

Better Data Quality for Better Data Science with the PI System

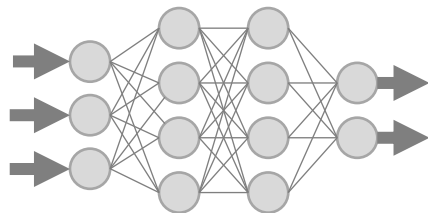
Brandon Perry



customer



$f(x)$



Symptom: losing money to shutdowns

Cause: unexpected equipment failure

Project: predict equipment failure

Symptom: many false alerts

Cause: poor data accuracy

Project: improve the data accuracy

Symptom: many false diagnoses

Cause: poor data interpretation

Data Quality

some common dimensions:

-Accuracy

-Believability

-Completeness

-Ease of
understanding

-Relevancy

-Timeliness

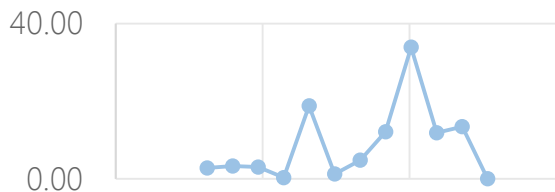
-Accessibility



Fermenter 13 bottom heater

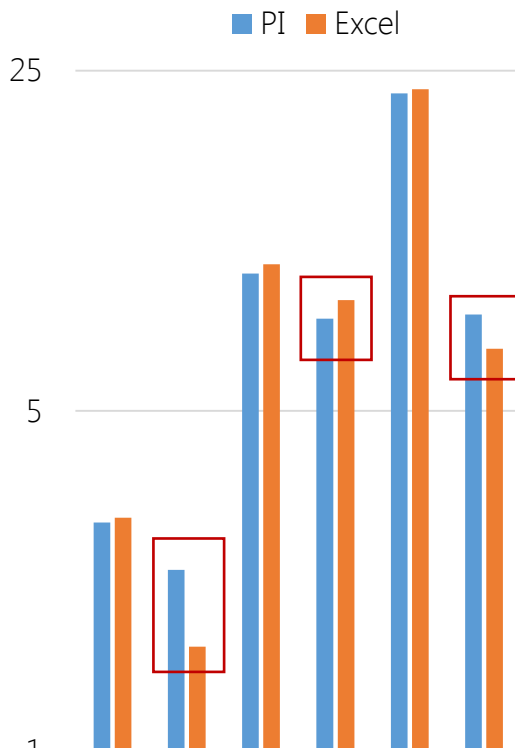
acsbrew.BREWERY.B2_CL_C1_FV13_TIC1550A/OUT.CV
Bottom TIC OUT [Control Value]

Raw data for 1h



Time	Value
8/13/18 18:03	2.77
8/13/18 18:08	3.28
8/13/18 18:13	3.00
8/13/18 18:18	0.28
8/13/18 18:23	18.78
8/13/18 18:28	1.23
8/13/18 18:33	4.79
8/13/18 18:38	12.10
8/13/18 18:43	33.90
8/13/18 18:48	11.84
8/13/18 18:53	13.42
8/13/18 18:58	0.00

Averages every 10 minutes



Time	Average (PI)	Average (Excel)	% Error
18:00	2.95	3.02	3
18:10	2.36	1.64	-30
18:20	9.58	10.00	4
18:30	7.73	8.44	9
18:40	22.45	22.87	2
18:50	7.89	6.71	-15

Data Quality

I. Why it matters

Impact

II. What it is

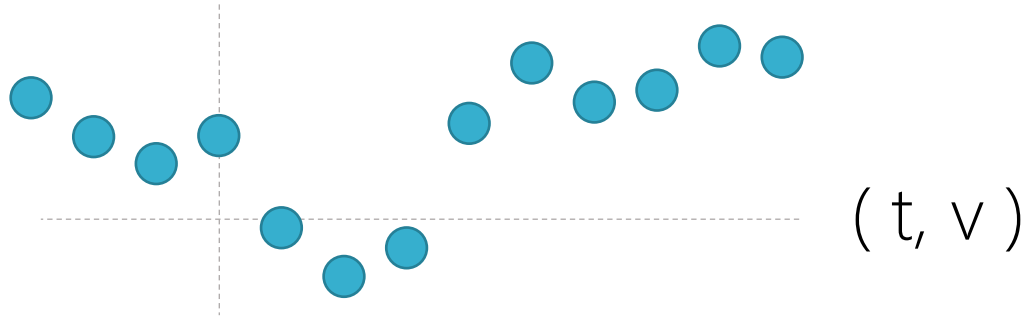
Understanding

III. What to do

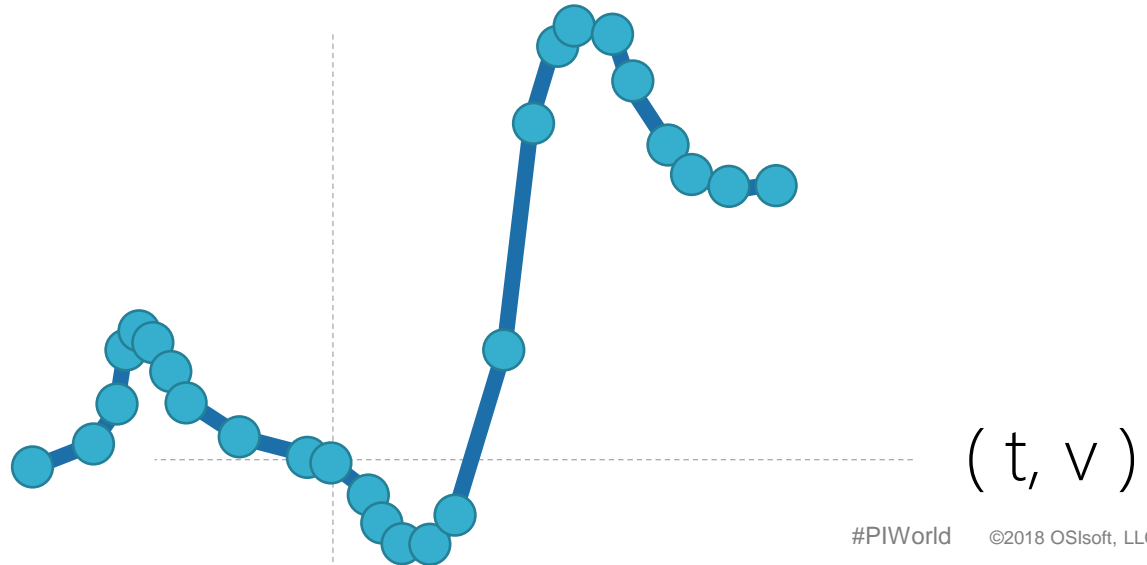
Action

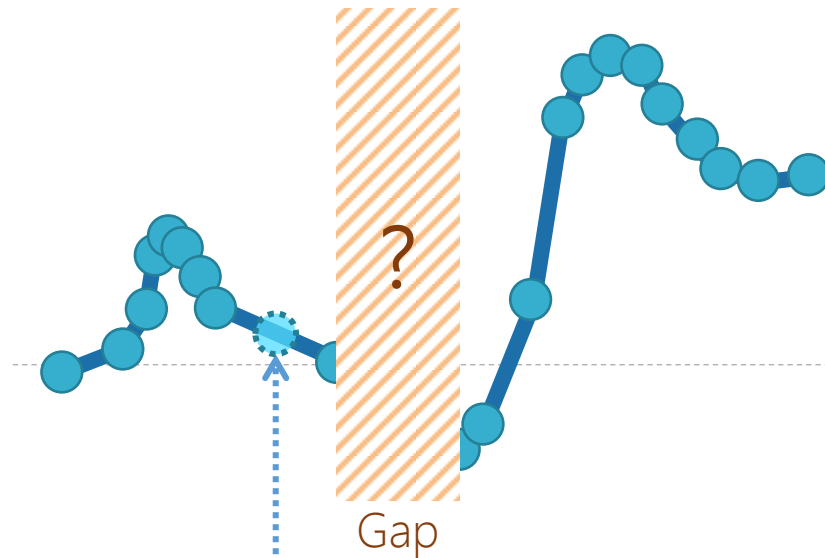
II. What it is

Time Series "Samples"



