Building the Uncertainty

Roger J-B Wets, Mathematics, University of California, Davis
& David Woodruff, G.S. Management, University of California, Davis

Workshop "Statistics for Optimization", Paris, Summer 2014

Collaborators: Ignacio Rios, Univ. de Chile, Santiago
Fabian Rüdel, Univ. Duisburg-Essen,
Julia Peyre, Ecole Polytechnique, Paris
Cesar Silva Monroy, Sandia National Labs, New Mexico
Sarah Ryan, Iowa State University, Iowa
Johannes Royset, Naval Post-Graduate School, California

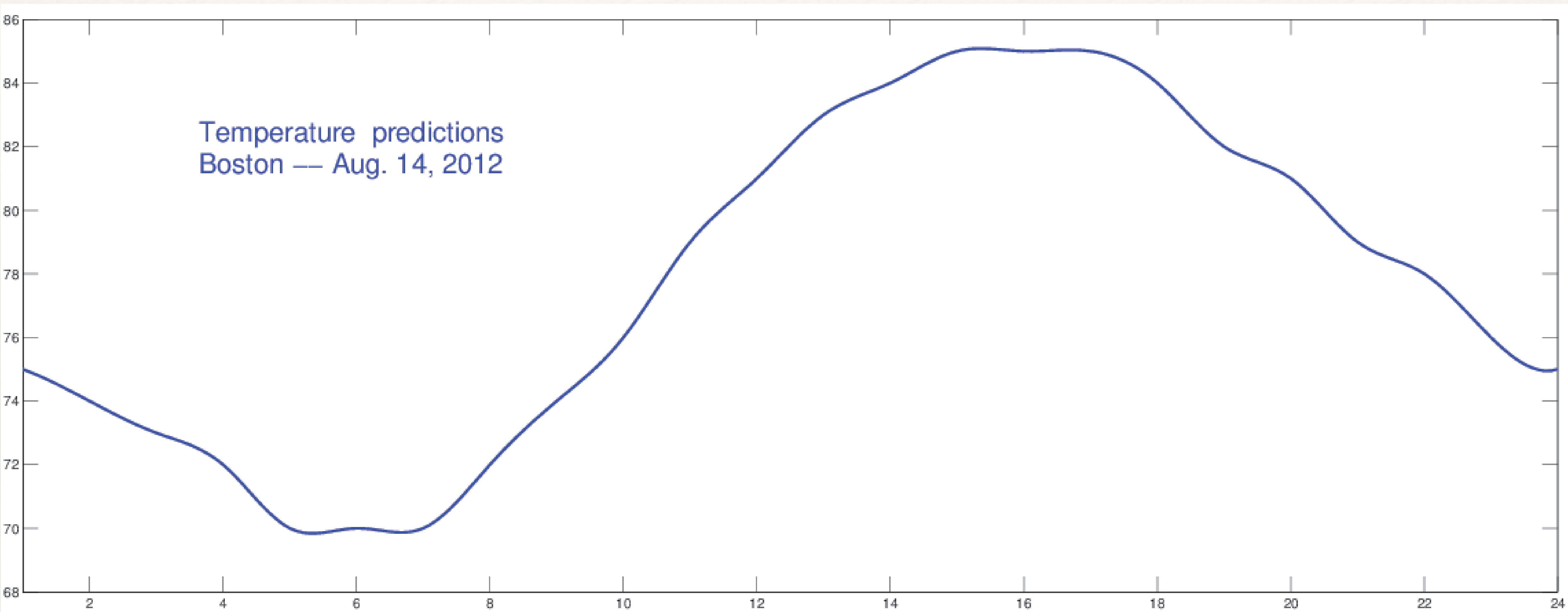
Building the Uncertainty

Roger J-B Wets, Mathematics, University of California, Davis
& David Woodruff, G.S. Management, University of California, Davis

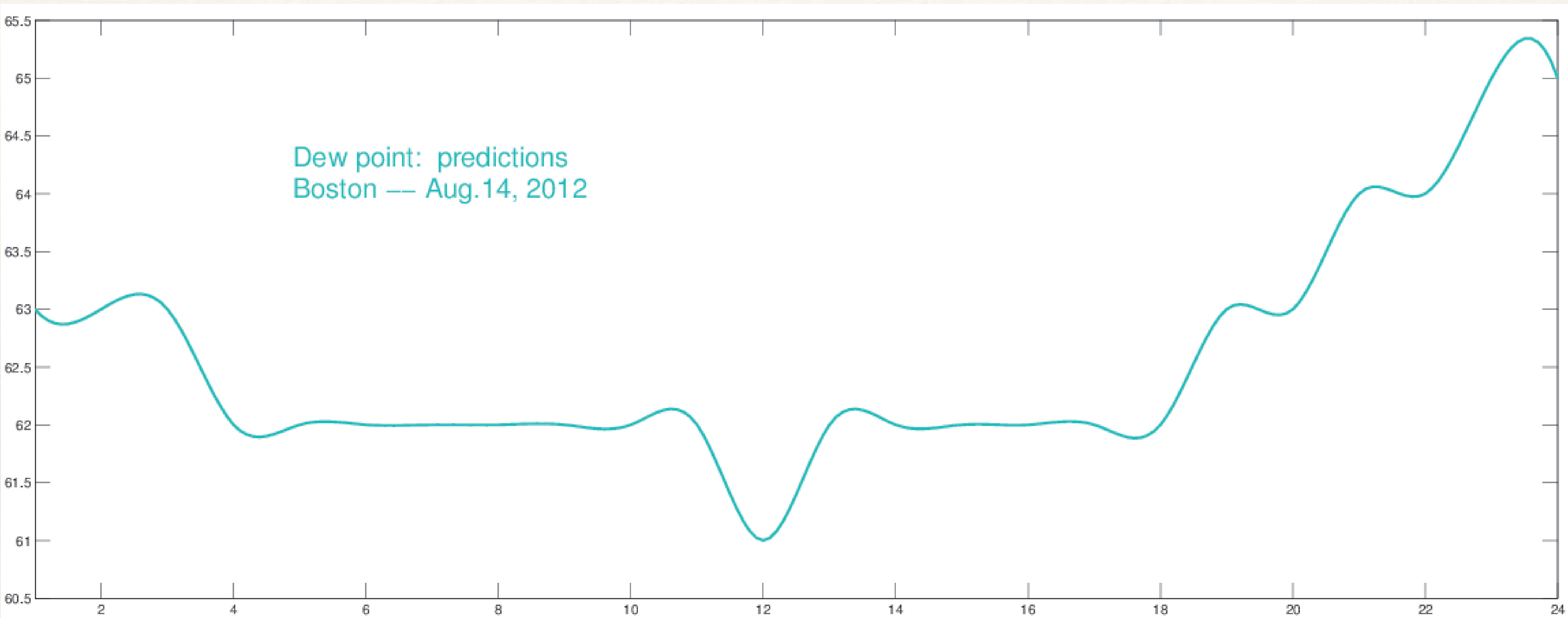
Workshop "Statistics for Optimization", Paris, Summer 2014

Forecasting Load & Renewables

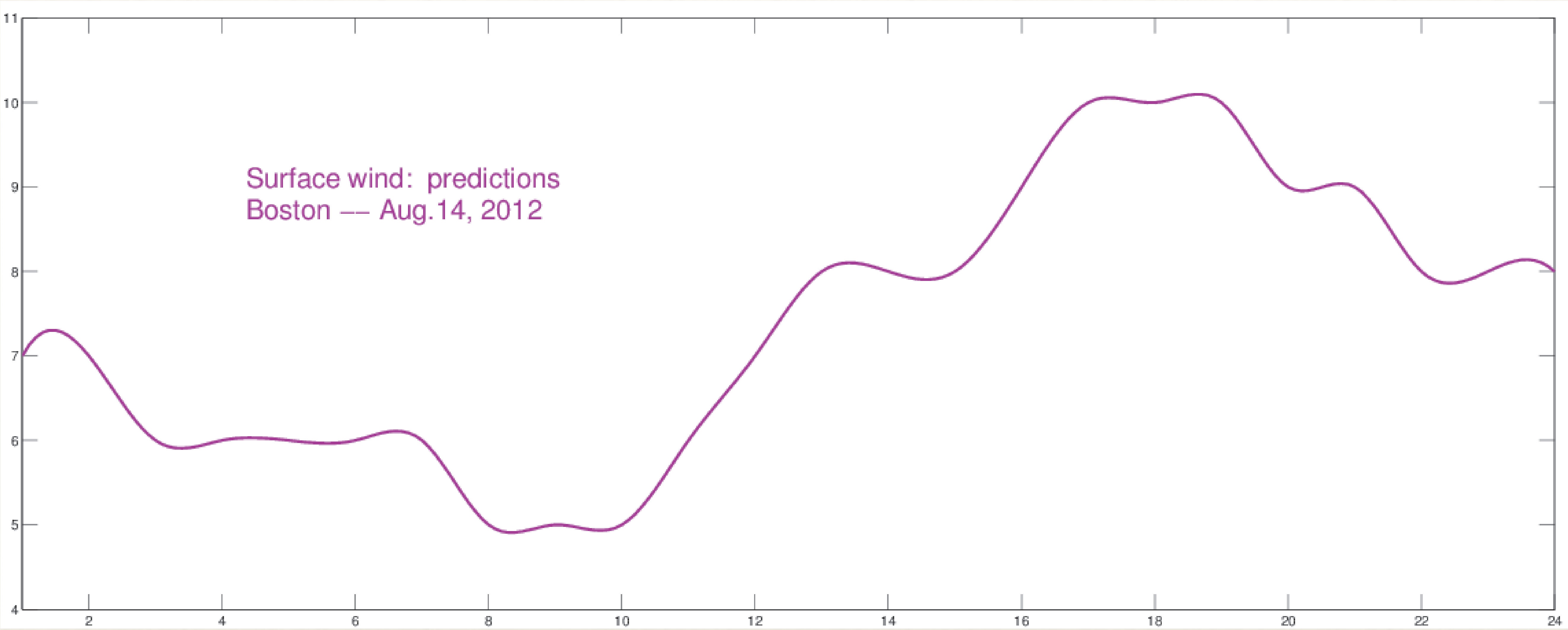
from predictions on day D-1 to load forecasts on day D



