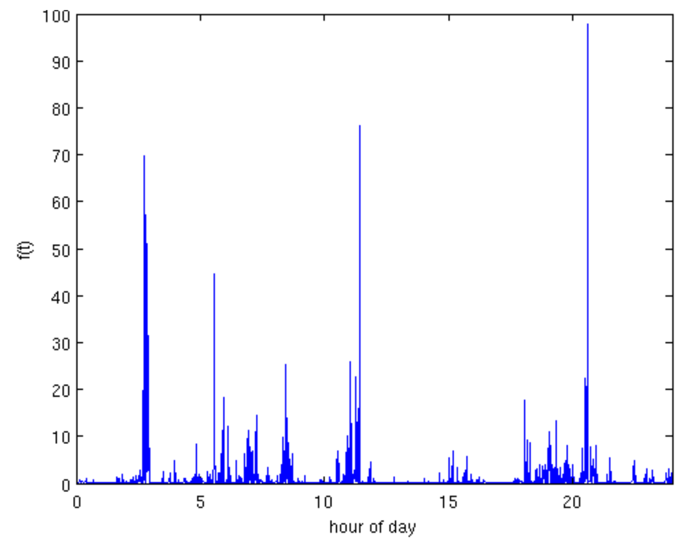


PDPM Synthesis/Visualization – Uniform

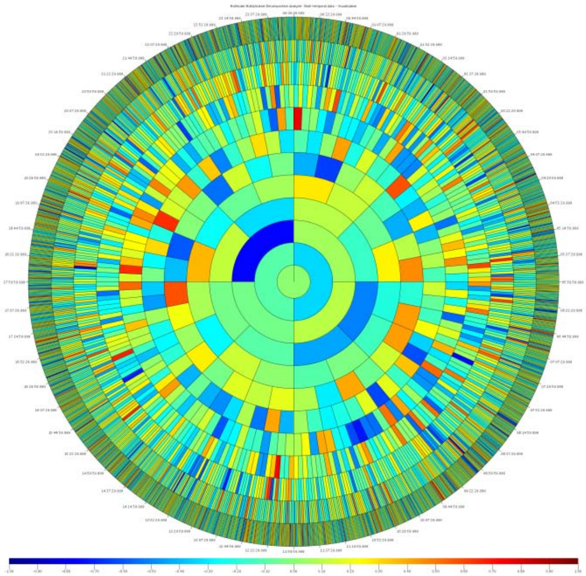


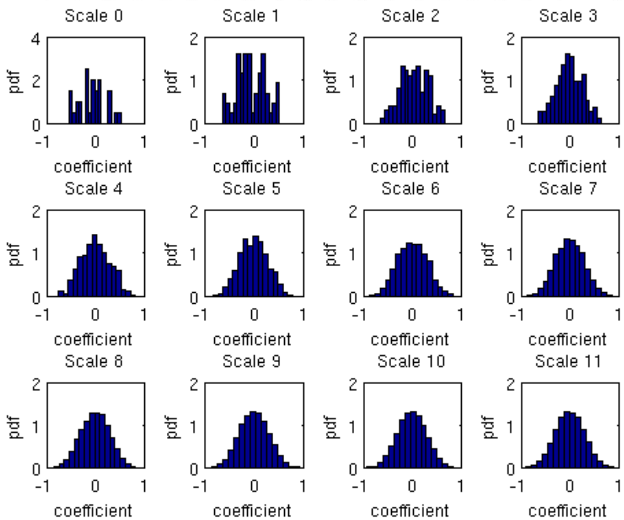


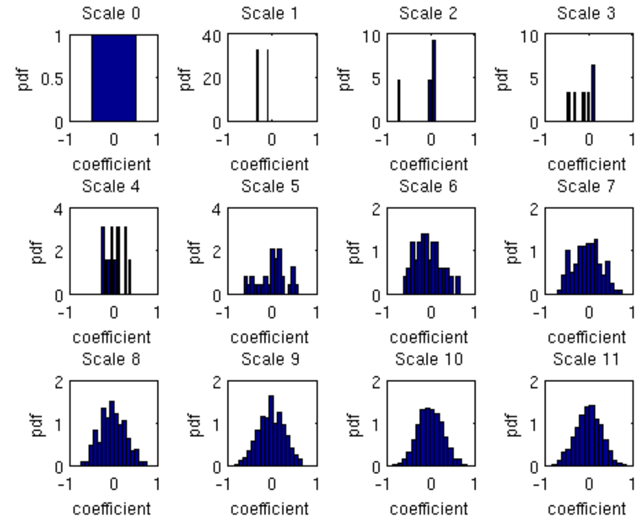