Time Sequence "Signal"





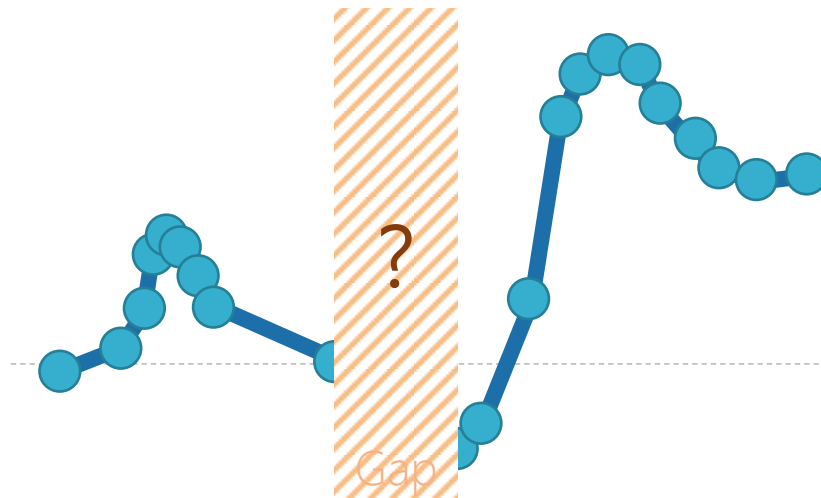
(t , $v_{\text{temperature}}$)

Interpolation

9/5/18 21:31 991.42

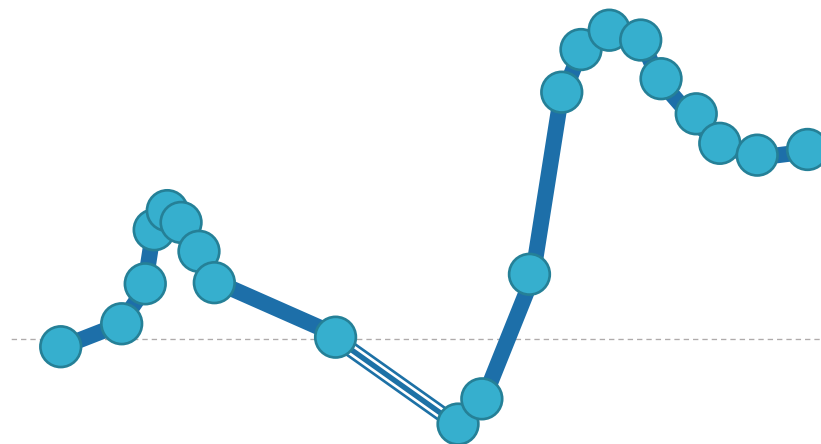
	A	B	C	D
2	9/5/18 21:22	990.34		
3	9/5/18 21:26	994.02		
4	9/5/18 21:35	989.96		
5	9/5/18 21:36	SOURCE OFFLINE		
6	9/5/18 22:01	986.92		
7	9/5/18 22:02	987.54		

Gaps
Known



9/5/18 21:35	989.96
9/5/18 21:36	SOURCE OFFLINE
9/5/18 22:01	986.92

Gaps
Unknown



9/5/18 21:35	989.96
9/5/18 22:01	986.92

☒ Questionable

this value might not be useful

☒ Substituted

this value was modified

☒ Annotated

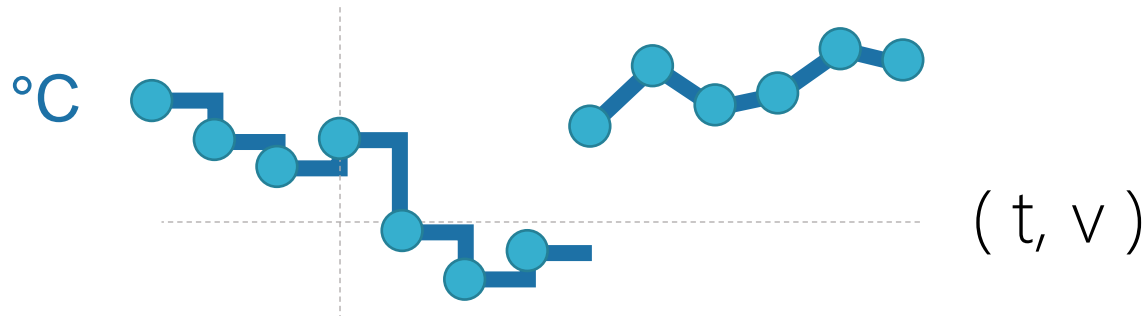
this value has a note attached

Complex
Quality

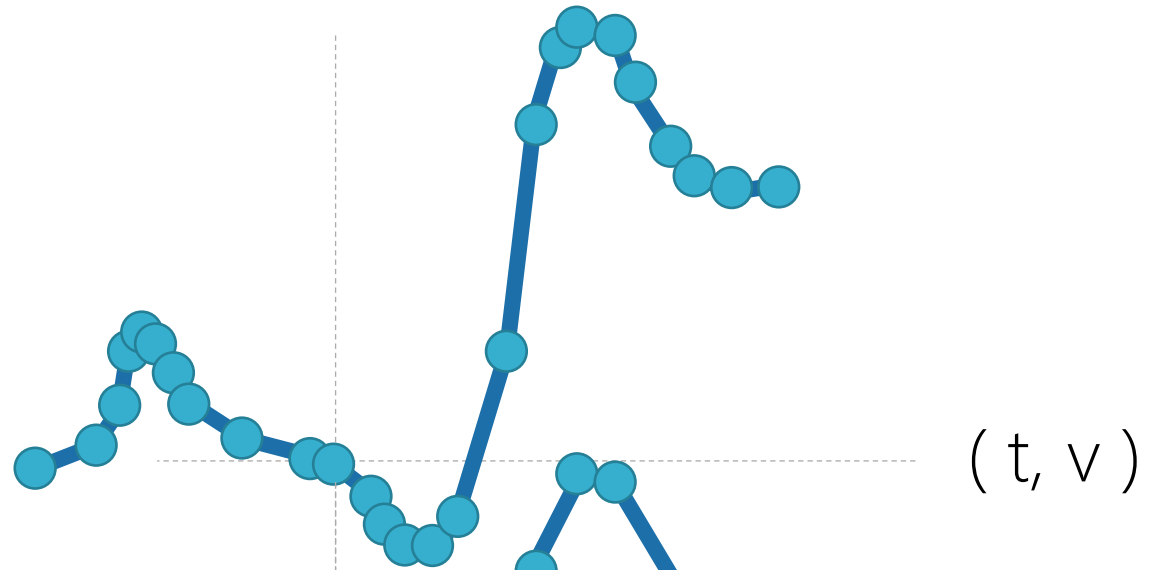
$\left[\begin{array}{l} \text{value: } 42.0 \\ \text{quality: Uncertain - Last Usable Value} \end{array} \right]$