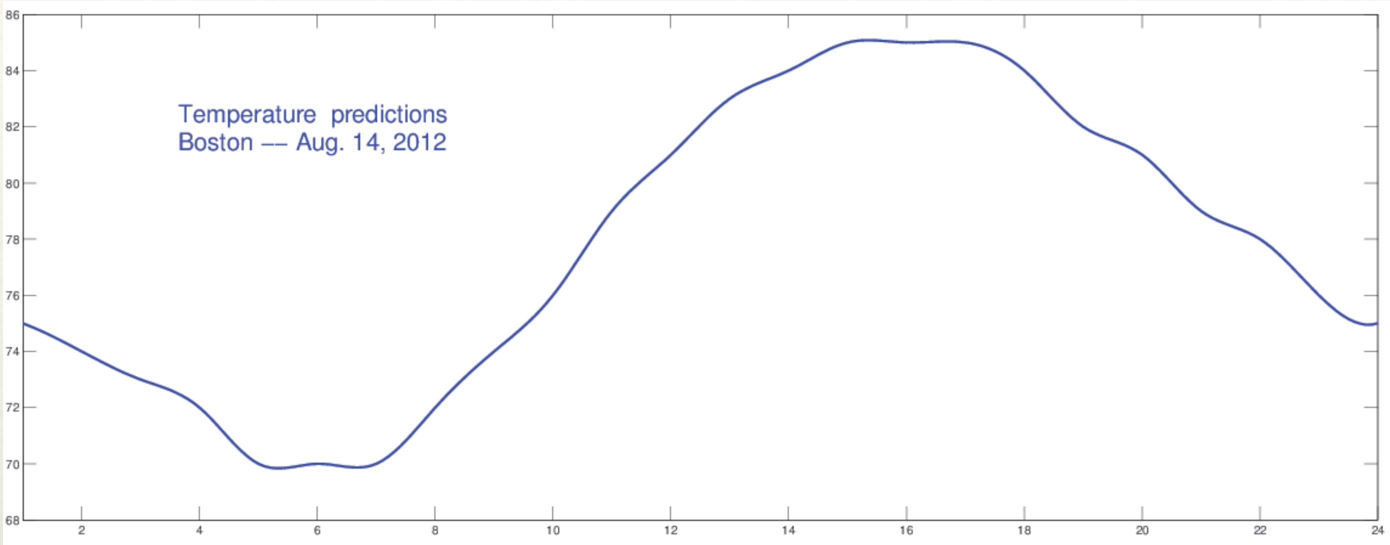
from predictions on day D-1 to load forecasts on day D



from predictions on day D-1 to load forecasts on day D



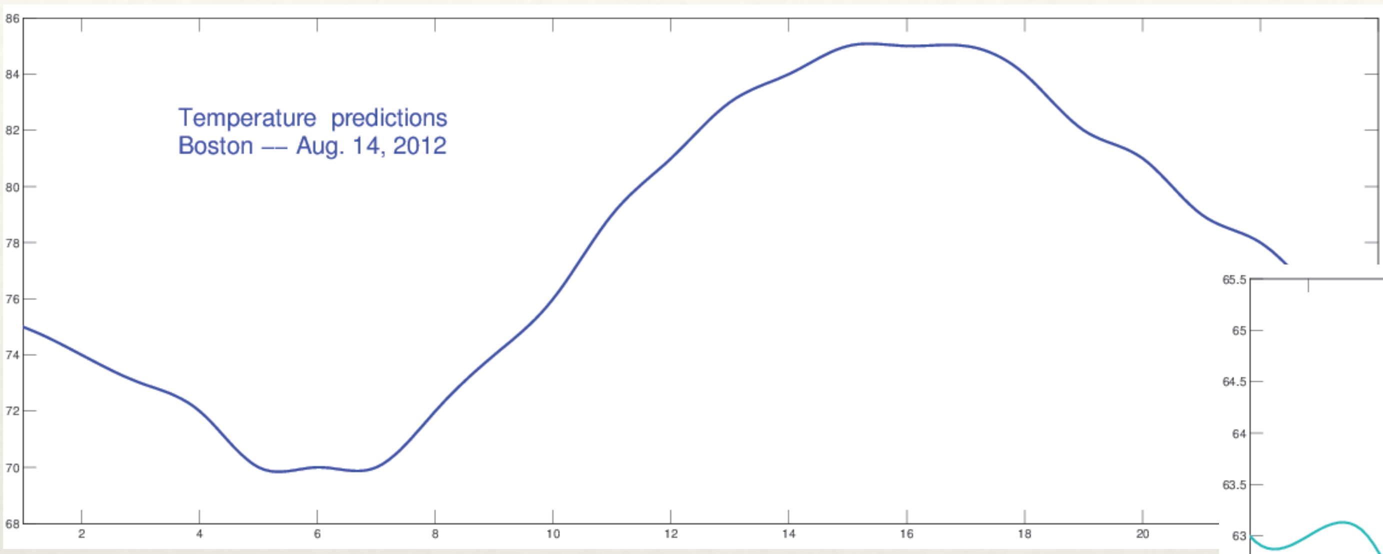
from predictions on day D-1 to load forecasts on day D