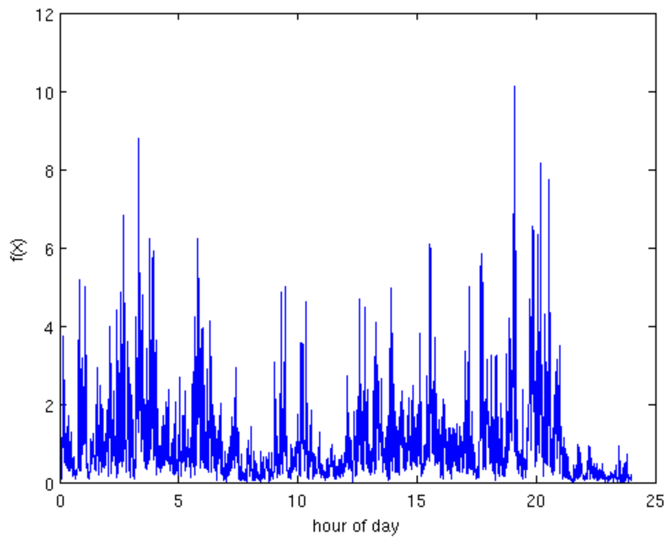


PDPM Synthesis/Visualization – Bell









Real Data from Weather Event

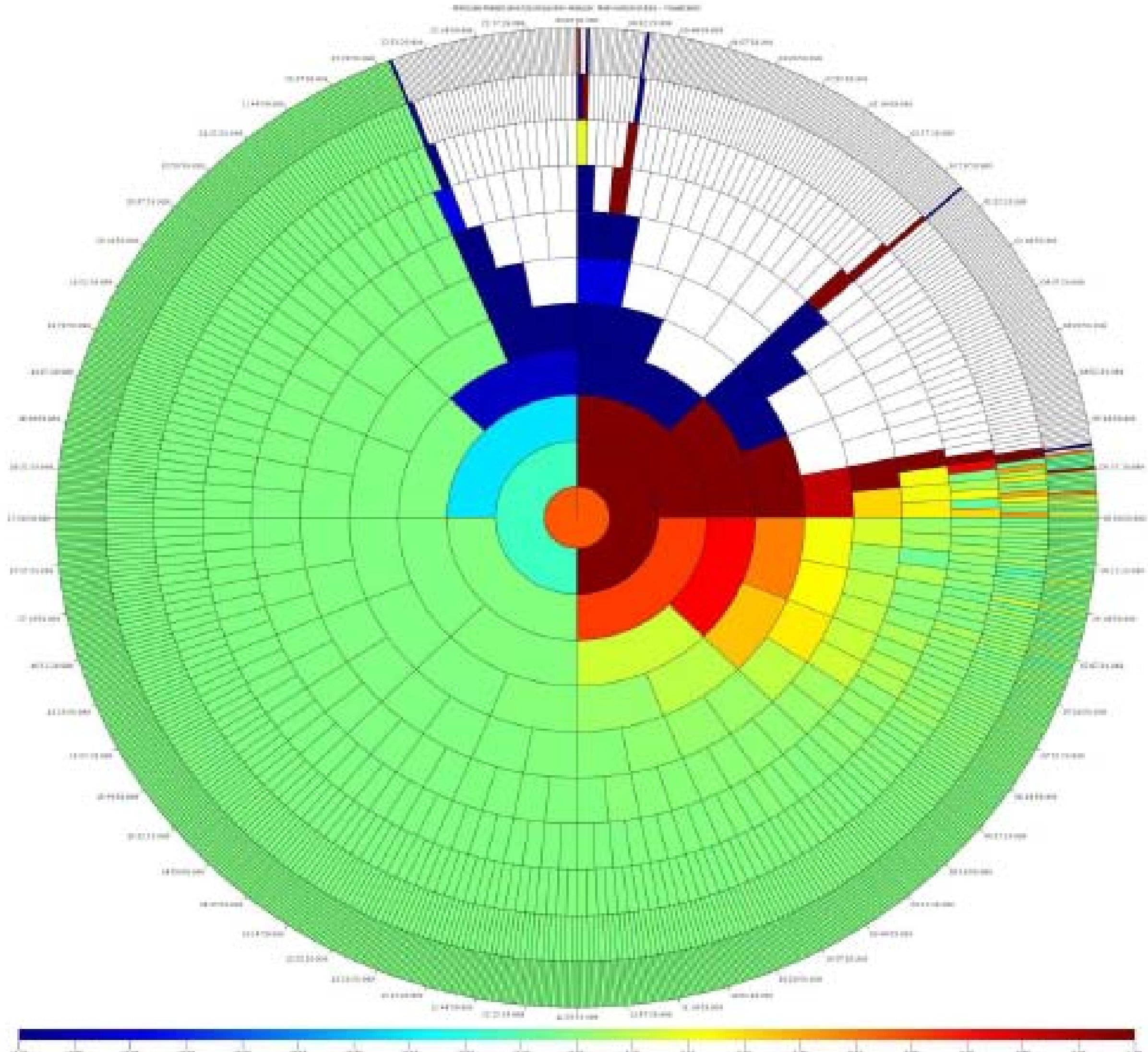
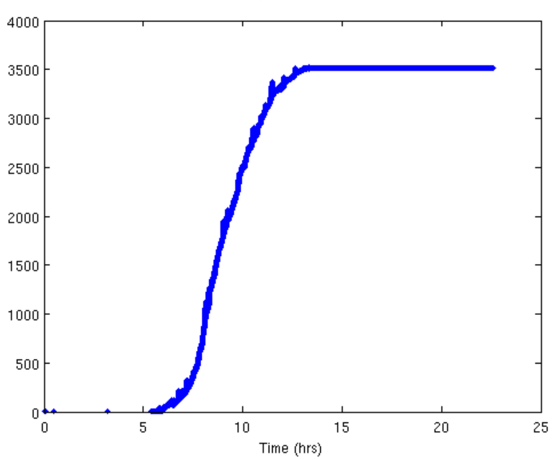