Quality as reported by some sources

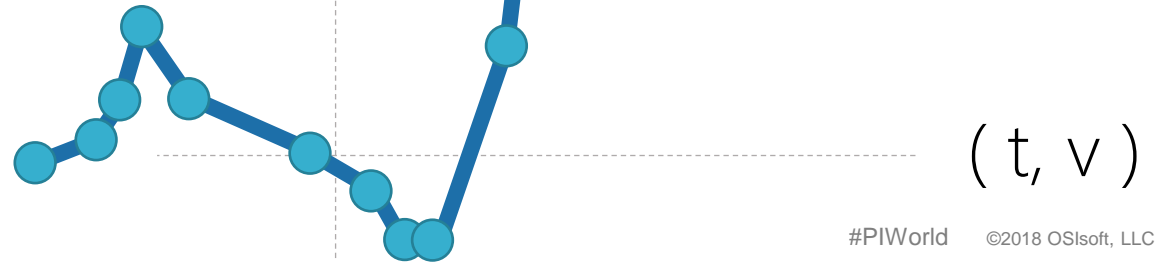
Metadata



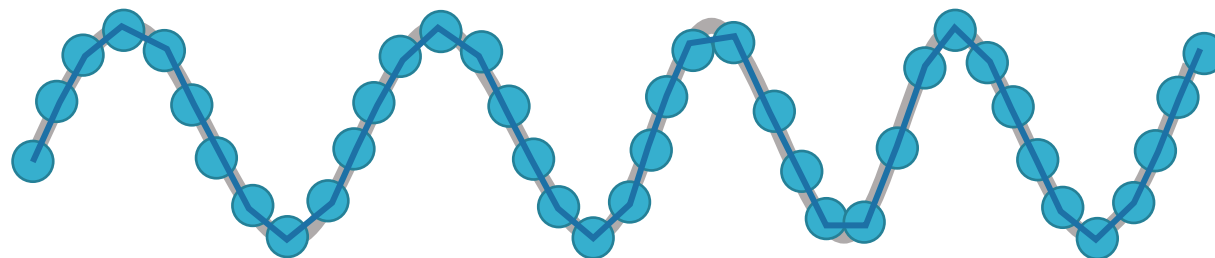
Uncompressed



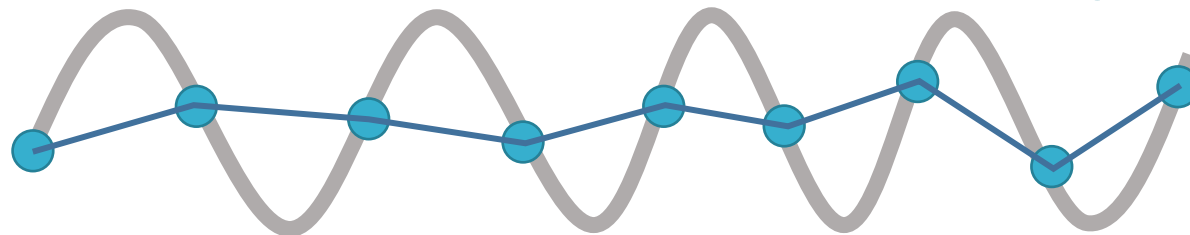
Compressed



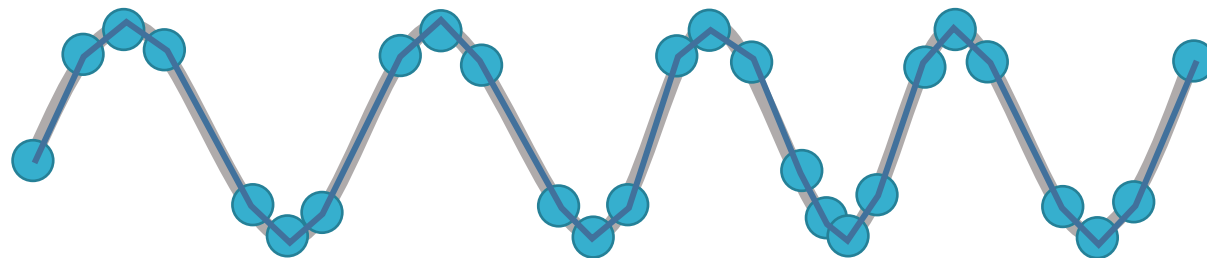
Well-sampled



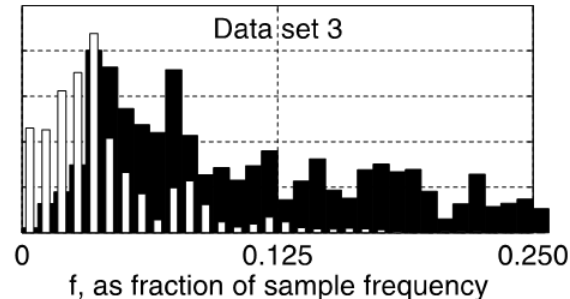
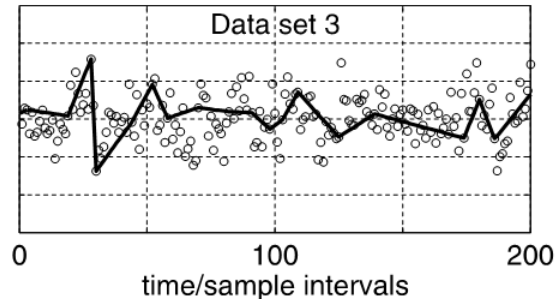
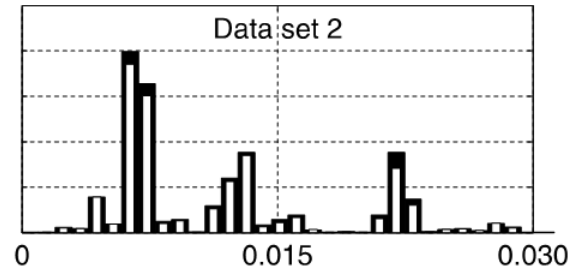
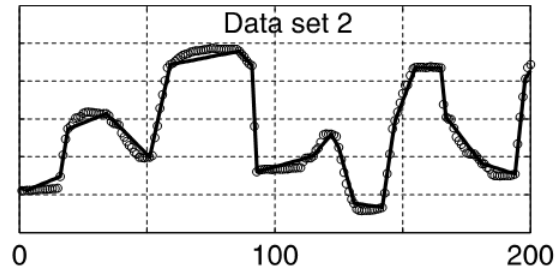
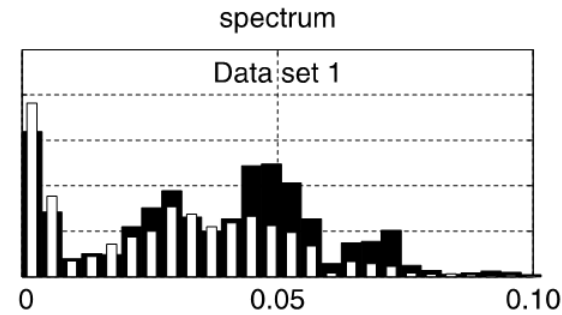
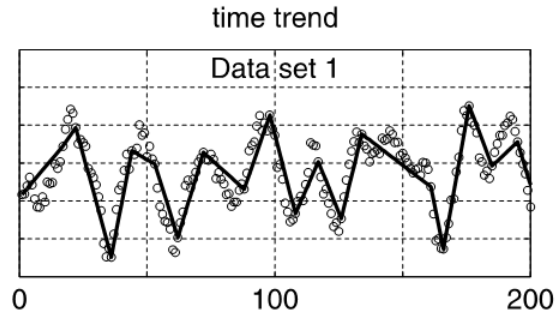
Under-sampled



