TYPICAL PATTERNS, ATYPICAL EVENTS, AND UNCERTAINTY IN COMPLEX SYSTEMS

Brett G. Amidan

Thomas A. Ferryman

Applied Statistics and Computational Modeling Group

Pacific Northwest National Laboratory
Richland, WA

Sept 15, 2011

Pacific Northwest National Laboratory: Battelle-managed and mission-driven

<u>Our vision</u>

PNNL will be recognized worldwide and valued nationally and regionally for our leadership in science and for rapidly translating discoveries into solutions for challenges in energy, the environment, and national security.

- DOE Office of Science Laboratory
- Operated by Battelle since 1965
- Outstanding science, impactful solutions
- Nearly 5,000 employees

Applied Statistics and Computational Modeling Group

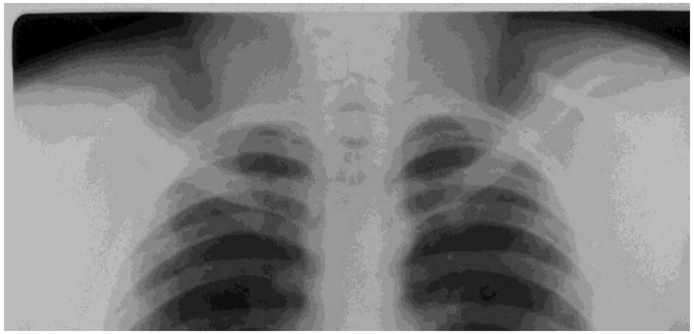
- *Applied Statistics* is a mathematical discipline dealing with methods of obtaining, analyzing and summarizing data.
- *Computational Modeling* utilizes extensive computational resources to generate models to study the behavior of complex systems.
- Our Group consists of 24 Statistical Scientists, 7 Operations Research Scientists, and 13 in other fields.

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

Diversity of Projects

- Our group works on MANY diverse projects, including –
  - Insider Threat
  - Chemometrics
  - Epidemiology
  - VSP (Visual Sample Plan)
  - POW clavicle identification
  - RPMP (Portal Monitoring)
  - . . .
  - Situational Awareness and Alerting



The Aviation Problem and Needs

- For many years, the aviation industry relied mainly on domain expertise to understand aviation safety.
- Gigabytes of data are now recorded daily.
  - On-board instrumentation records hundreds of variables for every flight (i.e. roll, pitch, airspeed, engine temperature, etc).
  - Thousands of flights daily.
- Aviation experts are in need of sophisticated, user-friendly software to rapidly and effectively drill into the data to find insight into possible safety issues.

Morning Report

The Morning Report was developed to help the aviation industry use mathematical methods to look at thousands of flights a day.

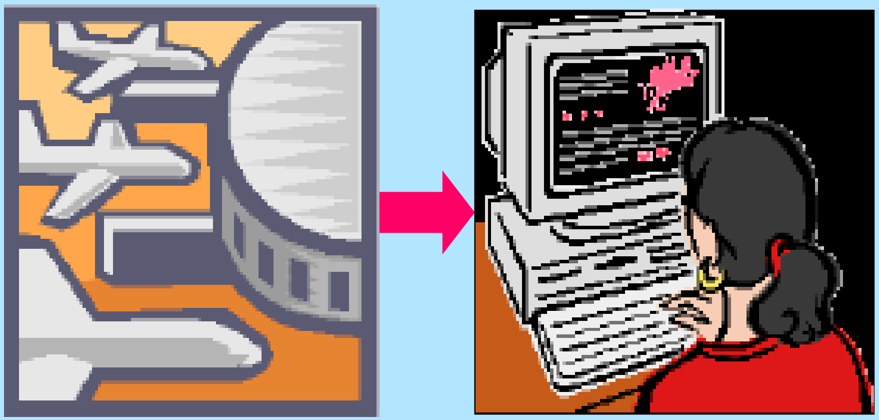
These analyses focused on -

- Typical patterns, that characterize >99% of the flights
- Atypical events, that are worthy of individual inspection

Pacific Northwest
NATIONAL LABORATORY

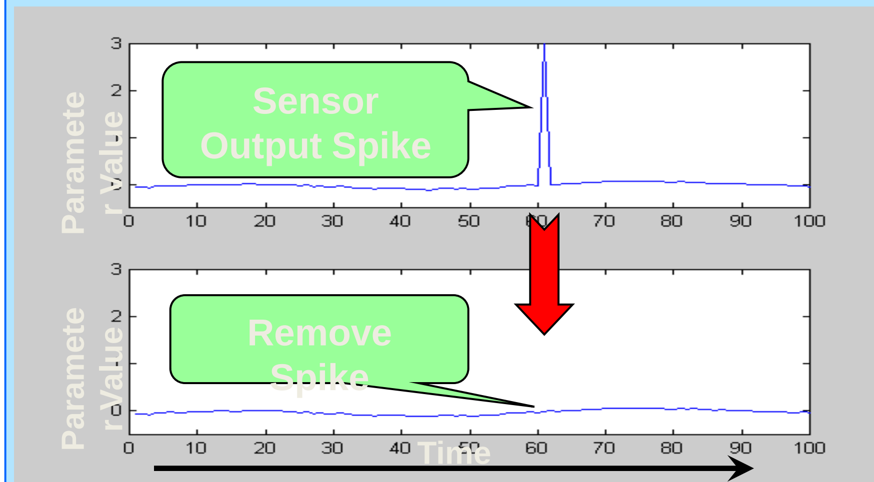
Proudly Operated by Battelle *Since 1965*

Step 1:
Download Data



- Download daily or weekly
- From tapes, disks, or solid state devices
- Use commercially available playback software
- Insert data into commercially vended database

Step 2:
Check the Data Quality



- Apply knowledge-based filters
- Identify “bad” data
- Remove the “bad” data
- Inform user of QA problems

Step 3:
Conduct Pre-defined Exceedance Checks

- Airline experts define specific data comparisons to be made at specific routine events
  - Are the gear down while altitude is above 18,000 ft?
  - Are the flaps extended while airspeed is greater than 300 knots?
  - Etc.

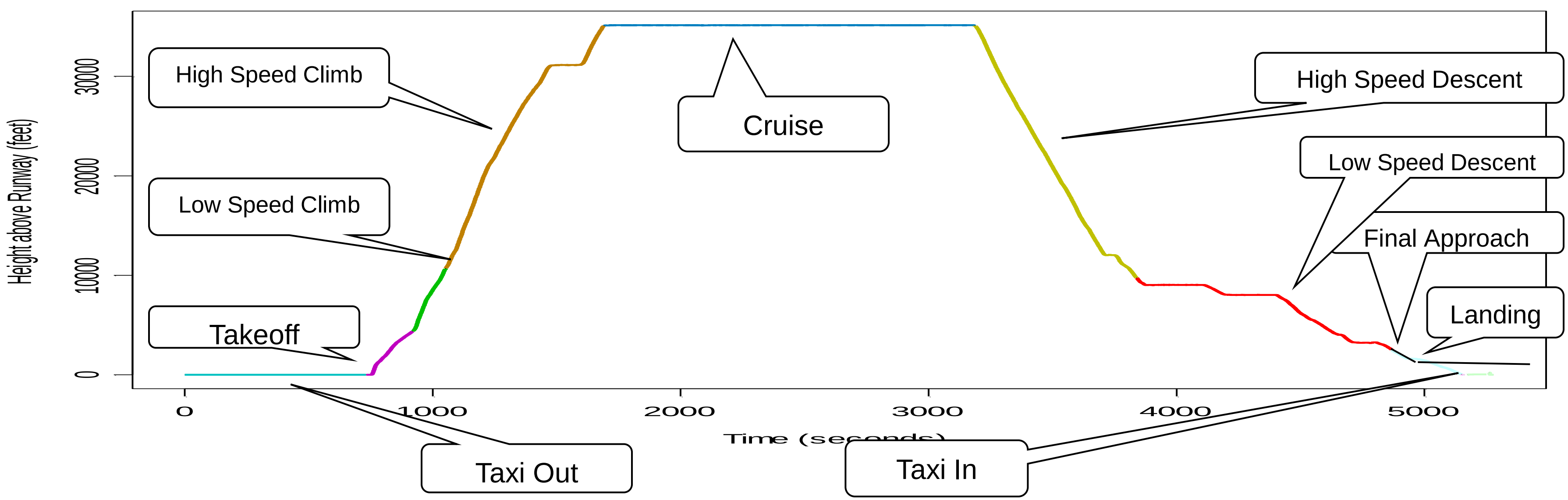
Time (secs)	Param 1	Param 2	. . .	Param P	Routine Events
1	103.40	1		277.40	Start Takeoff
2	103.70	1		266.30	
…	…	…	…	…	
126	104.49	1		267.31	
127	104.98	1		268.19	
128	[illegible]	1		[illegible]	
129	105.45	0		269.12	Gear Up
[illegible]	[illegible]	[illegible]		[illegible]	
131	106.39	0		269.78	
…	…	…	…	…	
4021	106.82	0		270.71	
4022	107.33	0		270.78	
4023	107.89	0		270.85	10000 ft AFE
4024	108.40	0		271.14	
4025	108.53	0		271.53	
4026	109.38	0		272.03	
. . .	…	…	…	…	
N	110.68	0		273.70	Touchdown

This requires that we envision the potential problems before they occur.

Step 4: Structure the Data

- Data are parsed into flight segments
- Flight Segments based on Event Markers, e.g.
  - Gear-up
  - Cross outer-marker
  - Descent through 1000 ft AFE
- Customizable to each air carrier phase definitions

Time (secs)	Param 1	Param 2	. . .	Param P	Event Marker	ACR Phase
151						
152	103.40	1		277.40	Rotate	Takeoff
153	103.70	1		103.70		
…	…	…	…	…		
335	105.13	1		105.13		
336	105.45	0		105.45	Gear Up	
337	105.73	0		105.73		
…	…	…	…	…		climb
1225	106.82	0		106.82		
1226	107.89	0		107.89	10000 ft AFE	
1227	108.10	0		108.07		
…	…	…	…	…		
3236	108.51	0		109.04		
3237	109.33	0		109.12	Max Altitude	Cruise
3238	110.25	0		109.74		
…	…	…	…	…		
6259	109.04	0		108.60		
6260	109.85	0		109.57	10000 ft AFE	
6261	109.87	0		110.39		
…	…	…	…	…		Approach
6673	110.70	0		110.53		
6674	111.19	0		110.68	Gear Down	
6675	111.90	1		111.29		
…	…	…	…	…		
7786	112.13	1		112.10		Landing
7787	112.91	1		112.43	Touchdown	
7788	113.63	1		112.90		



Step 5: Create Derived Parameters to Capture Physics Based Insights

- Aircraft heading with respect to runway
- Aircraft location with respect to runway
- Derived Energy Parameters
  - Total energy
  - Kinetic energy
- Others

Step 6: Calculate Preliminary Flight Parameter Signatures

- Continuous Variable
  - Air speed, roll, altitude, vibration, etc.
- Discrete Variables
  - Gear position, autopilot mode, reversers status, etc.
- Data Compression Signature
  - Lossy compression for continuous variables
  - Lossless compression for discrete variables

Step 7: Store the Signatures into the Database

Analysis

- The previous steps (Steps 1-7) are performed once for each flight.
- After many flights are collected, Steps 8-11 compare the flights to each other and identify
  - Typical patterns
  - Atypical events
- The relevant information is shared with the user.

Step 8: Select the Data

- Select a subset of data:
  - Aircraft type
  - Airports
  - Flight Phase
  - Time Frames
  - Other Parameters

Time (secs)	Param 1	Param 2	. . .	Param P	Event Marker
1	103.40	1		277.40	Start Takeoff
2	103.70	1		266.30	
…	…	…	…	…	
126	104.49	1		267.31	
127	104.98	1		268.19	
[illegible]	[illegible]	[illegible]		[illegible]	
129	105.45	0		269.12	Gear Up
130	105.73	0		269.73	
131	106.39	0		269.78	
…	…	…	…	…	
4021	106.82	0		270.71	
4022	107.33	0		270.78	
4023	107.89	0		270.85	10000 ft AFE
[illegible]	[illegible]	[illegible]		[illegible]	
4025	108.53	0		271.53	
4026	109.38	0		272.03	
. . .	…	…	…	…	
N	110.68	0		273.70	Touchdown

Time (secs)	Param 1	Param 2	. . .	Param P	Event Marker
129	105.45	0		269.12	Gear Up
130	105.73	0		269.73	
131	106.39	0		269.78	
…	…	…	…	…	
4021	106.82	0		270.71	
4022	107.33	0		270.78	
4023	107.89	0		270.85	10000 ft AFE

Step 9: Transform the Signatures

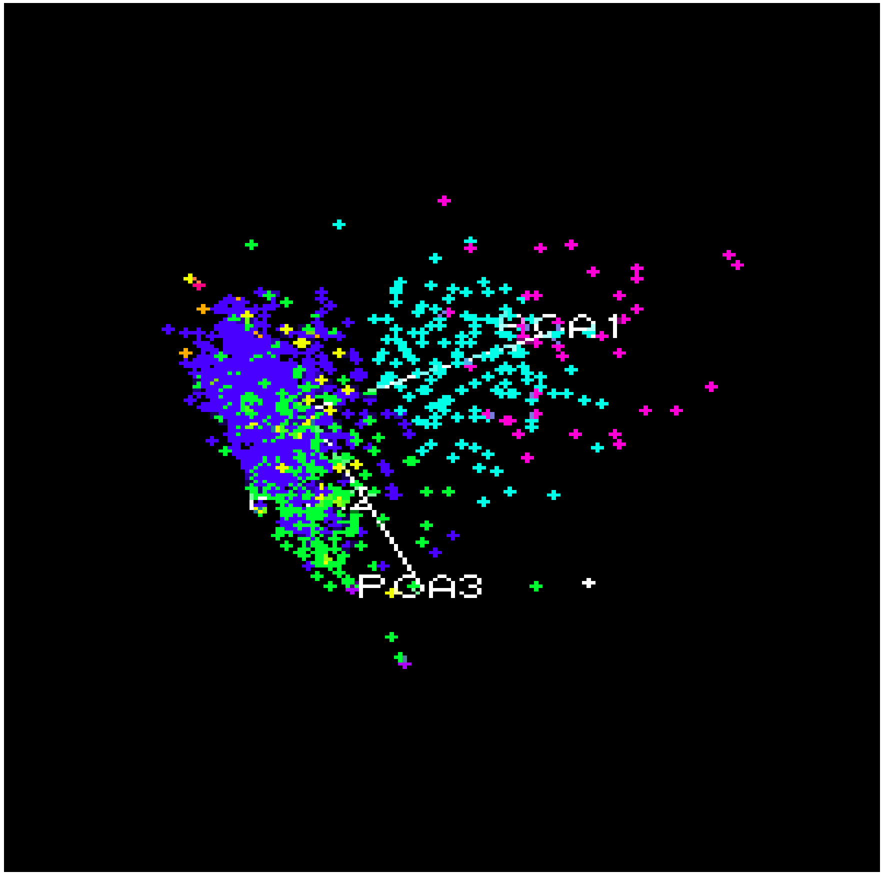
- Transform and summarize the signature from the desired subset of data.

Time (secs)	Param 1	Param 2	. . .	Param P	Event Marker	ACR Phase
151						
152	103.40	1		277.40	Rotate	Takeoff
153	103.70	1		103.70		
…	…	…	…	…		
335	105.13	1		105.13		
336	105.45	0		105.45	Gear Up	
337	105.73	0		105.73		
…	…	…	…	…		climb
1225	106.82	0		106.82		
1226	107.89	0		107.89	10000 ft AFE	
1227	108.10	0		108.07		
…	…	…	…	…		
3236	108.51	0		109.04		
3237	109.33	0		109.12	Max Altitude	Cruise
3238	110.25	0		109.74		
…	…	…	…	…		
6259	109.04	0		108.60		
6260	[illegible]	0		109.57	10000 ft AFE	
[illegible]	[illegible]	0		110.39		
[illegible]	[illegible]	[illegible]	…	…		Approach
[illegible]	[illegible]	0		110.53		
[illegible]	[illegible]	0		110.68	Gear Down	
[illegible]	[illegible]	1		111.29		
[illegible]	[illegible]	[illegible]	…	…		
[illegible]	[illegible]	1		112.10		Landing
[illegible]	[illegible]	1		112.43	Touchdown	
[illegible]	[illegible]	1		112.90		

e.g.; these flight segments combine to form the “Cruise” phase

Step 10:
Cluster the Transformed Signatures

- Typical patterns
  - Clusters of similar flights
  - Summarized in plain English
- Atypical flights
  - Singletons, clusters of 1 or 2
  - Summarized in plain English
- Performed for each user-defined and selected flight phase



Step 11: Find the Atypical Flights

- Atypical flights are defined to be -
  - Singletons
  - Very small clusters (atypical clusters)
- Differs from classic exceedance analysis which look for parameter values outside of pre-defined ranges within a flight phase
- Can be the impetus for further investigation

Finds the unenvisioned !!

End-users don't have to know what they are looking for !!