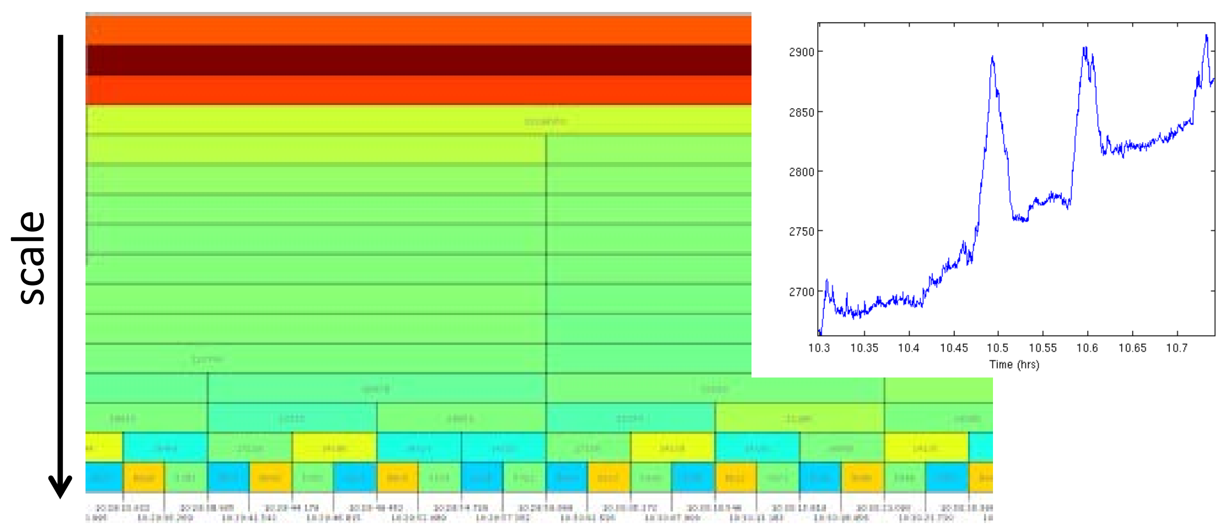


Figure 1

Dataset I: Local Information

- Network measurements over the span of ~24 hours; ~65k sources
- Overall ramping behavior
- Inherent bursty behavior revealed at select scales



