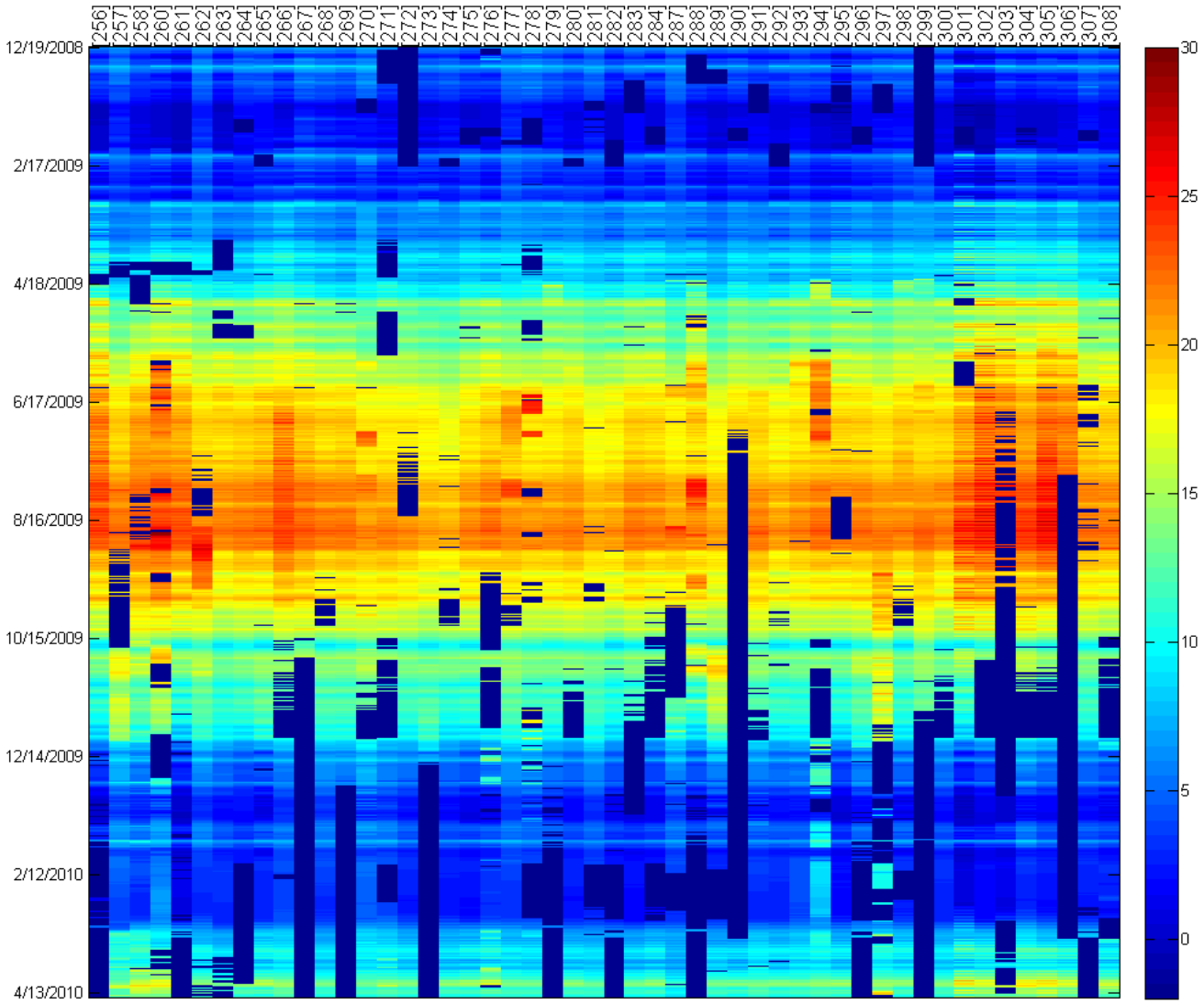
Locations


300
297
290
299
291
289
288
296
292
284
298
287
295
294
274
277
293
282
273
276
283
279
272
275
280
271
278
270
281
265
264
308
263
307
262
301
304
306
261
302
303
260
BS
286
258
305
256
269
268
267
257
266
0
15
30
60 Meters
N

Soil Temperature Dataset



Observations

- Data is
  - Correlated in time and space
  - Evolving over time (seasons)
  - Gappy (Due to failures)
  - Faulty (Noise, Jumps in values)

Temporal Gap Filling

- Data is correlated in the temporal domain
- Data shows
  - Diurnal patterns
  - Trends
- Trends are modeled using slope and intercept
- Correlations captured using functional PCA

Gap Filling (Connolly & Szalay)[+]

$$f_\lambda = \sum_i a_i e_{i\lambda}$$

Original signal representation as a linear combination of orthogonal vectors

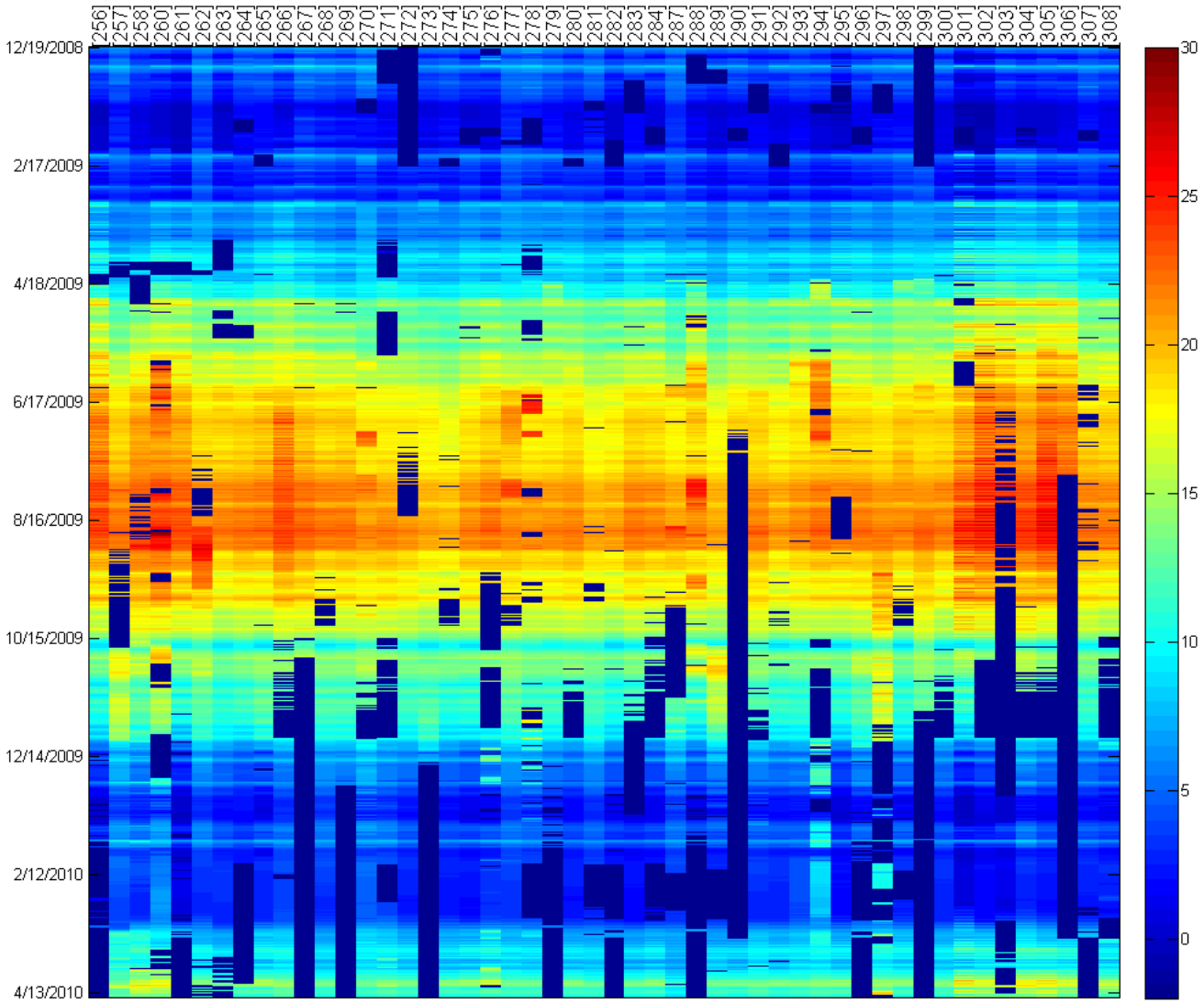
$$\chi^2 = \sum_\lambda w_\lambda (f_\lambda^o - \sum_i a_i e_{i\lambda})^2$$