Step 12: Present the Findings

- New flights processed overnight
- Analysis processing occurs overnight
- Morning report is ready by 7 am every morning
- Identifies most atypical flights
  - Excludes flights previously reviewed and dispositioned
  - Enables drill down to flight details
  - Allows capture images in Microsoft PowerPoint files for communication ease

APMS Morning Report

Morning Report Tab

Summary | Flight List

Significant parameters explaining reasons for atypicality

Morning Report Summary

New Flights : 3199

Fleet : B737-700

Flight Dates : 3/16/2004 - 5/30/2004

Morning Report Date : 5/24/2004

☑ Level 3 Flights

☑ Level 2 Flights

☐ Level 1 Flights

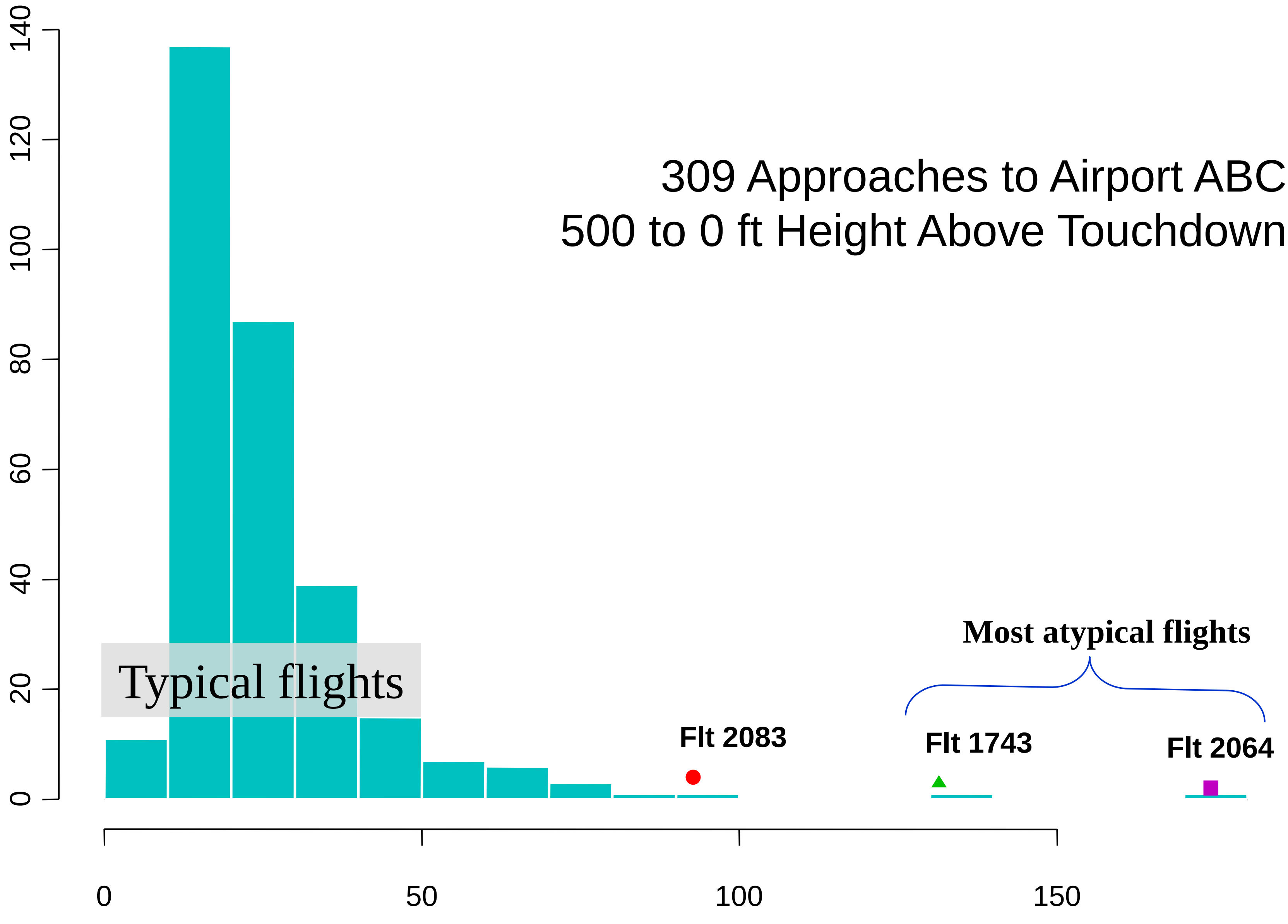
Explore Flight

Flight ID info intentionally blocked out.

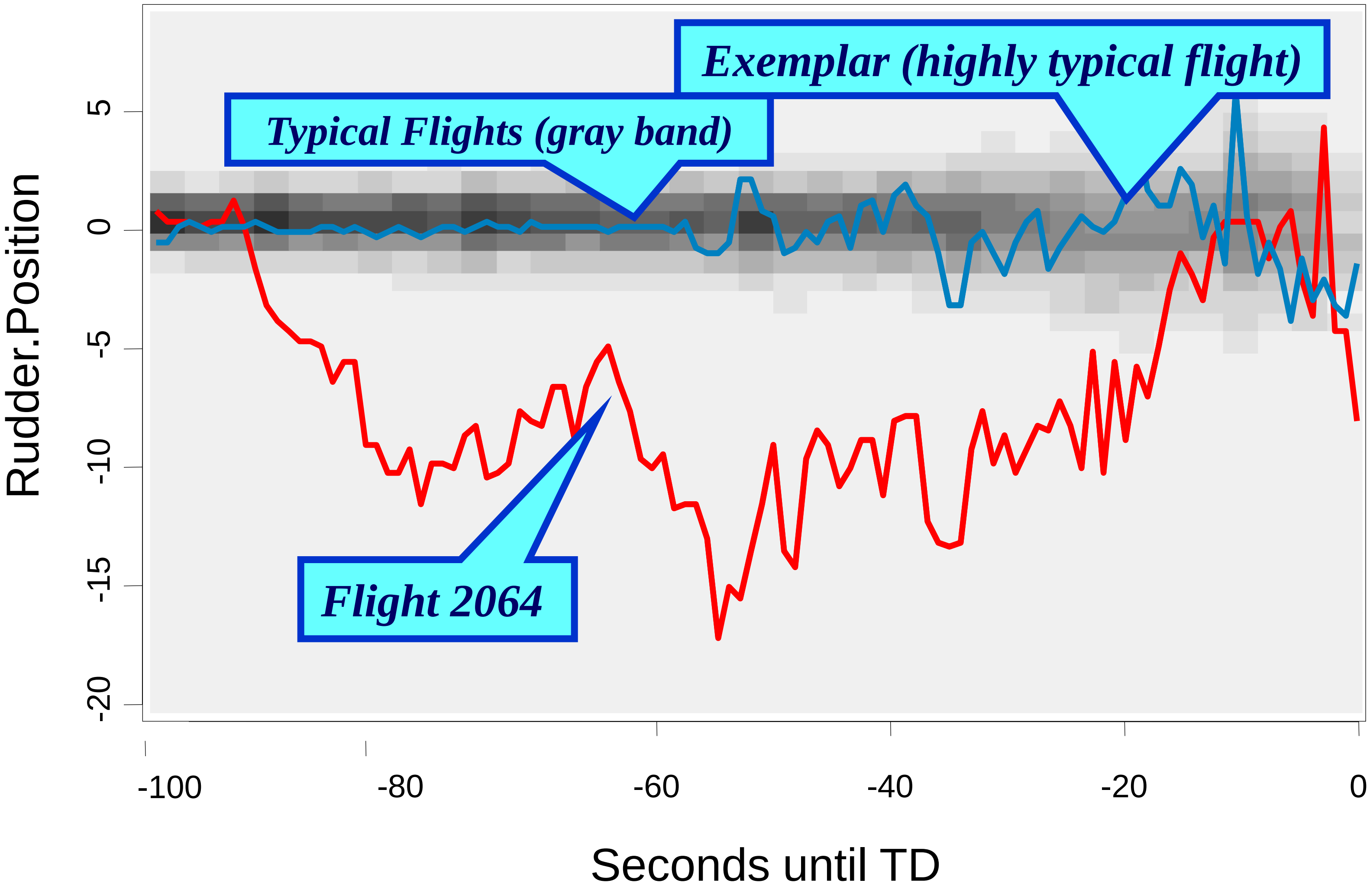
Level	Flight	Tail Number	Analysis_ID	Phase	Origin	Destination	Validation	Rationale
3	373[redacted]	[redacted]	[redacted]:24 AM	3 - Landing	MDW	ATL	Pending	(1)Angle_of_attack_L, (2)Long_Pres_
3	373[redacted]	[redacted]	[redacted]:50 AM	3 - Low Speed Climb	ATL	SAV	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att
3	371[redacted]	[redacted]	[redacted]:07 AM	3 - Takeoff	ATL	DEN	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att
3	370[redacted]	[redacted]	[redacted]:24 AM	3 - Landing	DAY	ATL	Pending	(1)Angle_of_attack_L, (2)Long_Pres_
3	370[redacted]	[redacted]	[redacted]:52 AM	3 - Final Approach	ABQ	ATL	Pending	(1)Glide_Slope_Dev_Dots, (2)Ldg_Gr_
3	375[redacted]	[redacted]	[redacted]:19 AM	3 - Low Speed Descent	IAD	ATL	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att
3	376[redacted]	[redacted]	[redacted]:24 AM	3 - Landing	SAV	ATL	Pending	(1)Angle_of_attack_L, (2)Long_Pres_
3	376[redacted]	[redacted]	[redacted]:29 AM	3 - Takeoff	SLC	MCO	Pending	(1)Lat_Pres_Pos_Corr, (2)Spd_Brake
3	370[redacted]	[redacted]	[redacted]:57 AM	3 - Landing	BOI	SLC	Pending	(1)Dwn_Adv, (2)Height_Above_LO, (
3	374[redacted]	[redacted]	[redacted]:24 AM	3 - Landing	*NA*	ATL	Pending	(1)Angle_of_attack_L, (2)Long_Pres_
3	375[redacted]	[redacted]	[redacted]:07 AM	3 - Takeoff	ATL	ORD	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att
3	373[redacted]	[redacted]	[redacted]:52 AM	3 - Final Approach	BHM	ATL	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att
3	376[redacted]	[redacted]	[redacted]:08 AM	3 - Final Approach	MCO	SLC	Pending	(1)Up_Advisory, (2)Elevator_Pos_L, (
3	371[redacted]	[redacted]	[redacted]:19 AM	3 - Low Speed Descent	OAK	ATL	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att
3	372[redacted]	[redacted]	[redacted]:24 AM	3 - Landing	DFW	ATL	Pending	(1)Angle_of_attack_L, (2)Long_Pres_
3	374[redacted]	[redacted]	[redacted]:07 AM	3 - Takeoff	ATL	BOS	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att
3	376[redacted]	[redacted]	[redacted]:29 AM	3 - Takeoff	SLC	LAX	Pending	(1)Lat_Pres_Pos_Corr, (2)Fuel_Flow_
3	375[redacted]	[redacted]	[redacted]:50 AM	3 - Low Speed Climb	ATL	TLH	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att
3	376[redacted]	[redacted]	[redacted]:57 AM	3 - Landing	SFO	SLC	Pending	(1)Alt_QNE_Corr, (2)Height_Above_L
3	373[redacted]	[redacted]	[redacted]:24 AM	3 - Landing	EWR	ATL	Pending	(1)Angle_of_attack_L, (2)Long_Pres_
3	374[redacted]	[redacted]	[redacted]:19 AM	3 - Low Speed Descent	SFO	ATL	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att
3	375[redacted]	[redacted]	[redacted]:24 AM	3 - Landing	DFW	ATL	Pending	(1)Angle_of_attack_L, (2)Long_Pres_
3	370[redacted]	[redacted]	[redacted]:29 AM	3 - Takeoff	SLC	BOS	Pending	(1)Height_Above_TD, (2)Alt_QNE_Co
3	375[redacted]	[redacted]	[redacted]:24 AM	3 - Landing	SEA	ATL	Pending	(1)Angle_of_attack_L, (2)Long_Pres_
3	376[redacted]	[redacted]	[redacted]:50 AM	3 - Low Speed Climb	ATL	GDL	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att
3	376[redacted]	[redacted]	[redacted]:24 AM	3 - Landing	MYR	ATL	Pending	(1)Angle_of_attack_L, (2)Long_Pres_
3	374[redacted]	[redacted]	[redacted]:20 AM	3 - Low Speed Climb	SLC	SLC	Pending	(1)Elevator_Pos_L, (2)Pitch_Angle, (3
3	371[redacted]	[redacted]	[redacted]:24 AM	2 - Landing	COS	ATL	Pending	(1)Angle_of_attack_L, (2)Long_Pres_
3	375[redacted]	[redacted]	[redacted]:14 AM	3 - Low Speed Descent	EWR	SLC	Pending	(1)Dwn_Adv, (2)Fuel_Burn_Hr_Avg, (
3	372[redacted]	[redacted]	[redacted]:50 AM	3 - Low Speed Climb	ATL	IAD	Pending	(1)Ldg_Gr_Sel_Dwn, (2)Angle_of_att

Finished.

Atypicality Histogram of Flights

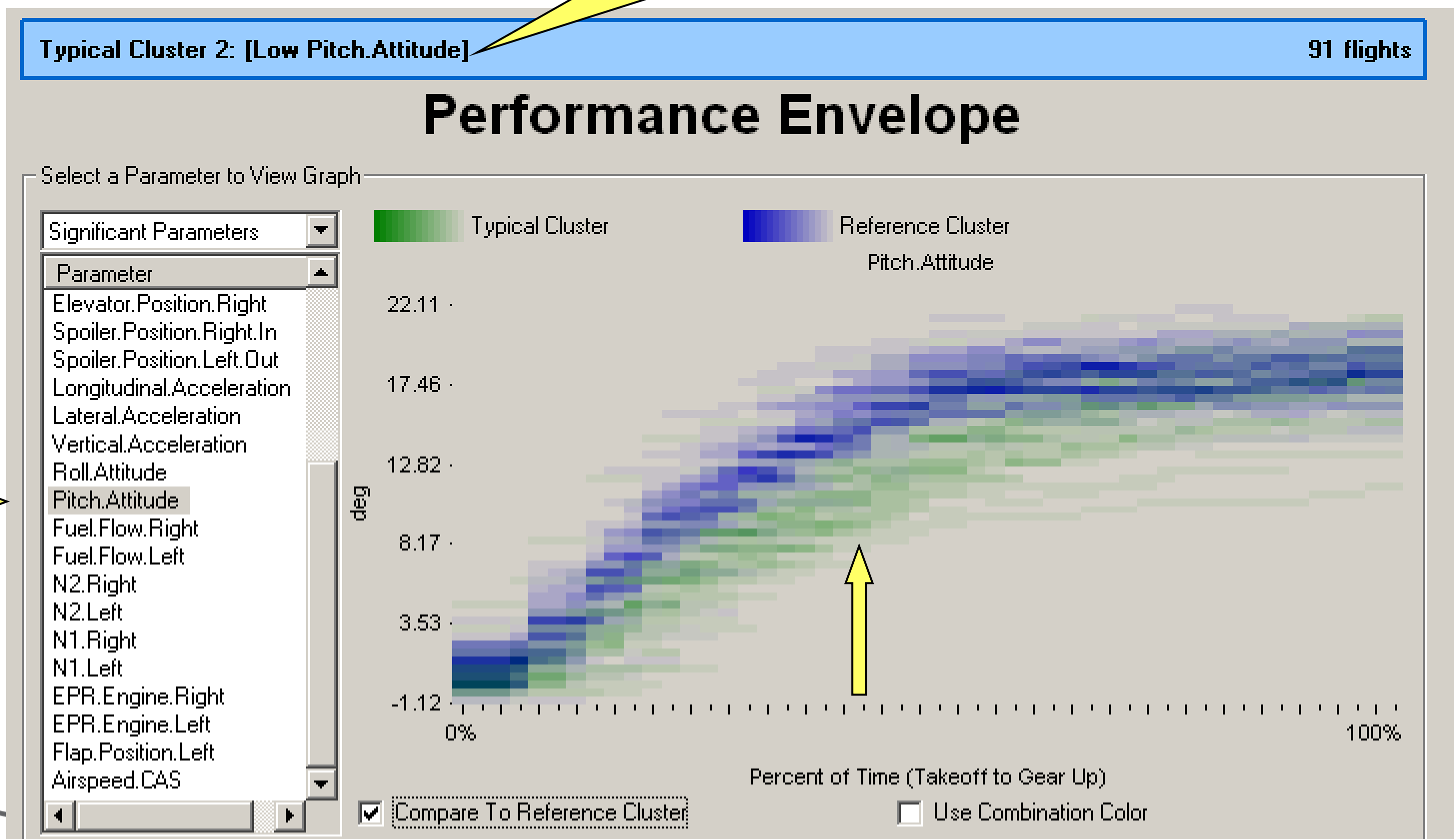


What made Flight 2064 Atypical?



Cluster Comparison

Automated cluster label

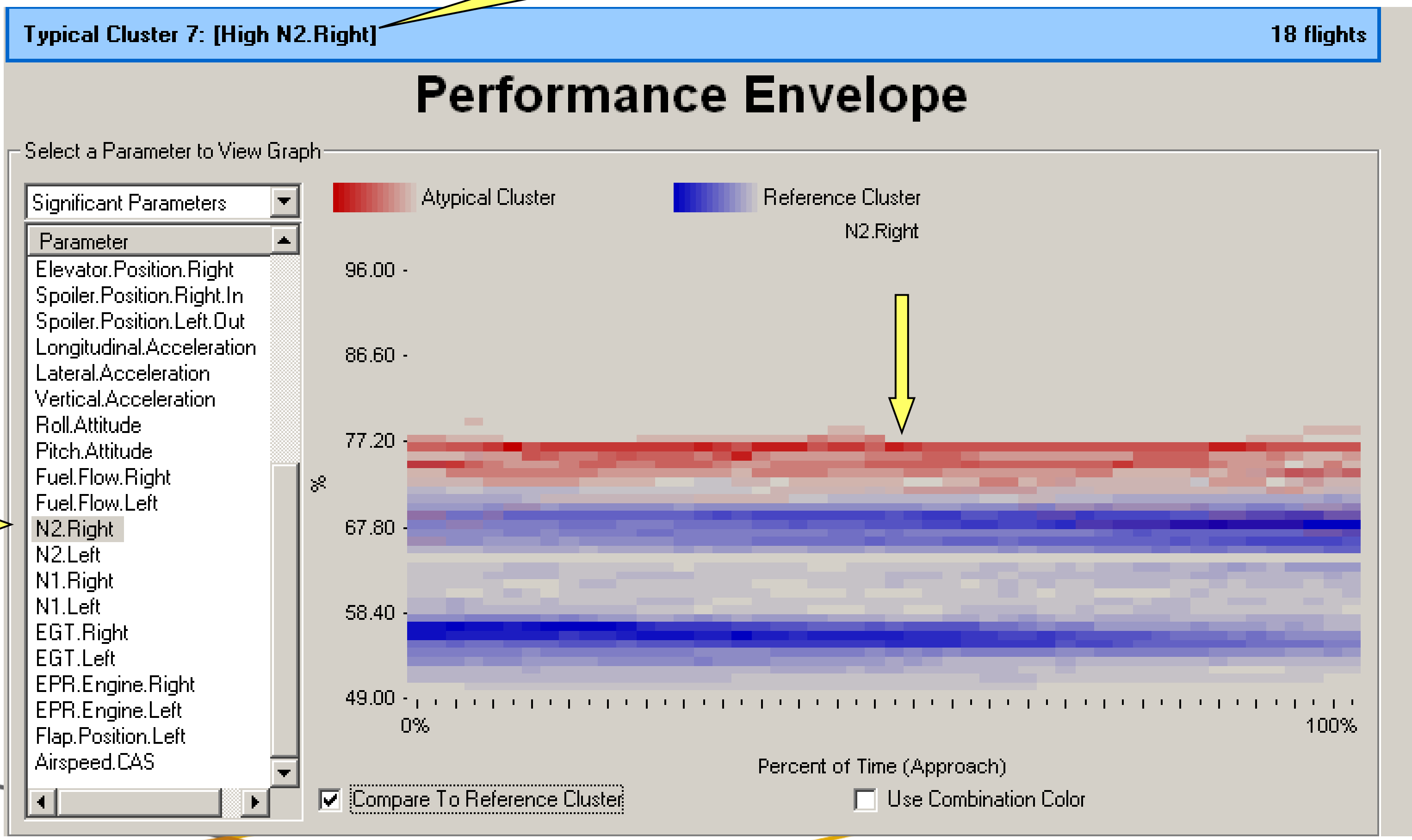


Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

Atypical Cluster Performance Envelope

Automated cluster label



Storymeister Example

- Cluster 8 contains 18 flights. It has highly unusual values in the engine parameter set during the 5000 ft to 2500 ft approach phase. It also has moderately unusual values in the flight controls parameter set during this phase.
- Cluster 8 has extremely large N1.Left (mean value of 96 PCT) and N1.Right (mean value of 97.1 PCT) values during the 5000 ft to 2500 ft approach phase. It also has unusually low flap.position.left (mean value of 3 degrees) values and extremely high noise in Airspeed.CAS (mean noise of 1.5 knots).   The Rudder.Position rate of change was moderately high (mean rate of change of 0.25 degrees).