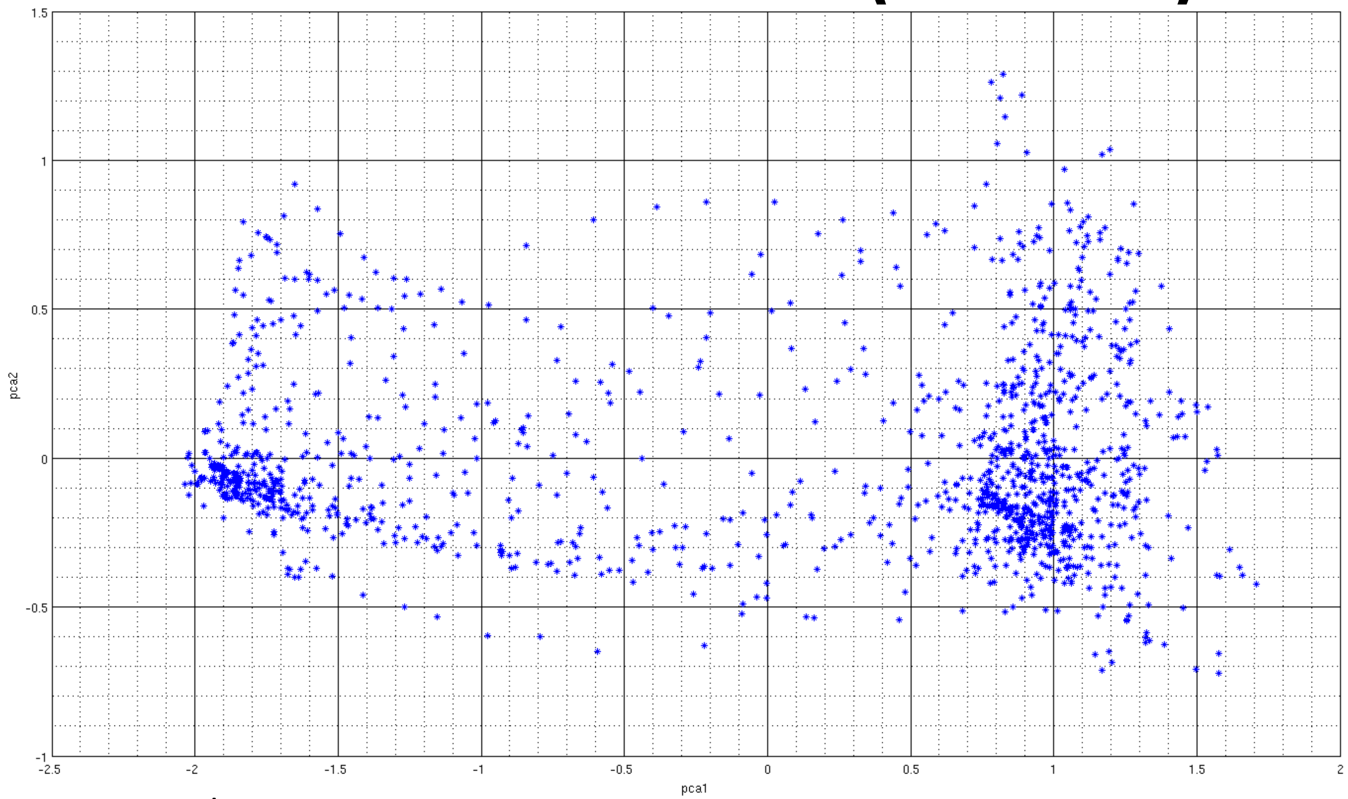


Note: Colors represents amount of function variation

Analysis of Another Data Set

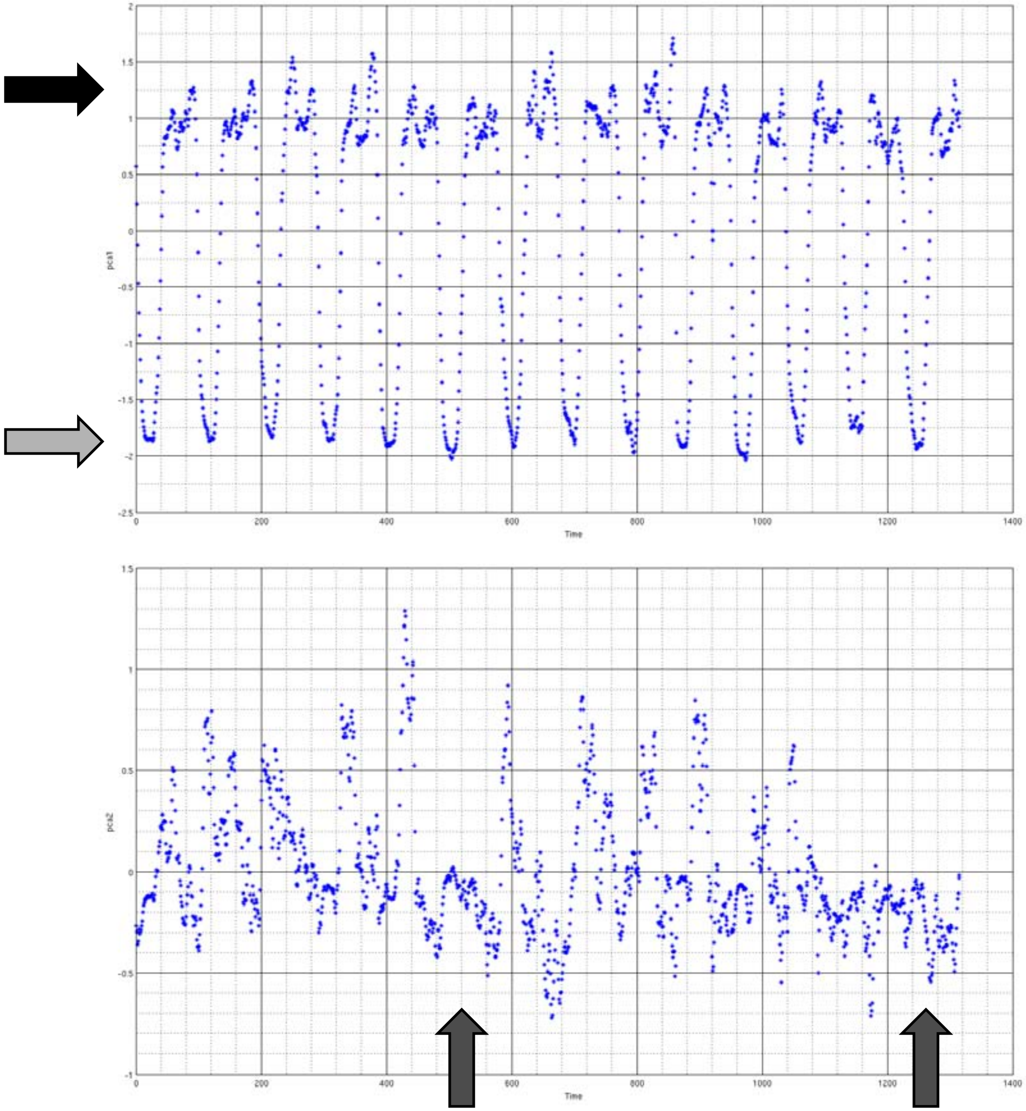
- In the first analysis to follow we will simply use the actual volume. A data point will be the volume at times t, t -1, t -2, t -3 (for each source). We then examine the data using first PCA ("Data Set 2: 1 and 2 of 2"), then Diffusion geometry ("Data Set 2: Diffusion Map"). Here there is no apparent advantage (over SVD) in using Diffusion Geometry.