Optimization function

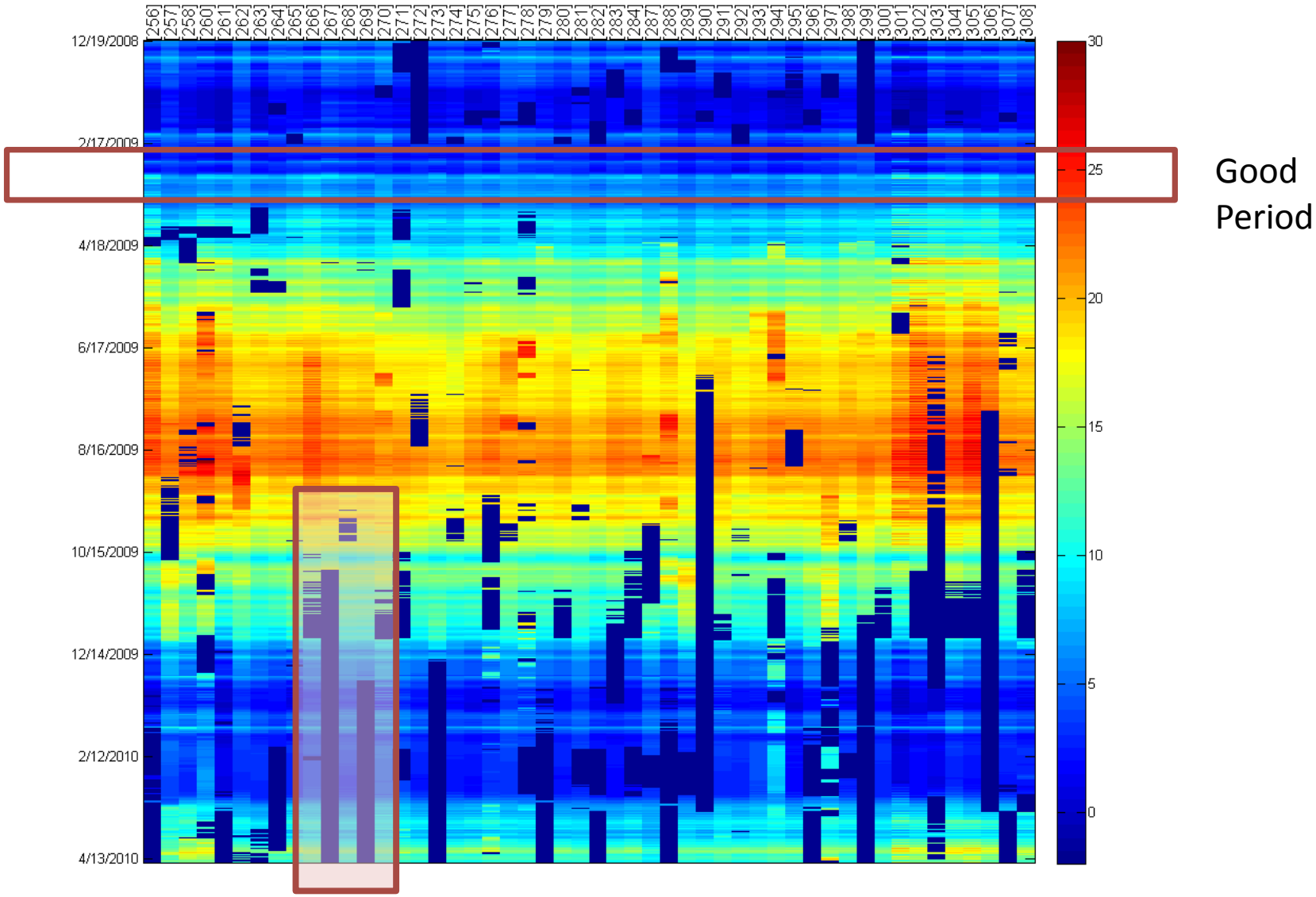
- $W_\lambda = 0$ if data is absent
- Find coefficients, a_i's, that minimize function.

[+] Connolly et al., **A Robust Classification of Galaxy Spectra: Dealing with Noisy and Incomplete Data,** http://arxiv.org/PS_cache/astro-ph/pdf/9901/9901300v1.pdf

Soil Temperature Dataset (same as slide 2)