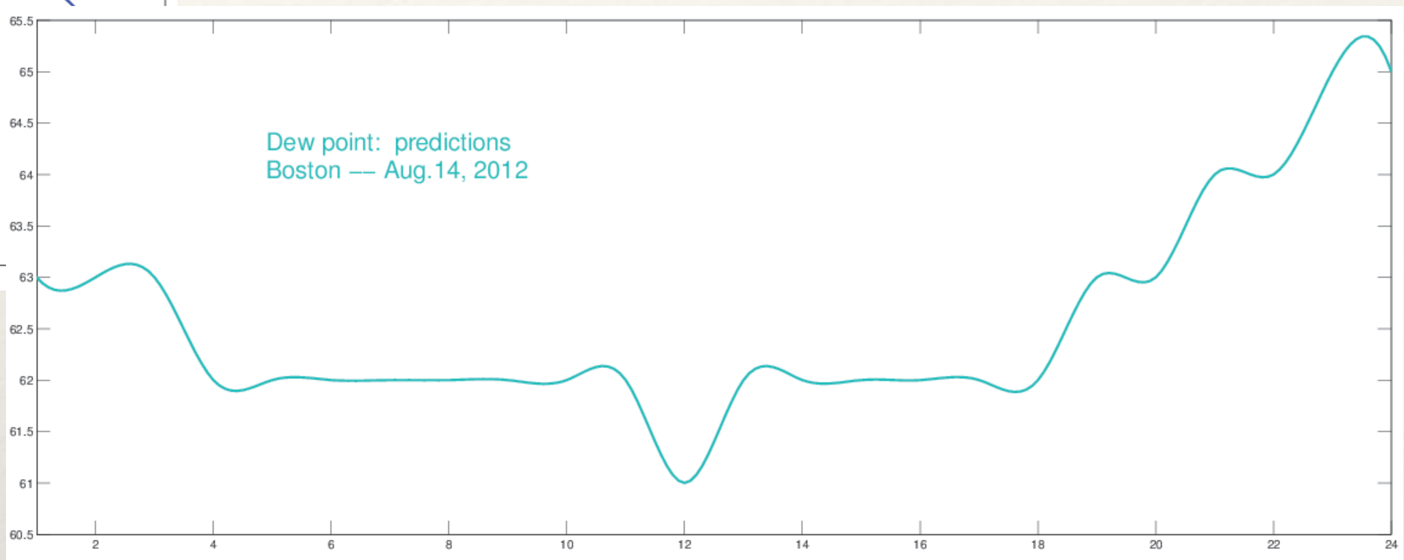


THE DATA

from predictions on day D-1 to load forecasts on day D

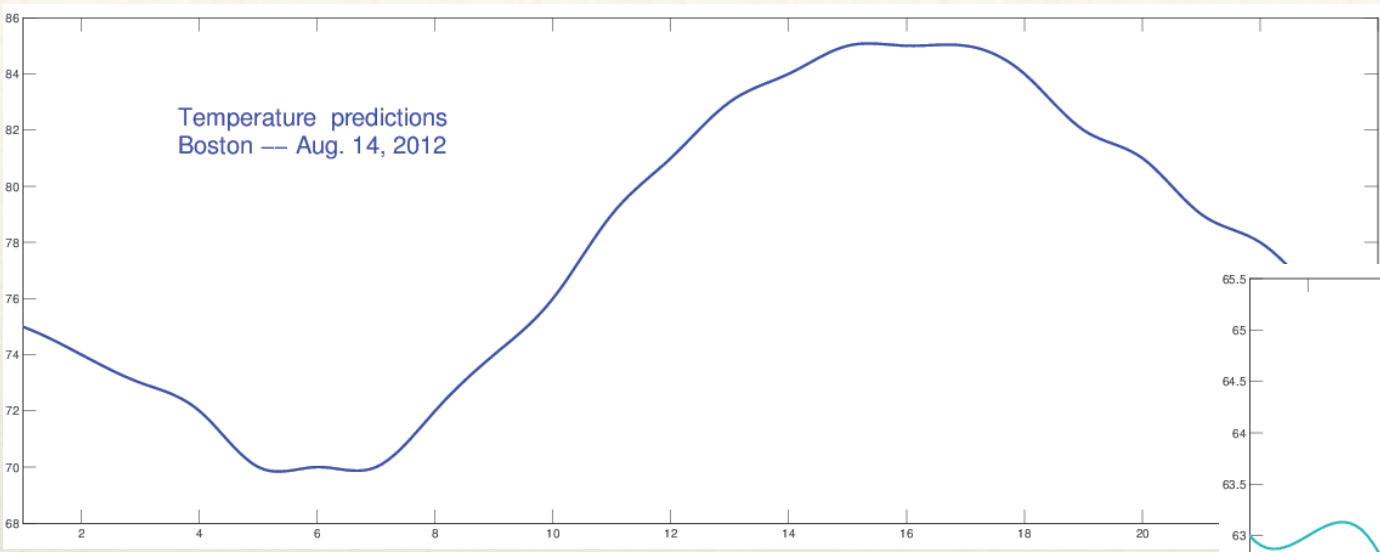
THE DATA

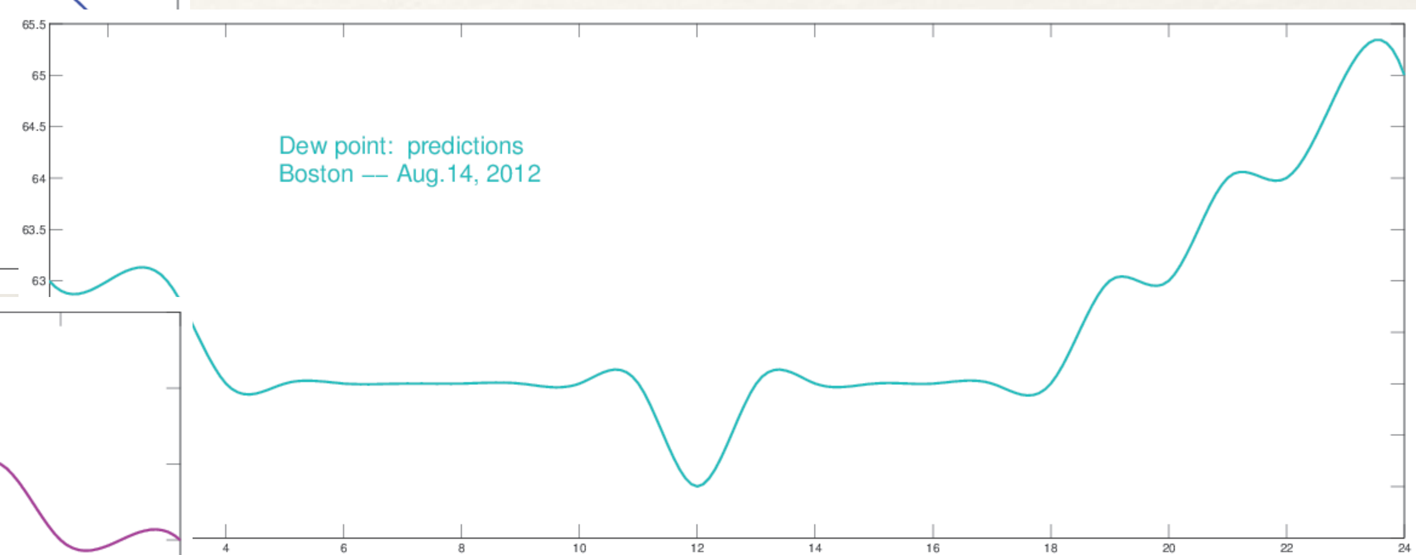


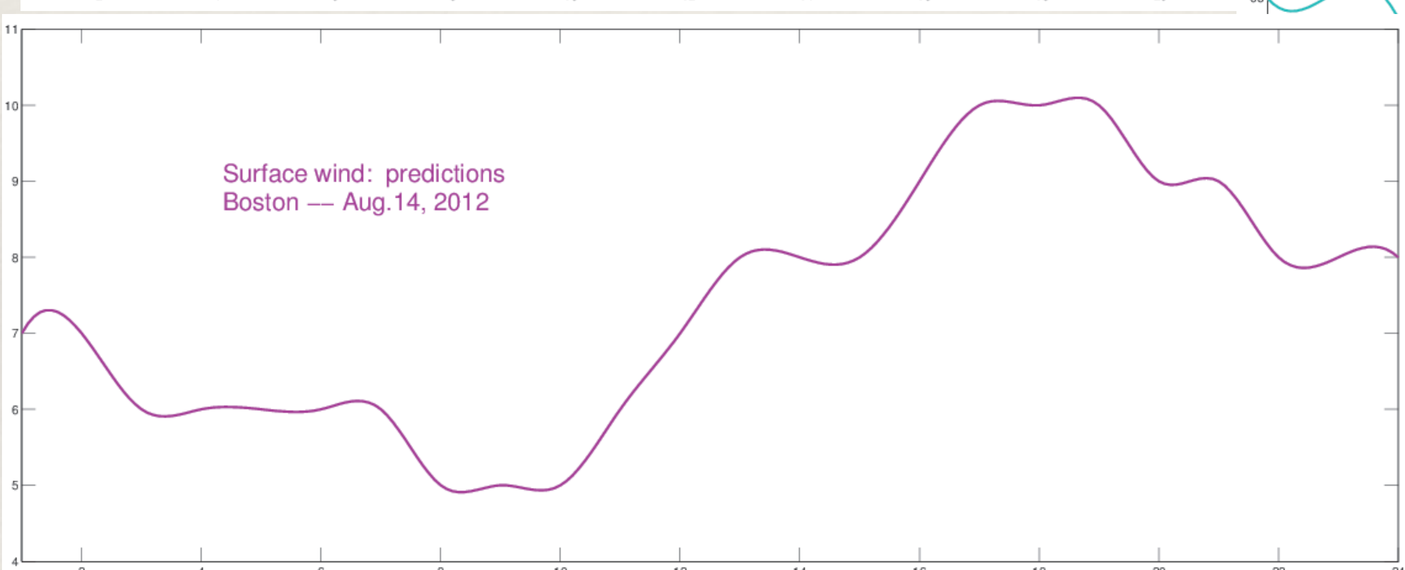


from predictions on day D-1 to load forecasts on day D

THE DATA

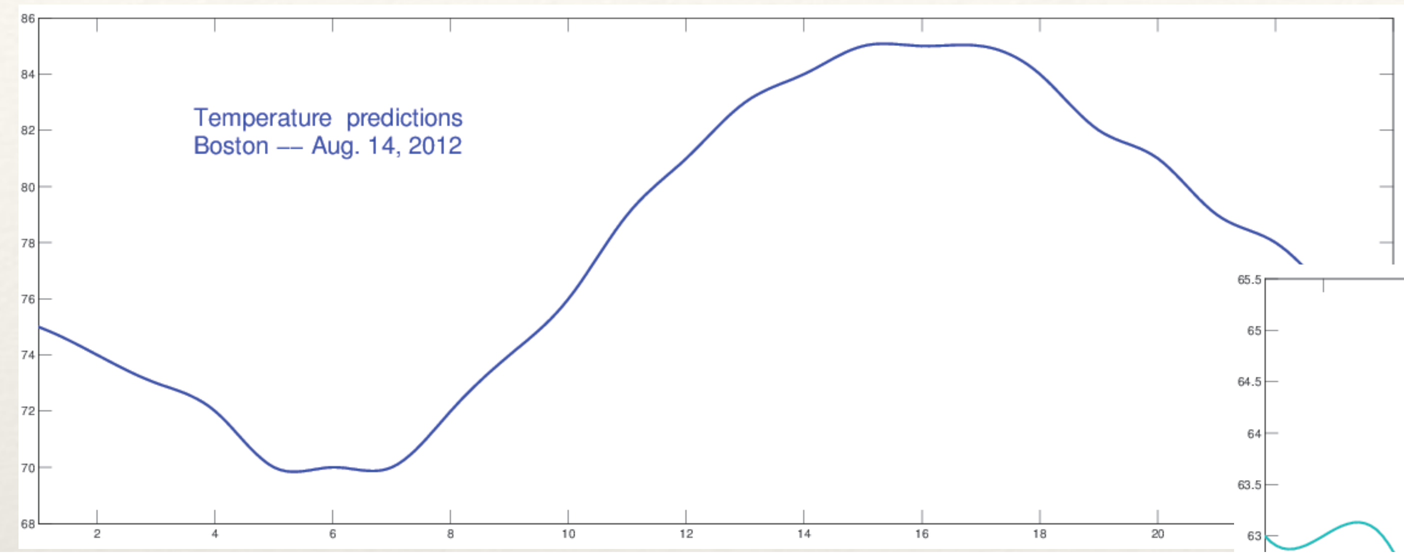


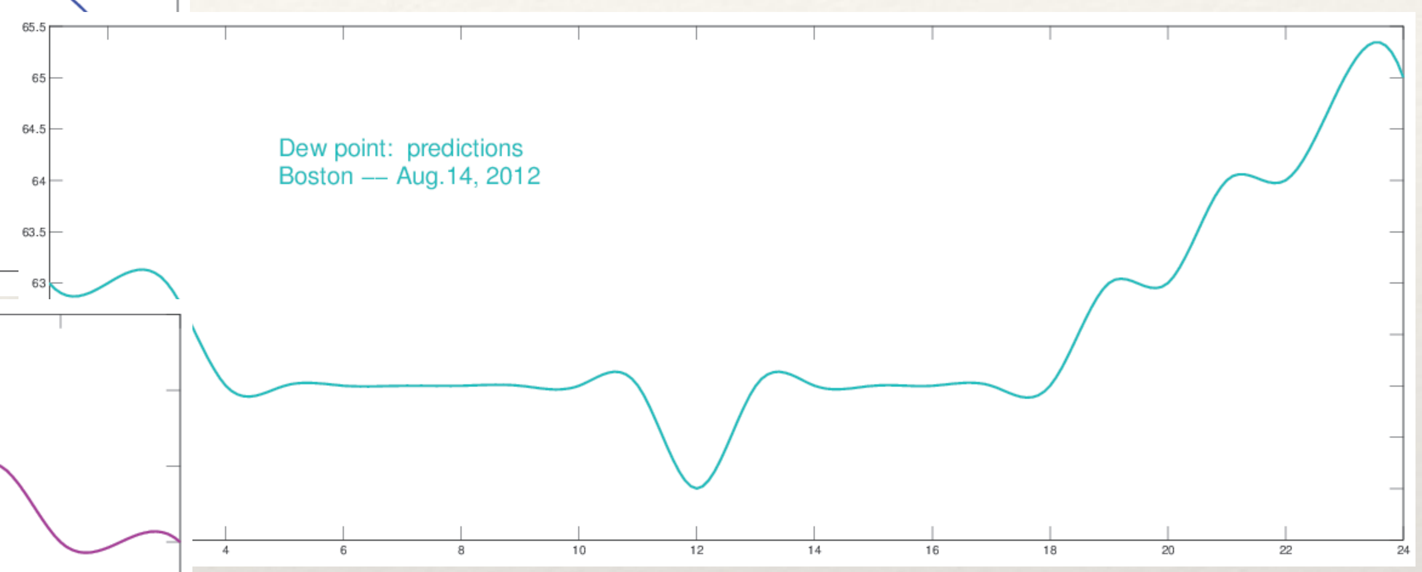