- In the second analysis we instead take a data point to be the coefficients of the product expansion of the day's volume (from each source), and display only the results from DG. Notice the excellent clustering that happened automatically (Data Set 2: PDPM + Diffusion Map").

Dataset II: SVD (1 of 2)



- Measurement data
  - 182 network entities
  - ~ 2 weeks of data @ 15 min intervals
  - 1315 data points
- PCA analysis
  - Window size: 1 hour
  - Dimensions: 728 (reduced 2-dim. representation above)

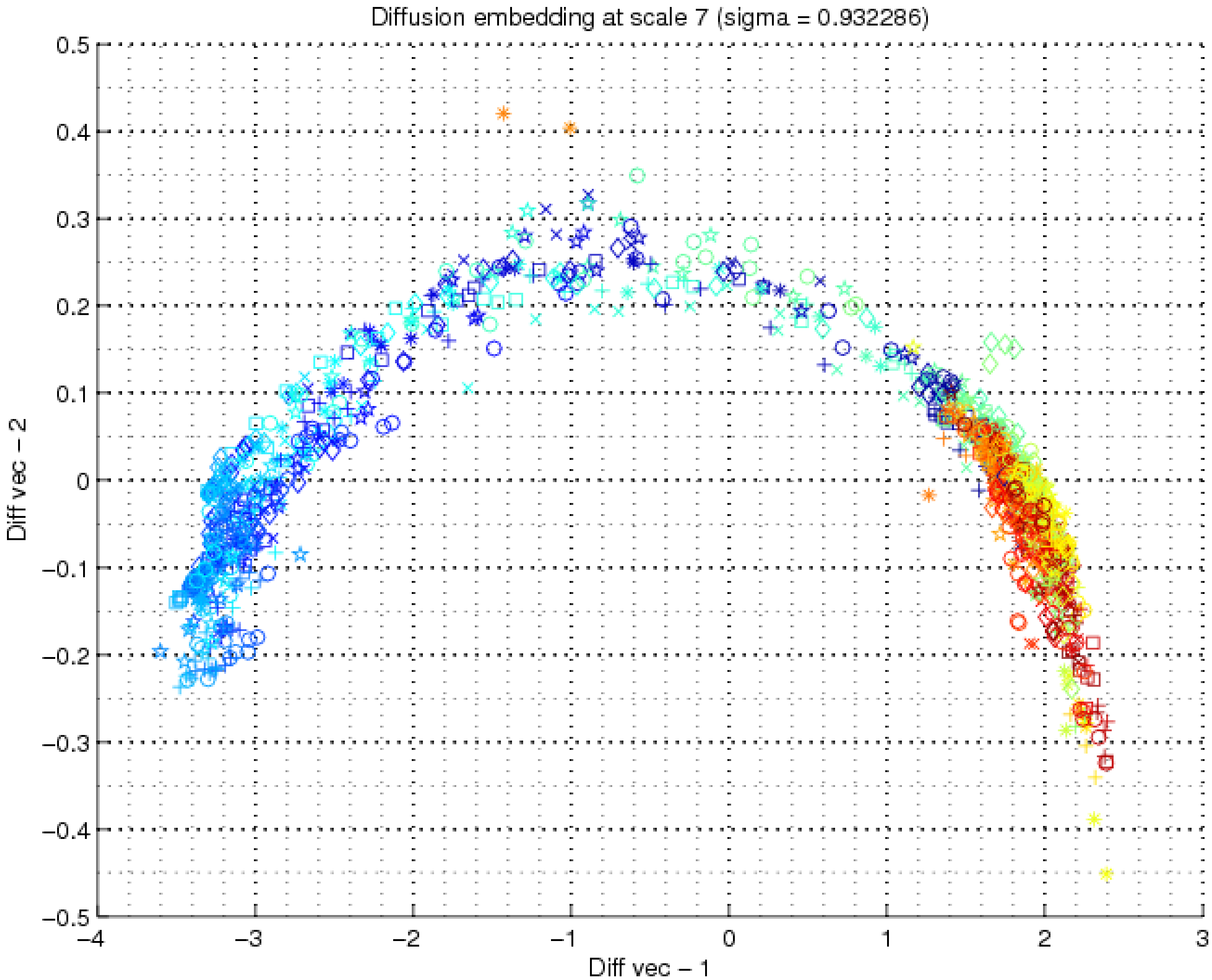
Dataset II: SVD (2 of 2)