Gap-filling using Daily Basis



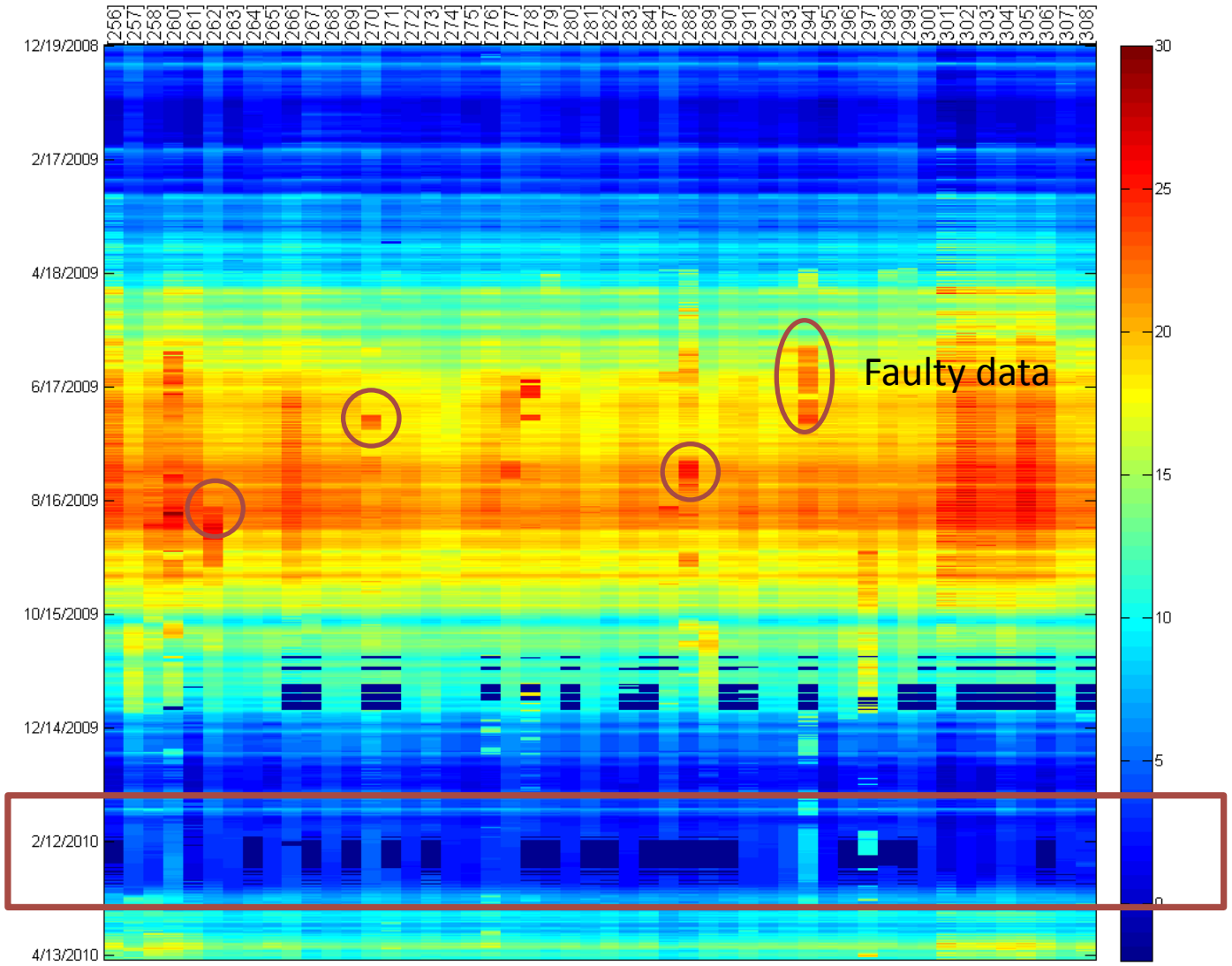
Only 2% of the data could be filled using daily model

Observations

- Hardware failures causes entire “days” of data to be missing
  - Temporal method (only using temporal correlation) is less effective

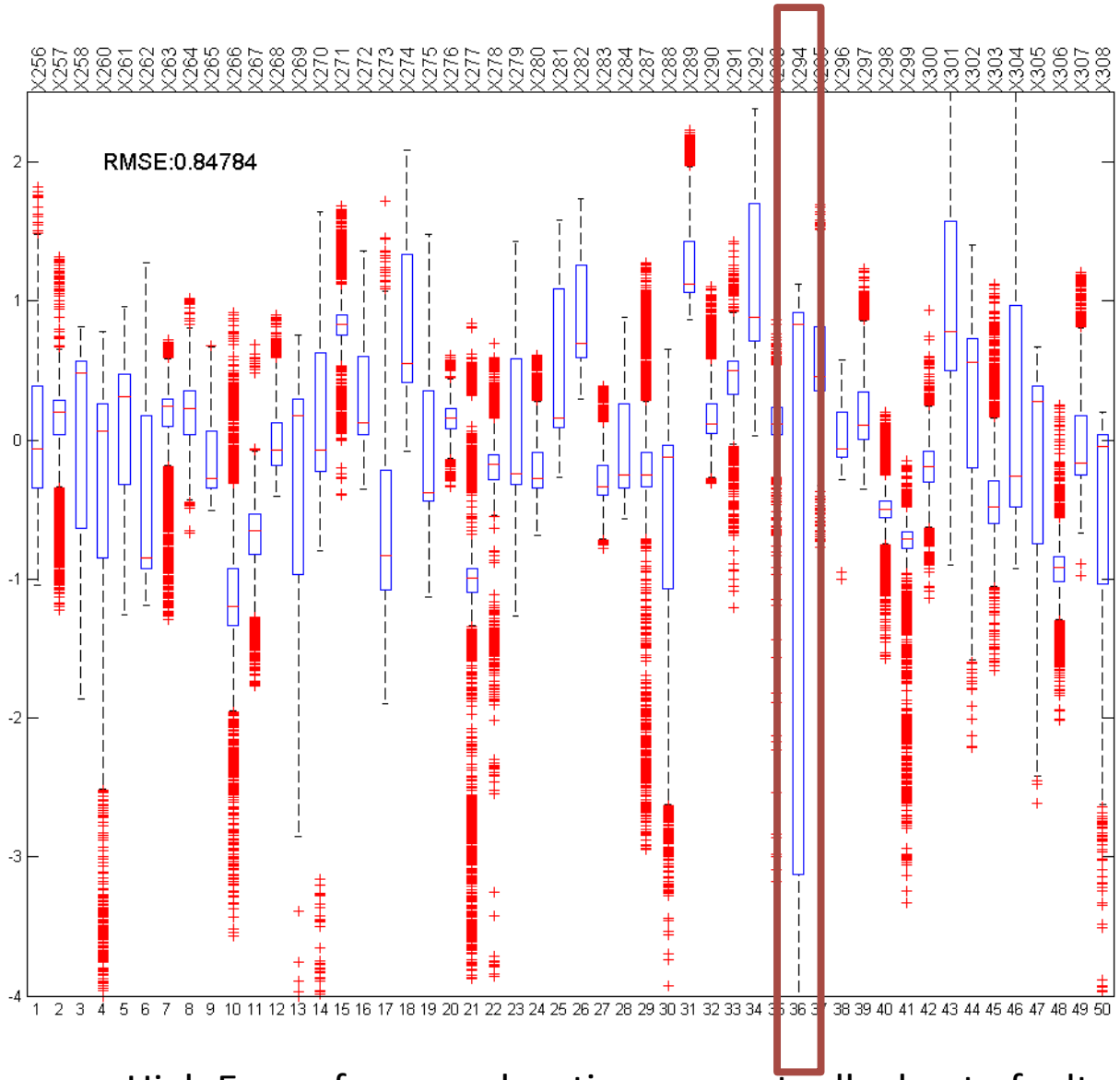
- If we look in the spatial domain,
  - For a given time, at least > 50% of the sensors are active
  - Use the spatial basis instead.
  - Initialize spatial basis using the period marked “good period” in previous slide

