DNV·GL

Techniques for Improving Leak Detection using Data Science and Machine Learning

Noorallah Rostamy / Boris Rodriguez

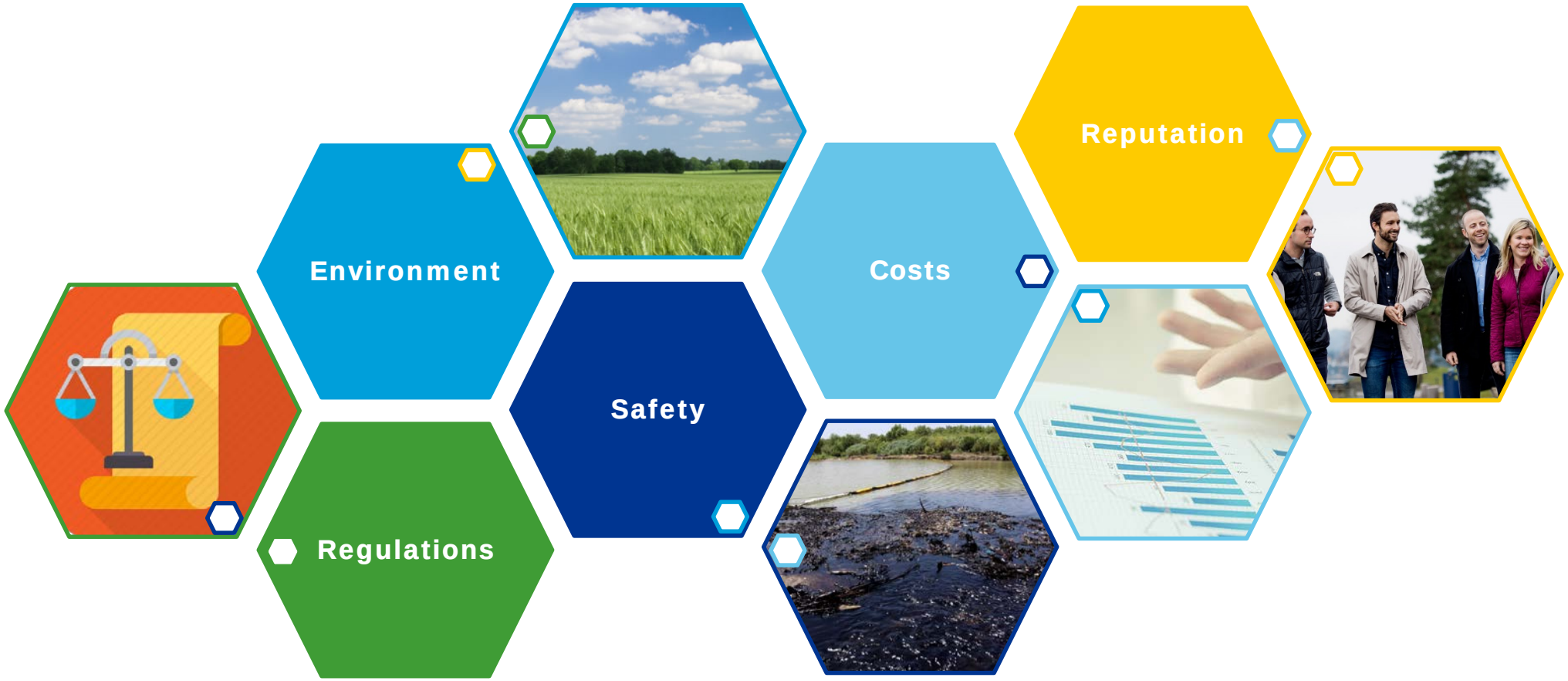
26 February 2020

Agenda

- Why Leak Detection?
- Regulatory requirements for Leak Detection
- Real-Time Transient (RTT) Model for Leak Detection
- Why Data matters on Leak Detection / RTT Models?
- How can we use ‘Data Science’ on RTT Models?
- Data selection using ‘Data Science’
- Results/Benefits and next steps

 DNV GL © 2019

Why Leak Detection?



Why Leak Detection?

Safety!

REQULATORY REQUIREMENTS & COMPLIANCE REVIEW

DNV GL © 2019
5

Why Regulations?

- To improve pipeline safety
- To improve environmental protection
- To provide assurance to the public concerning the safety of our nation's pipelines.

 DNV GL © 2019

U.S. Regulations and Standards

49 CFR Part 192:

Transportation of Natural and other gas by pipelines: Minimum Federal Safety Standards

49 CFR Part 195:

Transportation of Hazardous Liquids by Pipelines

 DNV GL © 2019

U.S. Regulations and Standards - 49 CFR p.195

Subpart C: Design Requirement

Section 195.134, *CPM Leak Detection*

It is applied to all hazardous liquid pipelines transporting single phase hazardous liquid (no gas in the liquid).

Each new computational pipeline monitoring (CPM) LD system and each replaced component of an existing CPM system **must** comply with section 4.2 of **API 1130** in its design and with any other design criteria addressed in API 1130 for components of the CPM leak detection system.

Subpart F: Operation and Maintenance

Section 195.444, *CPM Leak Detection*

Requires each CPM LD system installed on a hazardous liquid pipeline transporting single phase liquid to comply with API 1130 in operating, maintaining, testing, record keeping, and dispatcher training of the system.

New Rules by PHMSA (October 1, 2019)

- PHMSA proposed to amend§ 195.134 to require that all new lines be designed to have leak detection systems, including pipelines located in non-HCA areas.
- The new rule extends some of the existing requirements for high consequence area (HCA) pipelines currently subject to integrity management (IM) regulations to non-HCA pipelines, and strengthens assessment, repair and reporting requirements:
  - *New Reporting Requirements*
  - *Periodic Assessment*
  - *Integrity Management (IM) of Pipe*
  - ***Leak Detection***
  - *Inspections*
  - *Repair Criteria*

 DNV GL © 2019

Leak Detection Systems

Optimal Leak Detection Systems utilizes Recommended Practices and Standards for the following:

- Technologies
- Personnel
- Procedures

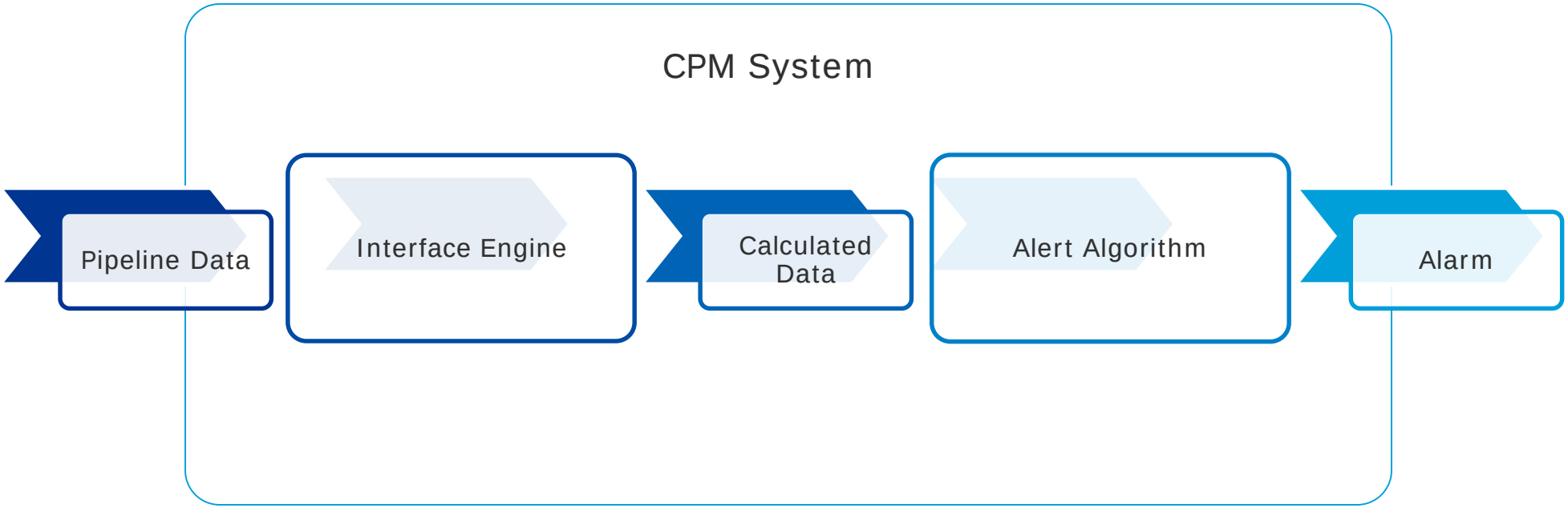
 DNV GL © 2019

CONTINUOUS LEAK DETECTION

DNV GL © 2019

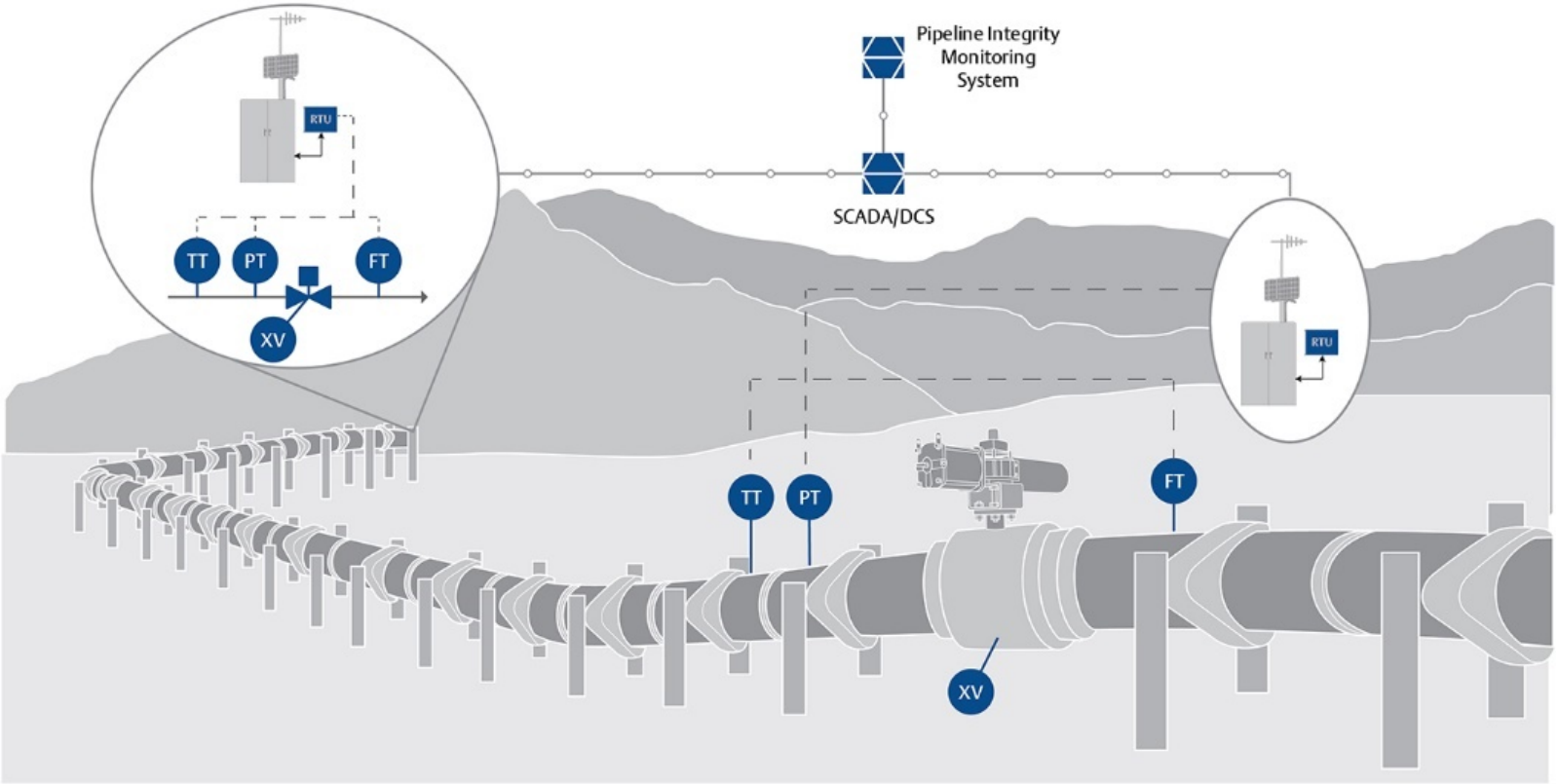
Leak Detection

Computational Pipeline Monitoring (CPM)