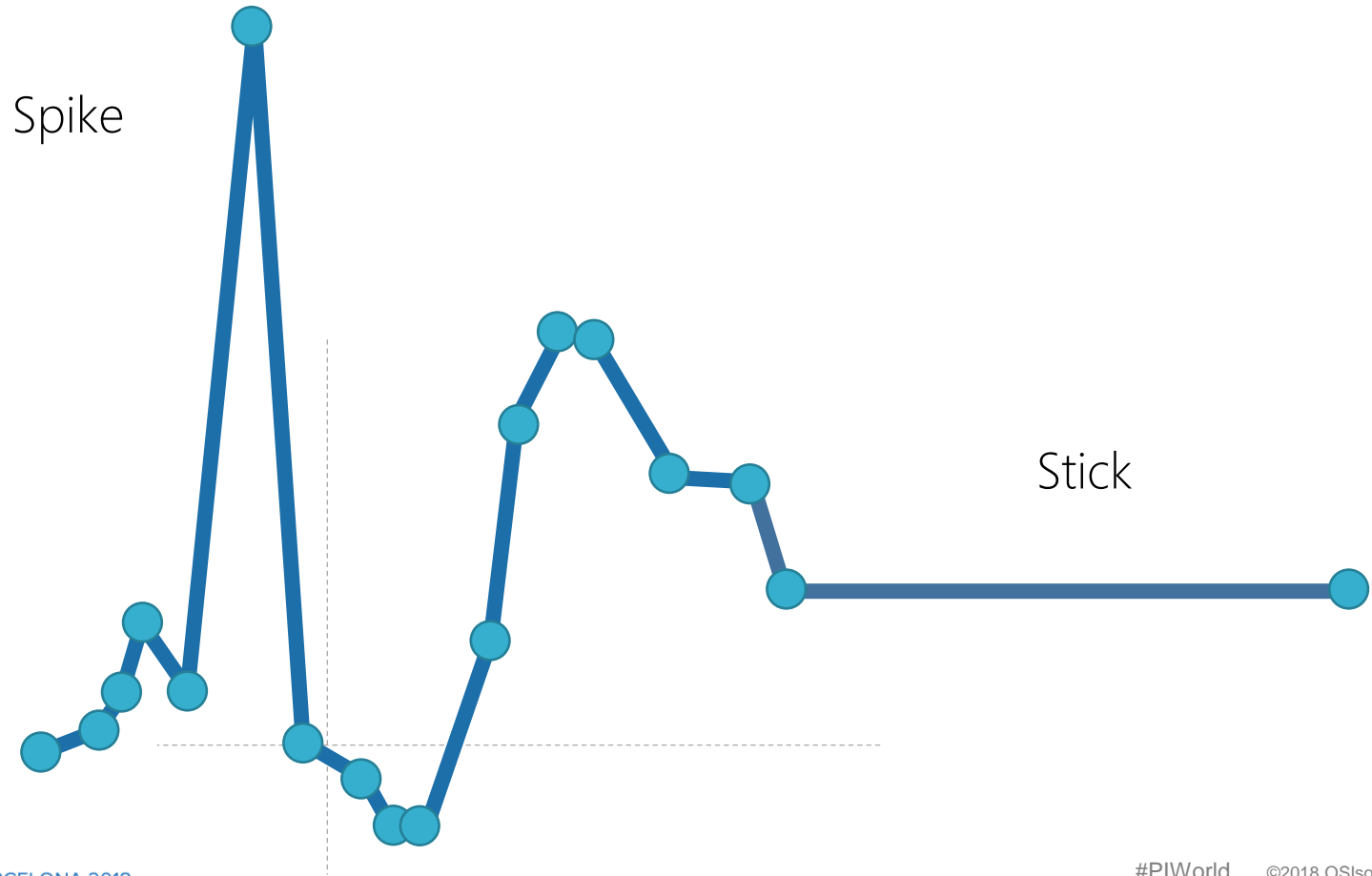
Well-sampled,
compressed



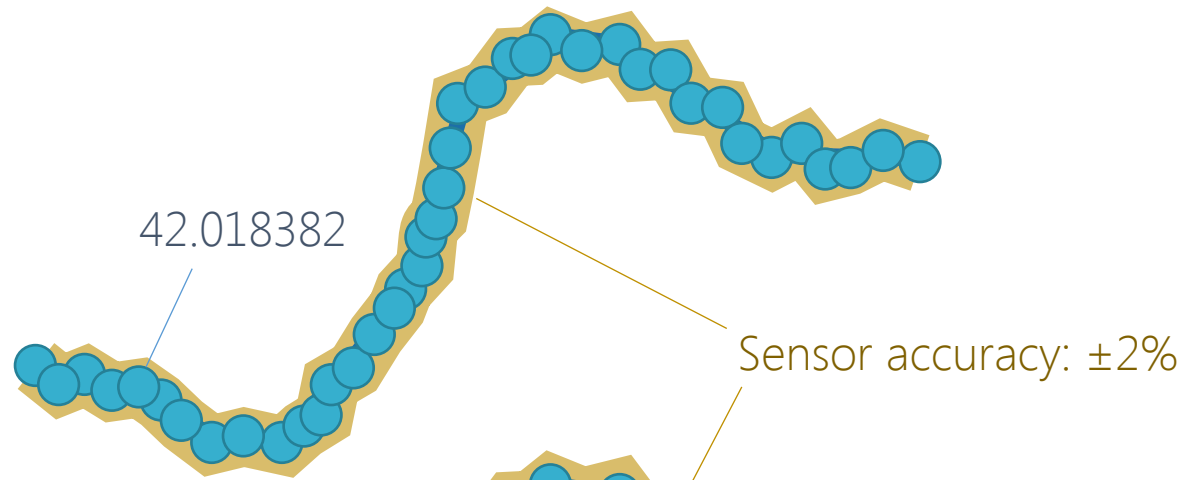
Trends @
10x compression



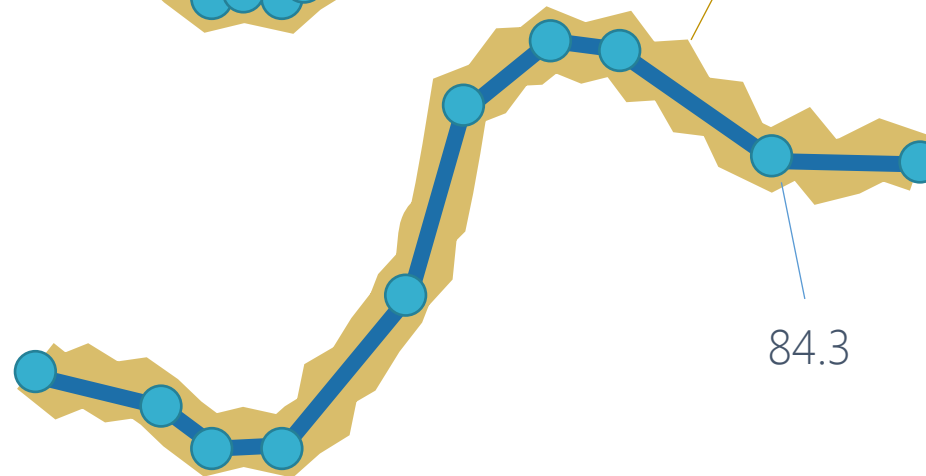
Frequency
spectra
(black original)



Falsely
precise



Realistic



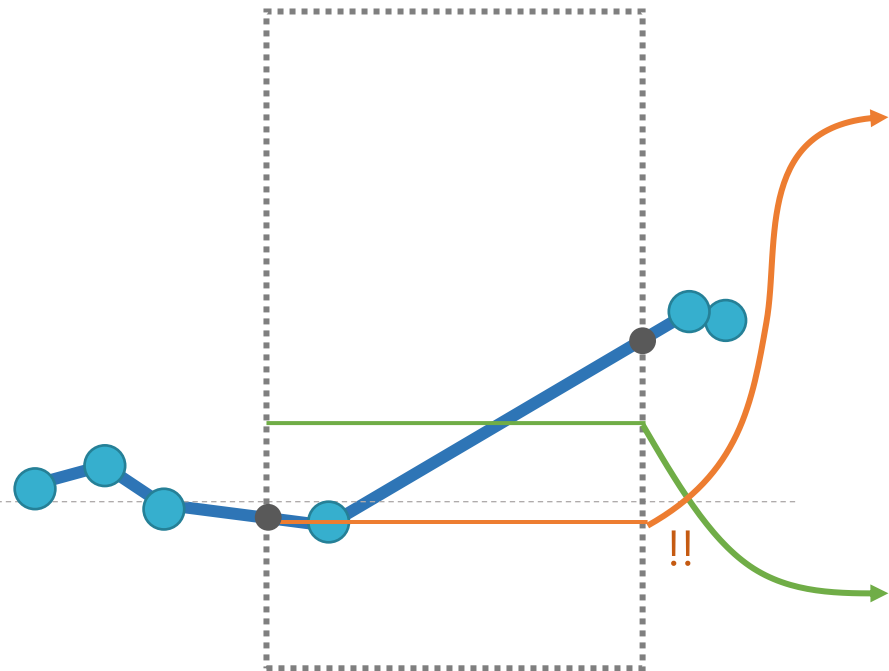
Raw

Time	Value A	Value B
8/21/18 17:50		78.751
8/21/18 17:52	33.899	
8/21/18 17:53		94.162
8/21/18 18:07		79.858
8/21/18 18:16	37.222	79.656
8/21/18 18:27	68.398	
8/21/18 18:30		97.063
8/21/18 18:41		35.461
8/21/18 18:50		42.960
8/21/18 19:00	72.527	

Interpolated together

Time	Value A	Value B
8/21/18 17:50	82.663	78.751
8/21/18 17:52	33.899	86.657
8/21/18 17:53	12.679	94.162
8/21/18 18:07	56.308	79.858
8/21/18 18:16	37.222	79.656
8/21/18 18:27	68.398	64.163
8/21/18 18:30	79.185	97.063
8/21/18 18:41	18.486	35.461
8/21/18 18:50	8.759	42.960
8/21/18 19:00	72.527	74.234

What is the average value
in this window?