Gap-filling using sensor correlations



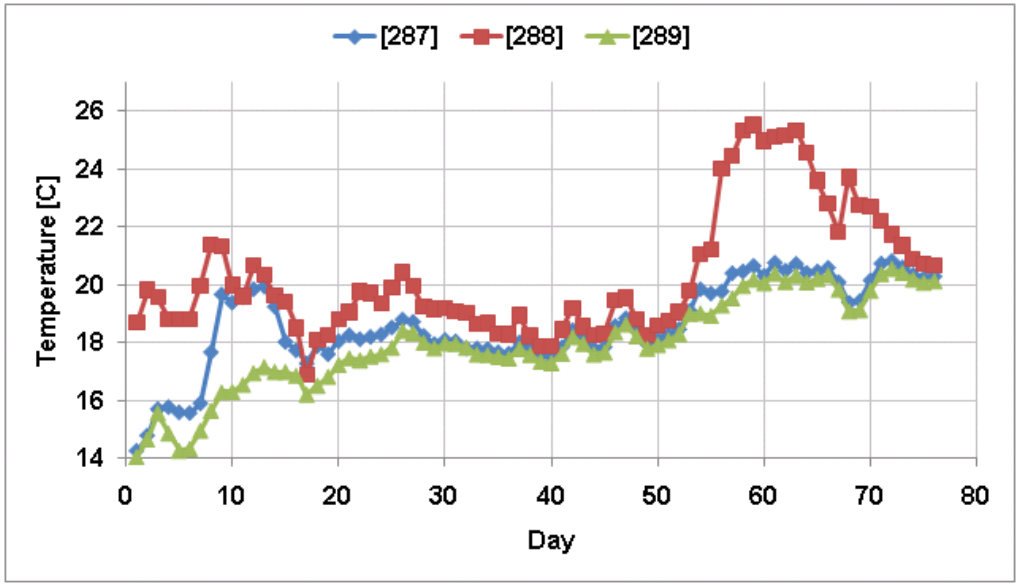
Too many missing readings for this period (Snow storm!)

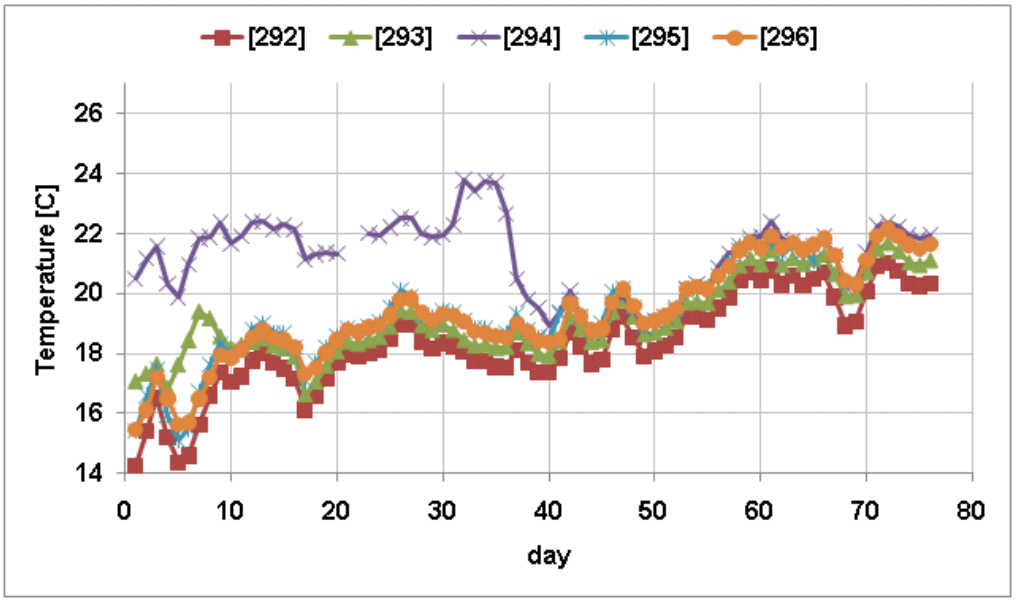
Accuracy of gap-filling



High Errors for some locations are actually due to faults
Not necessarily due to the gap-filling method