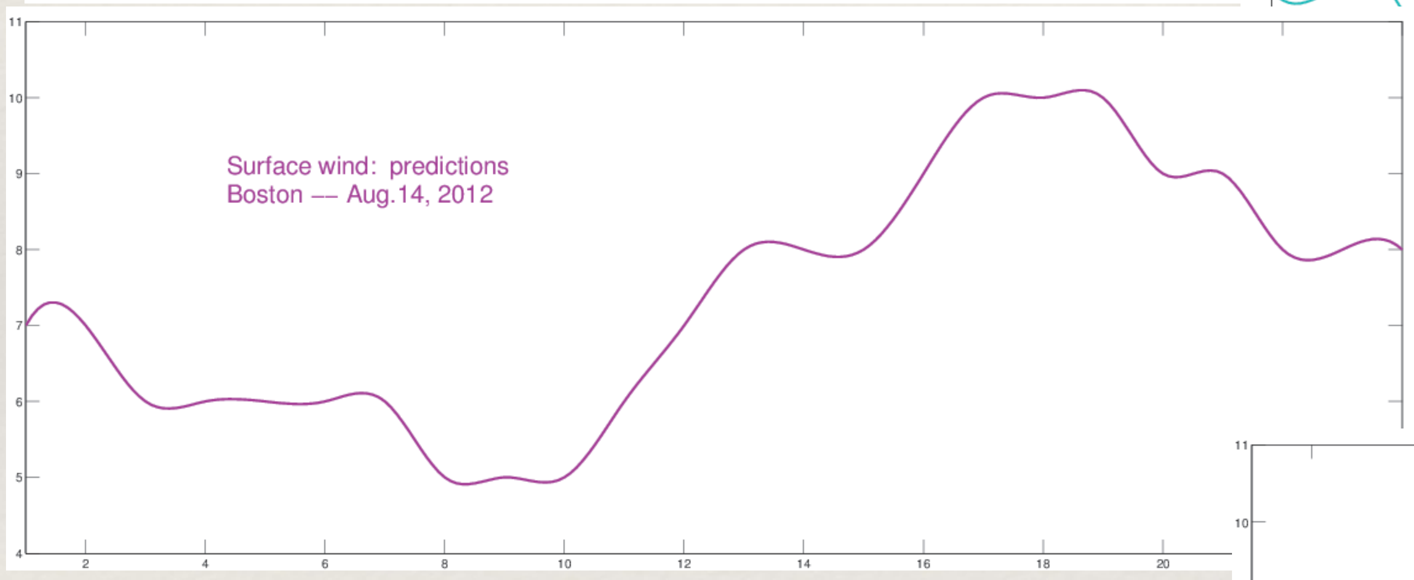


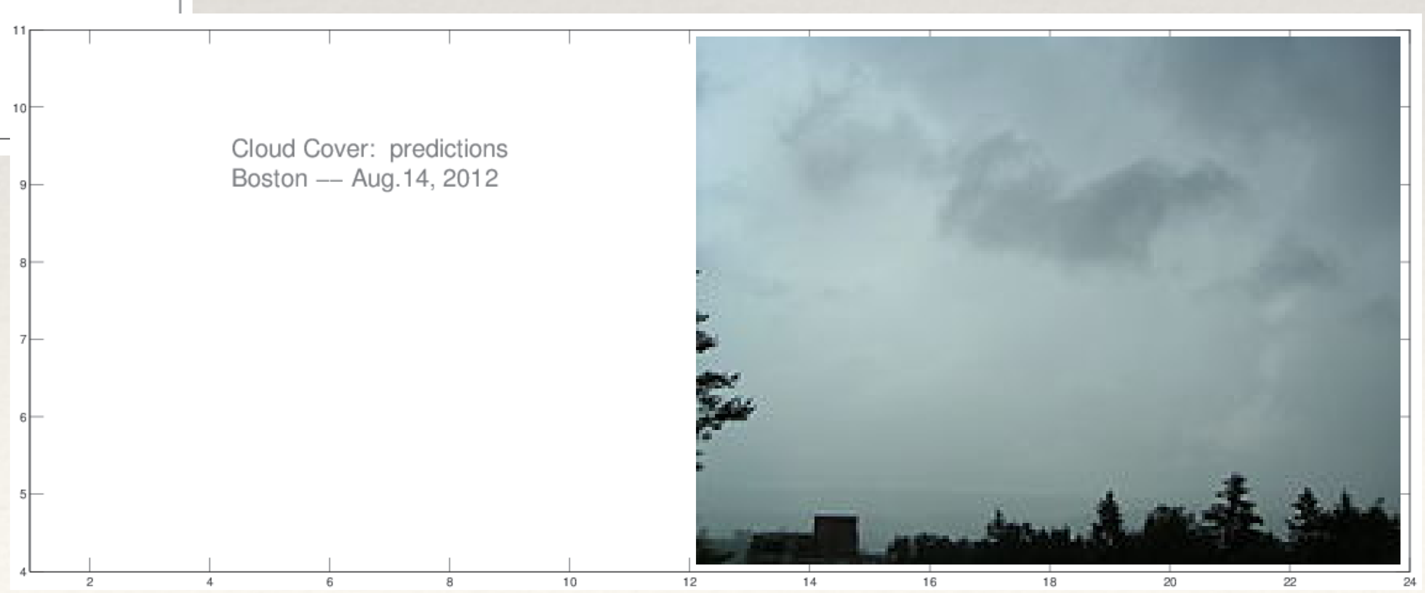
from predictions on day D-1 to load forecasts on day D

THE DATA







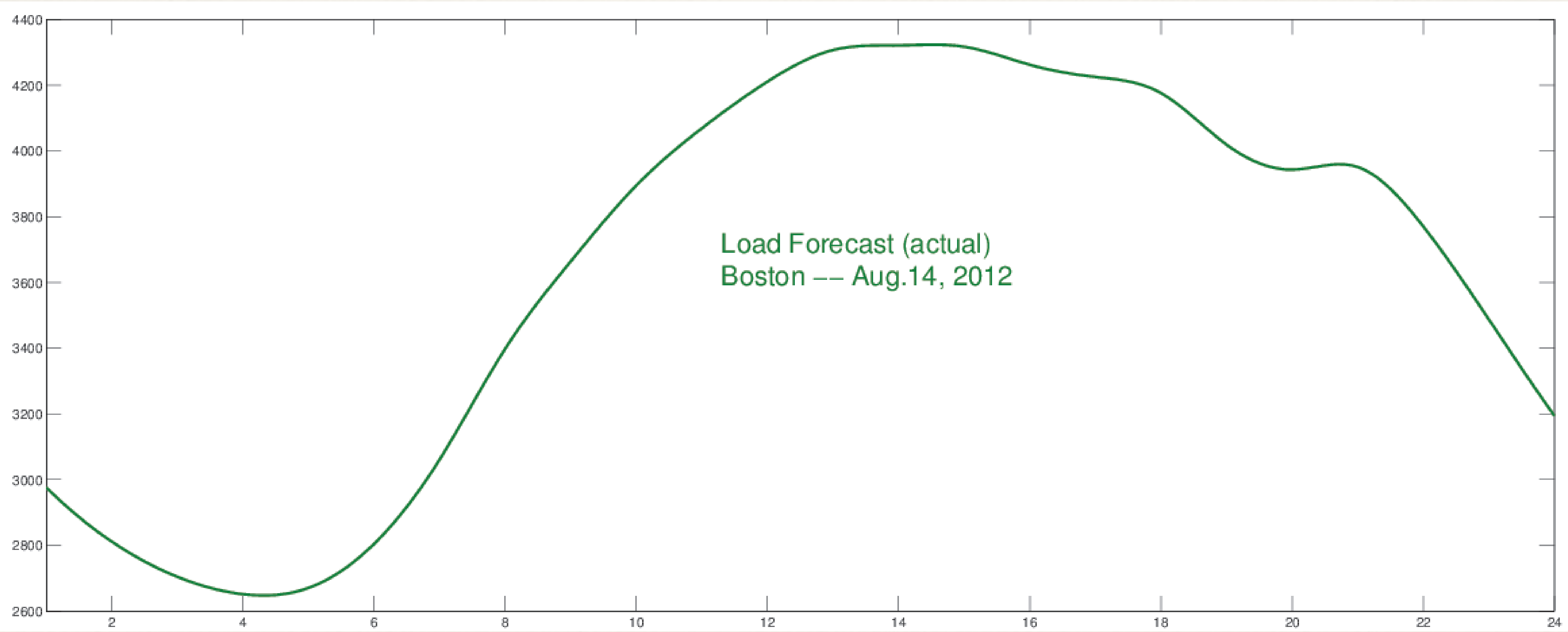


NOAA: ACTUALS

… to load on day D

to be delivered-load: $l(t)$

$$= \text{fcn}\Big(\text{temp}(\tau \le \text{t}), \text{dewpt}(\tau \le \text{t}), \text{clcover}(\tau \le \text{t}), \text{wind}(\tau \le \text{t})\Big), \ \ t \le 24$$