Naïve

e.g. **AVG()** in SQL or Excel

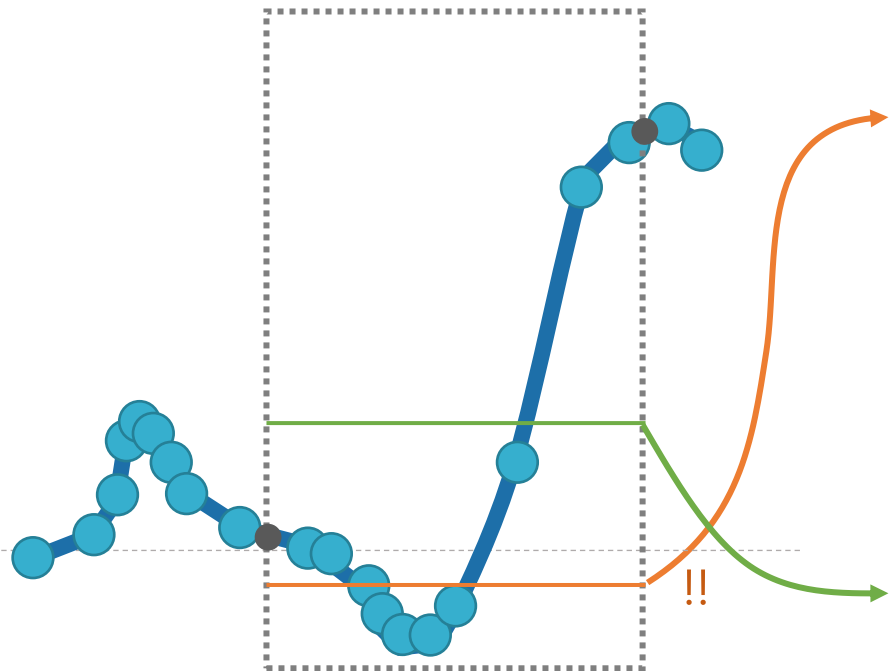
$$\frac{A + B + C + \dots}{n} = \boxed{-0.13}$$

*there are certainly times where event weighting is the right thing, but this choice should be made deliberately

Time-weighted

$$\frac{\overline{A} \cdot \Delta T_A + \overline{AB} \cdot \Delta T_{AB} + \dots}{\Delta T} = \boxed{4.38}$$

What is the average value in this window?



Naïve

e.g. **AVG()** in SQL or Excel

$$\frac{A + B + C + \dots}{n} = \boxed{-0.13}$$

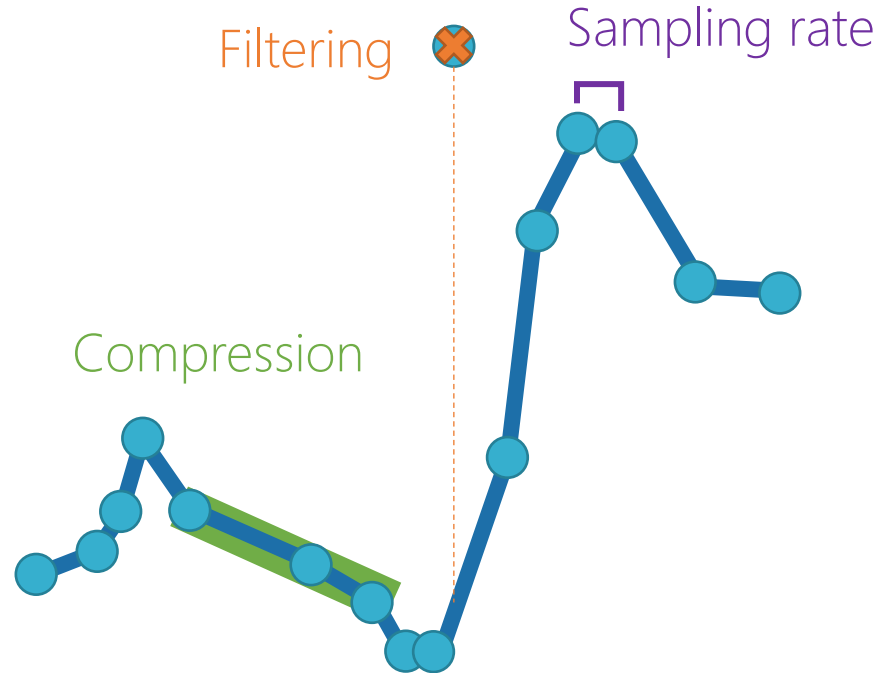
**there are certainly times where event weighting is the right thing, but this choice should be made deliberately*

Time-weighted
















$$\frac{\overline{A} \cdot \Delta T_A + \overline{AB} \cdot \Delta T_{AB} + \dots}{\Delta T} = \boxed{4.38}$$

III. What to do

Adjust your data collection settings



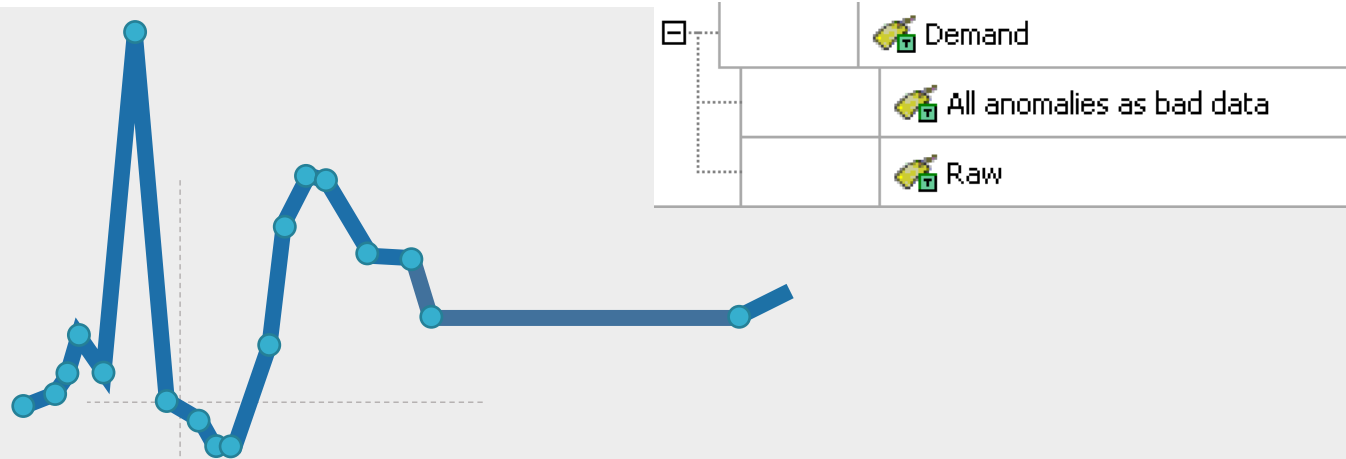
Add sensor metadata to your PI Assets

			Regenerant Temperature	172.9 °F
			Process maximum	220 °F
			Process minimum	150 °F
			Sensor accuracy	2 delta °F
			Sensor maximum	480 °F
			Sensor minimum	0 °F
			Sensor type	Type K Thermocouple