- PCA analysis
  - major discriminator is time of day
  - minor discriminator is day of week
  - Legend: morning, night, weekend

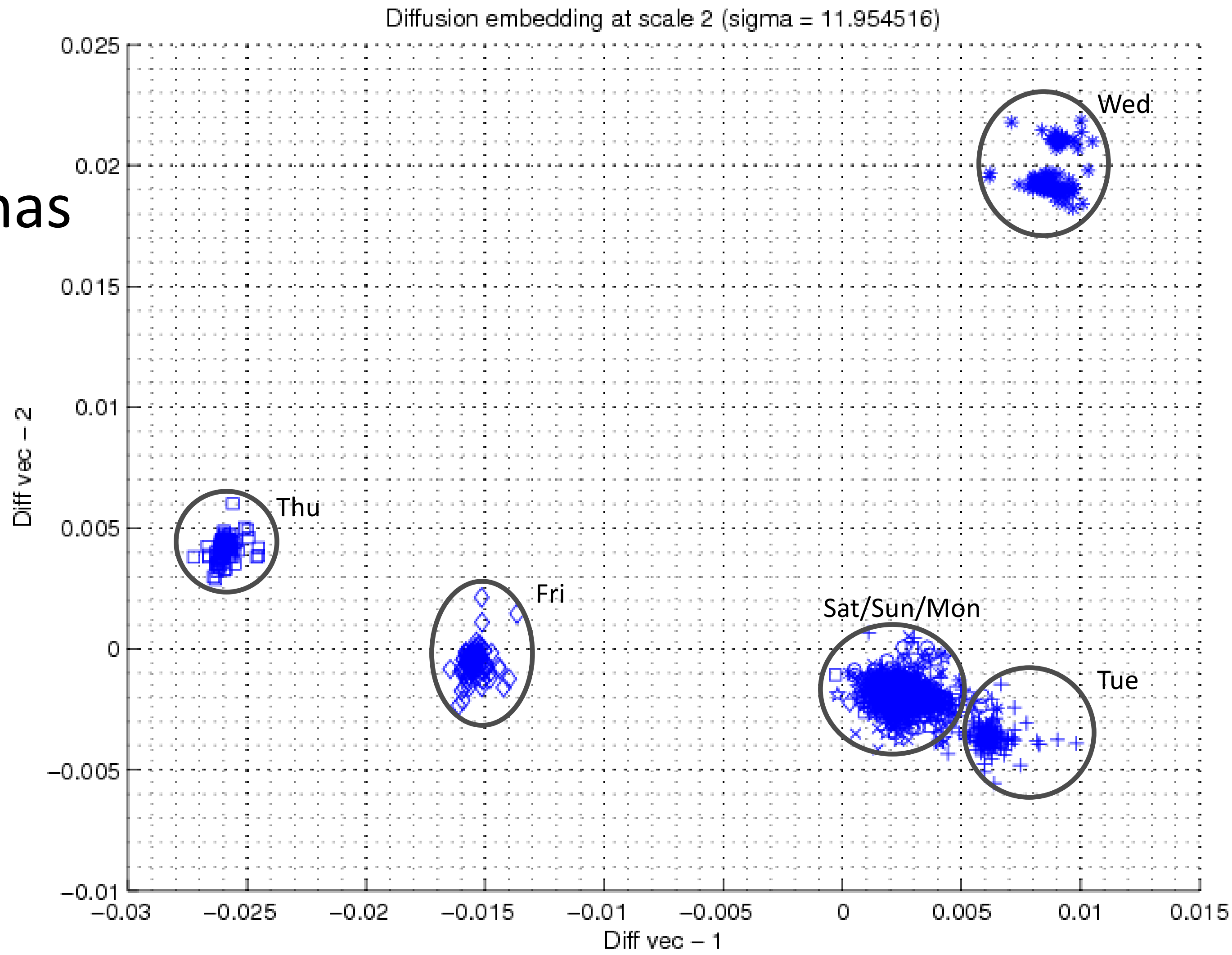
Dataset II: Diffusion Map

- Time of day
  - JET color map
- 1 hr window



Dataset II: PDPM + Diffusion Map

- Daily profile
  - 182 antennas
  - 14 days



Diffusion Geometry

The idea behind Diffusion Geometry is to put a new, nonlinearly defined representation af the data set. "Cartoon Version":

Step 1. Embed the data set in $\mathbb{R}^d$.

Step 2. Choose a value of σ to define a length scale. Build the data matrix

$$M = (\exp\{-|x_i - x_j|^2/\sigma\})$$

Step 3. Compute the Eigenvectors $\{\Phi_k\}$.

DG Continued

4. Carefully choose a small number of eigenfunctions, e.g. Φ_3 , Φ_4 , Φ_7 . The new data set representation is given by the image

$$x_i \longrightarrow \{\Phi_3(x_i), \Phi_4(x_i), \Phi_7(x_i)\}$$