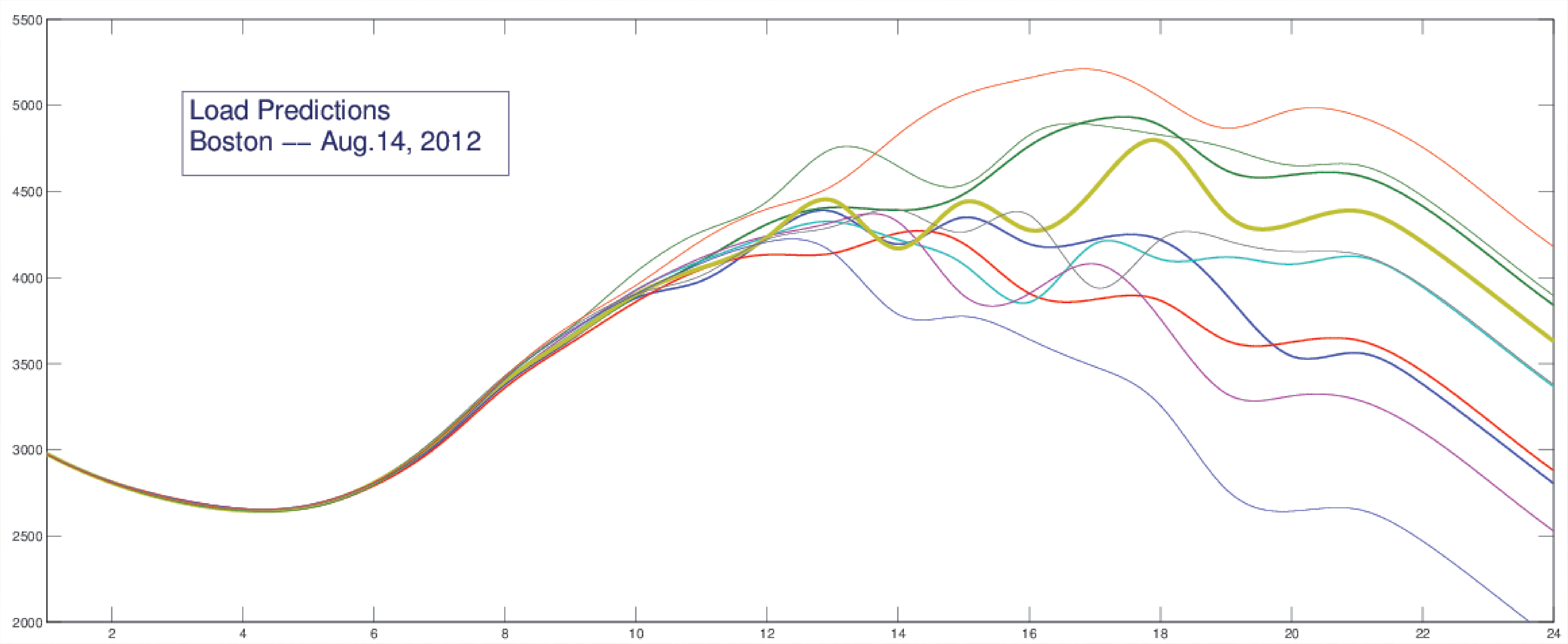
… to load on day D

to be delivered-load: $l(t)$

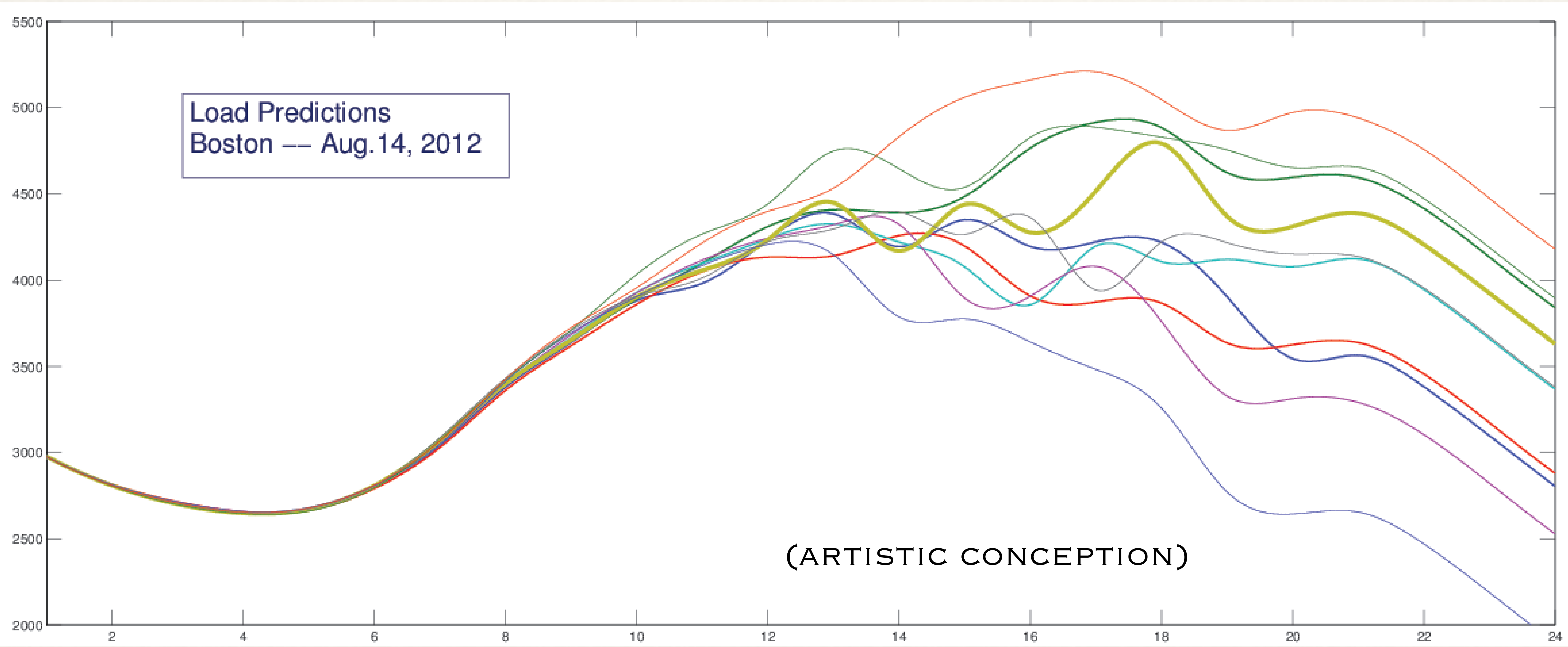
$$= \text{fcn}\Big(\text{temp}(\tau \leq \text{t}), \text{dewpt}(\tau \leq \text{t}), \text{clcover}(\tau \leq \text{t}), \text{wind}(\tau \leq \text{t})\Big), \;\; t \leq 24$$

BUT THAT WOULDN'T CAPTURE THE UNCERTAINTY!
ONE WOULD EXPECT:

“Realistic” Forecasts



“Realistic” Forecasts



Troubling Issues

Troubling Issues

- weather prediction @ 11 a.m. better @ 11 p.m. ... but too late!

Troubling Issues

- weather prediction @ 11 a.m. better @ 11 p.m. ... but too late!
- surface wind =>? power wind

Troubling Issues

- weather prediction @ 11 a.m. better @ 11 p.m. ... but too late!
- surface wind =>? power wind
- cloud cover (no historical prediction data) -- only actuals are available

Troubling Issues

- weather prediction @ 11 a.m. better @ 11 p.m. ... but too late!
- surface wind =>? power wind
- cloud cover (no historical prediction data) -- only actuals are available
- model to be used for the stochastic load predictions model: SDE, time series, ??? all inappropriate

Stochastic Load Process ⟵ Scenarios

Stochastic Load Process ⟵ Scenarios

a) segmentation: season + day characteristics

Stochastic Load Process ⟵ Scenarios

a) segmentation: season + day characteristics
b) functional regression for given segment

Stochastic Load Process ⟼ Scenarios

a) segmentation: season + day characteristics
b) functional regression for given segment
c) hourly distribution of errors per segment

Stochastic Load Process ⟵ Scenarios

a) segmentation: season + day characteristics
b) functional regression for given segment
c) hourly distribution of errors per segment
HOW THIS IS CARRIED OUT (next lecture)

Stochastic Load Process ⟼ Scenarios

a) segmentation: season + day characteristics
b) functional regression for given segment
c) hourly distribution of errors per segment
HOW THIS IS CARRIED OUT (next lecture)
d) conditional distribution of errors => process