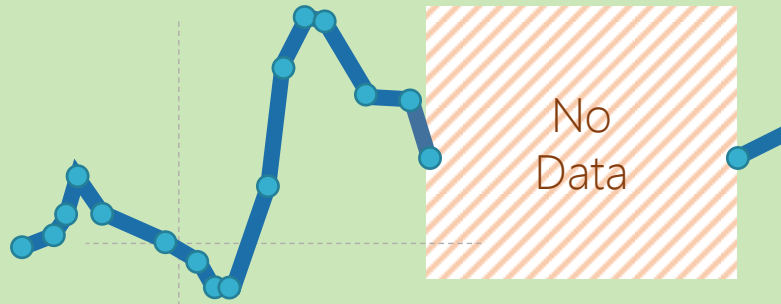
Cleanse your raw data

right in the PI System so others can benefit too

Original



Cleansed

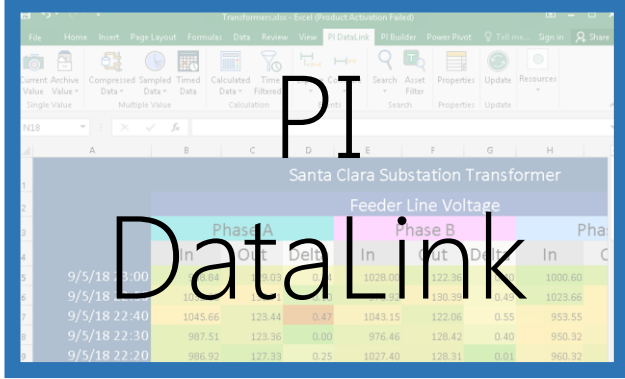


PI
Integrators

PI SQL

PI Web
API

PI
DataLink



The screenshot shows the PI DataLink application window. The title bar reads 'Transmittation - Local (Product Activation Period)'. The ribbon includes 'File', 'Home', 'Insert', 'Page Layout', 'Formulas', 'Data', 'Review', 'View', 'PI DataLink', 'PI Builder', 'Power Plant', 'Tools', 'Sign in', and 'Share'. The 'Data' tab is active, showing options for 'Current Value', 'Archive Value', 'Compressed Data', 'Sampled Data', 'Timed Data', 'Calculated Data', 'Trends', and 'Filter'. Below the ribbon, a data table is displayed with the title 'Santa Clara Substation Transformer'. The table has columns for 'Phase A', 'Phase B', and 'Phase C', each with 'In', 'Out', and 'Delta' sub-columns. The data rows show timestamps and numerical values for each phase.

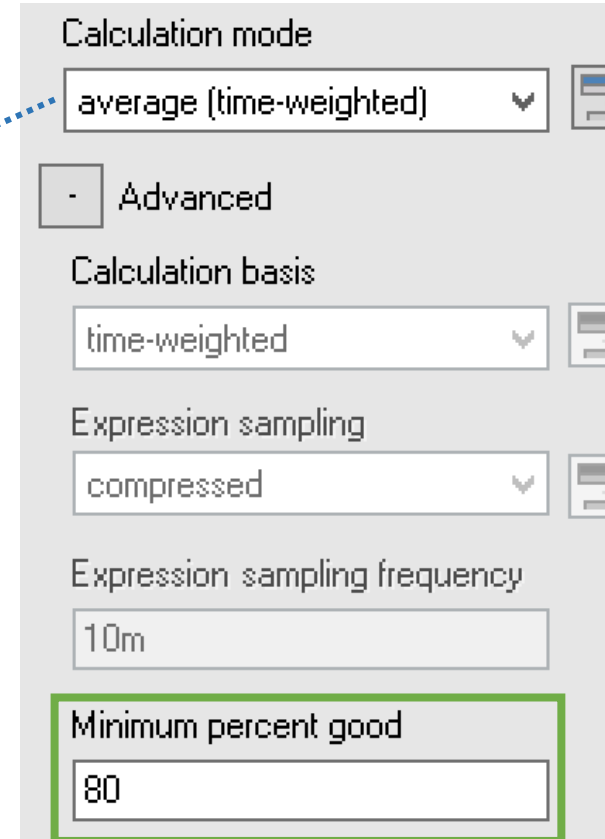
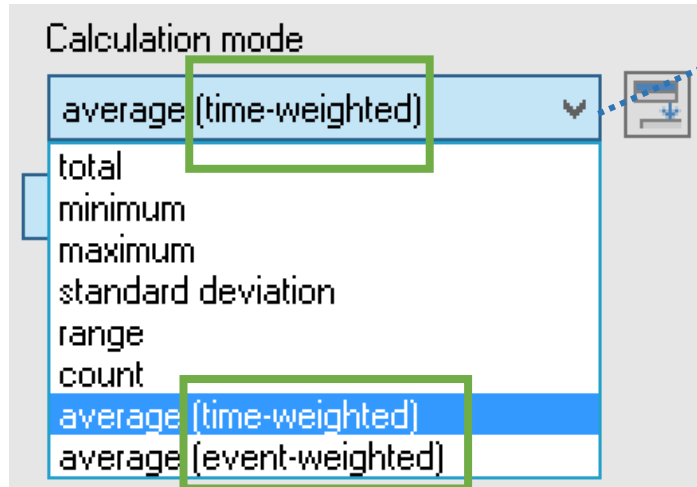
Phase A			Phase B			Phase C		
In	Out	Delta	In	Out	Delta	In	Out	Delta
1028.00	1028.00	0.00	1028.00	1028.00	0.00	1028.00	1028.00	0.00
1028.00	1028.00	0.00	1028.00	1028.00	0.00	1028.00	1028.00	0.00
1028.00	1028.00	0.00	1028.00	1028.00	0.00	1028.00	1028.00	0.00
1028.00	1028.00	0.00	1028.00	1028.00	0.00	1028.00	1028.00	0.00
1028.00	1028.00	0.00	1028.00	1028.00	0.00	1028.00	1028.00	0.00
1028.00	1028.00	0.00	1028.00	1028.00	0.00	1028.00	1028.00	0.00
1028.00	1028.00	0.00	1028.00	1028.00	0.00	1028.00	1028.00	0.00
1028.00	1028.00	0.00	1028.00	1028.00	0.00	1028.00	1028.00	0.00
1028.00	1028.00	0.00	1028.00	1028.00	0.00	1028.00	1028.00	0.00