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

How Does This Apply to Other Domains?

- Aviation has many flights with each flight containing many variables being recorded over time. Industry is interested in atypical events and typical patterns.
- Many other domains have many variables being recorded over time. Examples include air traffic control, cyber security, finance, weather, and the electric power grid.
- The following slides show a few examples of how this has been applied to weather and power grid data.

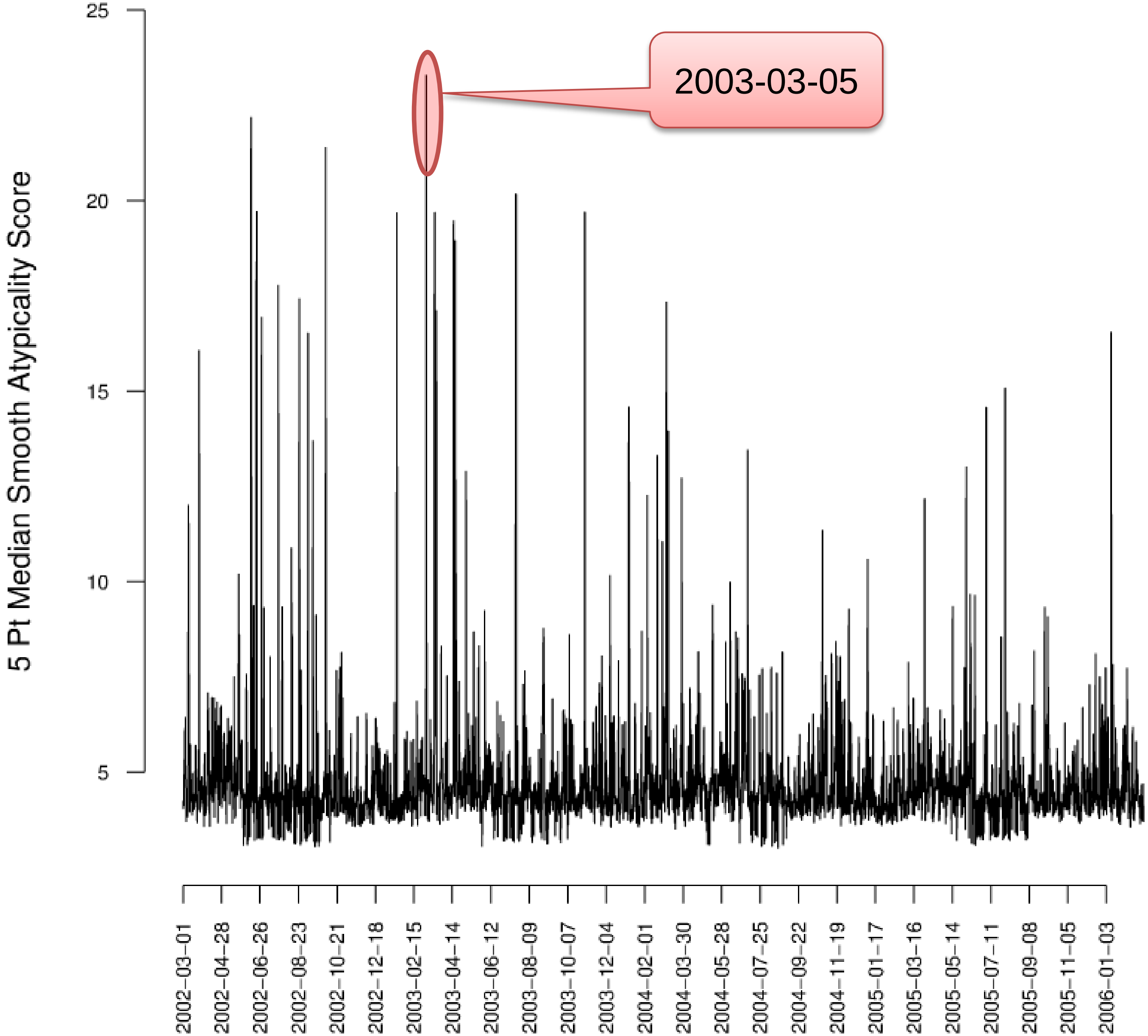
Changes to the Morning Report System to Accommodate Other Domains

- Focus on time-related events (a little different than aircraft).
- Output converted to html web pages, so that most output can now be viewed in a simple web browser.
- All coding done in R.

Weather Data Analyses

- Two different data streams included in the analyses.
  - Stream 1 has a few thousand variables measured every hour for about 10 years.
  - Stream 2 has a few thousand variables measured every minute for 3 years.

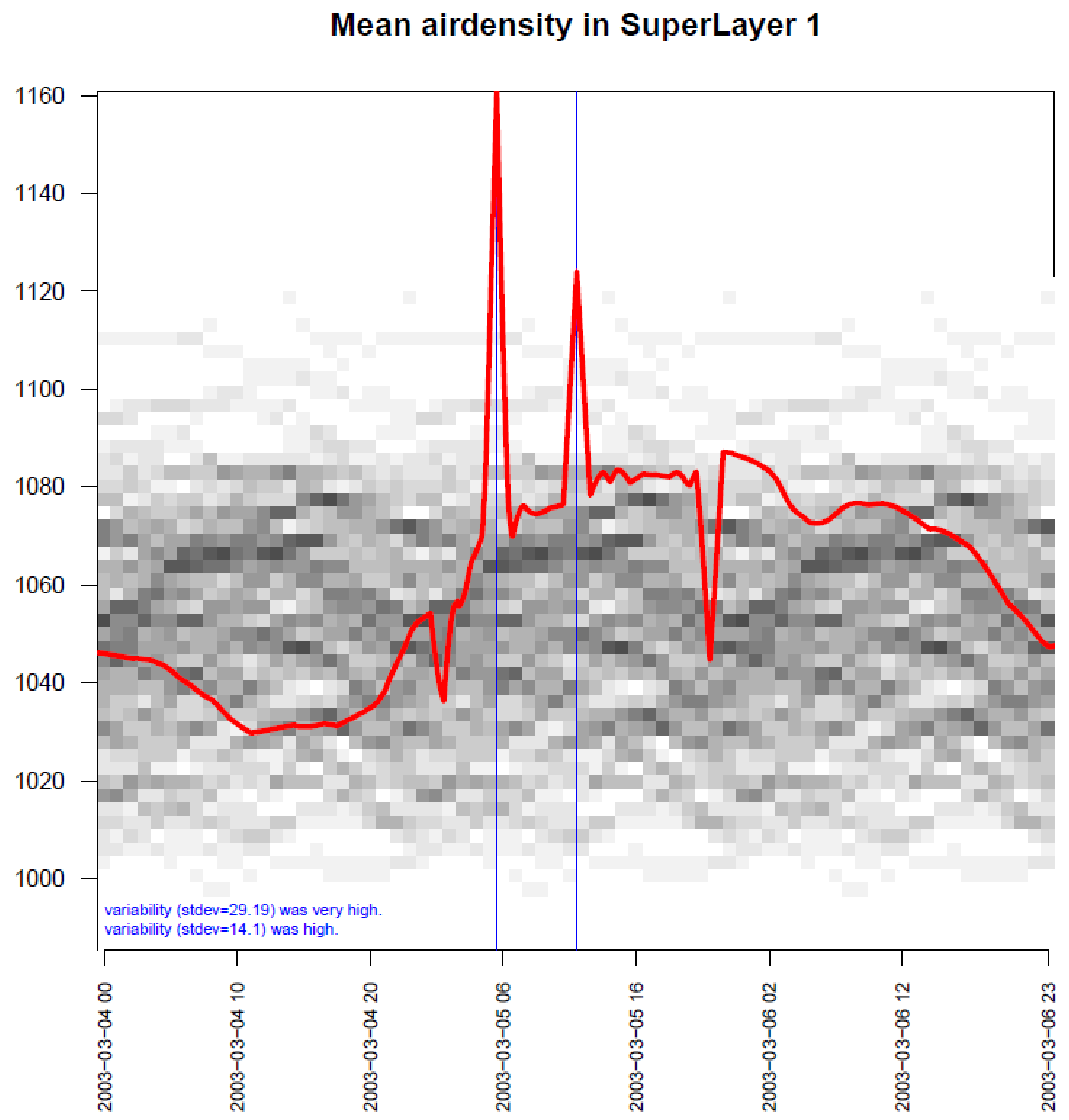
2002-03-01 to 2006-02-28



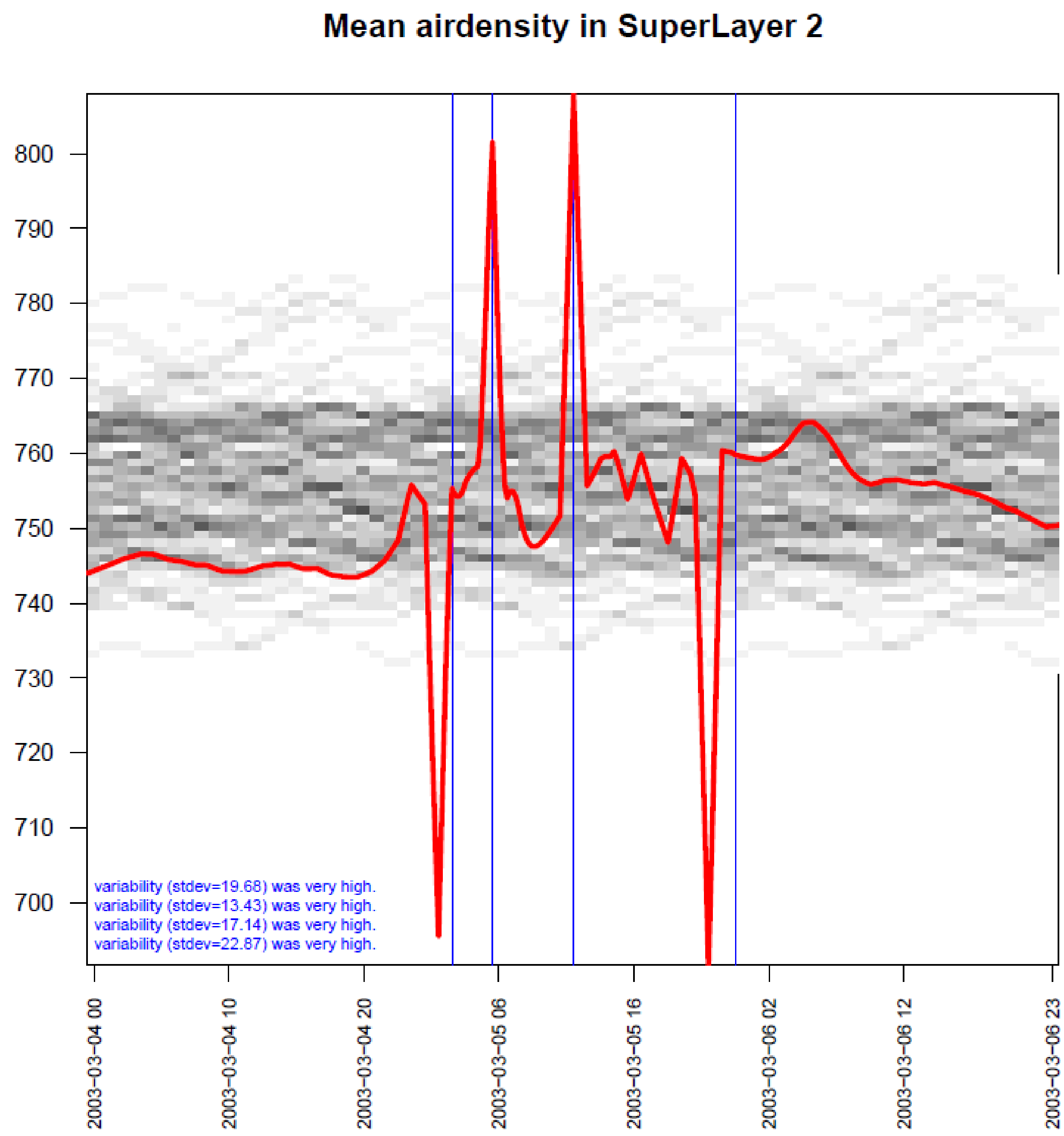
2003-01-01 to 2003-03-31

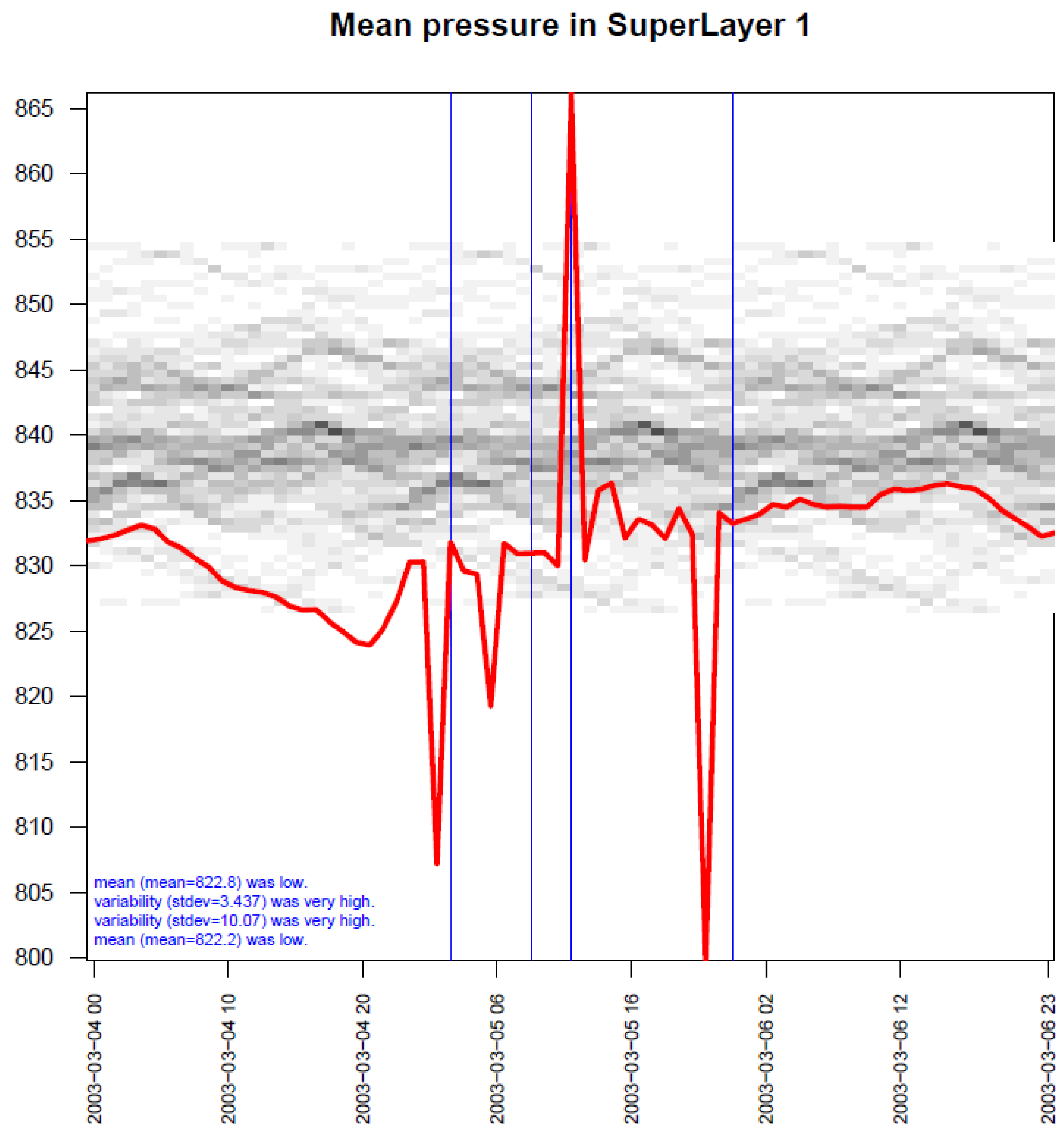
Atypicality Report (Sorted by Global Atypicality Score)