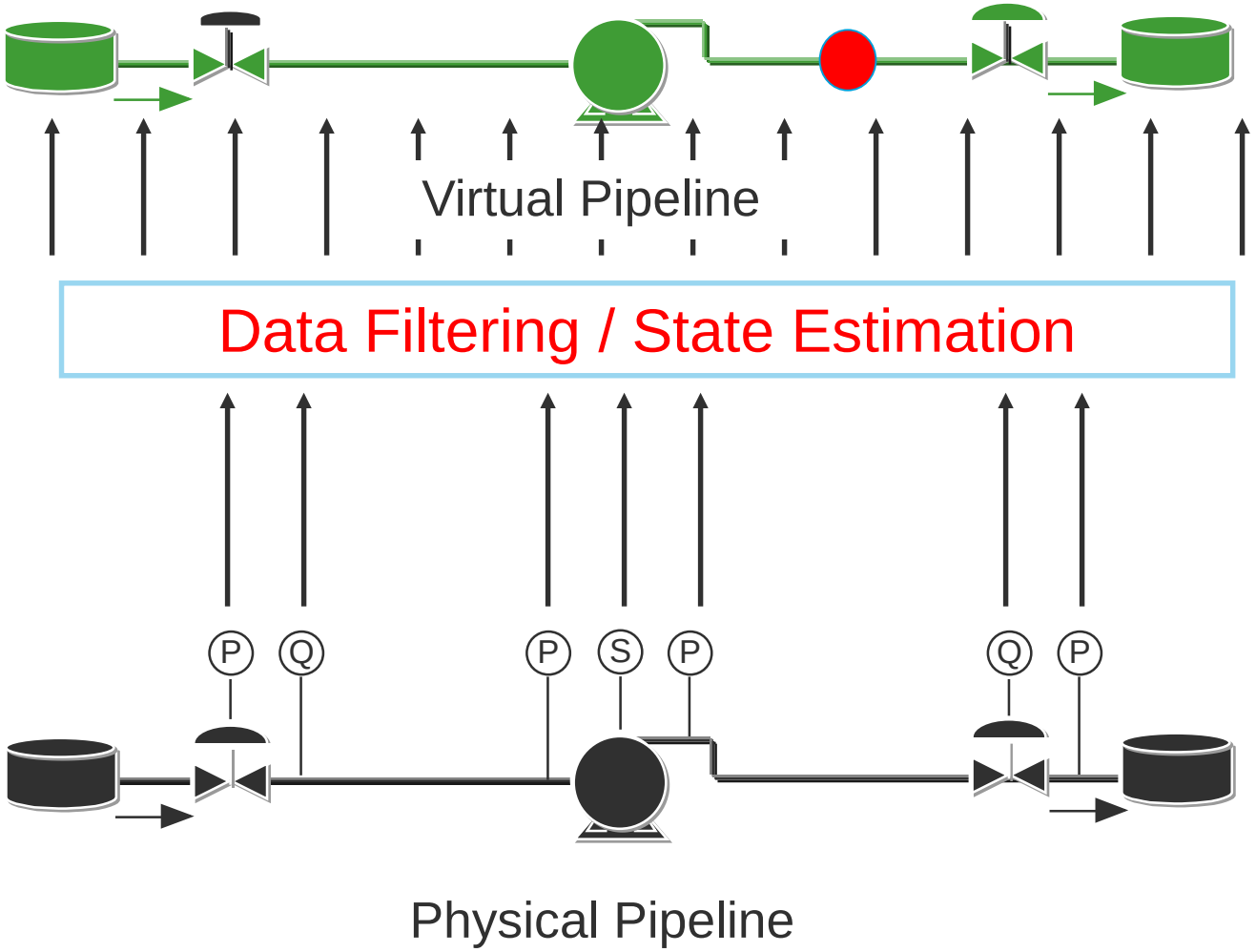


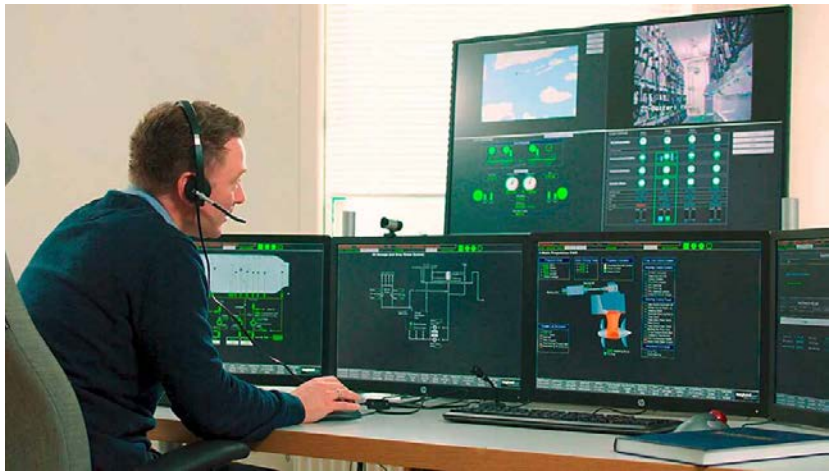
Leak Detection

- Computational Pipeline Monitoring (CPM)
  - Mass balance with line pack correction
  - Real time transient (RTT) modeling



Real-Time Transient (RTT) Model





Real-Time Transient (RTT) Model

Model of hydraulic behavior

Physical pipeline data

Product characteristics

Instrumentation/Data issues

Effort: implement, tuning, maintenance

 DNV GL © 2019

DNV·GL

Leak Detection System – Performance Criteria

Sensitive

- Minimum detectable Leak Rate
- Detection Time

Reliable

- Avoid False Alarms
- Reliable detect leaks

Accurate

- Leak Size
- Leak Location

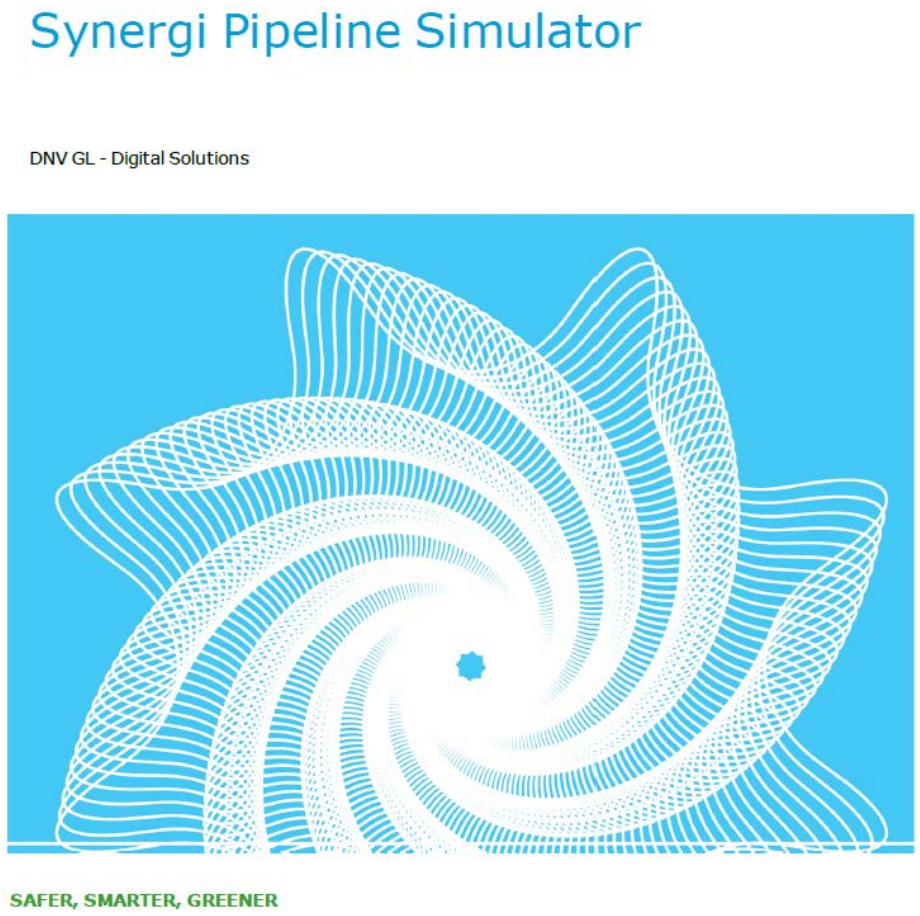
Robust

- Detect failing sensors
- Strategies for sensor failure

 DNV GL © 2019

Leak Detection – Synergi Pipeline Simulator (SPS)

- SPS Statefinder / Leakfinder
- State estimation of the pipeline using advanced fluid mechanics and hydraulic modelling.
- SCADA data is used for Hydraulic State estimation.
- Custom leak detection methods can be scripted.
- Leakfinder: off-the-shelf leak detection module.

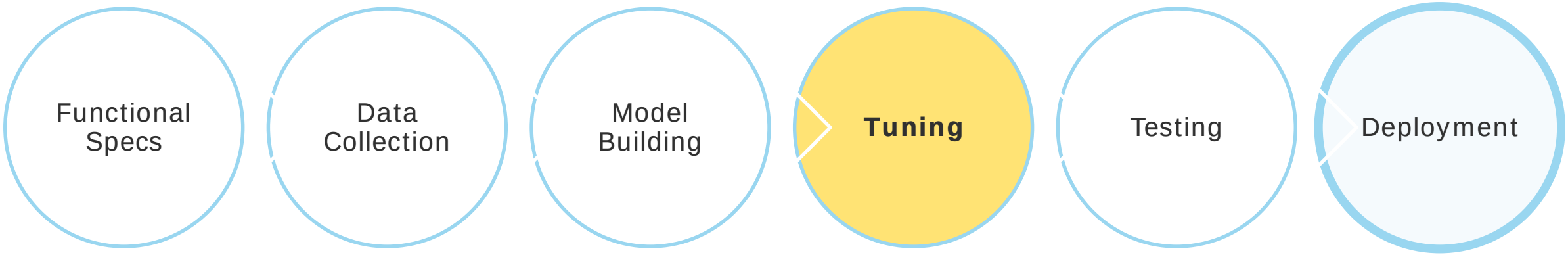


Synergi Pipeline Simulator - Model Implementation

× Cost associated with **implementation, tuning, maintenance, complexity of the system** and training.