Interpolate when you need regularity

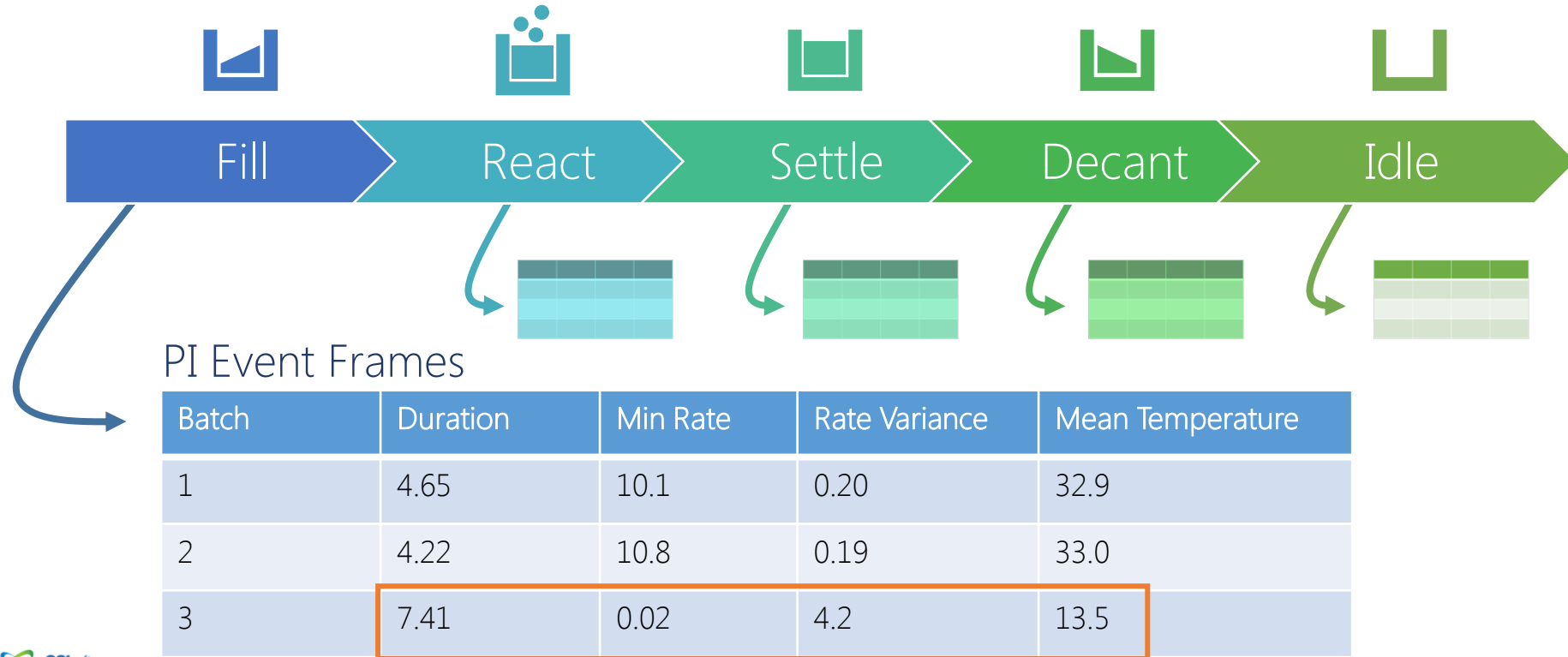
10-minute
samples

	A	B	C	D	E	F	G	H	I	J	K
1	Santa Clara Substation Transformer										
2		Feeder Line Voltage									
3		Phase A			Phase B			Phase C			
4		In	Out	Delta	In	Out	Delta	In	Out	Delta	
5	9/5/18 23:00	988.84	129.03	0.34	1028.00	122.36	0.40	1000.60	127.96	0.18	
6	9/5/18 22:50	1032.25	130.71	0.10	978.92	130.39	0.49	1023.66	119.19	0.59	
7	9/5/18 22:40	1045.66	123.44	0.47	1043.15	122.06	0.55	953.55	118.79	0.03	
8	9/5/18 22:30	987.51	123.36	0.00	976.46	128.42	0.40	950.32	120.04	0.08	
9	9/5/18 22:20	986.92	127.33	0.25	1027.40	128.31	0.01	960.32	129.08	0.56	
10	9/5/18 22:10	1018.62	128.71	0.09	1026.48	124.79	0.23	1032.61	129.70	0.04	
11	9/5/18 22:00	1029.66	123.74	0.32	998.33	122.17	0.17	1037.60	131.02	0.08	
12	9/5/18 21:50	989.96	124.25	0.03	977.33	124.61	0.16	1048.14	126.98	0.25	
13	9/5/18 21:40	994.02	120.92	0.22	996.89	124.66	0.00	1015.86	122.23	0.31	
14	9/5/18 21:30	967.34	128.68	0.48	997.28	121.06	0.24	977.83	129.73	0.46	
15	9/5/18 21:20	1029.46	121.18	0.50	968.50	130.60	0.58	1037.84	121.45	0.55	

Use time-weighted aggregates when appropriate, and set a minimum quality



Aggregate on phases or states



and now...

Contact Information



Brandon Perry

Research

OSIssoft

bperry@osisoft.com

Data Quality at TransCanada

Keary Rogers & Ionuț Buse



TransCanada Corporation (TSX/NYSE: TRP)

One of North America's Largest Natural Gas Pipeline Networks

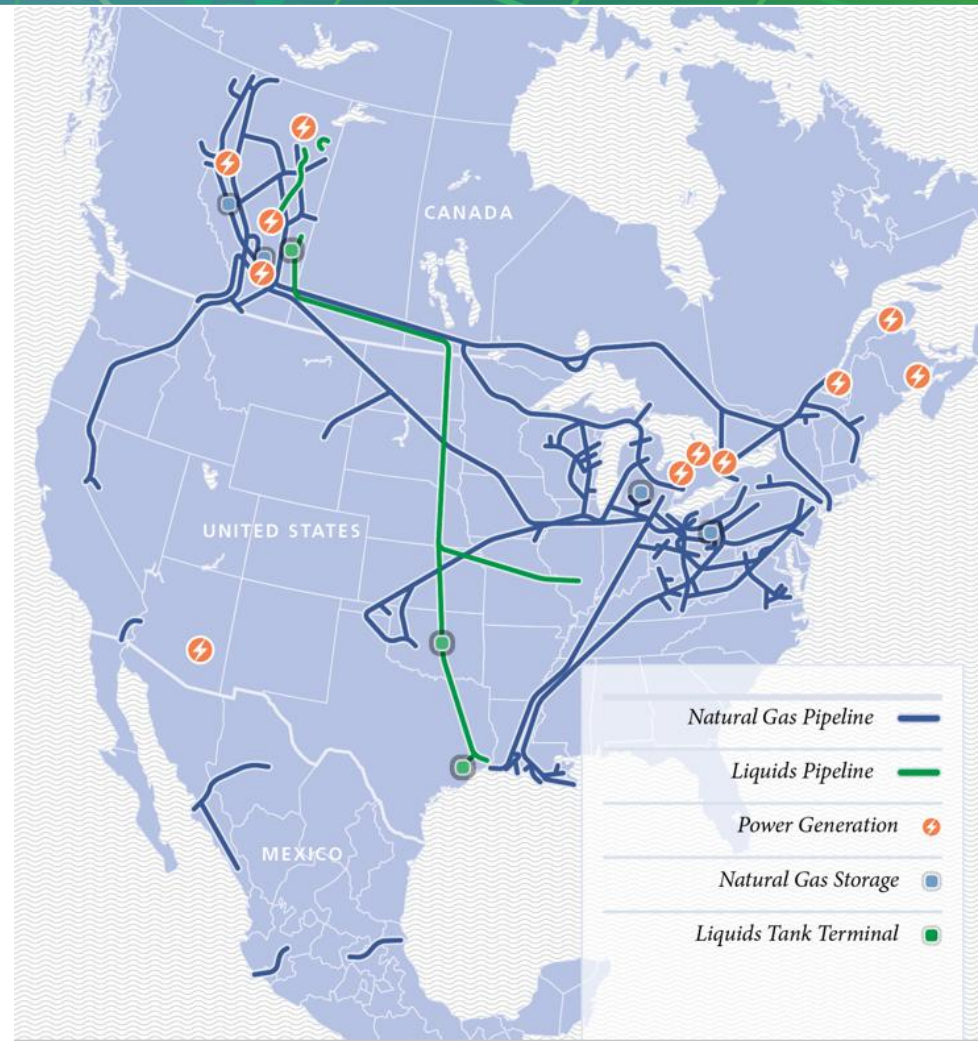
- Operate 91,900 km (57,100 mi.) of pipelines
- Transport ~25 per cent of continental demand
- Over 650 Bcf of gas storage capacity

One of Canada's Largest Private Sector Power Generators

- 11 power facilities, approximately 6,100 MW
- Diversified portfolio including wind, nuclear and natural gas

Premier Liquids Pipeline System

- 4,900 km (3,000 mi.)
- Keystone System transports ~20 per cent of Western Canadian exports
- Safely delivered more than 1.9 billion barrels of Canadian oil to U.S. markets



North America Natural Gas Demand Growth



City Centers



Universities



Schools



Our Children



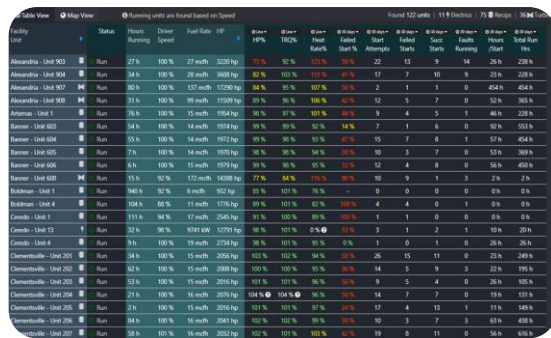
Medical Facilities



Elderly



Fleet Optimization



The screenshot displays the PI Vision software interface for monitoring 'Eagle Unit 1'. The top navigation bar includes the 'PI Vision' logo and a dropdown menu currently set to 'Unit01'. The main interface is divided into several sections:

- Unit Overview:** This section provides a high-level view of the unit's status. It includes a process flow diagram showing the 'Driver' connected to the 'Gas Compressor'. Key data points are displayed in a grid:
 - Ambient Temp: 64°F
 - Engine Speed: 329 rpm
 - Horsepower: 1,275 HP
 - Fuel Rate: 8.4 MCFH
 - Discharge Press: 958 psi
 - Discharge Temp: 105°F
 - Comp Rate: 1.29
 - Flow: 88,626 MCFD
 - Suction Press: 741 psi
 - Suction Temp: 67°F
- Driver Trend:** This section shows a multi-line graph representing the trend of various parameters over time. The graph displays multiple colored lines (red, green, blue, yellow, etc.) representing different data series.