	GAS	Cluster#	Values Rationale	Slope Rationale
2003-03-05 03:00:00	19.70	63.00	pressure.std.1 variability (stdev=3.442) was very high. airdensity.std.5 variability (stdev=2.094) was very high. airdensity.std.4 variability (stdev=2.1) was very high. pressure.std.2 variability (stdev=1.249) was very high. airdensity.std.1 variability (stdev=4.631) was very high. airdensity.mean.2 variability (stdev=19.68) was very high. pressure.mean.5 mean (mean=61.31) was very low. v_nwp_p.std.1 mean (mean=9.516) was very high. pressure.mean.2 mean (mean=543.4) was very low. pressure.mean.4 mean (mean=157.1) was very low. pressure.std.5 mean (mean=20.37) was very low. pressure.mean.3 mean (mean=312.5) was very low. airdensity.mean.3 mean (mean=468) was very low. airdensity.std.2 variability (stdev=2.728) was very high. pressure.std.3 mean (mean=55.18) was very low. airdensity.mean.5 mean (mean=100.8) was very low. cld_frac.std.1 mean (mean=46.38) was high. cld_reliq.std.1 mean (mean=2.678) was high. pressure.mean.1 mean (mean=822.8) was low. airdensity.mean.4 mean (mean=250.2) was low.	airdensity.mean.1 slope variability (stdev=0.2762) was very high. temperature.mean.1 slope variability (stdev=0.05126) was very high. cld_dgeice.mean.1 slope mean (mean=0.09512) was very high. T_sfc slope variability (stdev=0.9033) was very high. p_sfc slope mean (mean=1.546) was high. watervapor_mr.std.1 slope mean (mean=-4.211e-06) was low. stdev_lwup slope mean (mean=-1.33) was low. wdir_sfc slope variability (stdev=25.11) was high. temperature.mean.2 slope variability (stdev=0.04004) was high. stdev_swdn slope variability (stdev=13.14) was high. T_nwp_p.mean.1 slope mean (mean=-1.048) was low. T_nwp_p.std.1 slope variability (stdev=0.2154) was high. cld_mid slope variability (stdev=8.391) was high. aerosol_ext_500.mean.2 slope mean (mean=-3.379e-05) was low. cld_reliq.mean.1 slope variability (stdev=0.1218) was high. temperature.std.2 slope variability (stdev=0.007794) was high. omega_nwp_p.std.1 slope mean (mean=-0.06577) was low. stdev_swup slope variability (stdev=2.949) was high. stdev_swdif slope mean (mean=-9.622) was low.
2003-03-05 15:00:00	19.30	88.00	pressure.std.1 variability (stdev=4.291) was very high. pressure.std.3 variability (stdev=1.413) was very high. airdensity.std.1 variability (stdev=4.647) was very high. cld_thick variability (stdev=1.102) was very high. cld_frac.std.3 variability (stdev=12.58) was very high. airdensity.std.3 variability (stdev=2.181) was very high. cld_dgeice.std.1 mean (mean=27.22) was high. cld_dgeice.mean.1 mean (mean=24.71) was high. cld_frac.mean.3 variability (stdev=14.91) was high. cld_tot variability (stdev=30.95) was high. cld_frac.std.1 mean (mean=41.88) was high. cld_reliq.std.1 mean (mean=2.41) was high.	pressure.std.2 slope mean (mean=-0.06119) was very low. pressure.mean.5 slope mean (mean=-0.08412) was very low. pressure.std.5 slope mean (mean=-0.02439) was very low. airdensity.mean.5 slope mean (mean=-0.1395) was very low. pressure.mean.4 slope mean (mean=-0.1541) was very low. pressure.mean.1 slope mean (mean=-0.1592) was very low. pressure.mean.2 slope mean (mean=-0.1891) was very low. airdensity.std.5 slope mean (mean=-0.03939) was very low. airdensity.mean.4 slope mean (mean=-0.2424) was very low. pressure.mean.3 slope mean (mean=-0.1151) was very low. airdensity.std.2 slope mean (mean=-0.07682) was very low. pressure.std.4 slope mean (mean=-0.007687) was very low. airdensity.mean.3 slope mean (mean=-0.1682) was very low. airdensity.mean.2 slope mean (mean=-0.2551) was very low. airdensity.std.4 slope variability (stdev=0.02119) was high. rh_nwp_p.mean.1 slope variability (stdev=1.801) was high. temperature.std.5 slope variability (stdev=0.007362) was high. temperature.mean.5 slope variability (stdev=0.01235) was high. cld_top slope mean (mean=2.402) was high. lw_net_TOA slope mean (mean=-10.57) was low.
2003-03-18 00:00:00	18.79	84.00	watervapor_mr.mean.1 variability (stdev=0.00236) was very high. temperature.mean.4 variability (stdev=6.453) was very high. watervapor_mr.mean.2 variability (stdev=0.0008035) was very high. temperature.mean.3 variability (stdev=7.672) was very high. temperature.mean.2 variability (stdev=9.162) was very high. temperature.mean.1 variability (stdev=9.526) was very high. watervapor_rh.std.3 variability (stdev=12.16) was very high. watervapor_mr.mean.5 mean (mean=2.657e-05) was very high. stdev_lwp variability (stdev=98.01) was very high. watervapor_mr.mean.3 variability (stdev=8.288e-05) was very high. watervapor_mr.std.4 variability (stdev=4.27e-06) was very high. aerosol_g_500.mean.3 mean (mean=0.6487) was very high. cld_reliq.std.1 variability (stdev=1.303) was very high. cld_lwp.std.1 variability (stdev=10.82) was very	watervapor_rh.mean.5 slope mean (mean=-0.4544) was very low. watervapor_rh.mean.4 slope mean (mean=-0.4964) was very low. airdensity.mean.3 slope mean (mean=-0.3045) was very low. airdensity.std.4 slope mean (mean=-0.04578) was very low. airdensity.mean.2 slope mean (mean=-0.5141) was very low. airdensity.mean.1 slope mean (mean=-0.6978) was very low. airdensity.mean.4 slope mean (mean=-0.1479) was very low. watervapor_mr.std.1 slope mean (mean=-8.086e-06) was very low. stdev_swdn slope mean (mean=-45.34) was very low. airdensity.std.3 slope mean (mean=-0.06688) was very low. watervapor_rh.mean.3 slope mean (mean=-0.3191) was very low. stdev_swdir slope mean (mean=-40.39) was very low. stdev_swup slope mean (mean=-10.71) was very low. stdev_lwup slope mean (mean=-2.248) was very low. cld_lwp.mean.1 slope variability (stdev=1.494) was very high. airdensity.std.2 slope mean

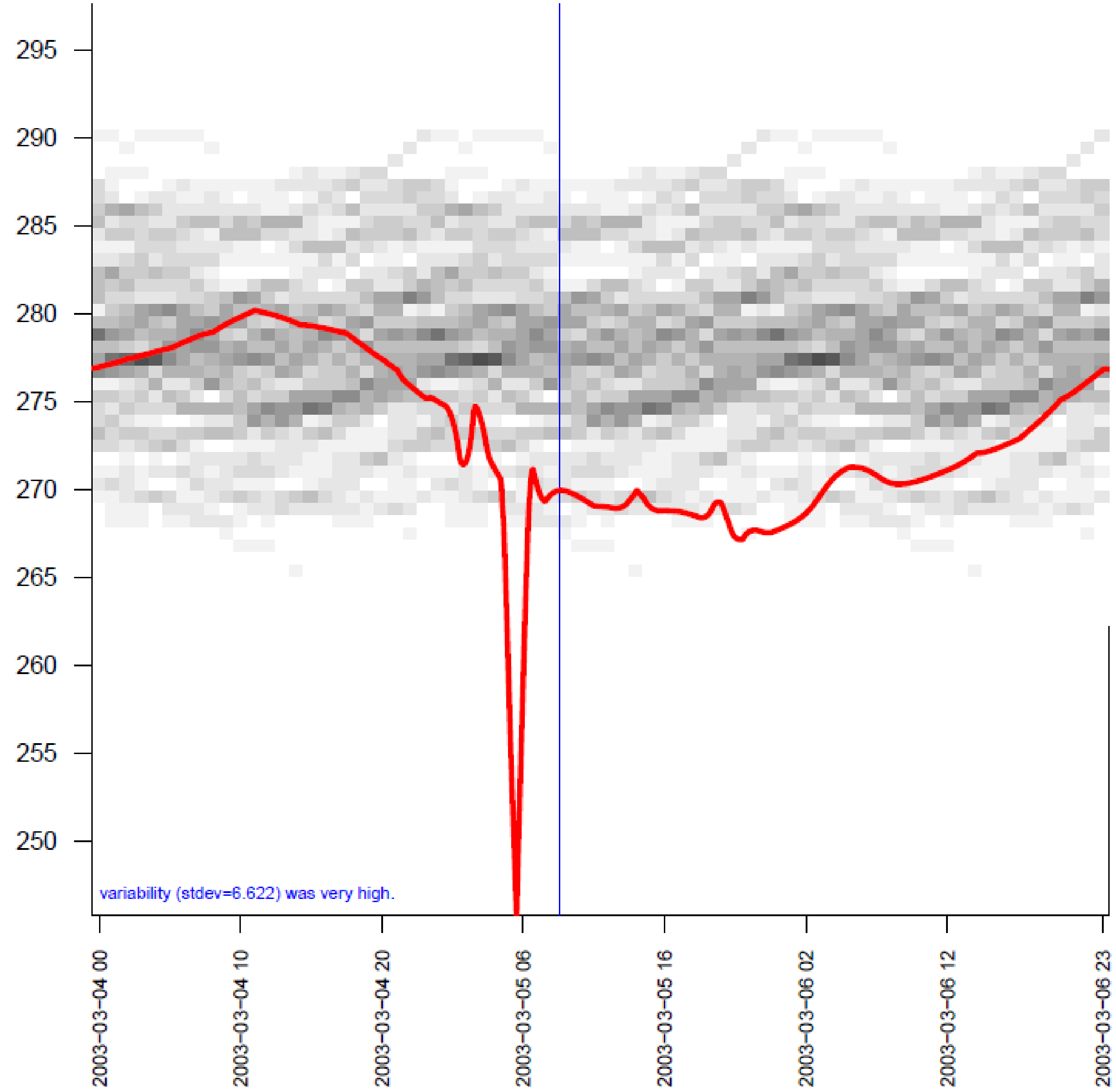


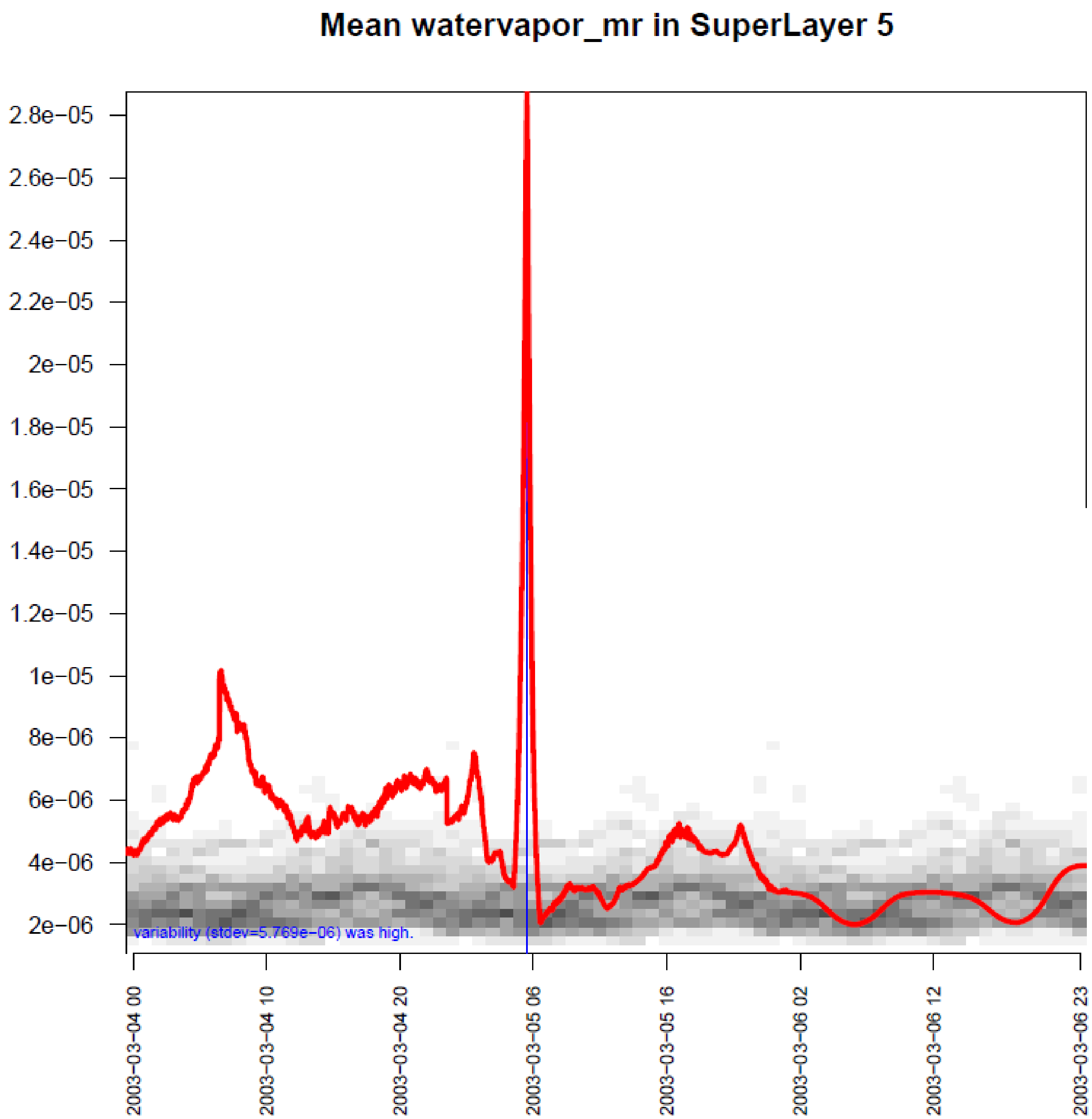
Mean airdensity in SuperLayer 1
1160
1140
1120
1100
1080
1060
1040
1020
1000
variability (stdev=29.19) was very high.
variability (stdev=14.1) was high.
2003−03−04 00
2003−03−04 10
2003−03−04 20
2003−03−05 06
2003−03−05 16
2003−03−06 02
2003−03−06 12
2003−03−06 23


Mean airdensity in SuperLayer 2
800
790
780
770
760
750
740
730
720
710
700
variability (stdev=19.68) was very high.
variability (stdev=13.43) was very high.
variability (stdev=17.14) was very high.
variability (stdev=22.87) was very high.
2003−03−04 00
2003−03−04 10
2003−03−04 20
2003−03−05 06
2003−03−05 16
2003−03−06 02
2003−03−06 12
2003−03−06 23

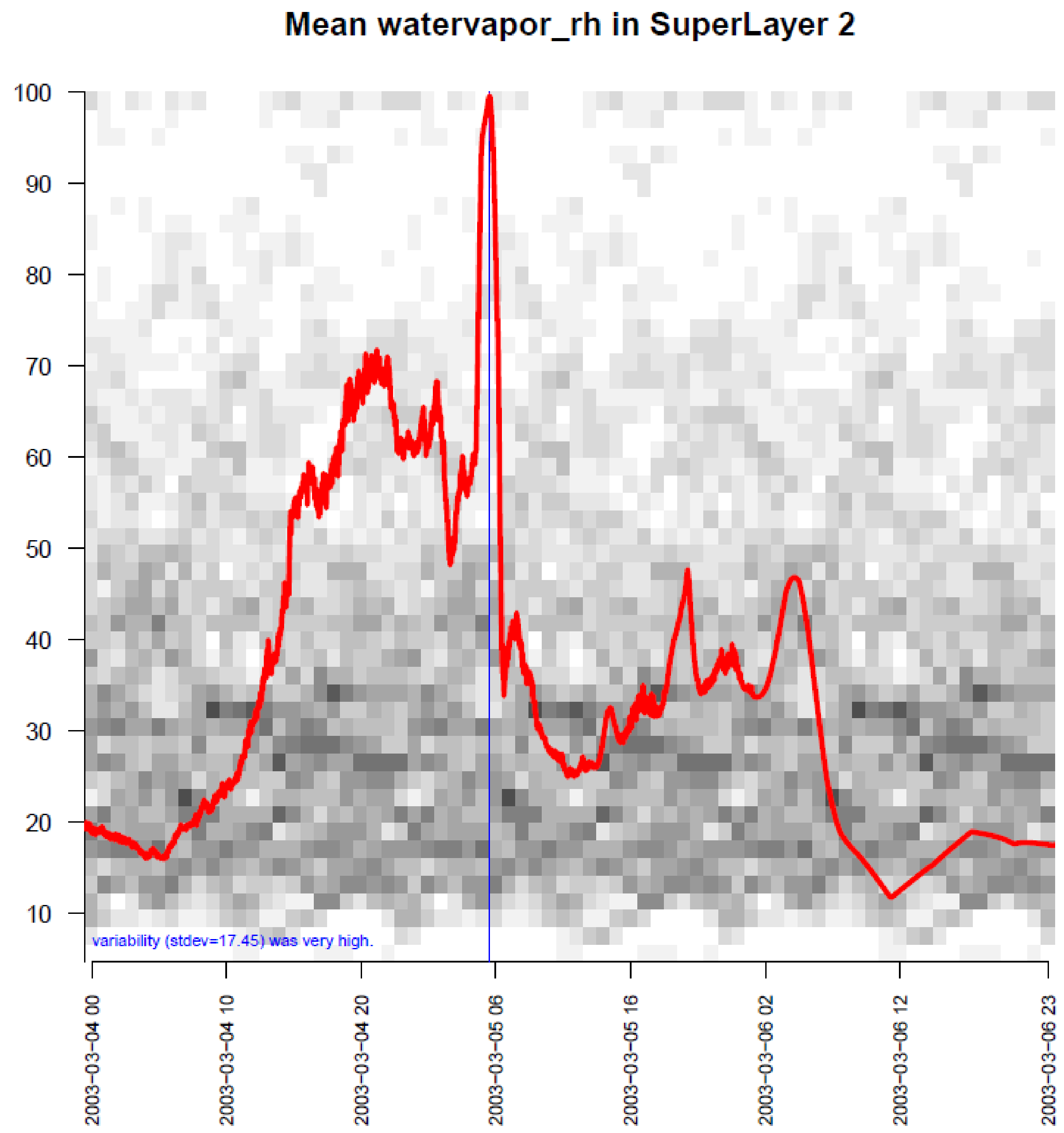

Mean pressure in SuperLayer 1
865
860
855
850
845
840
835
830
825
820
815
810
805
800
mean (mean=822.8) was low.
variability (stdev=3.437) was very high.
variability (stdev=10.07) was very high.
mean (mean=822.2) was low.
2003−03−04 00
2003−03−04 10
2003−03−04 20
2003−03−05 06
2003−03−05 16
2003−03−06 02
2003−03−06 12
2003−03−06 23