SPS's based Real Time Model project success is heavily dependent on the real time SCADA measurements (time series data.)

SPS Project timeline:



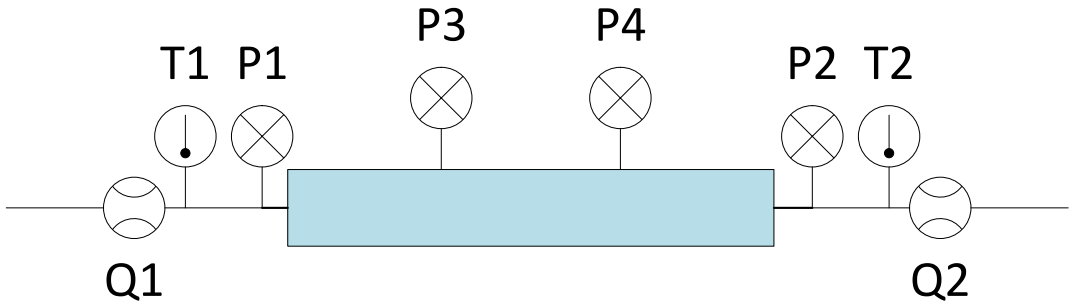
Challenge – SPS's Leak Detection Model Tuning

- SPS's Real Time Model project success is heavily dependent on the real time SCADA measurements (time series data)
- Time series data is fed into model to determine hydraulic state and distinguish between normal hydraulic behavior versus anomalous (leak) behavior
- Model is tuned against this data
- Tuning is a big effort – and tuning over the wrong data set can be disastrous from a project standpoint

 DNV GL © 2019

A little more about Model Tuning

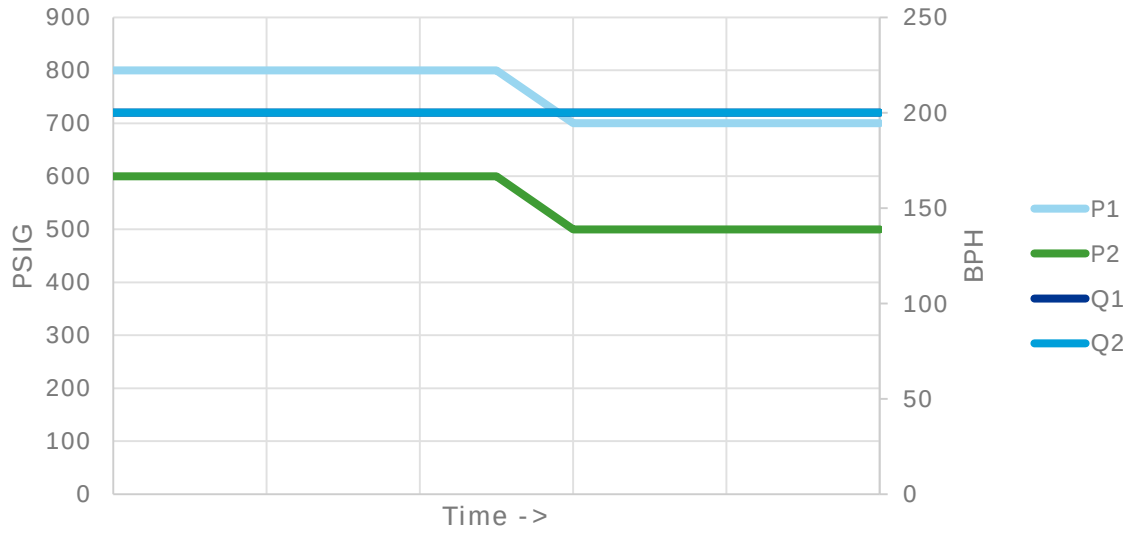
- The time series data is applied at various locations to the pipeline's Virtual Twin.
  - Data is typically Pressures, Flow rates, Temperatures, and various fluid properties
  - The Virtual Twin consists of physical pipeline parameters – length, elevation, fluid density, etc.
- In a perfect world, the model results would agree with the SCADA measurements…
  - A 10 PSI pressure drop in a segment of pipe results in a measured flowrate of 100 BPH and in a model estimated flow rate of 100 BPH.
  - But that is not always the case.
- Model Tuning looks at:
  - How good is the time series data?
  - How good are the physical pipe parameters?
  - How good are the fluid properties understood?
  - The model is adjusted to best match the 'good' time series data

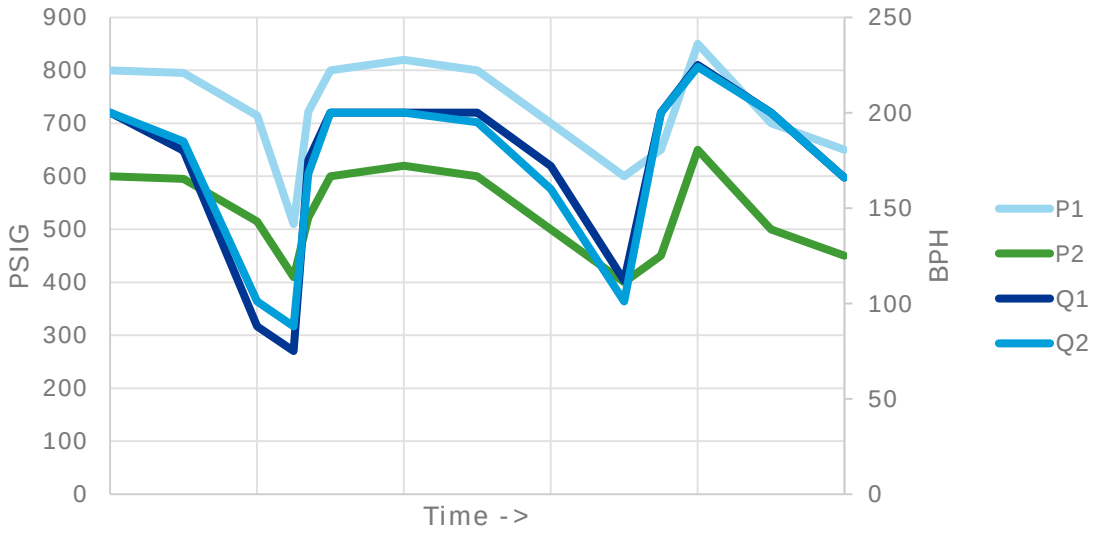


DNV GL © 2019

Time series variations

- Take a closer look at this time series data
  - Large part of effort
  - Not much standardization across team members about how we look at it
  - Time series can be for weeks, months, or years
  - Can tune over months of minimally transient data – only to be shocked when major transients occur





 DNV GL © 2019

The approach to a solution

- Can we get smarter about picking the time series data that we use?
- If we do, we should reduce effort AND produce a more reliable, accurate, sensitive Real Time Model system