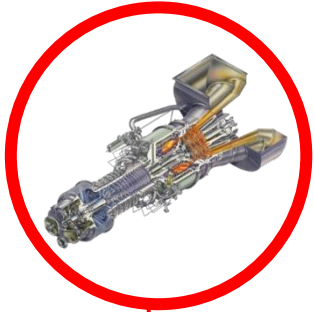
Performance indicators at the top right show 'Engine Power %' at 94% and 'Engine %' at 98%, both with circular progress indicators.

Early Detection of Functional Degradation

Expose Data to Operations Personnel

Asset Performance & Efficiency

How Real-time Data Impacts Our Business?



Functional degradation starts occurring on the gas producer bearing drain packing



Abnormal Oil Tank Pressure increase is flagged through SQC anomaly detection



Reliability Analyst performs data analysis & communicates to Maintenance Lead



Unit is taken offline planned, controlled & safely. The drain packing is replaced



Unit is back in service. Failure was mitigated without any customer impact

Real-time Data | Technology & People

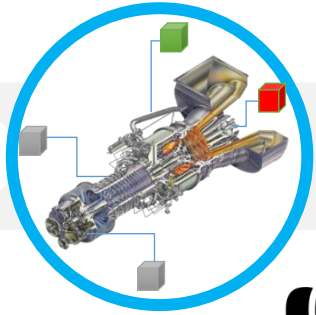
Sensors
& PLC

Network

PI
Interfaces

PI Data
Archive

PI Asset
Framework



Automation
& Control



Network
Support



Real-time
Systems



Core
Reliability



Hardware

+



Software

+



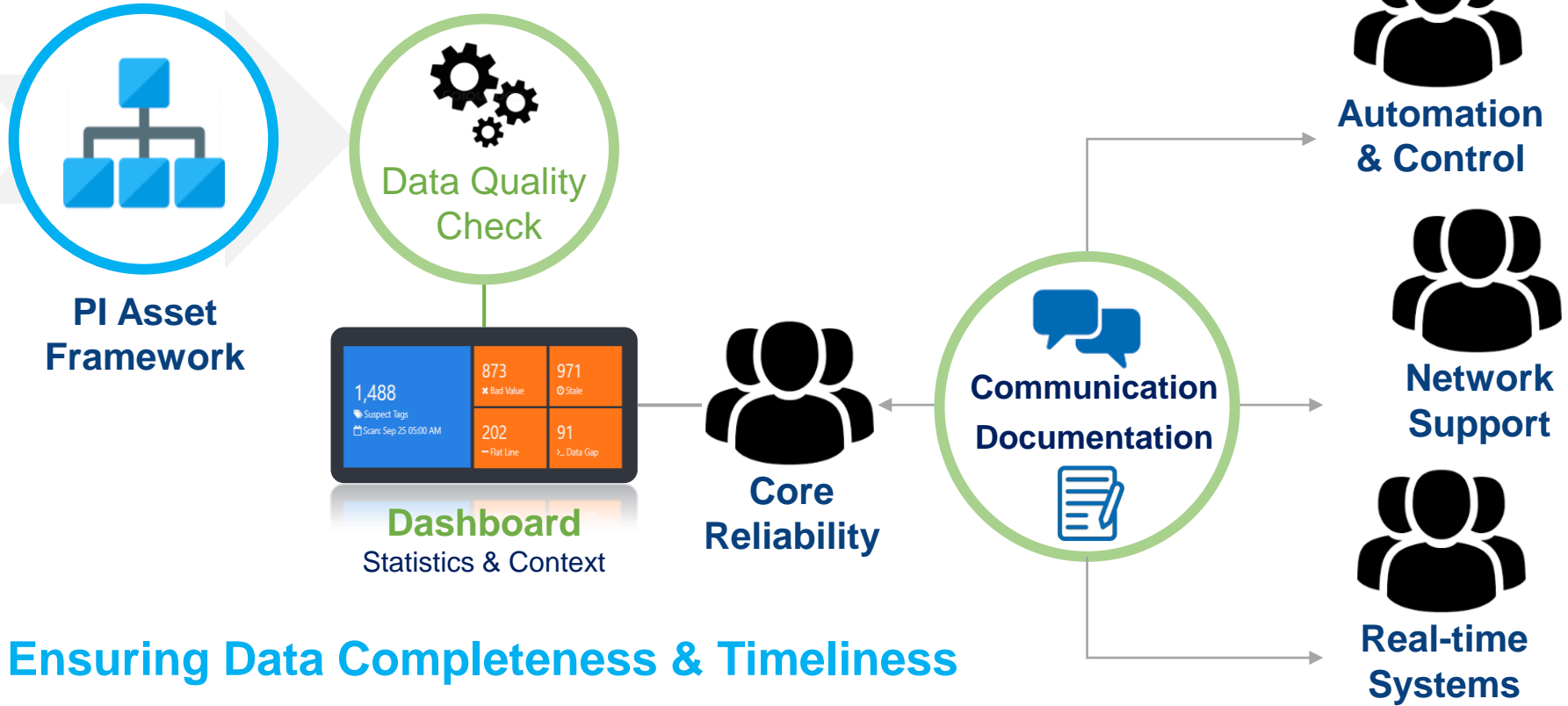
People



24/7

365

Real-time Data | Process



Ensuring Data Completeness & Timeliness

Real-time Data | Process Management Dashboard

Reliability Analysis Platform

Data Integrity

TransCanada

in business to deliver

USGO

North East

East

Central East

South East

Heartland

Mid America

Borders West

Great Lakes

Data Quality

Sensor Coverage

Process Management

5:00 AM

18,464

1,226 | 6.6%

707 | 3.8%

358 | 1.9%

338 | 1.8%

54 | 0.3%

Reload

Download

Mid America

Eunice - Unit 2

Suspect Tags: 127 Total | 127 Under Investigation

Central East

Lost River - Unit 12

Suspect Tags: 80 Total | 80 Under Investigation

Manage Pending

Attribute	Investigation	Tag	Issue(s)	Message	Last Run
Ambient Temperature	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.T1_Air_Inlet_Temperature	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
BHP	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.ABHP	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
Fuel Rate	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.Fuel_Rate	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
Gas Producer Speed	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.NGP_Speed	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
PCD	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.PCD	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
Power Turbine Speed	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.NPT_Speed	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
T5 Average	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.T5_Average	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
BAM Band 0	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.BAM_Band_0_Max_Peak_Amplitude	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
BAM Band 1	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.BAM_Band_1_Max_Peak_Amplitude	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
BAM Band 2	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.BAM_Band_2_Max_Peak_Amplitude	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
BAM Band 3	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.BAM_Band_3_Max_Peak_Amplitude	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
Drain Temp Bearing 1	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.Bearing_1_Drain_Temp	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
Drain Temp Bearing 2 & 3	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.Bearing_2_3_Drain_Temp	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018
Drain Temp Bearing 4 & 5	Inv (TCPL brendan_bell)	LostRiver_CS.AME.Unit_12.Bearing_4_5_Drain_Temp	Bad Value Stale (540.9 hrs)	Bad current value. Over 24 hours stale.	04/16/2018

Path: | Driver | Bearing Temperatures

Run Hours Scanned: 2,002

Total Count: 887

Distinct Count: 242

Min: 84.800

Max: 188.200

Mean: 155.639

Std Dev: 18.716

Central East

Strasburg - Unit 3


Suspect Tags: 79 Total | 1 Pending | 78 Under Investigation

Heartland

Saint John - Unit 7

Suspect Tags: 77 Total | 20 Pending | 57 Under Investigation

Real-time Data Quality | Failure Scenarios



Bad Value

Unexpected system state is written to the current value


✓ Accomplished



Stale

Data has stopped updating and the last timestamp is older than exception max


✓ Accomplished



Flat Line

Data is updating but same value gets written Identified by leveraging the asset structure

✓ Accomplished



Granularity

Data is not collected at adequate granularity to be used in statistical and machine learning methods In-depth data analysis is required to address this issue

✓ Future Work

↑ Complexity

Contact Information



Keary Rogers

Manager, Core Reliability

TransCanada US Gas Operations

keary_rogers@transcanada.com



Ionuț Buse

Team Leader, Enterprise Analytics

TransCanada US Gas Operations

ionut_buse@transcanada.com

Questions?

Please wait for
the **microphone**

State your
name & company



Please rate this session in the mobile app!