5. Why do it? It could be helpful where PCA works poorly. It computes "local affinities" and builds coordinates from that information.

References

- For a tutorial see: Mauro Maggioni's Homepage. Click on "Diffusion Geometry".
- Why can it work? See:

"Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels". (Joint with Mauro Maggioni and Raanan Schul) PNAS, vol. 105 no. 6 (2008), pages 1803-1808

(Plus full paper on Raanan Schul's Homepage)

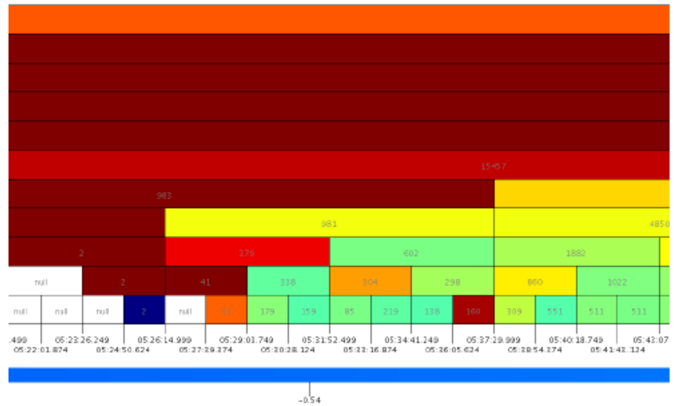


Figure 2

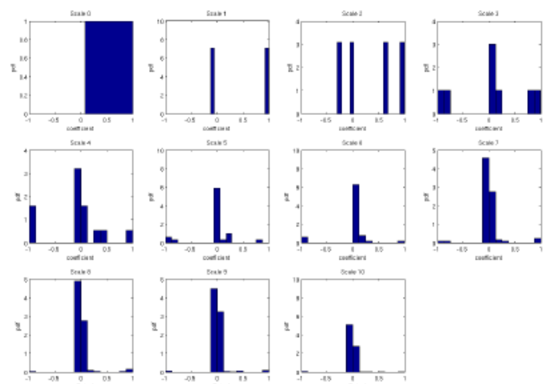


Figure 3

Conformal Field Theory without CFT

An analogous object appears in CFT. The toy model:

For each interval I choose iid Gaussians, $B_I(t)$ ~ $N(o,t)$, i.e. variance = t. Use “Feynman-Kac”:

$$\mu = \prod \exp\{B_I(t)\, h_I(x) - t/2\}$$

$$= \exp\{\Sigma\, (B_I(t)\, h_I(x) - t/2)\}$$

If $t < t_{critical}$, μ is (a.s.) a nozero, finite Borel measure. (Appears in “SLE”.)