Mean temperature in SuperLayer 1

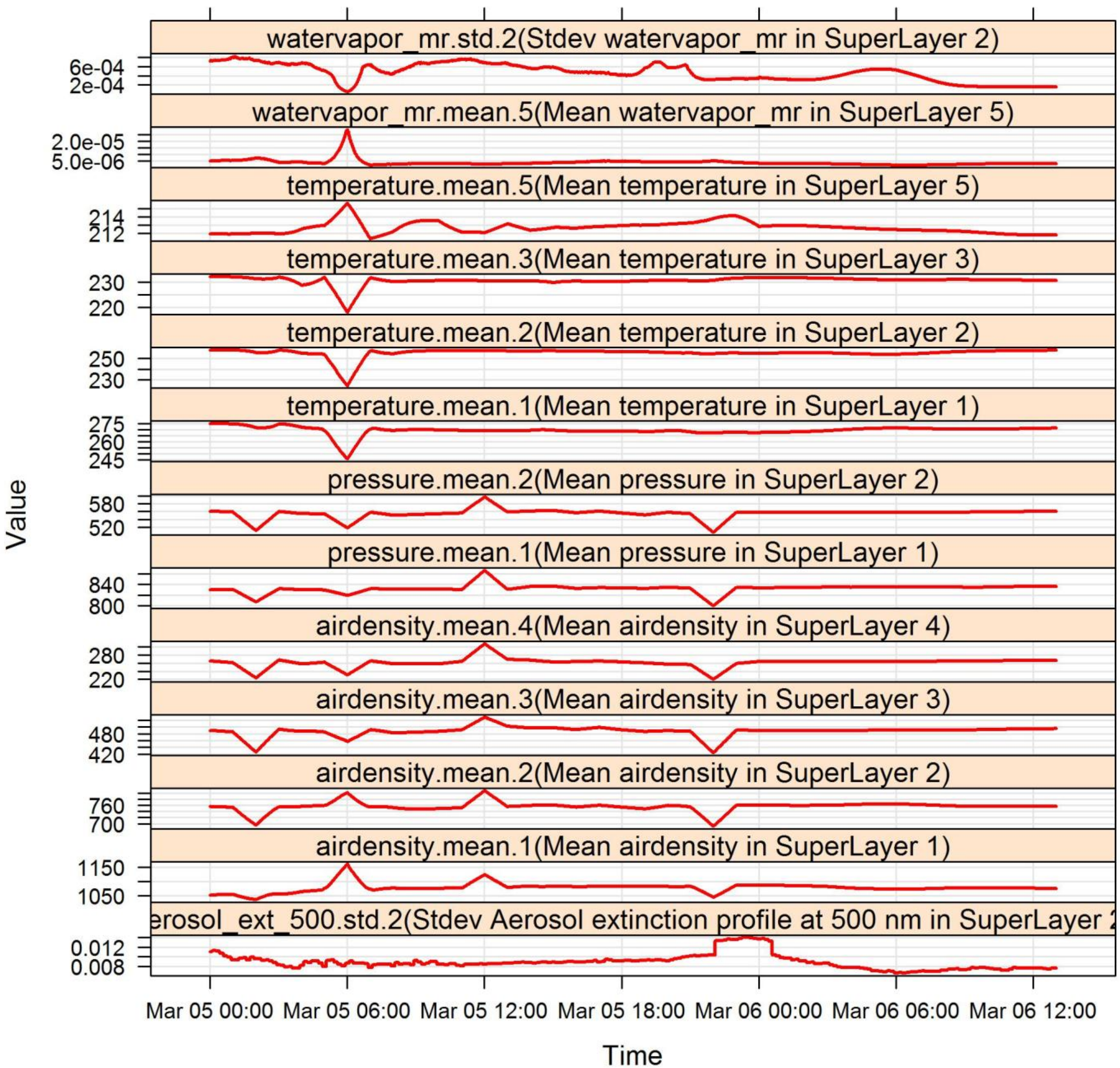




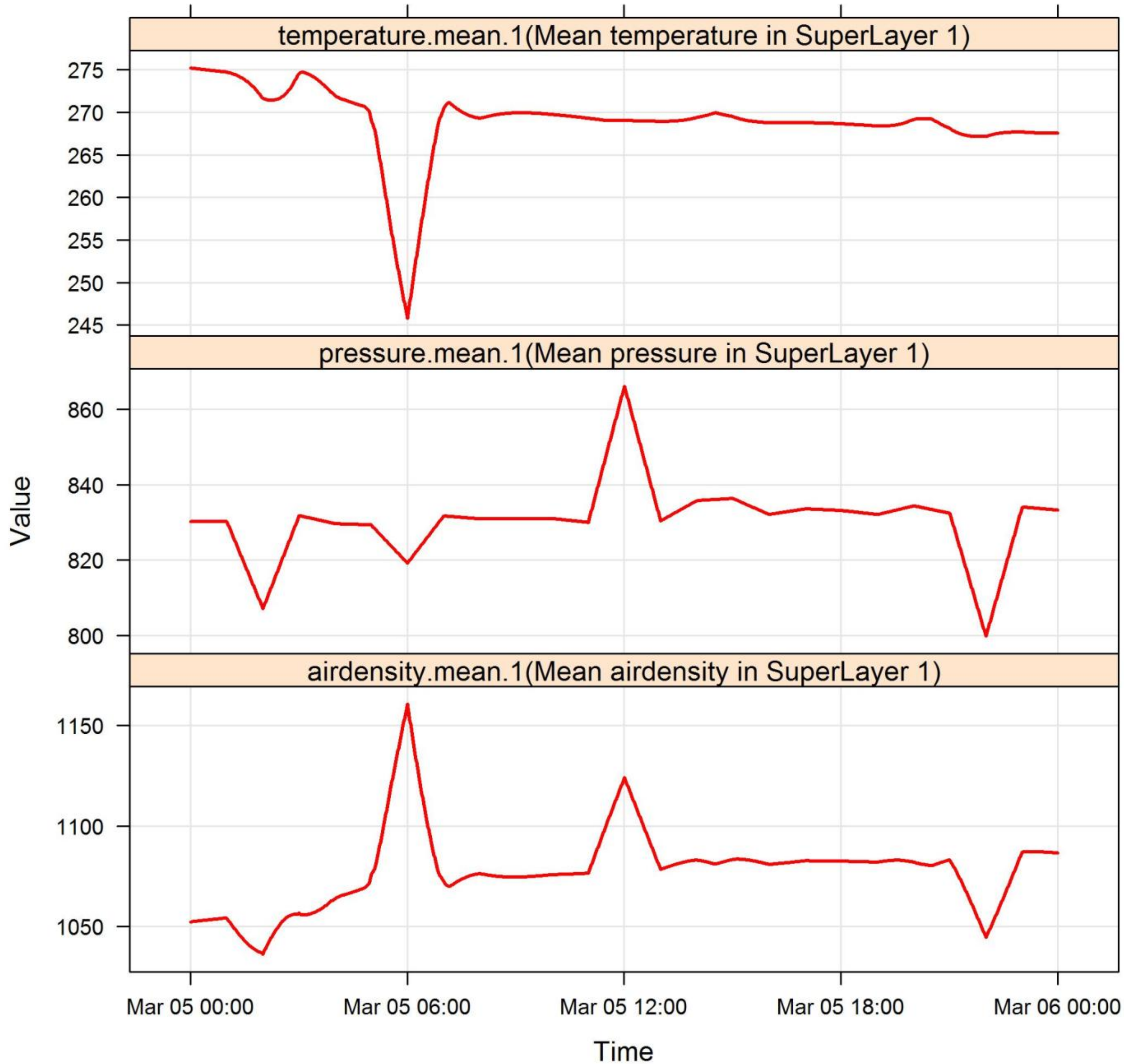
Mean watervapor_mr in SuperLayer 5
2.8e−05
2.6e−05
2.4e−05
2.2e−05
2e−05
1.8e−05
1.6e−05
1.4e−05
1.2e−05
1e−05
8e−06
6e−06
4e−06
2e−06
variability (stdev=5.769e−06) was high.
2003−03−04 00
2003−03−04 10
2003−03−04 20
2003−03−05 06
2003−03−05 16
2003−03−06 02
2003−03−06 12
2003−03−06 23



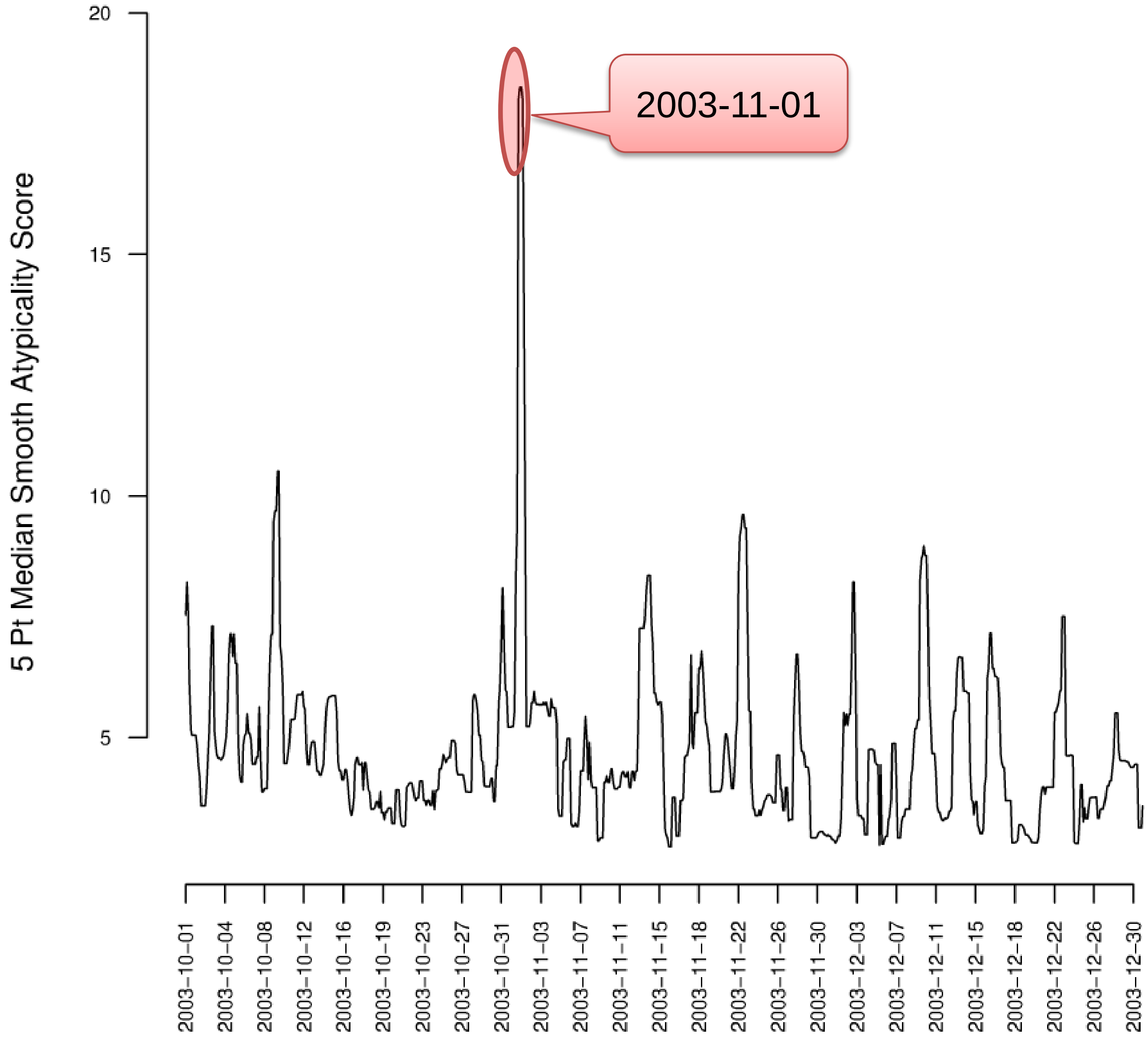
Mean watervapor_rh in SuperLayer 2
100
90
80
70
60
50
40
30
20
10
variability (stdev=17.45) was very high.
2003-03-04 00
2003-03-04 10
2003-03-04 20
2003-03-05 06
2003-03-05 16
2003-03-06 02
2003-03-06 12
2003-03-06 23



watervapor_mr.std.2(Stdev watervapor_mr in SuperLayer 2)
6e-04
2e-04
watervapor_mr.mean.5(Mean watervapor_mr in SuperLayer 5)
2.0e-05
5.0e-06
temperature.mean.5(Mean temperature in SuperLayer 5)
214
212
temperature.mean.3(Mean temperature in SuperLayer 3)
230
220
temperature.mean.2(Mean temperature in SuperLayer 2)
250
230
temperature.mean.1(Mean temperature in SuperLayer 1)
275
260
245
pressure.mean.2(Mean pressure in SuperLayer 2)
580
520
pressure.mean.1(Mean pressure in SuperLayer 1)
840
800
airdensity.mean.4(Mean airdensity in SuperLayer 4)
280
220
airdensity.mean.3(Mean airdensity in SuperLayer 3)
480
420
airdensity.mean.2(Mean airdensity in SuperLayer 2)
760
700
airdensity.mean.1(Mean airdensity in SuperLayer 1)
1150
1050
erosol_ext_500.std.2(Stdev Aerosol extinction profile at 500 nm in SuperLayer 2
0.012
0.008
Value
Mar 05 00:00
Mar 05 06:00
Mar 05 12:00
Mar 05 18:00
Mar 06 00:00
Mar 06 06:00
Mar 06 12:00
Time


temperature.mean.1(Mean temperature in SuperLayer 1)
275
270
265
260
255
250
245
pressure.mean.1(Mean pressure in SuperLayer 1)
860
840
820
800
airdensity.mean.1(Mean airdensity in SuperLayer 1)
1150
1100
1050
Value
Mar 05 00:00
Mar 05 06:00
Mar 05 12:00
Mar 05 18:00
Mar 06 00:00
Time

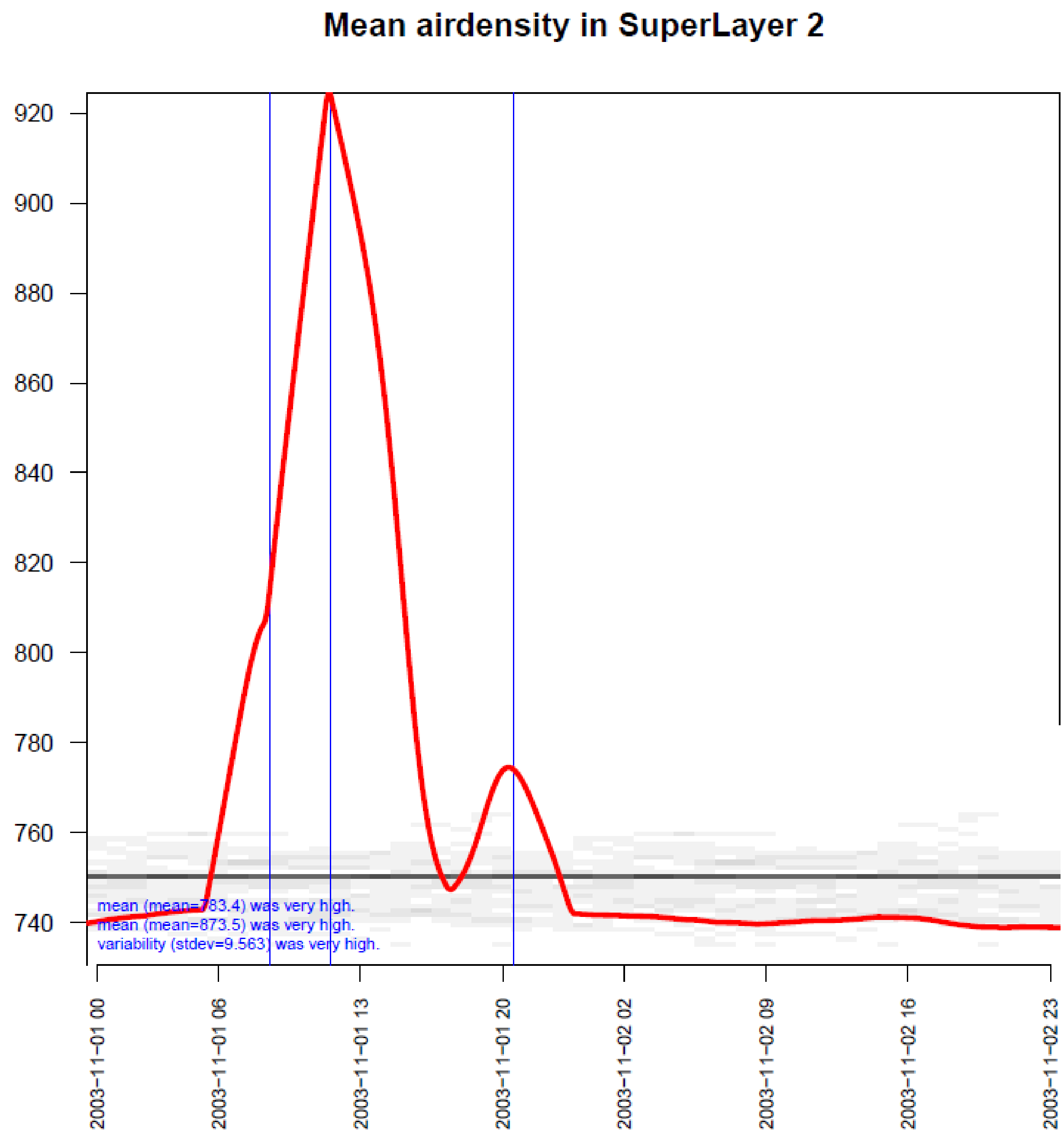
2003-10-01 to 2003-12-31



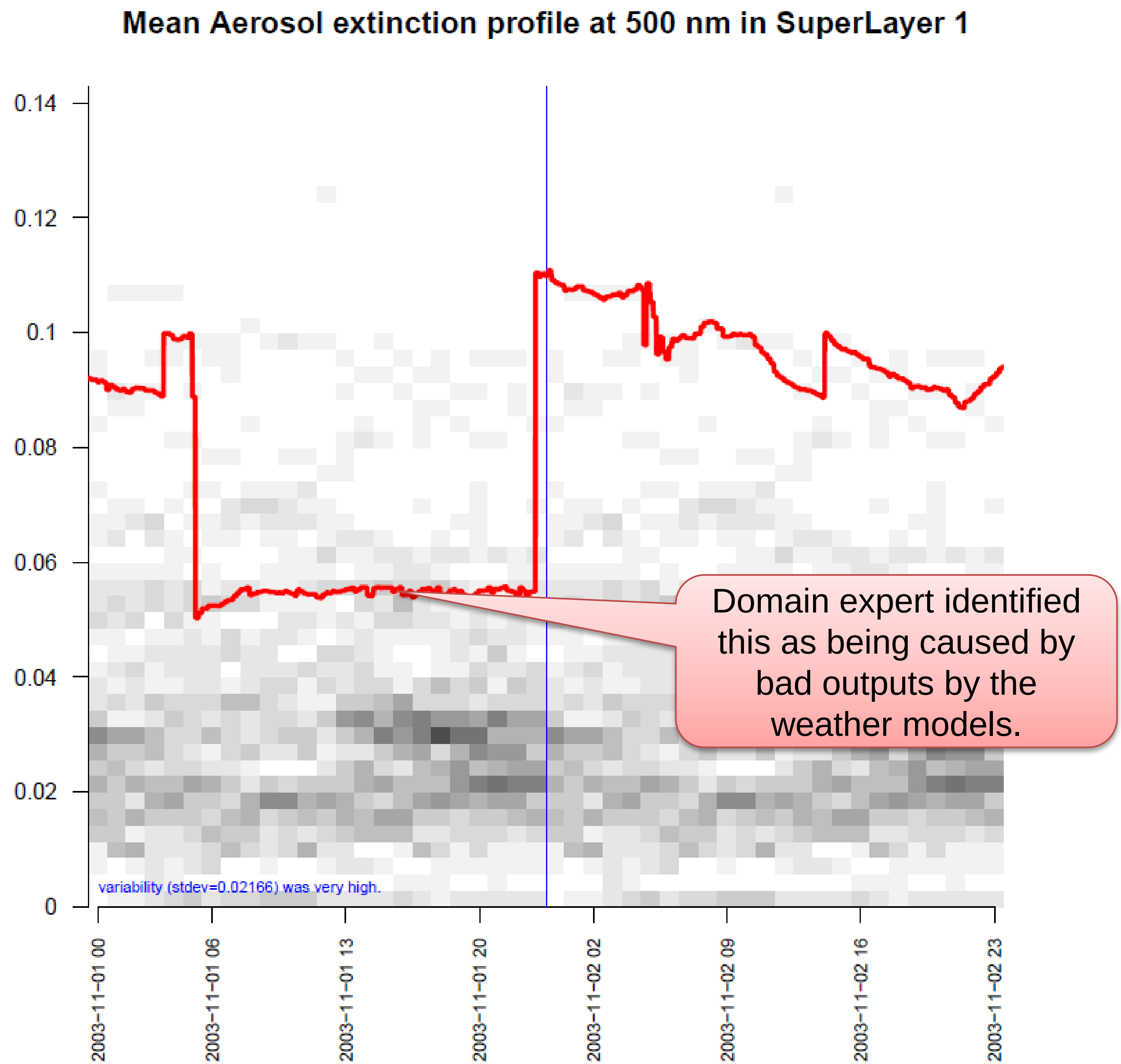
2003-10-01 to 2003-12-31

Atypicality Report (Sorted by Global Atypicality Score)