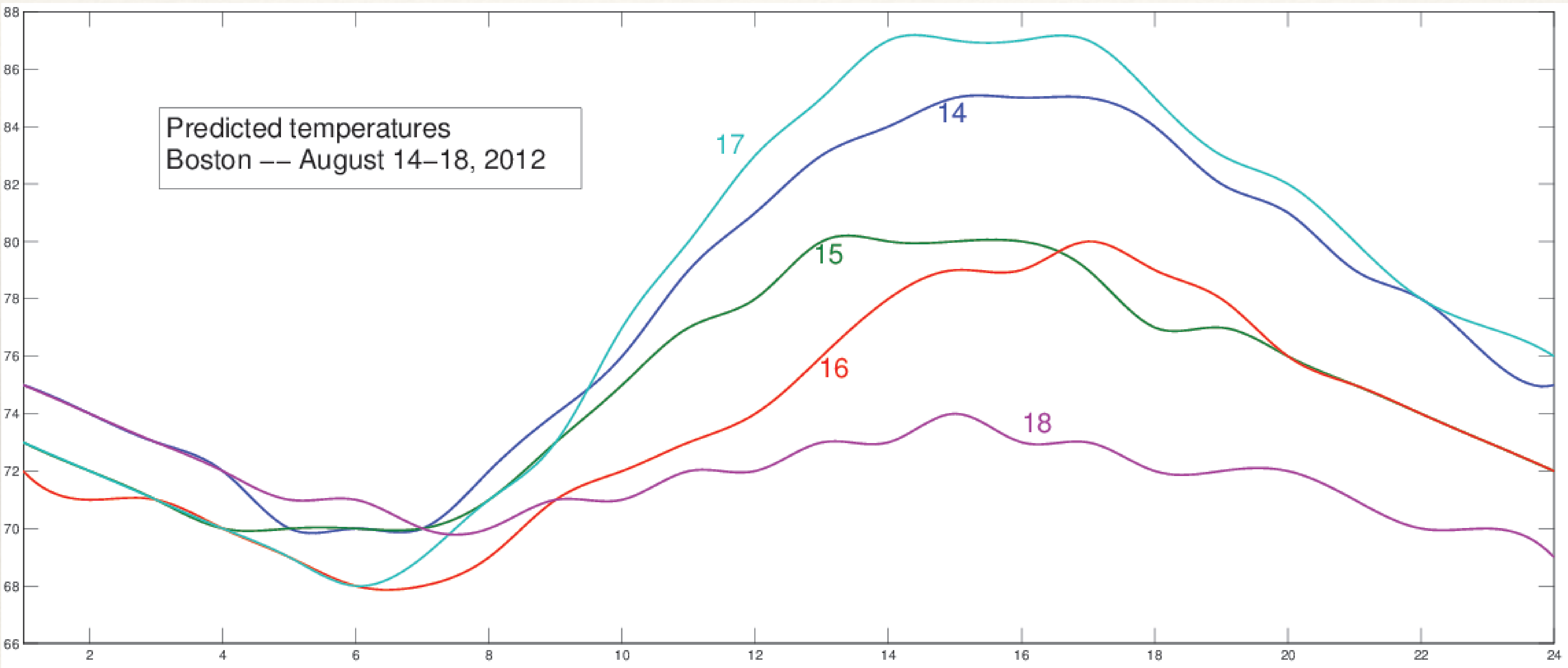
Stochastic Load Process ⟵ Scenarios

a) segmentation: season + day characteristics
b) functional regression for given segment
c) hourly distribution of errors per segment
HOW THIS IS CARRIED OUT (next lecture)
d) conditional distribution of errors => process
e) discretization of the process => scenarios

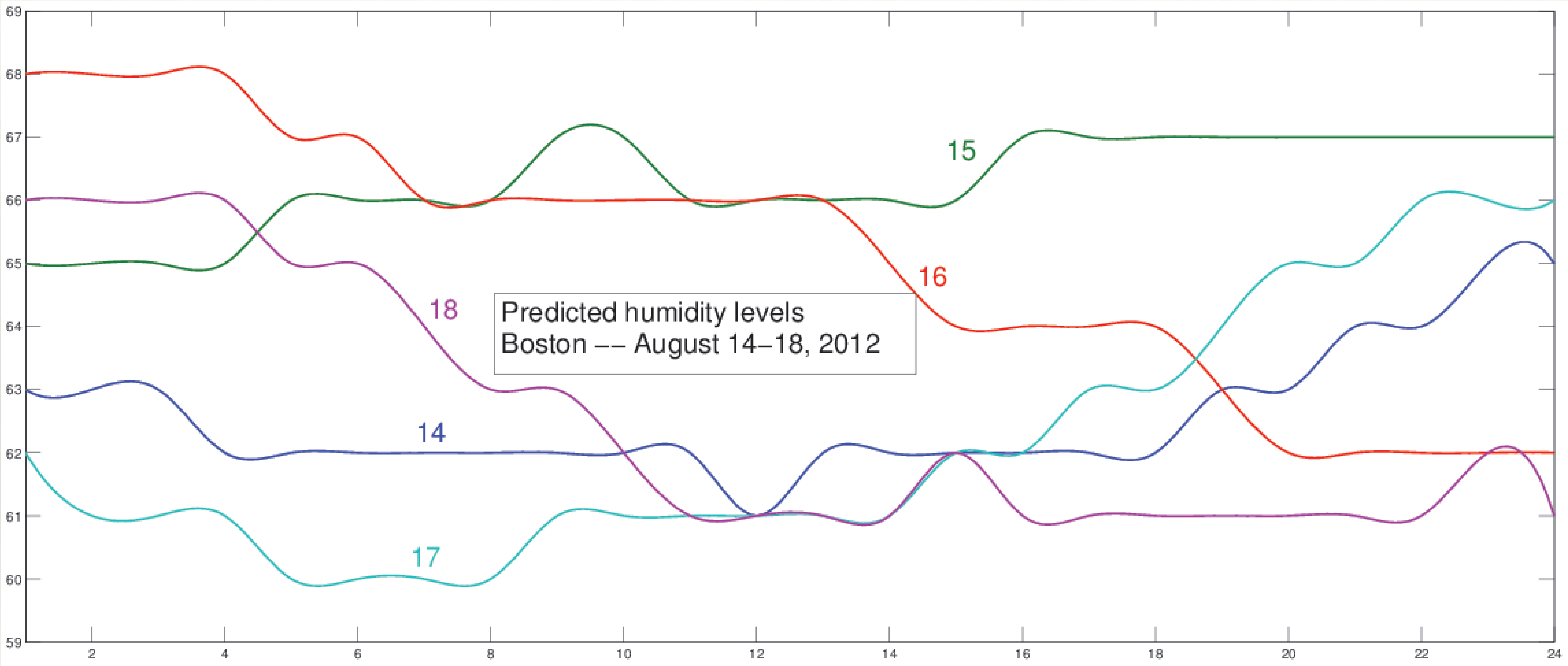
Segmentation

- ~ similars, analogs (± standard)
  to enrich data: Wednesday rule, zone rule?
- seasons: (factor analysis, ‘heuristics’)
  - ± spring & fall : temperature
  - winter: temperature & cloud cover
  - summer: temperature & dew point
- wind power (at present): handled independently
  based on 3TIER analogs
  total load ≈ load scenario - wind power scenario

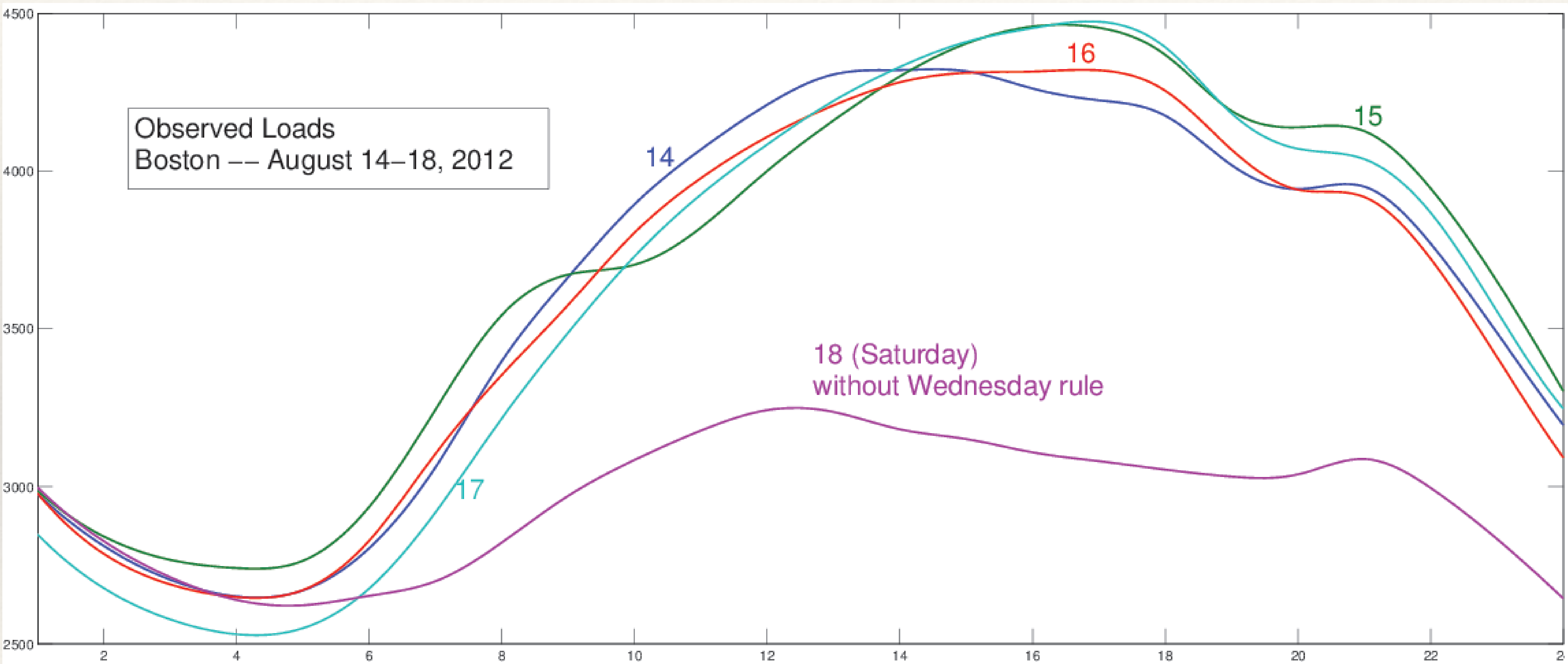
Summer segment “#1”