Examples of Faults



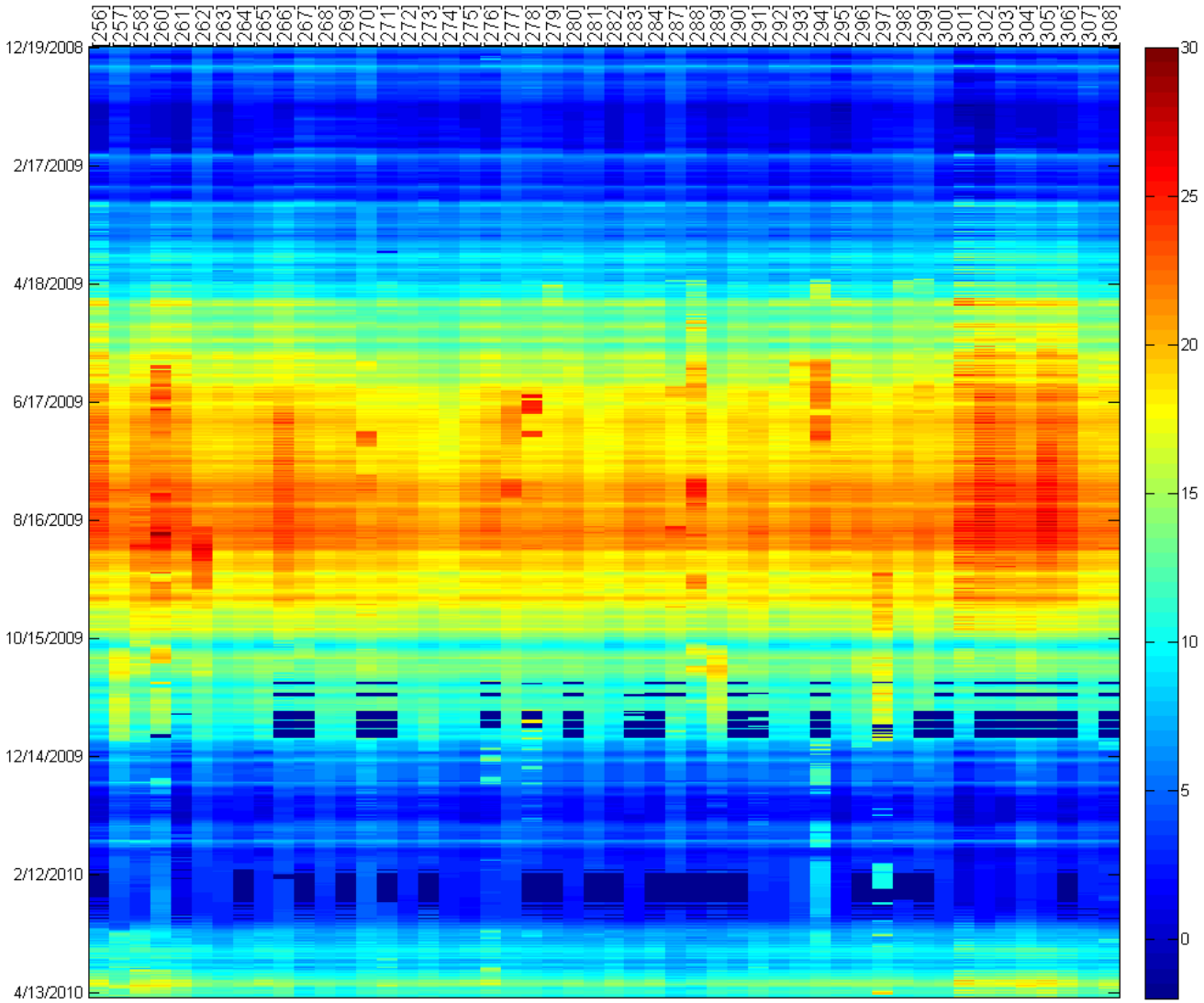


Observations & Incremental Robust PCA *

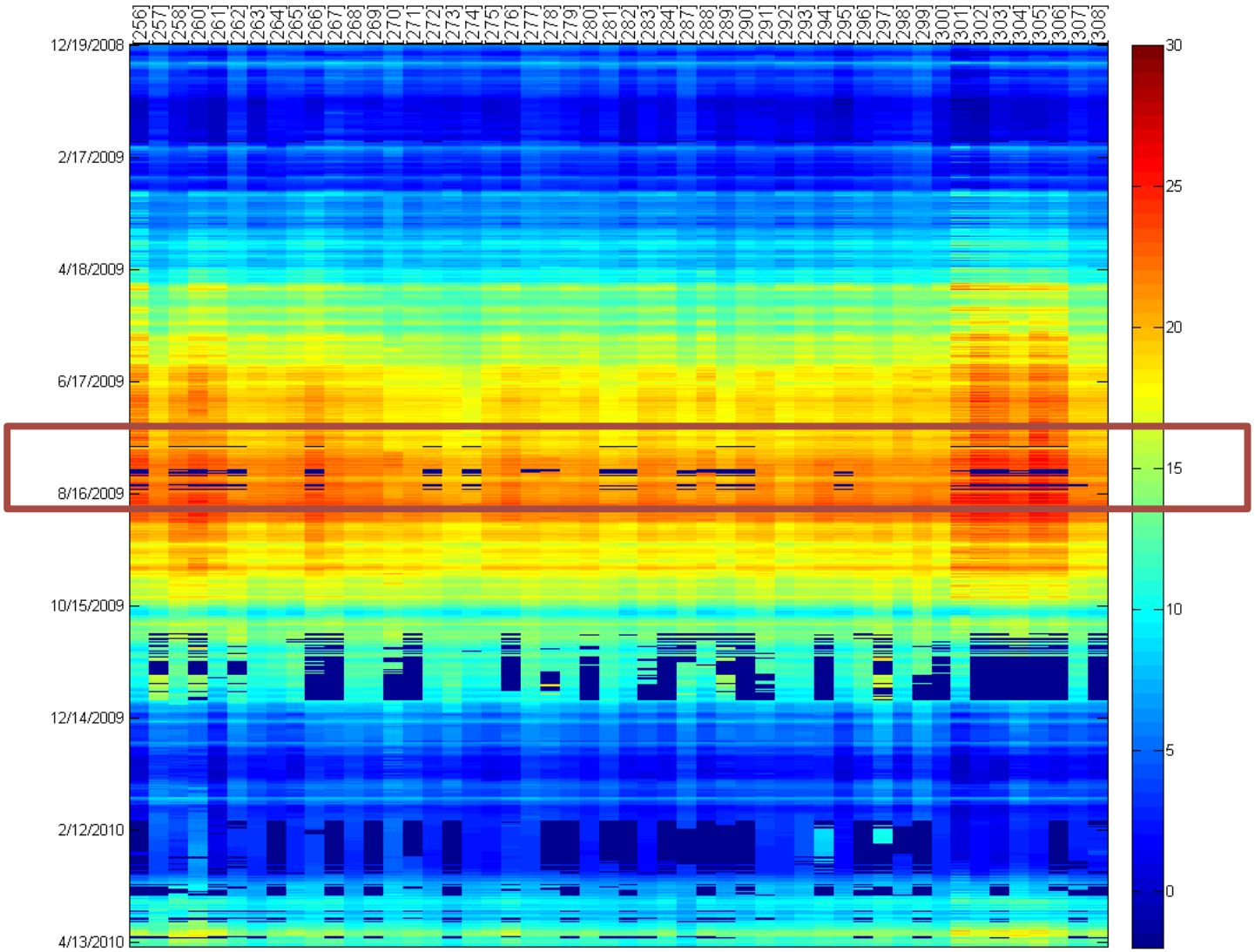
- Data are faulty
  - Pollutes the basis
  - Impacts the quality of gap-filling
- Basis is time-dependent
  - Correlations are changing over time
- Simultaneous fault-detection & gap-filling
  - Initialize basis using reliable data
  - Using basis at time t,
    - Detect and remove faulty readings
    - Weight new vectors depending on residuals
    - Update thresholds for declaring faults
  - Iterate over data to improve estimates

[+] Budavári, T., Wild, V., Szalay, A. S., Dobos, L., Yip, C.-W. 2009 *Monthly Notices of the Royal Astronomical Society*, 394, 1496, http://adsabs.harvard.edu/abs/2009MNRAS.394.1496B

Gap-filling using sensor correlations (same as slide 8)