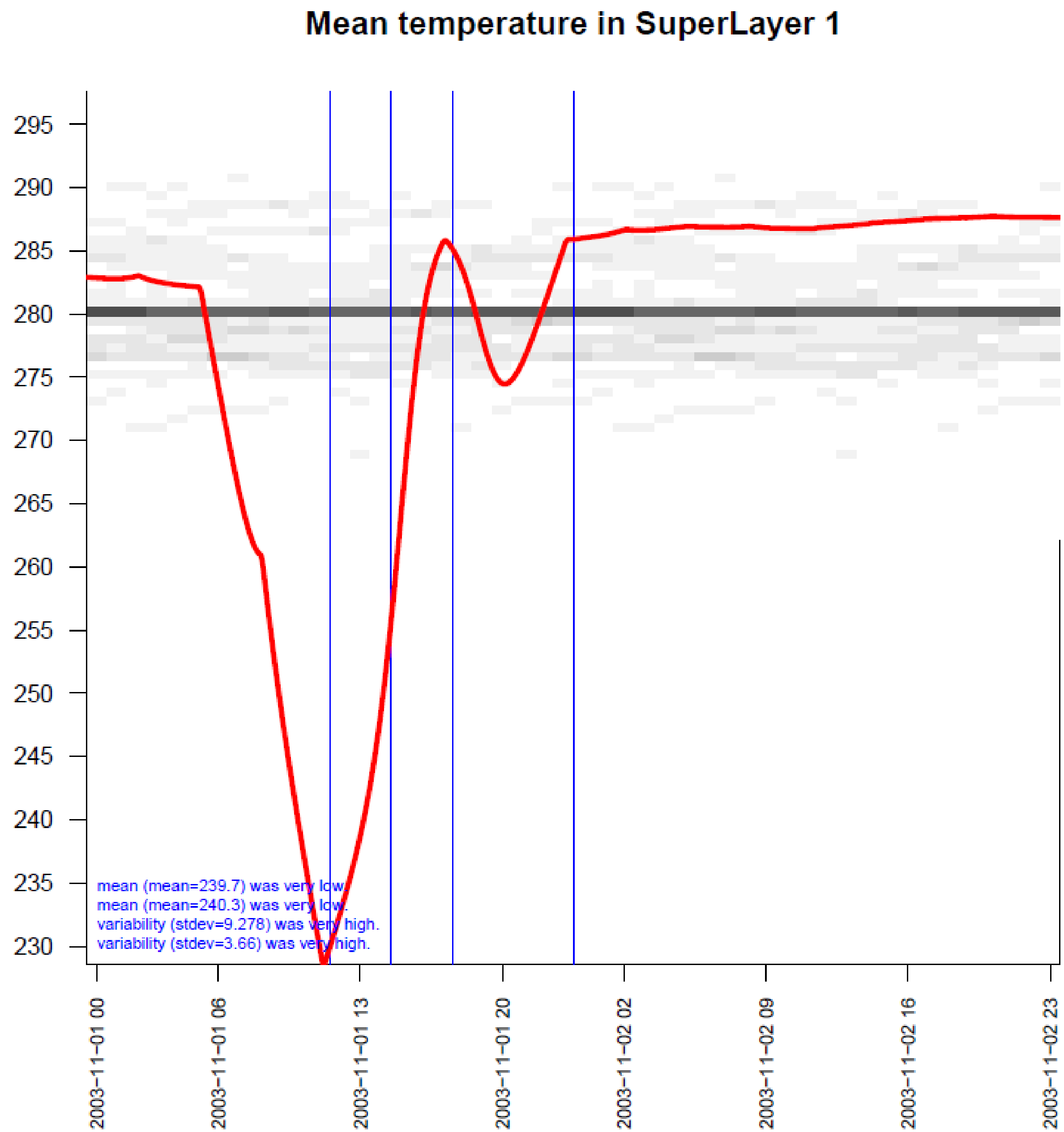
	GAS	Cluster#	Values Rationale	Slope Rationale
2003-11-01 18:00:00	20.41	89.00	temperature.mean.3 variability (stdev=9.901) was very high. temperature.mean.2 variability (stdev=9.542) was very high. temperature.mean.1 variability (stdev=9.278) was very high. watervapor_mr.mean.1 variability (stdev=0.003138) was very high. temperature.std.4 variability (stdev=1.674) was very high. watervapor_mr.mean.4 variability (stdev=1.295e-05) was very high. watervapor_mr.std.4 variability (stdev=1.303e-05) was very high. temperature.mean.5 mean (mean=196.5) was very low. watervapor_rh.std.1 variability (stdev=6.196) was very high. temperature.mean.4 mean (mean=199.1) was very low. cld_reliq.std.1 mean (mean=3.493) was very high. watervapor_rh.std.2 variability (stdev=4.728) was high. cld_dgeice.mean.4 variability (stdev=3.588) was high. watervapor_rh.mean.1 mean (mean=97.1) was high. watervapor_rh.std.4 variability (stdev=2.738) was high. watervapor_mr.mean.2 variability (stdev=0.0002479) was high. temperature.std.5 mean (mean=5.904) was high.	airdensity.mean.4 slope mean (mean=-0.2292) was very low. airdensity.mean.3 slope mean (mean=-0.401) was very low. airdensity.mean.5 slope mean (mean=-0.04399) was very low. watervapor_rh.mean.5 slope mean (mean=-0.2711) was very low. airdensity.mean.2 slope mean (mean=-0.5226) was very low. airdensity.std.4 slope mean (mean=-0.08265) was very low. airdensity.mean.1 slope mean (mean=-0.6624) was very low. watervapor_rh.std.5 slope mean (mean=-0.1514) was very low. watervapor_rh.mean.2 slope mean (mean=-0.4346) was very low. watervapor_rh.mean.4 slope mean (mean=-0.2387) was very low. temperature.std.2 slope variability (stdev=0.01017) was very high. airdensity.std.5 slope mean (mean=-0.01033) was very low. cld_dgeice.std.1 slope mean (mean=-0.1417) was very low. temperature.std.3 slope variability (stdev=0.007972) was very high. cld_dgeice.mean.1 slope mean (mean=-0.1343) was very low. watervapor_rh.mean.3 slope mean (mean=-0.3147) was very low. airdensity.std.2 slope mean (mean=-0.03818) was very low. airdensity.std.3 slope mean (mean=-0.02787) was very low. cld_iwp.std.4 slope mean (mean=-0.001321) was very low. temperature.std.1 slope mean (mean=-0.01115) was very low.
2003-10-09 12:00:00	19.31	96.00	cld_lwp.std.1 mean (mean=27.39) was very high. cld_reliq.std.1 variability (stdev=2.209) was very high. temperature.std.4 variability (stdev=0.5971) was very high. watervapor_mr.mean.1 mean (mean=0.01116) was high. watervapor_mr.std.2 mean (mean=0.001776) was high. cld_lwp.mean.1 variability (stdev=8.735) was high. watervapor_rh.mean.1 mean (mean=92.24) was high.	cld_lwp.std.3 slope mean (mean=-0.01284) was very low. cld_lwp.mean.3 slope mean (mean=-0.002278) was very low. cld_reliq.std.3 slope mean (mean=-0.007742) was very low. cld_iwp.mean.3 slope mean (mean=-0.09366) was very low. cld_iwp.std.3 slope mean (mean=-0.08683) was very low. cld_reliq.mean.2 slope mean (mean=-0.04169) was very low. cld_lwp.mean.2 slope mean (mean=-0.09878) was very low. cld_reliq.mean.3 slope mean (mean=-0.001422) was very low. watervapor_mr.mean.3 slope mean (mean=-3.375e-06) was very low. watervapor_mr.mean.2 slope mean (mean=-1.298e-05) was very low. watervapor_mr.std.4 slope mean (mean=-2.235e-07) was very low. cld_lwp.std.2 slope mean (mean=-0.08157) was very low. cld_reliq.std.2 slope mean (mean=-0.01961) was very low. watervapor_mr.mean.4 slope mean (mean=-1.416e-07) was very low. watervapor_mr.std.3 slope mean (mean=-1.951e-06) was very low. watervapor_rh.mean.3 slope mean (mean=-0.2592) was very low. cld_iwp.std.2 slope mean (mean=-0.06593) was very low. watervapor_rh.mean.2 slope mean (mean=-0.2515) was very low. cld_dgeice.mean.2 slope mean (mean=-0.2021) was very low. cld_dgeice.mean.3 slope mean (mean=-0.2067) was very low.
			watervapor_rh.mean.5 mean (mean=80.66) was very high. watervapor_mr.std.5 mean (mean=1.678e-05) was very high. airdensity.mean.3 mean (mean=576.2) was very high. airdensity.mean.2 mean (mean=873.5) was very high.	



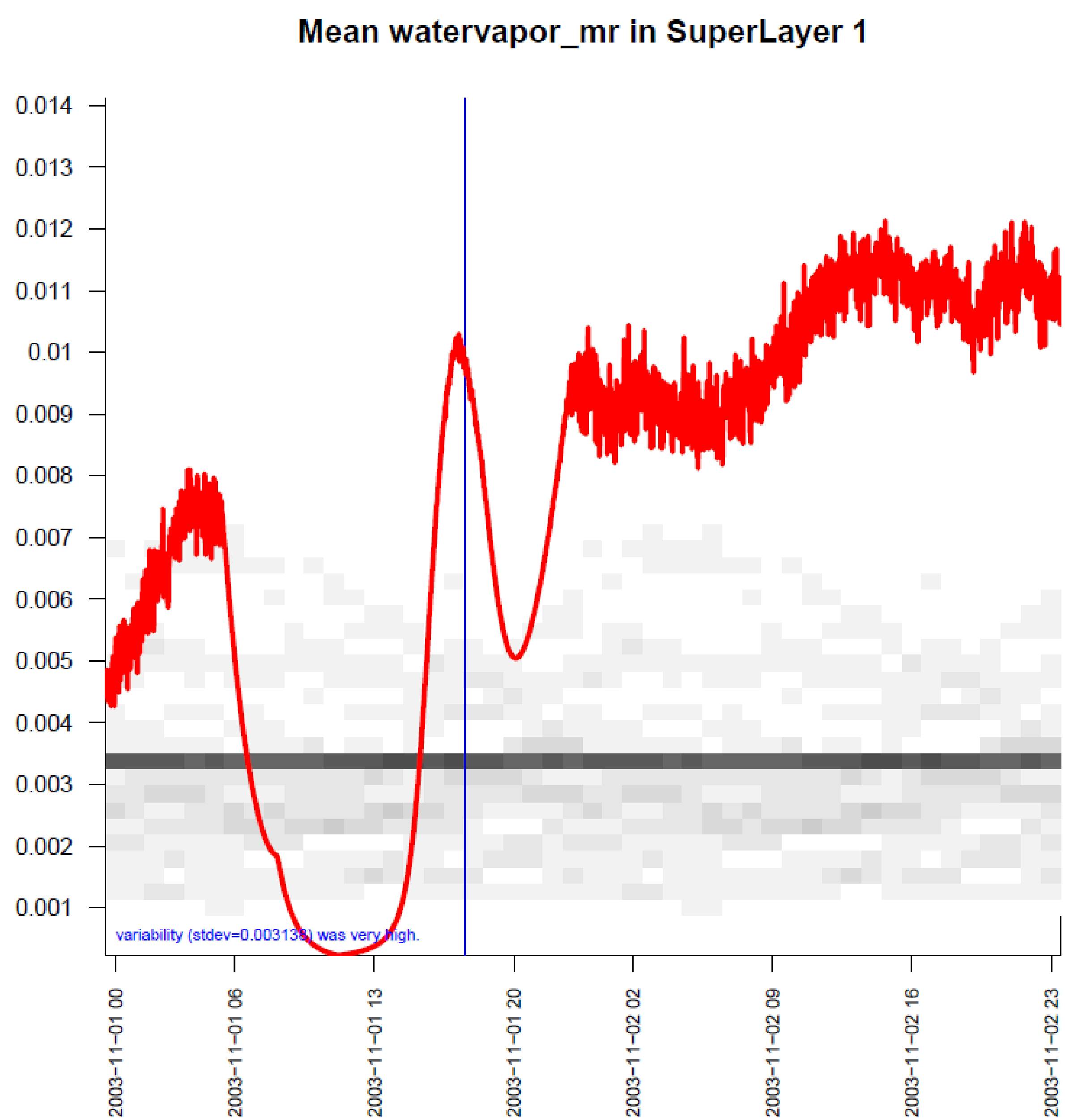
Mean airdensity in SuperLayer 2
920
900
880
860
840
820
800
780
760
740
mean (mean=783.4) was very high.
mean (mean=873.5) was very high.
variability (stdev=9.563) was very high.
2003−11−01 00
2003−11−01 06
2003−11−01 13
2003−11−01 20
2003−11−02 02
2003−11−02 09
2003−11−02 16
2003−11−02 23



Mean Aerosol extinction profile at 500 nm in SuperLayer 1
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
variability (stdev=0.02166) was very high.
2003−11−01 00
2003−11−01 06
2003−11−01 13
2003−11−01 20
2003−11−02 02
2003−11−02 09
2003−11−02 16
2003−11−02 23
Domain expert identified this as being caused by bad outputs by the weather models.



Mean temperature in SuperLayer 1
295
290
285
280
275
270
265
260
255
250
245
240
235
230
mean (mean=239.7) was very low.
mean (mean=240.3) was very low.
variability (stdev=9.278) was very high.
variability (stdev=3.66) was very high.
2003−11−01 00
2003−11−01 06
2003−11−01 13
2003−11−01 20
2003−11−02 02
2003−11−02 09
2003−11−02 16
2003−11−02 23



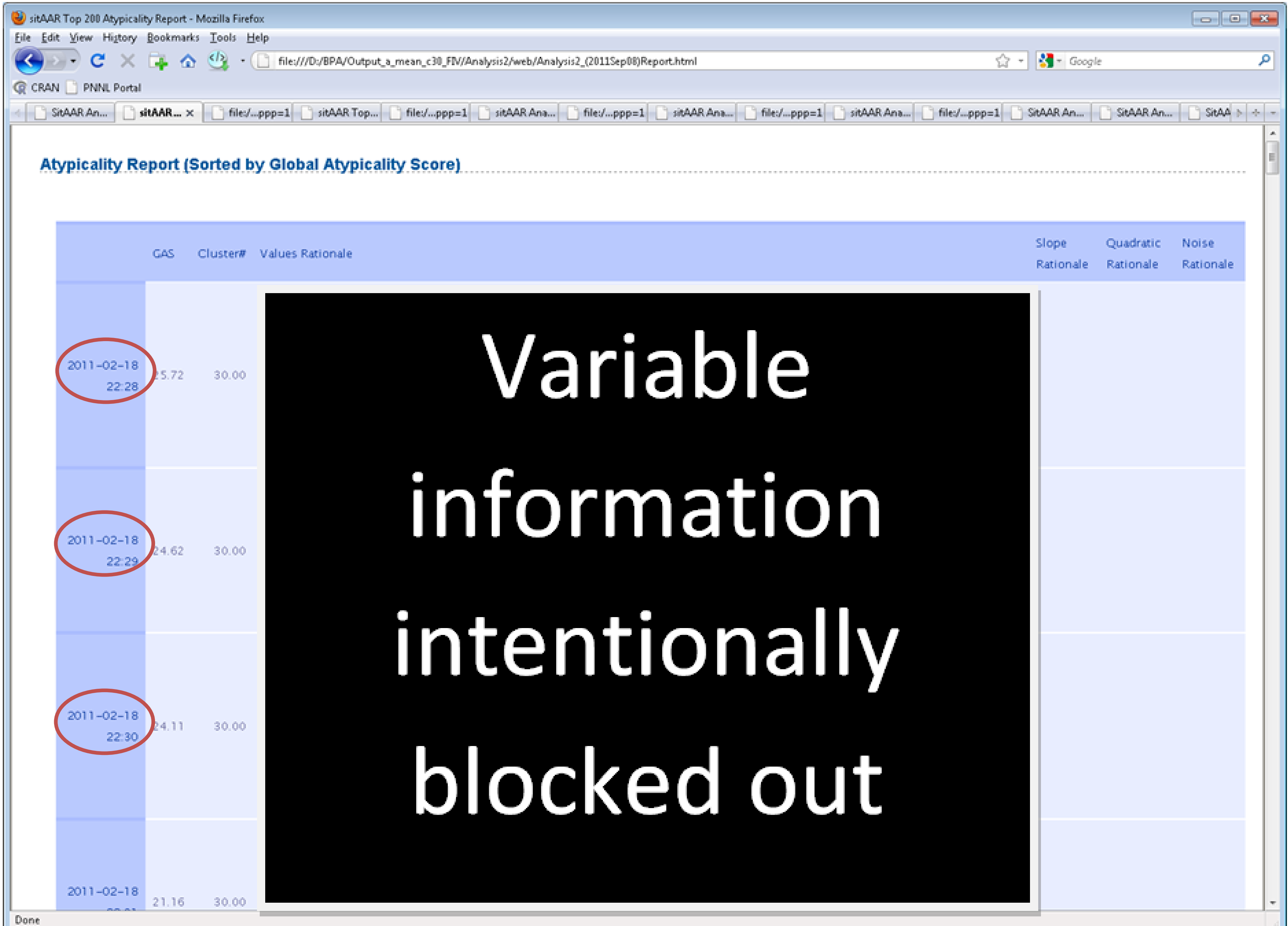
Mean watervapor_mr in SuperLayer 1
0.014
0.013
0.012
0.011
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
variability (stdev=0.003138) was very high.
2003−11−01 00
2003−11−01 06
2003−11−01 13
2003−11−01 20
2003−11−02 02
2003−11−02 09
2003−11−02 16
2003−11−02 23