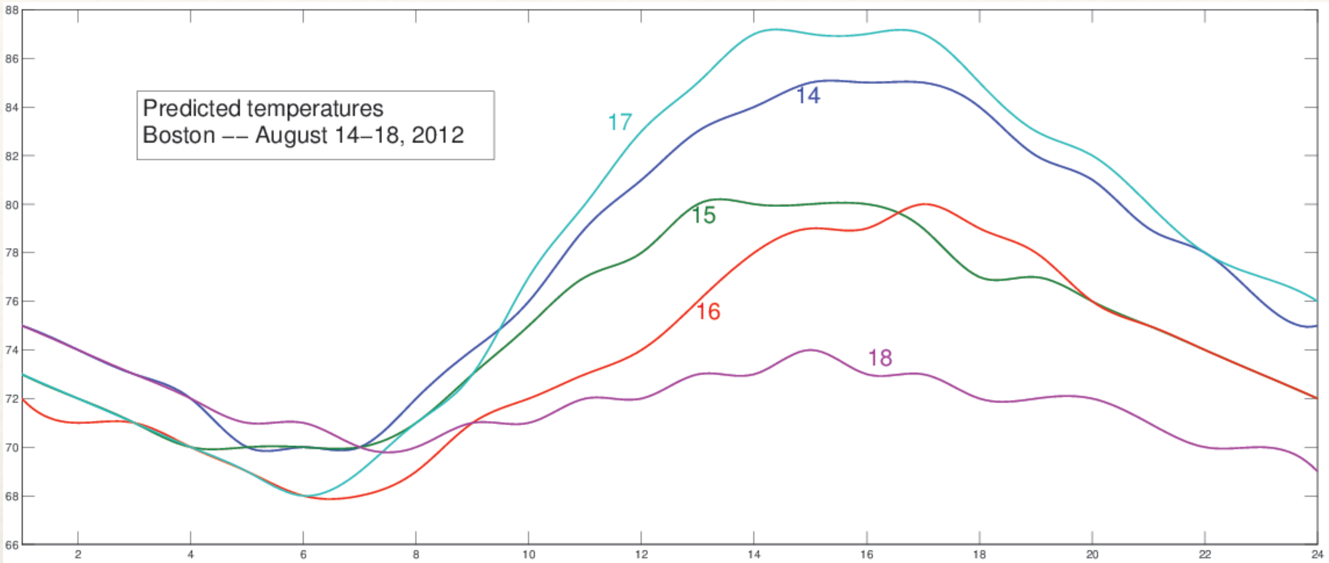


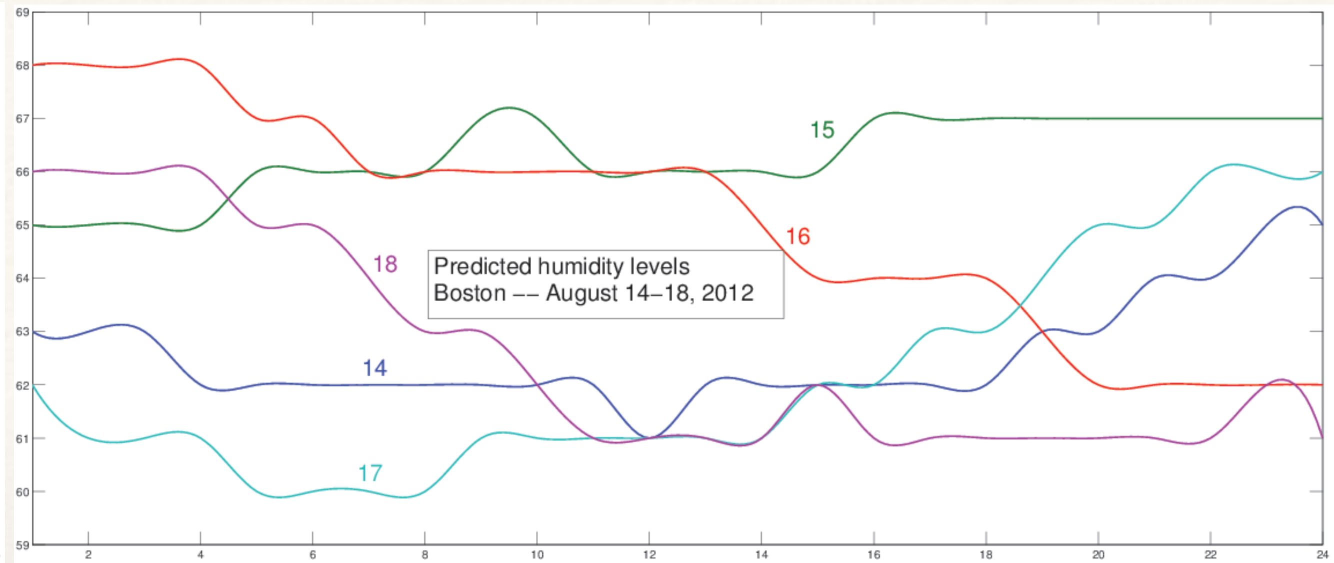
Summer segment “#1”



Summer segment “#1”

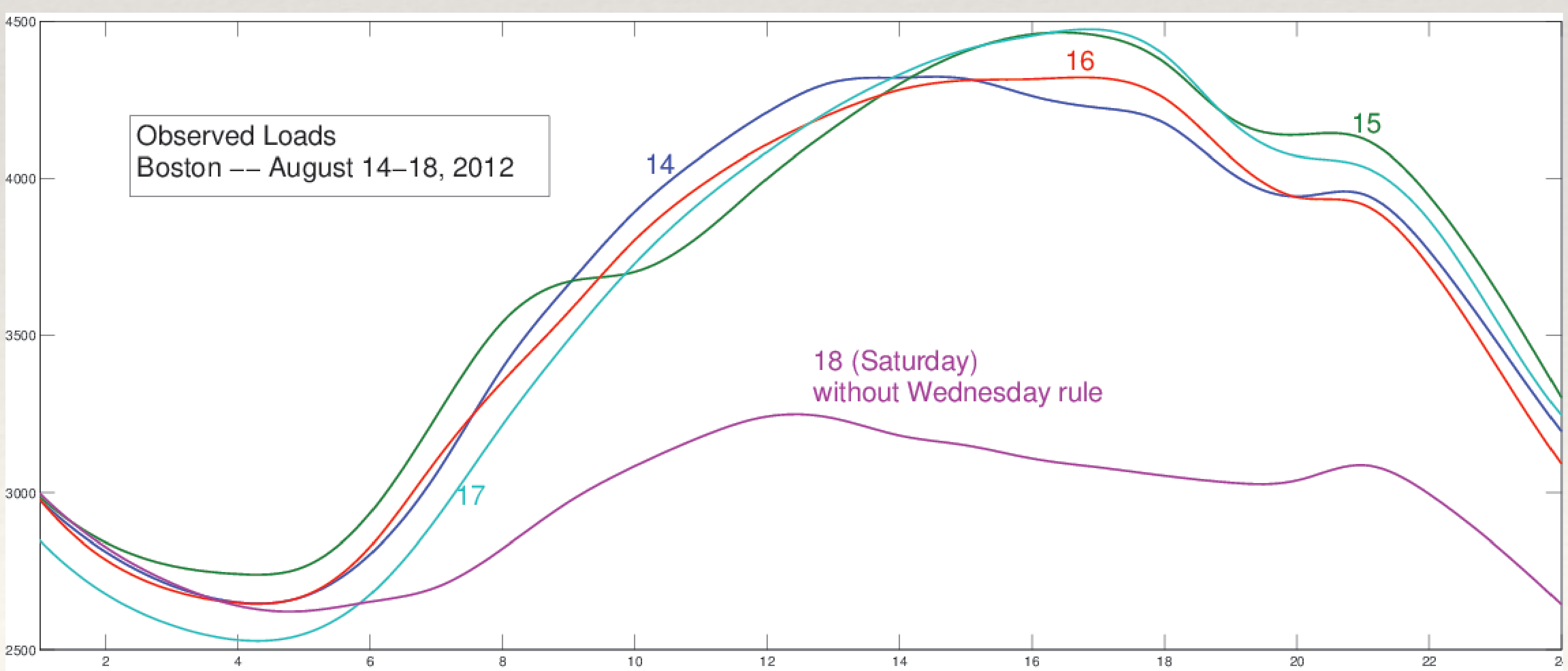






from day d-1 $\Rightarrow$ possible load on day d $\quad d = 14, ...,18$

1. regression(temp. curve, humid. curve) $\Rightarrow$ 'expected' load curve
2. get distribution of the errors (hourly, ... at any time)



The Regression Problem

find a function r that minimizes errors (with respect to $\|\square\|$)

$$\sum\nolimits_{\text{days d in segment}} \sum\nolimits_{\text{hours h in day}} \left\| r\left((tmp_{d,h}, hum_{d,h}) \right) - \text{load}_{d,h} \right\|$$

an infinite dimensional problem!