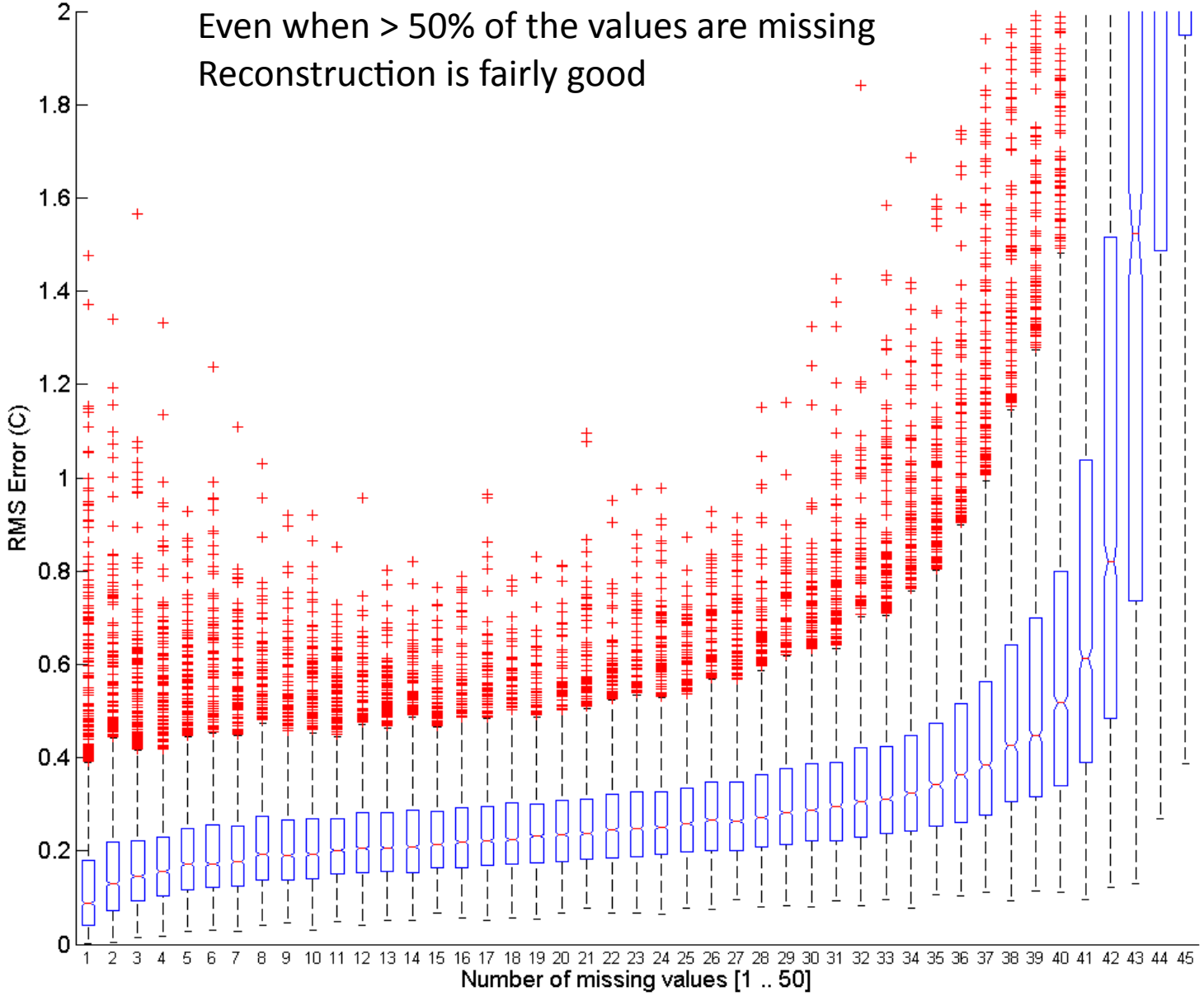
Iterative and Robust gap-filling



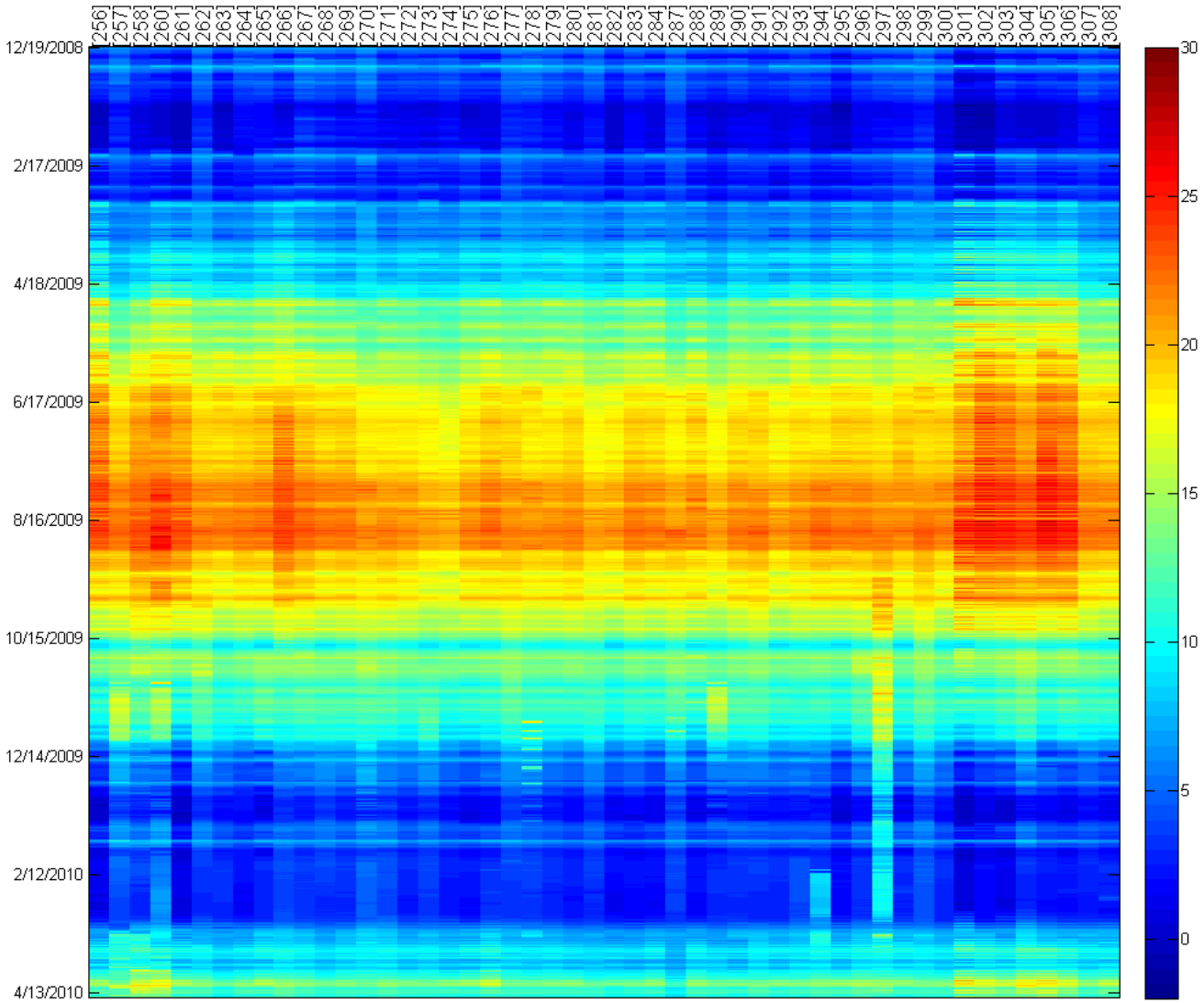
-Fault removal leads to more gaps and less data to work with

-However, data looks much cleaner.