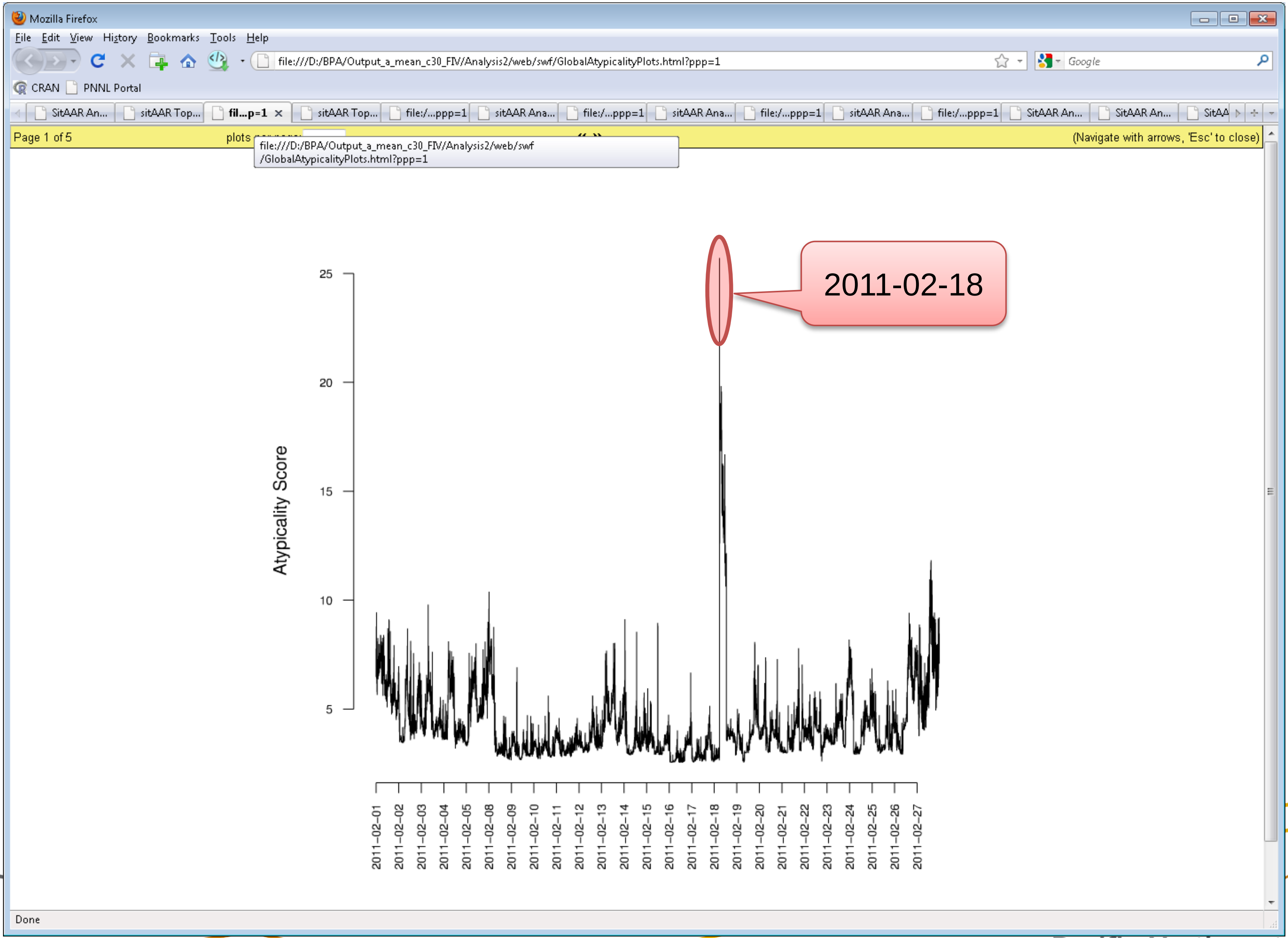
Weather Conclusions

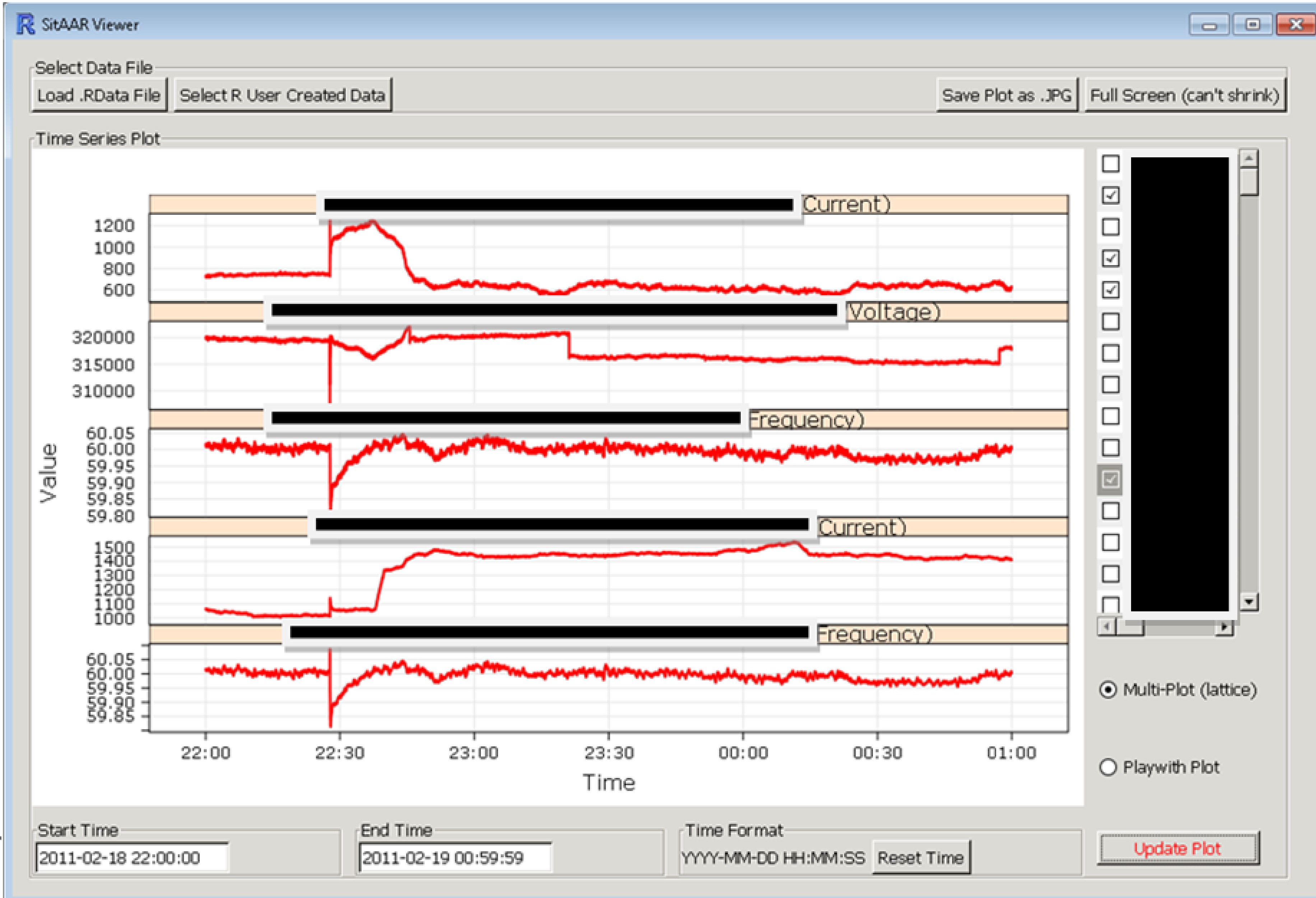
- It's hard to analyze weather data with only a few years of data. More data will better define the patterns.
- Some of the identified atypical events could have been detected by walking outside (i.e. storm front coming).
- Some of the identified atypical events are issues in the data, or in the output of the models. This process is very good at finding bad data.

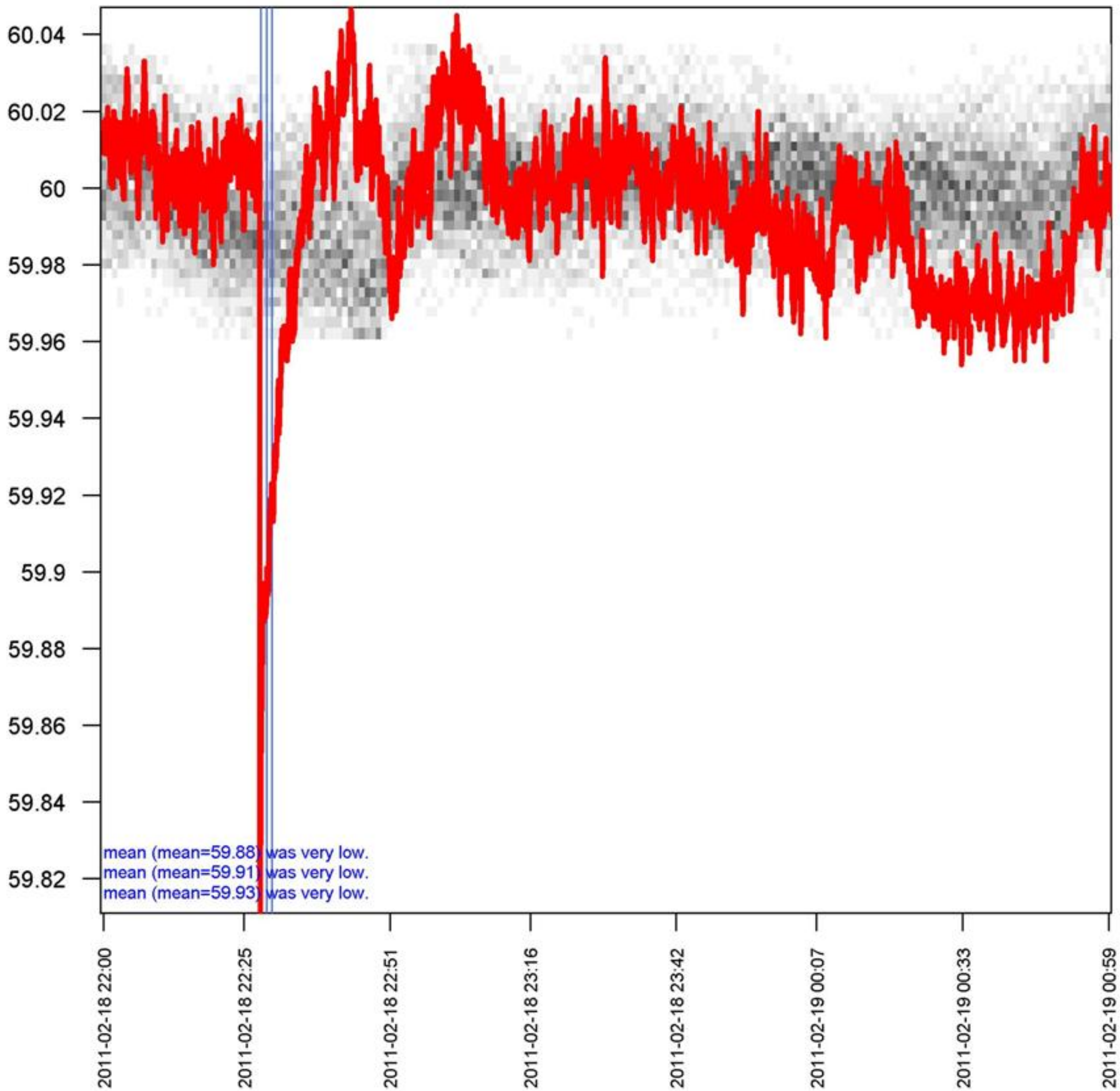
Power Grid Data

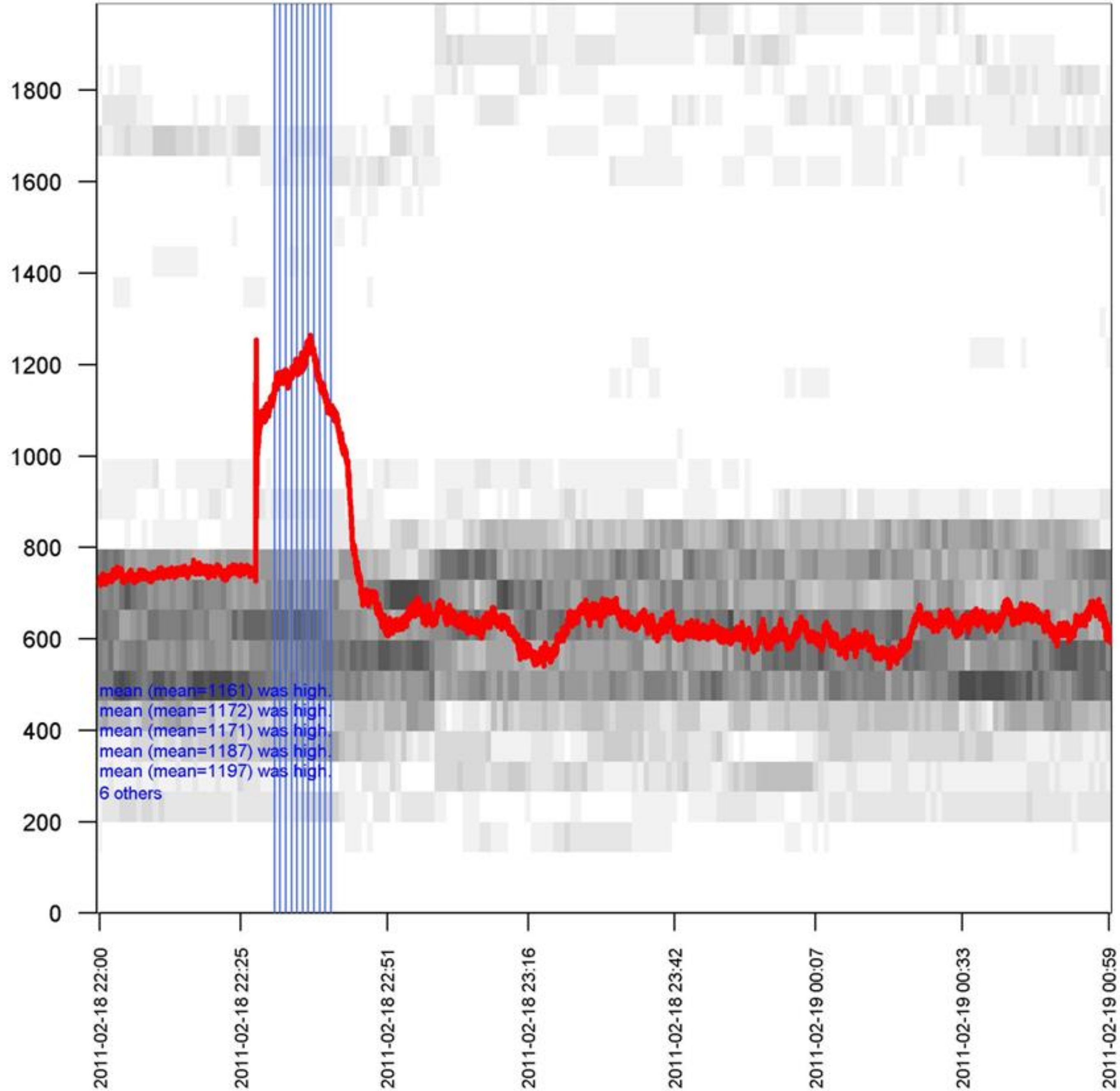
- Six months of PMU data was analyzed. The data consists of hundreds of variables recorded 30 times a second.
- Variables are related to frequency, current, and voltage and are taken from many locations.
- Variable names are blocked out due to non-disclosure agreements.


sitAAR Top 200 Atypicality Report - Mozilla Firefox
File Edit View History Bookmarks Tools Help
file:///D:/BPA/Output_a_mean_c30_FIV/Analysis2/web/Analysis2_(2011Sep08)Report.html
Google
CRAN PNNL Portal
SitAAR An... sitAAR... file:/...ppp=1 sitAAR Top... file:/...ppp=1 sitAAR Ana... file:/...ppp=1 sitAAR Ana... file:/...ppp=1 sitAAR Ana... file:/...ppp=1 SitAAR An... SitAAR An... SitAA
Atypicality Report (Sorted by Global Atypicality Score)
GAS
Cluster#
Values Rationale
Slope Rationale
Quadratic Rationale
Noise Rationale
2011-02-18 22:28
25.72
30.00
2011-02-18 22:29
24.62
30.00
2011-02-18 22:30
24.11
30.00
2011-02-18
21.16
30.00
Variable information intentionally blocked out
Done