Our approach: rely on 2-dimensional epi-splines ("innovation")

- epi-splines approximate with arbitrary accuracy 'any' function
- epi-splines are completely determined by a finite # of parameters
- allows (via constraints) to include 'soft' (non-data) information

The Errors Distributions

Given segment # and associated r, for fixed hour h

$$e_{d,h} = \text{load}_{d,h} - r\big((tmp_{d,h}, hum_{d,h})\big),\ d \in \text{segment } \#$$

$\Rightarrow$ estimate the density f_h of the errors (at h in segment #)

yields an overall estimate of the 'volatility' (in fact, more)

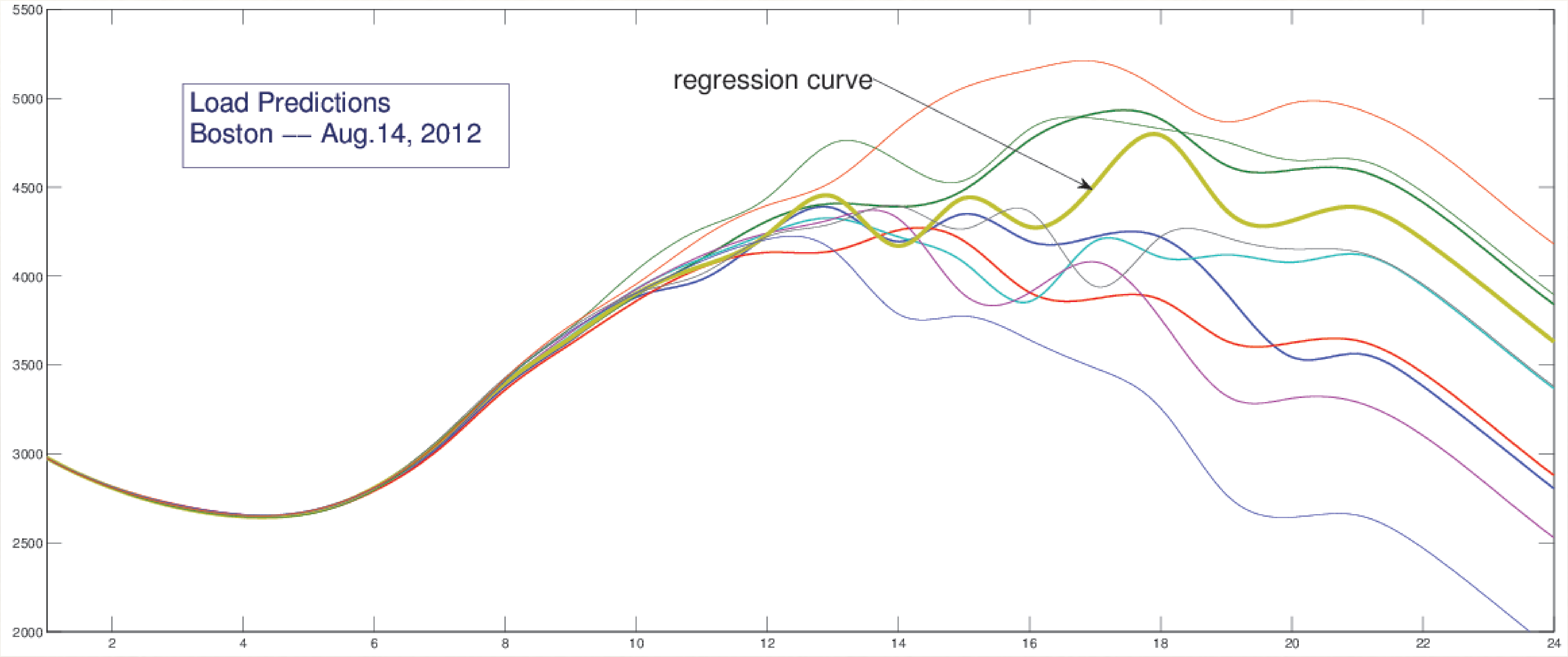
another infinite dimensional problem & data might be scarce

Our approach: estimation via exponential epi-spline (novel):

- $f_h = \exp(-s_h)$, s_h an epi-spline $\quad (\Rightarrow f_h \geq 0)$
- same properties as epi-spline, could include unimodality restriction

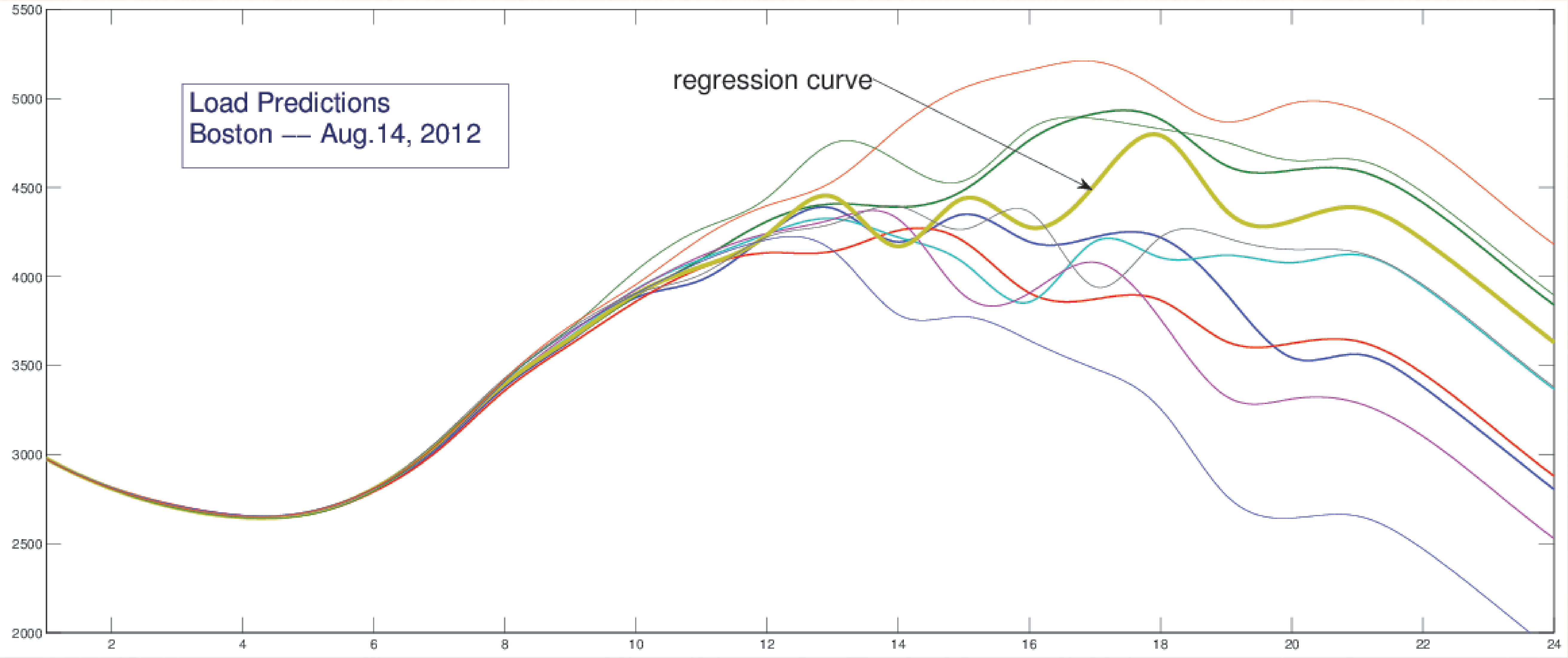
… et voilà!

regression curve & sampling from errors distribution



... et voilà!

regression curve & sampling from errors distribution

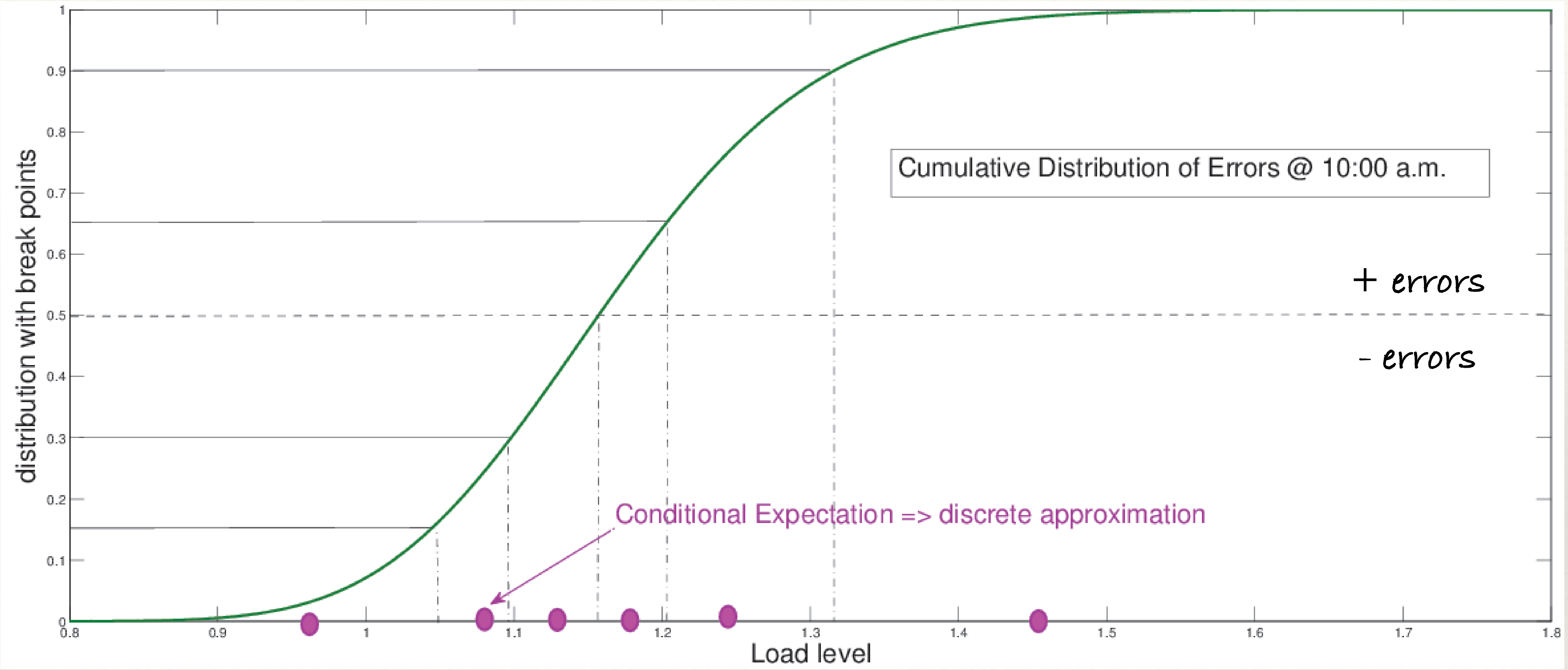


a. how many samples? $10^3, 10^5, \ldots$?

b. conditioning: @10 o'clock above or below the regression curve

actually: Building Scenario Trees

Conditioning & Discretization



a. identify all observed load curves in each sub-segment

b. for each sub-segment: re-calculate regression and errors distribution

c. repeat for each sub-segment @ (say, 1 p.m.) $\Rightarrow$ sub-sub-segment