Two CFT References

1. arXiv:0912.3423 Title: Random Curves by Conformal Welding
   Authors: K. Astala, P. Jones, A. Kupiainen, E. Saksman
   Comments: 5 pages
   Subjects: Complex Variables (math.CV); Mathematical Physics (math-ph)
2. arXiv:0909.1003
   Title: Random Conformal Weldings
   Authors: K. Astala, P. Jones, A. Kupiainen, E. Saksman
   Comments: 36 pages, 2 figures
   Subjects: Complex Variables (math.CV); Mathematical Physics (math-ph)

Measures and Curves

- Take a positive measure μ on the circle and normalize its mass to be 2π. Let $\Phi' = \mu$. If μ puts positive mass on each open interval and has no Dirac masses, Φ is a homeo of the circle.
- Amazing Fact: Under mild hypotheses, there is a curve corresponding to μ (the "welding curve"). An example is on the next page.

Jeff Brock's
Image of a
QuasiFuchsian
Limit Set

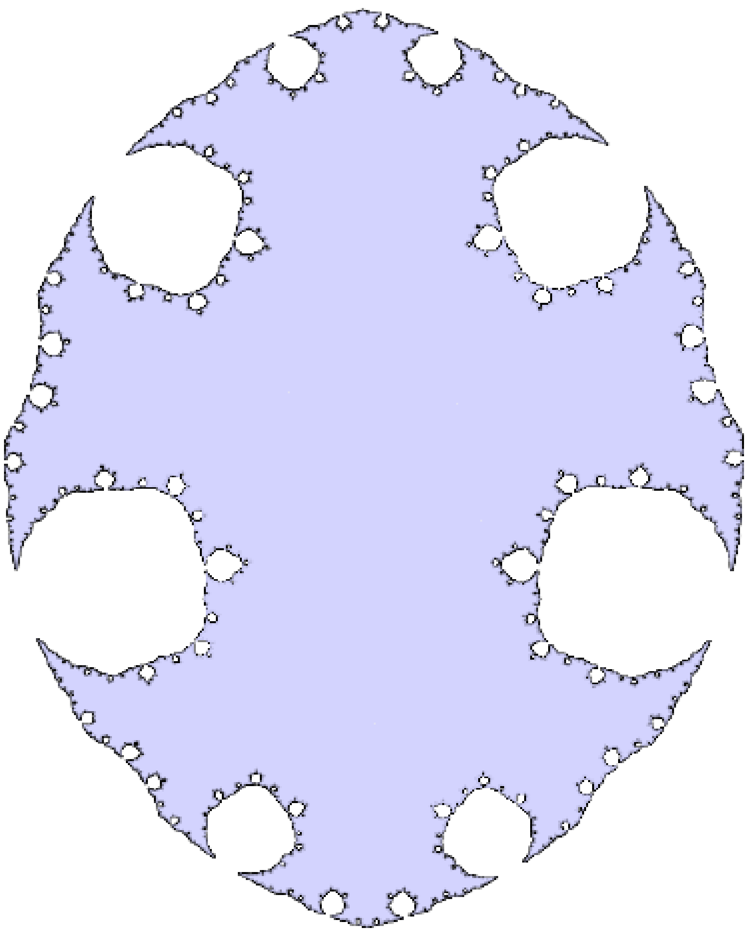
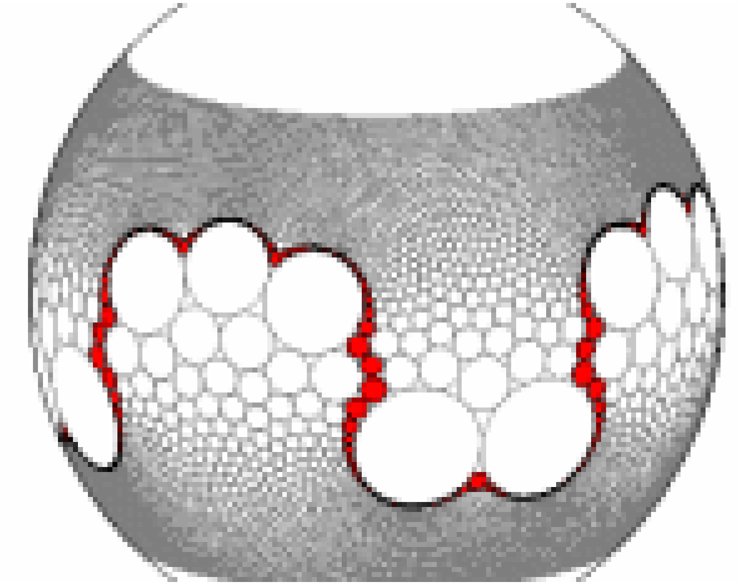


Figure 4



A Welding Curve (Circle Packing)