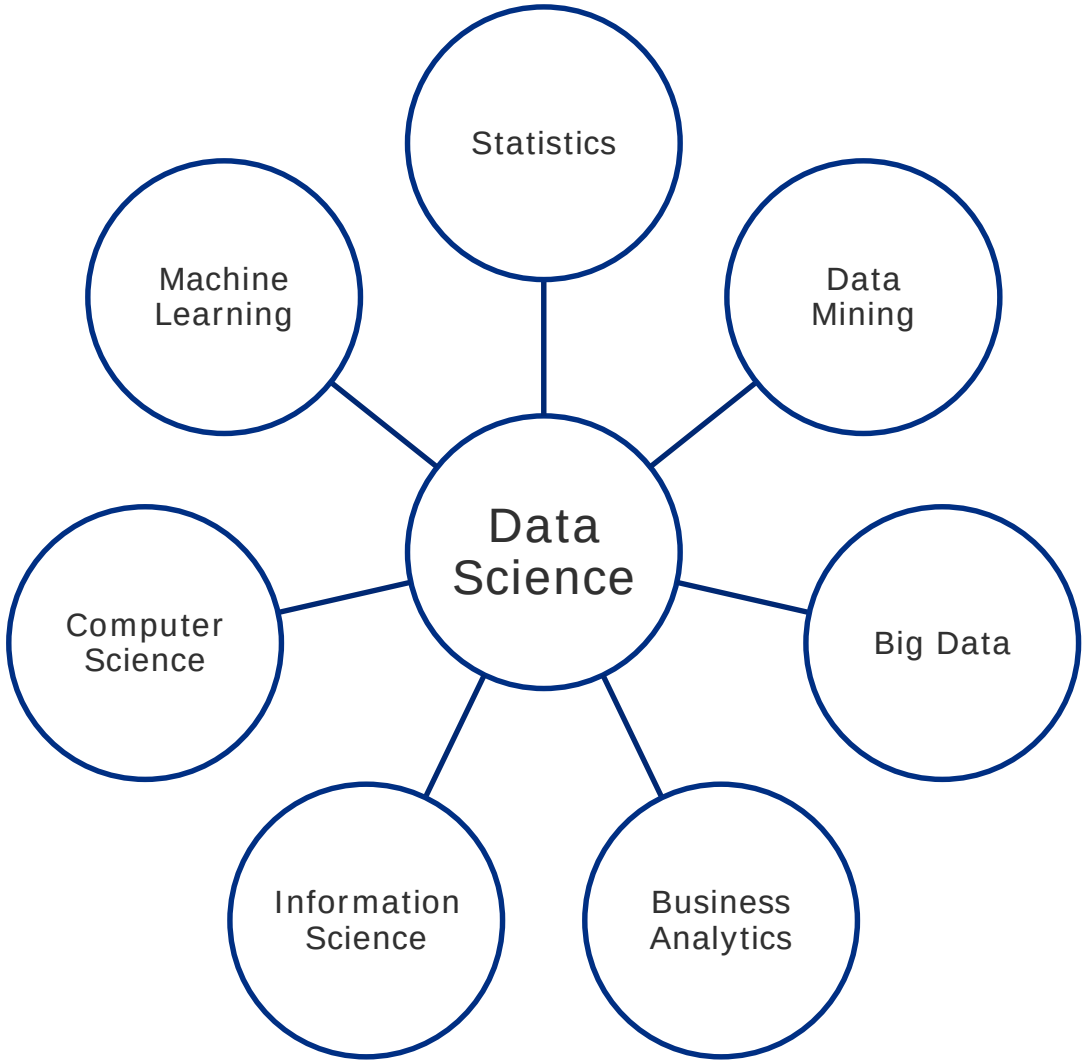
- The idea:
  - look through lots of data, quickly, to determine where interesting events are or are not.
  - Then be able to focus on those interesting events.
  - The old way of doing this is very manual and piecemeal and therefore laborious.

- We turned to **Data Science** to help us try and streamline and improve the tuning process.

 DNV GL © 2019

What is Data Science

- From Wikipedia, **Data science** is:
  - a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data
  - a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data
- In 2012, when Harvard Business Review called it "The Sexiest Job of the 21st Century", the term "data science" became a buzzword.
- Many data-science and big-data projects fail to deliver useful results



How can we use ‘Data Science’ on RTT Models?

- The idea – look through lots of data, quickly, to determine when interesting events are or are not.
  - Started small…
  - Started by looking at individually points…
  - Then combine some statistics to see what stands out.
- Utilized:
  - **DRTU**
    - Command line SPS tool that converts our rtu data file (.DT) from binary to text
  - **Python**
    - An interpreted, object-oriented, high-level programming language with dynamic semantics
    - Available without charge for all major platforms, and can be freely distributed
  - **Pandas** (Python library)
    - open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python

 DNV GL © 2019

Quick benefits

- By running through all of the data, can quickly find things like:

```
180,19/08/03,22:01:56,STAA_PT_004,4
274,19/08/03,22:02:08,STAA_PT_004,4(
308,19/08/03,22:02:18,STAA_PT_004
347,19/08/03,22:02:18,STAA_PT_004
400,19/08/03,22:02:28,STAA_PT_004,4
```

Duplicate Data

```
:03882,19/08/05,19:26:15,STAA_PT_004,4099.5117,GOC
:03975,19/08/05,19:26:26,STAA_PT_004,4099.5117,GOC
:04005,19/08/05,19:26:36,STAA_PT_004,4099.5117,BAD
:04071,19/08/05,19:26:46,STAA_PT_004,4099.5117,BAD
:04134,19/08/05,19:26:56,STAA_PT_004,4096.459,GOOD
:04227,19/08/05,19:27:07,STAA_PT_004,4096.459,GOOD
```

Quality = BAD Data

```
5067265,19/08/05,13:14:49,STAA_PT_004,4087.3015,GOOD
5067295,19/08/05,13:14:59,STAA_PT_004,4087.3015,GOOD
5067361,19/08/05,13:15:09,STAA_PT_004,4087.3015,GOOD
5067421,19/08/05,13:15:19,STAA_PT_004,4087.3015,GOOD
5067517,19/08/05,13:15:31,STAA_PT_004,4087.3015,GOOD
5067550,19/08/05,13:15:40,STAA_PT_004,4087.3015,GOOD
```

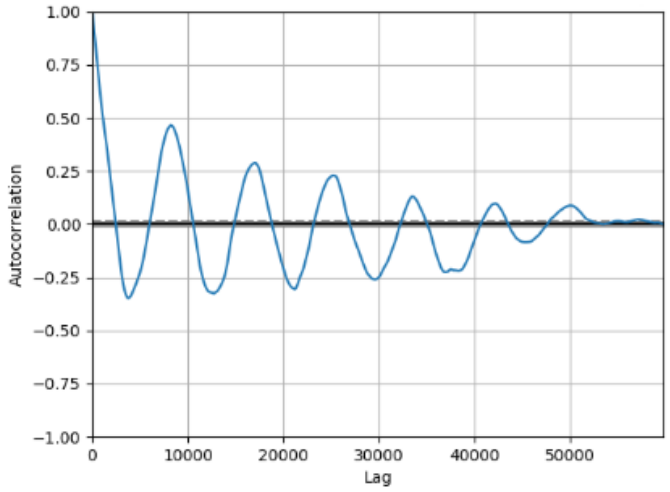
Flat Data

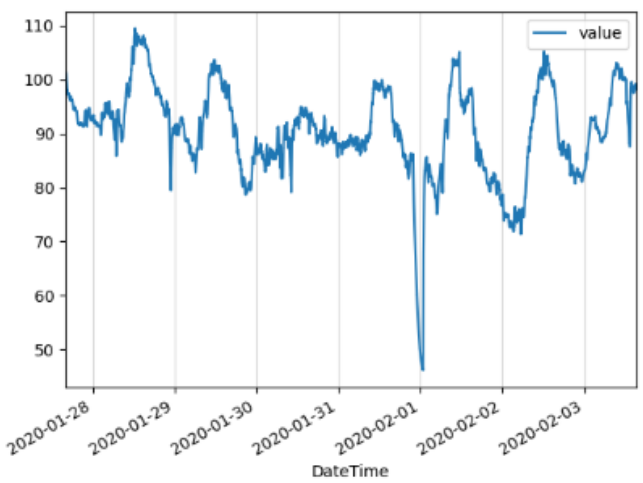
- Benefit:
  - Quickly look through all data, find issues, investigate, correct…then future data is improved