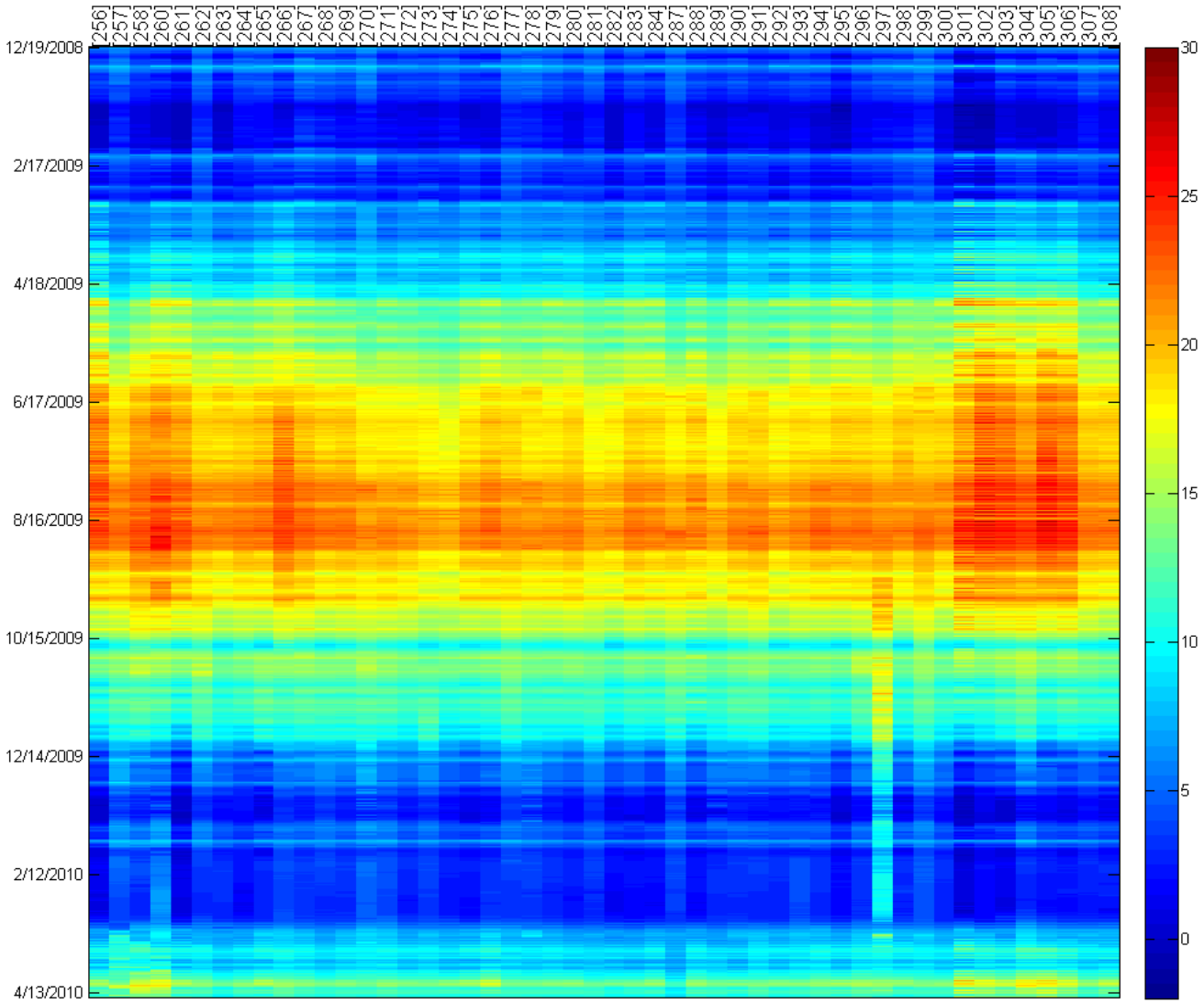
Impact of number of missing values



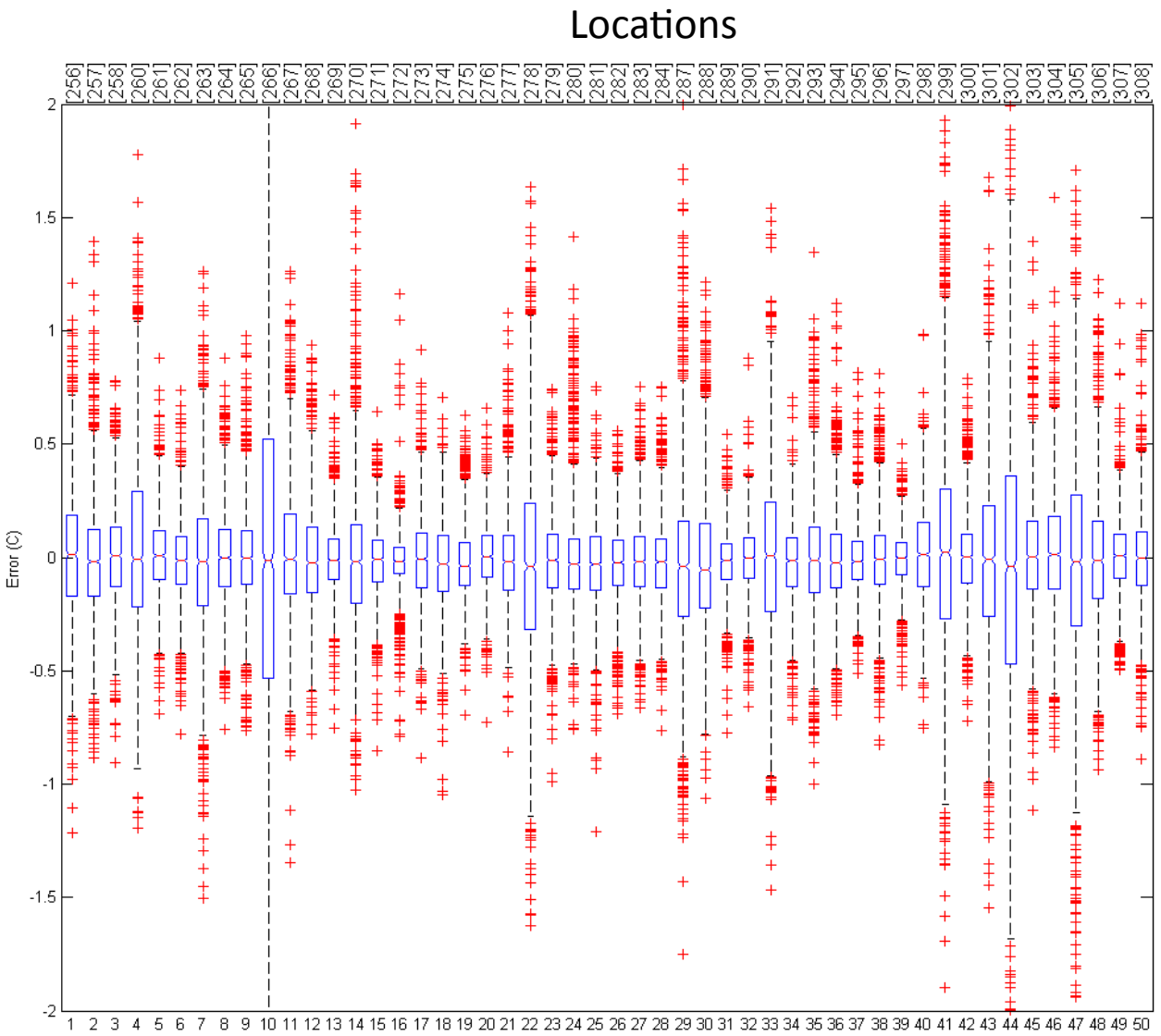
Iteration - I


256
257
258
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
12/19/2008
2/17/2009
4/18/2009
6/17/2009
8/16/2009
10/15/2009
12/14/2009
2/12/2010
4/13/2010
30
25
20
15
10
5
0