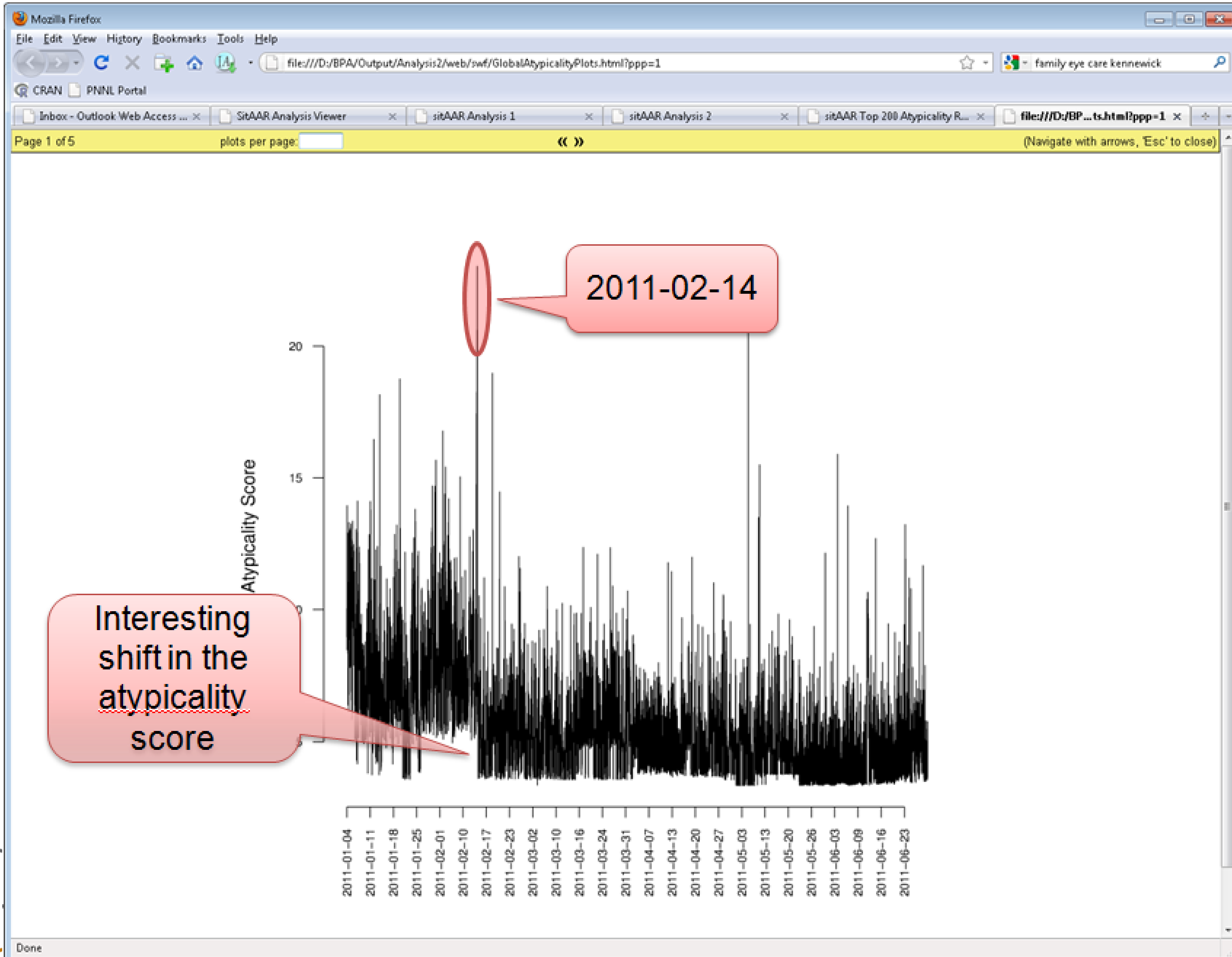

Mozilla Firefox
File Edit View History Bookmarks Tools Help
file:///D:/BPA/Output_a_mean_c30_FIV/Analysis2/web/swf/GlobalAtypicalityPlots.html?ppp=1
CRAN PNNL Portal
Page 1 of 5
(Navigate with arrows, 'Esc' to close)
2011-02-18
Atypicality Score
25
20
15
10
5
2011-02-01
2011-02-02
2011-02-03
2011-02-04
2011-02-05
2011-02-08
2011-02-09
2011-02-10
2011-02-11
2011-02-12
2011-02-13
2011-02-14
2011-02-15
2011-02-16
2011-02-17
2011-02-18
2011-02-19
2011-02-20
2011-02-21
2011-02-22
2011-02-23
2011-02-24
2011-02-25
2011-02-26
2011-02-27
Done

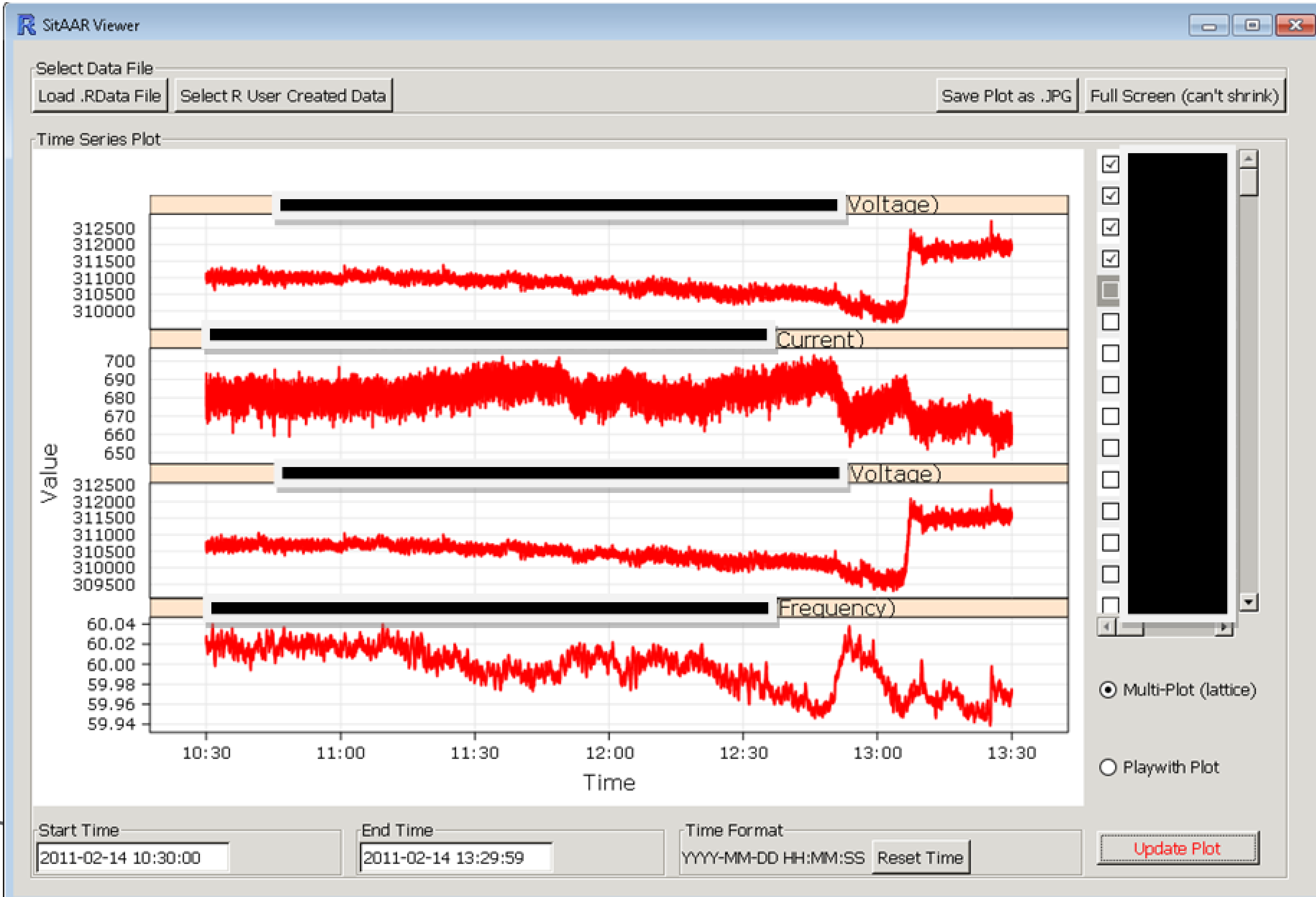

SitAAR Viewer
Select Data File
Load .RData File
Select R User Created Data
Save Plot as .JPG
Full Screen (can't shrink)
Time Series Plot
[redacted] Current)
1200
1000
800
600
[redacted] Voltage)
320000
315000
310000
[redacted] Frequency)
60.05
60.00
59.95
59.90
59.85
59.80
[redacted] Current)
1500
1400
1300
1200
1100
1000
[redacted] Frequency)
60.05
60.00
59.95
59.90
59.85
Value
22:00
22:30
23:00
23:30
00:00
00:30
01:00
Time
[redacted]
Multi-Plot (lattice)
Playwith Plot
Start Time
2011-02-18 22:00:00
End Time
2011-02-19 00:59:59
Time Format
YYYY-MM-DD HH:MM:SS
Reset Time
Update Plot


60.04
60.02
60
59.98
59.96
59.94
59.92
59.9
59.88
59.86
59.84
59.82
mean (mean=59.88) was very low.
mean (mean=59.91) was very low.
mean (mean=59.93) was very low.
2011-02-18 22:00
2011-02-18 22:25
2011-02-18 22:51
2011-02-18 23:16
2011-02-18 23:42
2011-02-19 00:07
2011-02-19 00:33
2011-02-19 00:59

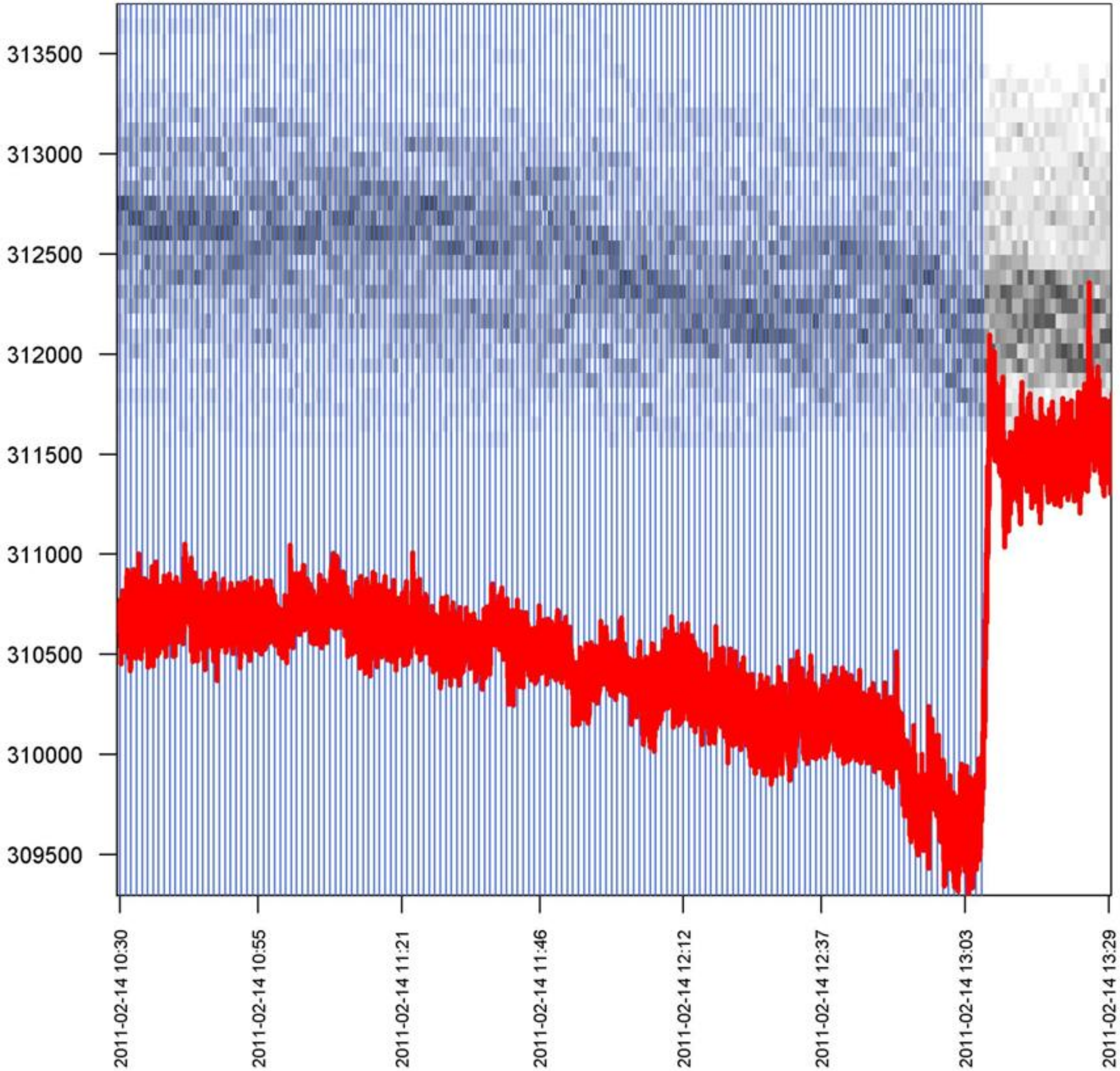


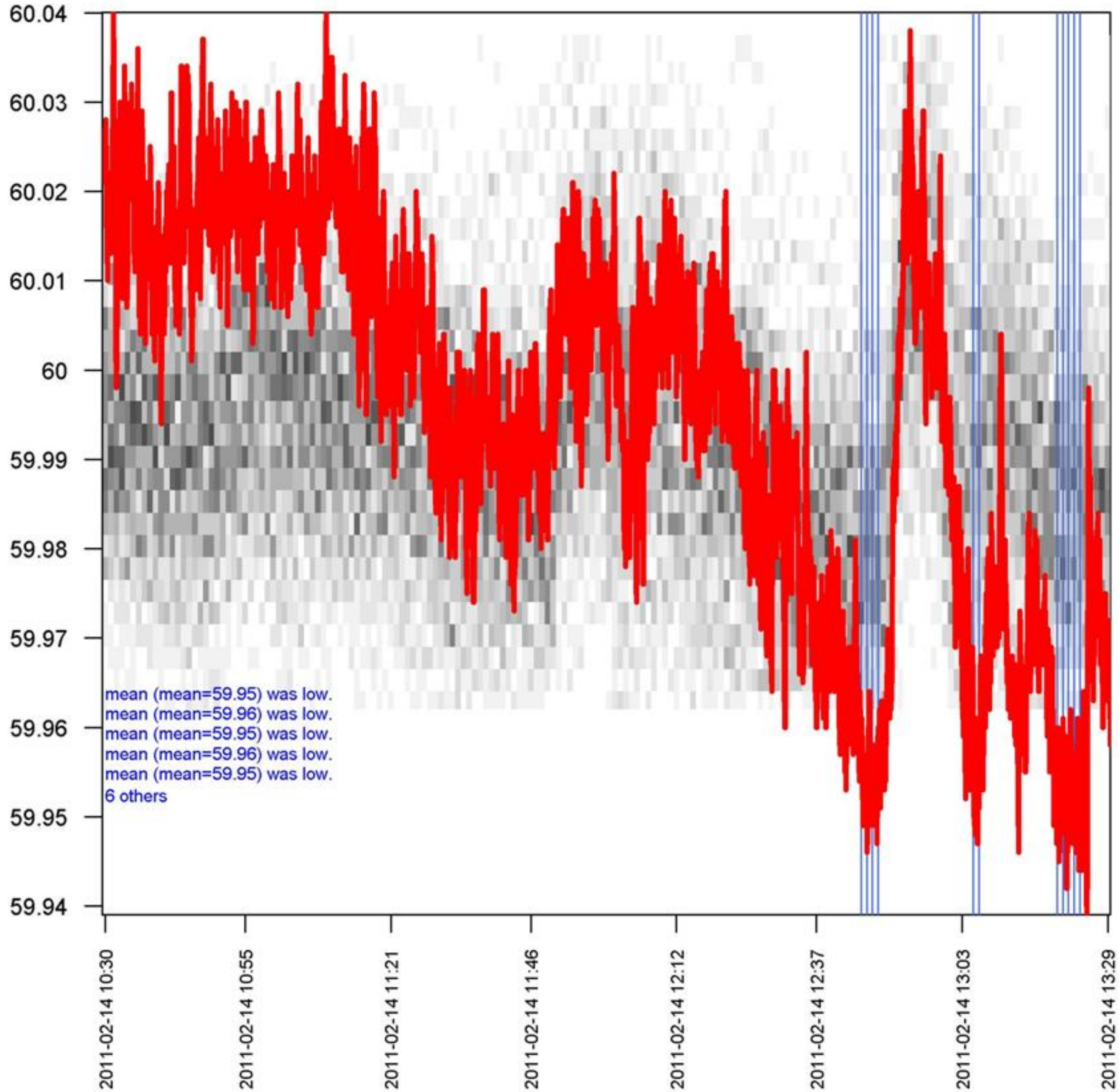
1800
1600
1400
1200
1000
800
600
400
200
0
mean (mean=1161) was high.
mean (mean=1172) was high.
mean (mean=1171) was high.
mean (mean=1187) was high.
mean (mean=1197) was high.
6 others
2011-02-18 22:00
2011-02-18 22:25
2011-02-18 22:51
2011-02-18 23:16
2011-02-18 23:42
2011-02-19 00:07
2011-02-19 00:33
2011-02-19 00:59


Mozilla Firefox
File Edit View History Bookmarks Tools Help
file:///D:/BPA/Output/Analysis2/web/swf/GlobalAtypicalityPlots.html?ppp=1
family eye care kennewick
CRAN
PNNL Portal
Inbox - Outlook Web Access ...
SitAAR Analysis Viewer
sitAAR Analysis 1
sitAAR Analysis 2
sitAAR Top 200 Atypicality R...
file:///D:/BP...ts.html?ppp=1
Page 1 of 5
plots per page:
(Navigate with arrows, 'Esc' to close)
2011-02-14
Interesting shift in the atypicality score
Atypicality Score
20
15
2011-01-04
2011-01-11
2011-01-18
2011-01-25
2011-02-01
2011-02-10
2011-02-17
2011-02-23
2011-03-02
2011-03-10
2011-03-16
2011-03-24
2011-03-31
2011-04-07
2011-04-13
2011-04-20
2011-04-27
2011-05-03
2011-05-13
2011-05-20
2011-05-26
2011-06-03
2011-06-09
2011-06-16
2011-06-23
Done



SitAAR Viewer
Select Data File
Load .RData File
Select R User Created Data
Save Plot as .JPG
Full Screen (can't shrink)
Time Series Plot
[redacted] Voltage)
312500
312000
311500
311000
310500
310000
[redacted] Current)
700
690
680
670
660
650
[redacted] Voltage)
312500
312000
311500
311000
310500
310000
309500
[redacted] Frequency)
60.04
60.02
60.00
59.98
59.96
59.94
Value
10:30
11:00
11:30
12:00
12:30
13:00
13:30
Time
[redacted]
Multi-Plot (lattice)
Playwith Plot
Start Time
2011-02-14 10:30:00
End Time
2011-02-14 13:29:59
Time Format
YYYY-MM-DD HH:MM:SS
Reset Time
Update Plot


313500
313000
312500
312000
311500
311000
310500
310000
309500
2011-02-14 10:30
2011-02-14 10:55
2011-02-14 11:21
2011-02-14 11:46
2011-02-14 12:12
2011-02-14 12:37
2011-02-14 13:03
2011-02-14 13:29


60.04
60.03
60.02
60.01
60
59.99
59.98
59.97
59.96
59.95
59.94
mean (mean=59.95) was low.
mean (mean=59.96) was low.
mean (mean=59.95) was low.
mean (mean=59.96) was low.
mean (mean=59.95) was low.
6 others
2011-02-14 10:30
2011-02-14 10:55
2011-02-14 11:21
2011-02-14 11:46
2011-02-14 12:12
2011-02-14 12:37
2011-02-14 13:03
2011-02-14 13:29

Next Steps
SitAAR
(Situation Awareness Alerts in Real-time)

- We have demonstrated a reporting system which finds atypicalities and typical patterns. Next step is to convert this to a real-time process.
  - Process enough data within the domain to be able to establish typical patterns.
  - Use "active learning" to refine the patterns with domain input.
  - Convert to a classification system, using classification rules to identify the patterns.
  - Allow for the creation of new patterns, as they develop.
  - Add Shewhart and other control chart alerts.

Contact Information

- Brett Amidan
  - b.amidan@pnl.gov
  - 509 375-3692
- Tom Ferryman
  - tom.ferryman@pnl.gov
  - 509 375-3888