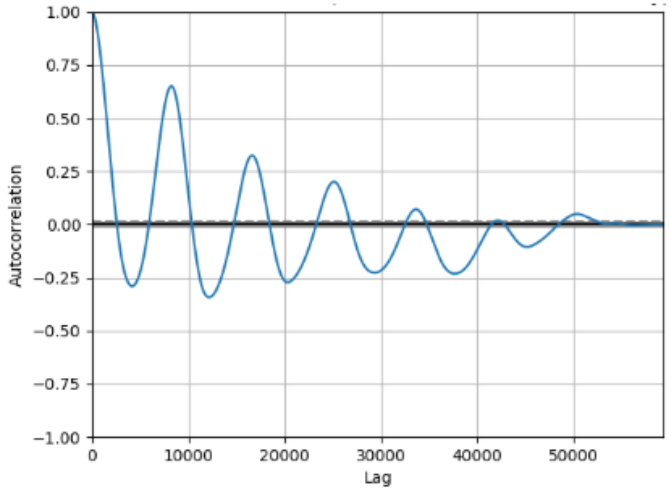
DNV GL © 2019

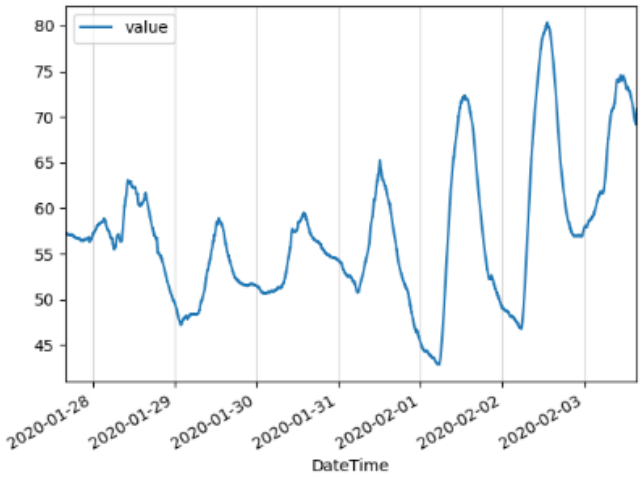
Detection Cyclical data

- Auto-detect tags that have a daily cycle
- Most likely due to daily heating/cooling influence









Data Summary

- A table of results can be generated for all individual points
  - Max Rate of Change (RB)
  - Mean Rate of Change (RDR)
  - Scan Period (SP)
  - Mean Change (REP)

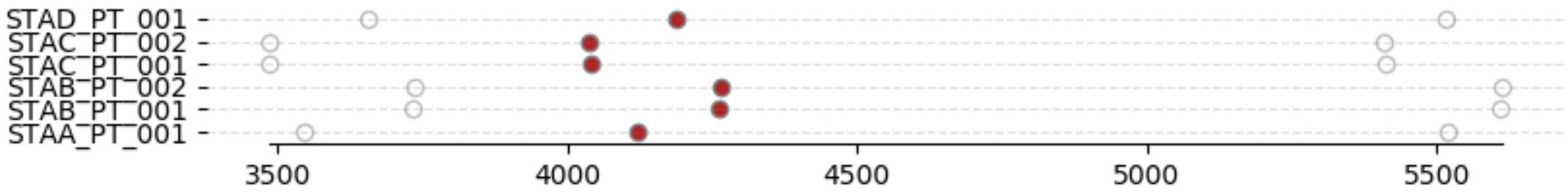
Tag	min	max	mean	sp	REP	maxROC	meanROC
STAA_PT_001	3546.25	5520.62	4120.59	0.17	1.21	-1104.55	7.09
STAA_QX_001	0.00	246.77	152.40	0.17	0.18	474.90	1.06
STAA_QX_002	0.00	0.00	0.00	0.17	0.00	0.00	0.00
STAA_QX_003	0.00	0.00	0.00	0.17	0.00	0.00	0.00
STAB_PT_001	3733.00	5610.00	4263.21	0.17	1.41	2200.00	8.22
STAB_PT_002	3735.00	5613.00	4265.30	0.17	1.41	2004.00	8.25
STAC_PT_001	3486.00	5412.00	4040.74	0.17	1.35	-1584.00	7.88
STAC_PT_002	3485.00	5409.00	4038.97	0.17	1.36	-1578.00	7.97
STAD_PT_001	3656.90	5515.87	4188.24	0.17	1.47	2124.54	8.58
STAD_QX_001	-0.11	287.82	124.79	0.17	0.30	287.88	1.75
STAD_QX_002	-0.11	268.91	27.54	0.17	0.16	846.15	0.95

Data Summary

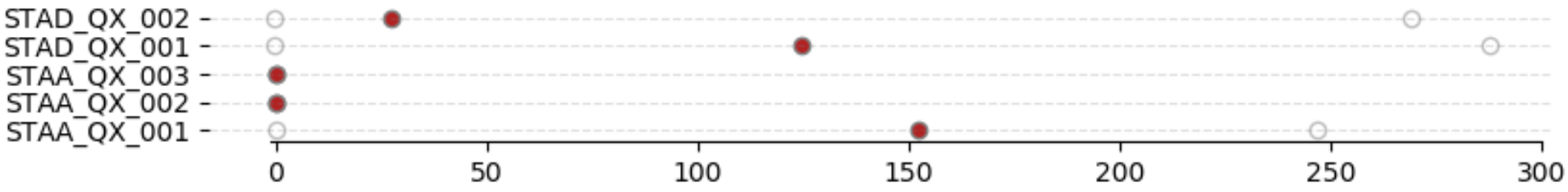
- Quick look at Pressures and Flow data ranges:

  Min / Max / Mean

Pressures



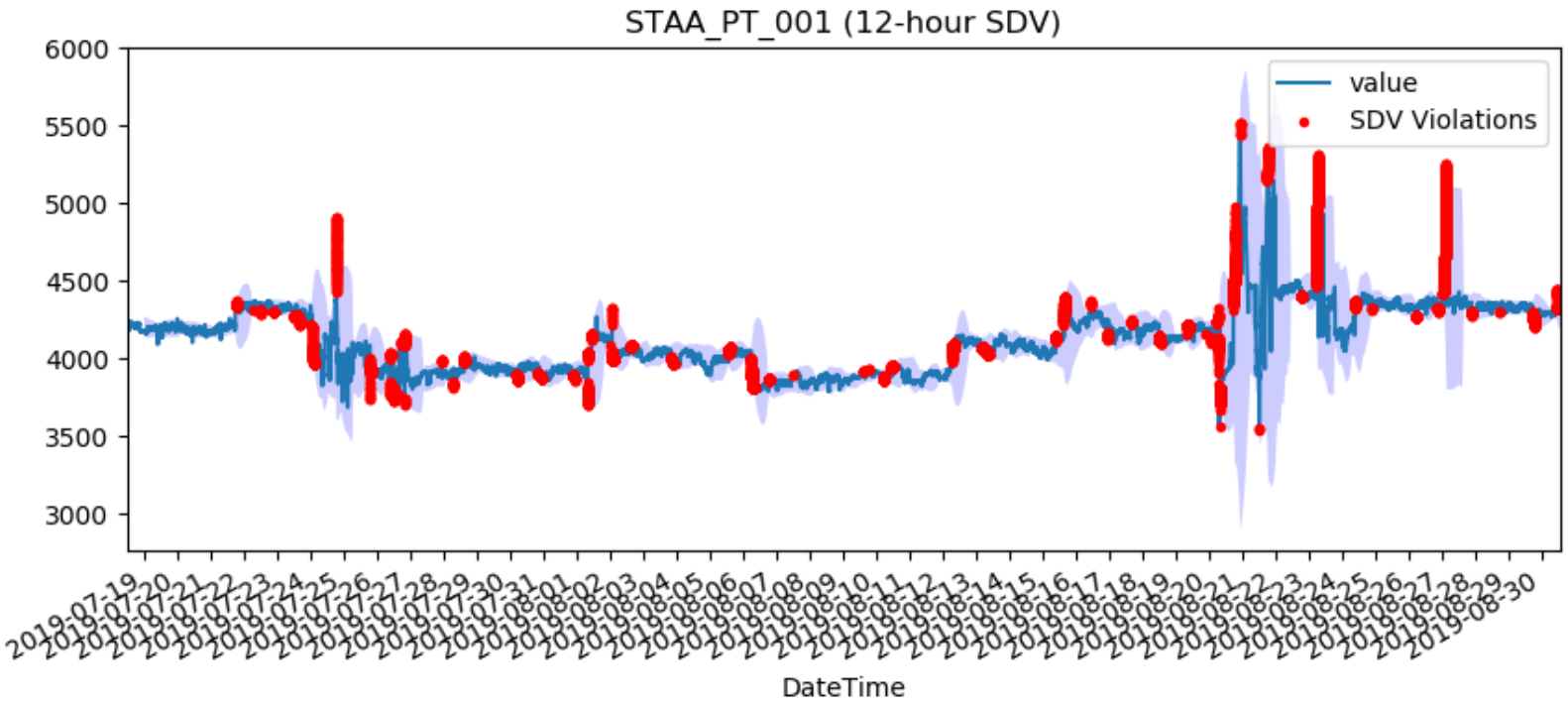
Flows



DNV GL © 2019

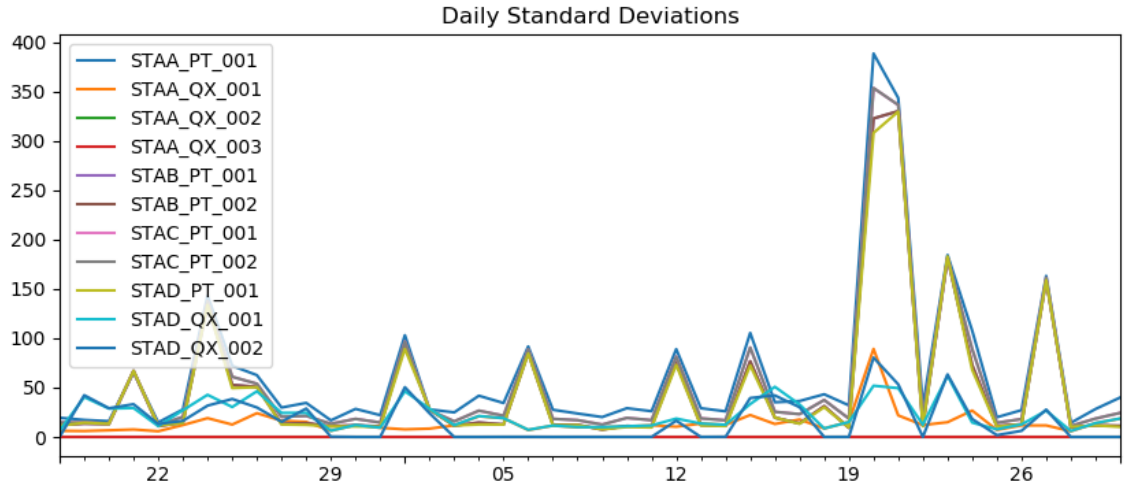
Quick trends

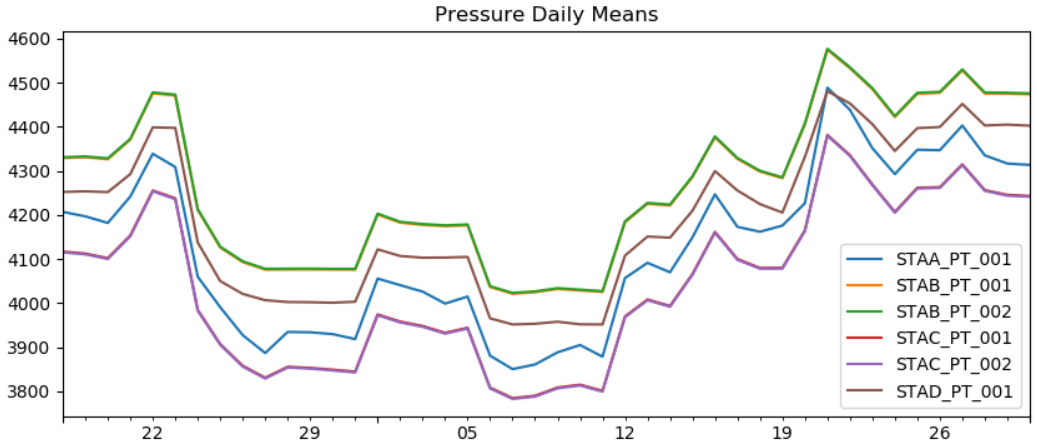
- Quickly identify non-conforming pressures
- Plotting a rolling Standard Deviation envelope and highlight points outside the envelope
  - For this plot, using 12 hour period and +/- 3 SDV
  - In the finance/stock trading world similar to Bollinger Bands
- The period, the number of SDV and the way that the rolling number is calculated varies these results

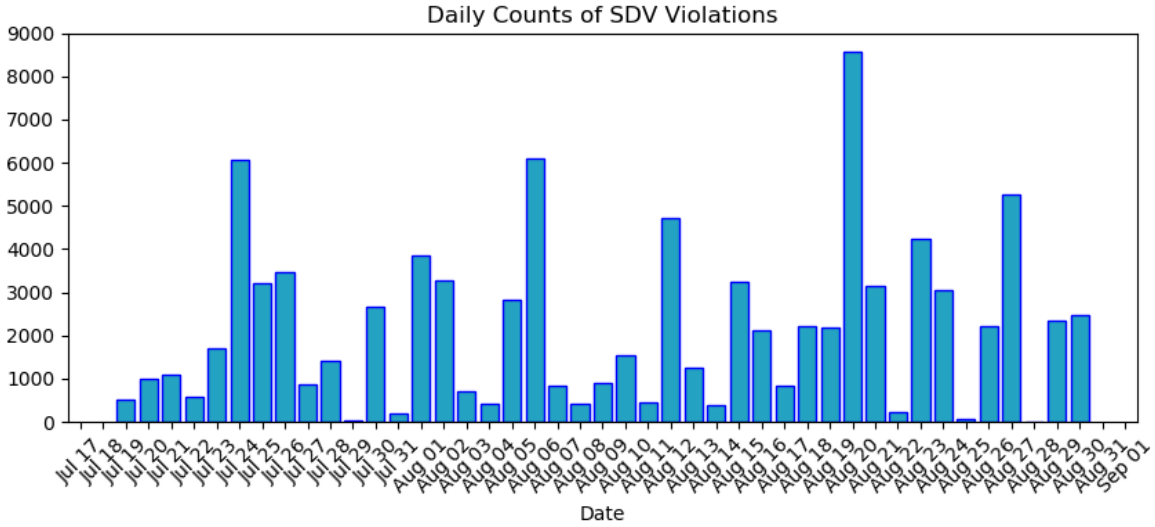


Daily results

- Combining a few charts, trends in the data are then more apparent.
- The charts show us that the most varying data is around August 20th, and July 23th.
- I want to choose those periods to focus the tuning efforts. The other dates are 'boring'.