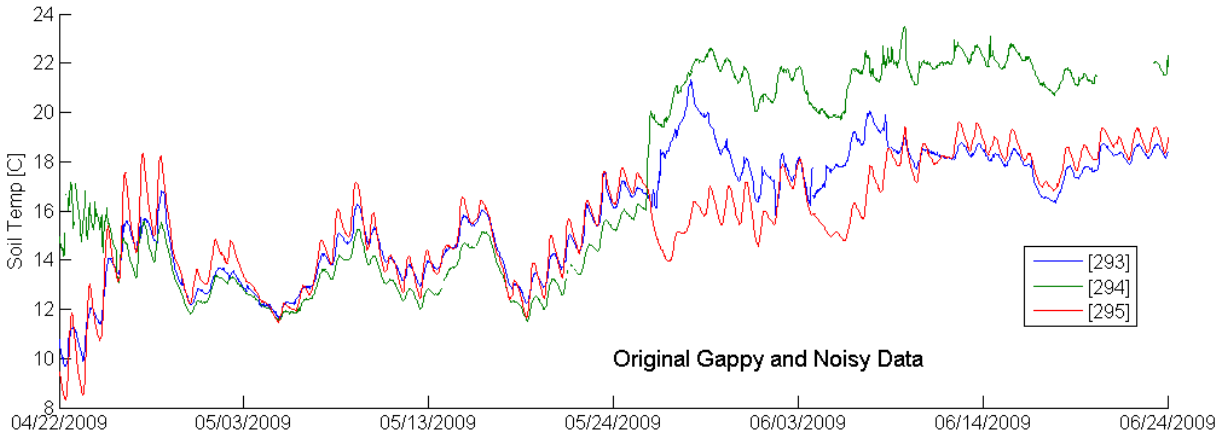
Iteration - II

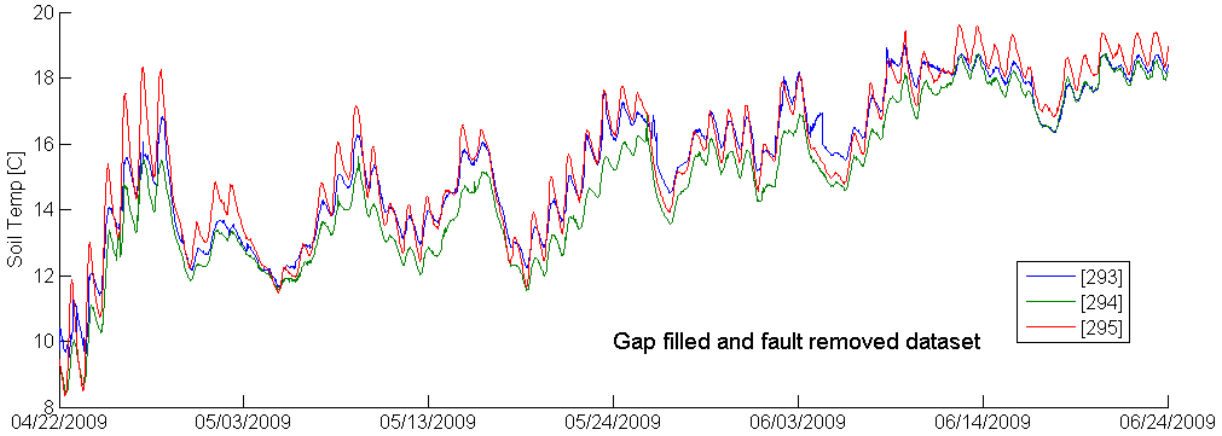


Accuracy using spatiotemporal gap-filling



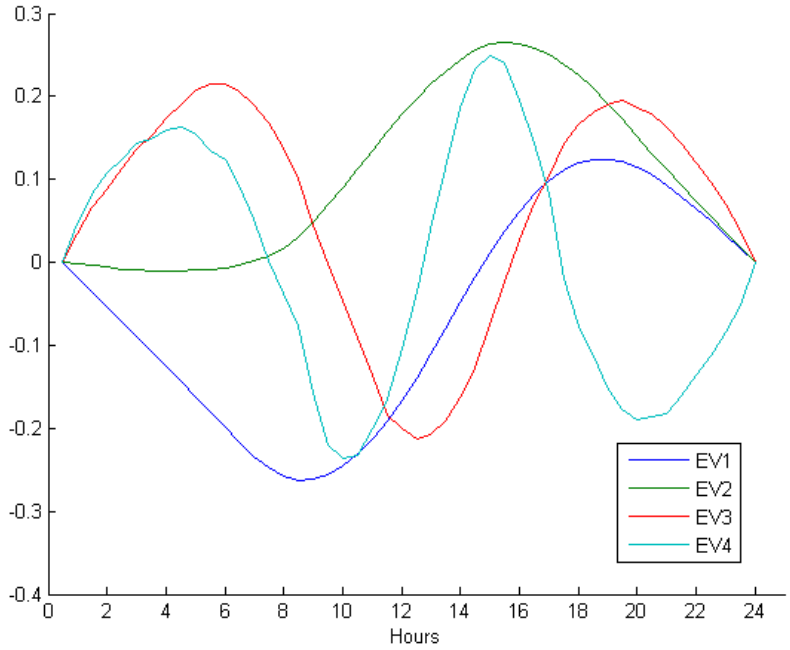
Example of running algorithm

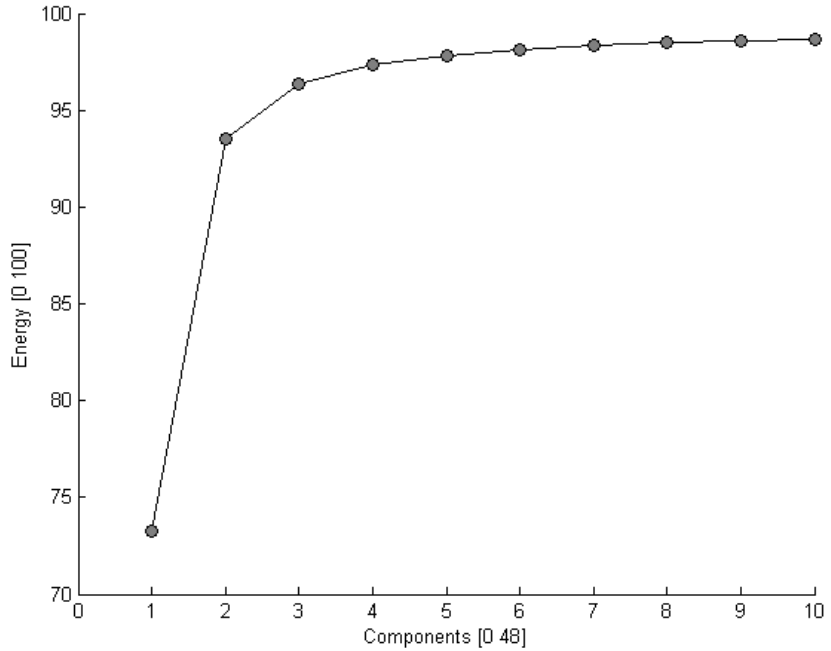




ETC

Daily Basis


0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
2
4
6
8
10
12
14
16
18
20
22
24
Hours
EV1
EV2
EV3
EV4


100
95
90
85
80
75
70
Energy [0 100]
0
1
2
3
4
5
6
7
8
9
10
Components [0 48]

Collected measurements