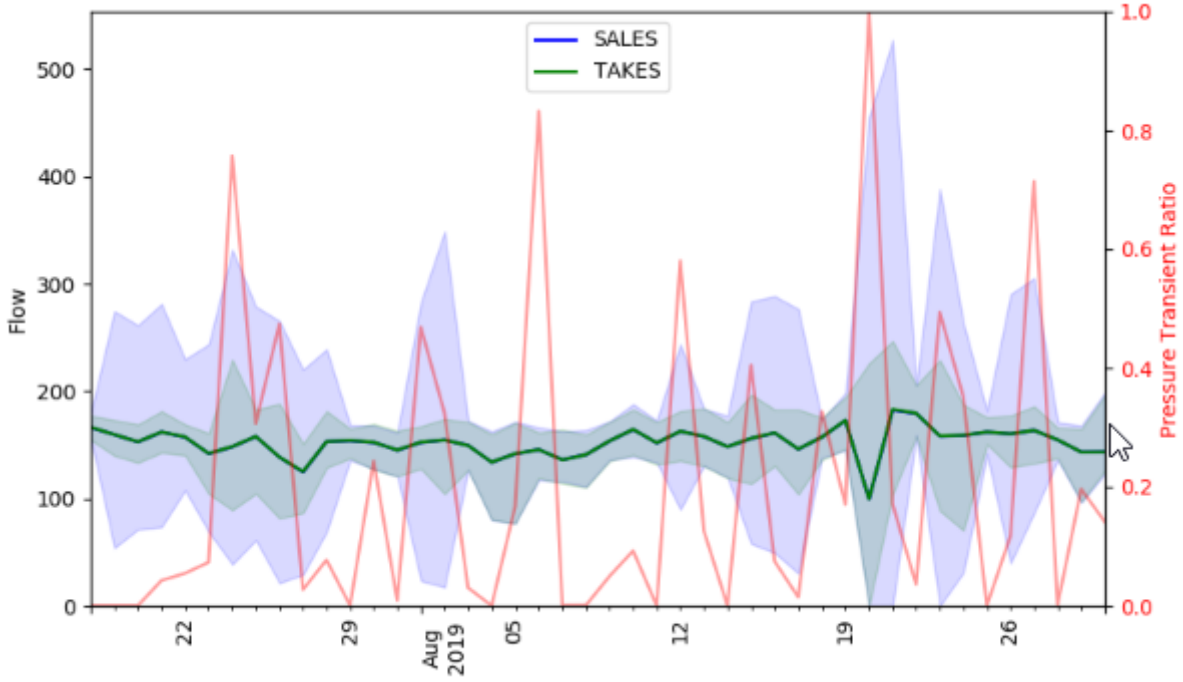


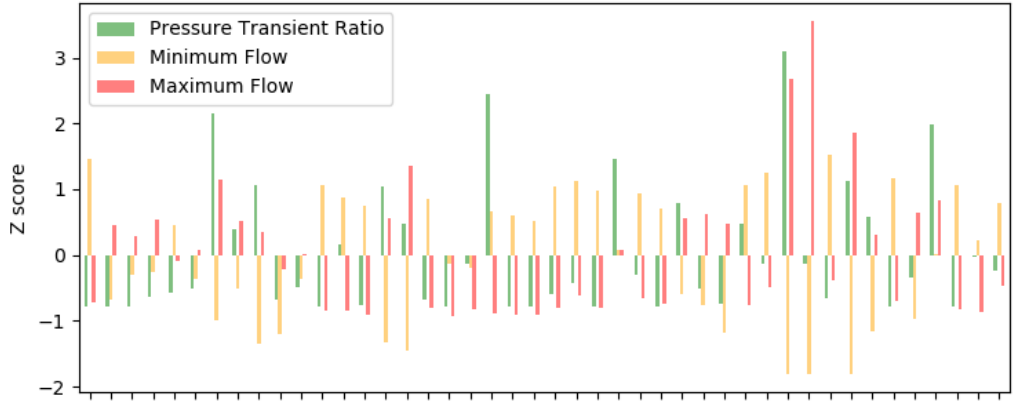


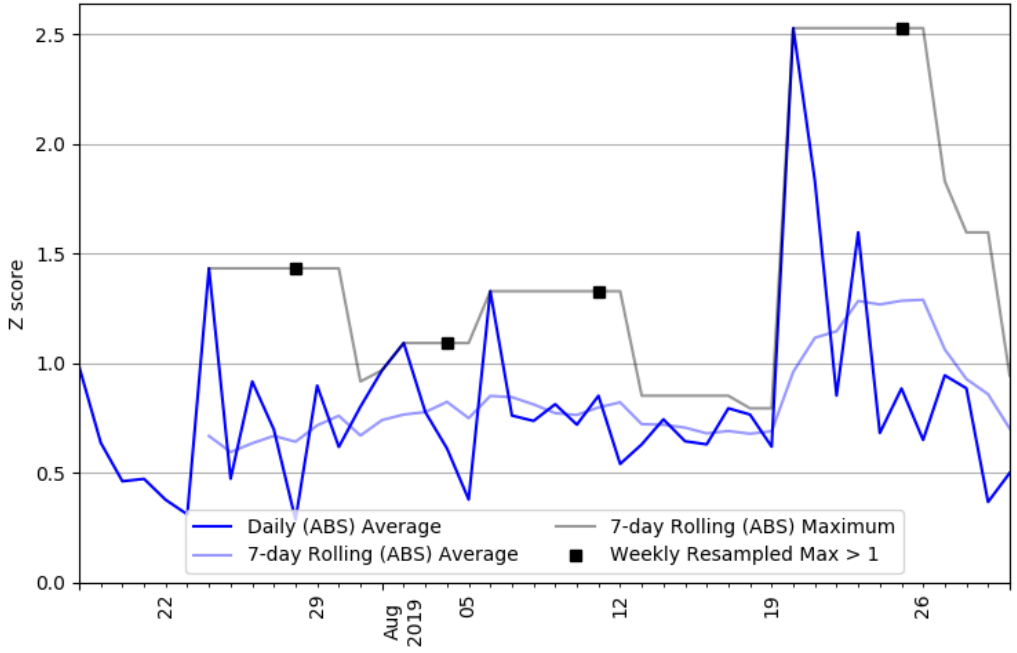


Daily results

- Flow data analysis
- Z score: $z = \frac{|x_i - \overline{x}|}{\sigma}$







Putting it all together …

- The script spits out a webpage that can then be shared with the end client are used for review …

DT file Analysis

This analysis is automatically generated.

Analysis v0.4: Processed on 2020/02/13 @ 22:38:01

Completed on 2020/02/13 @ 22:39:02

Input Data	Details
DT file name	testdata.dt
DRTU output name	testdata_ALL.txt
DT Start Date	2019-07-18
DT End Date	2019-08-30

Recommended Weeks
2019-07-28
2019-08-04
2019-08-11
2019-08-25

Final Summary

DT file name	Start Date	End Date	Total Tags	>10% BAD
testdata.dt	2019-07-18	2019-08-30	11	0

Lesson learned

- Looking at all data does not provide clear indicators
  - Equipment status, temperatures, ‘other’ provide noise
- Looking at flow and pressure data together does not provide clear overall indicators
  - Inject flows at same site can switch and seem dynamic, but in reality, the net flow and pressures stay constant
- Best to take a refined approach and pressures seem to be the best indicators
  - Run analysis on all pressures
  - Remove pressures that clearly are not relevant
  - Re-run analysis on remaining pressures
- Still in the trial/error phase of the SDV analysis
- Python 32-bit can handle about 10,000,000 records at a time. 64-bit version improves performance.

 DNV GL © 2019

Wrap up

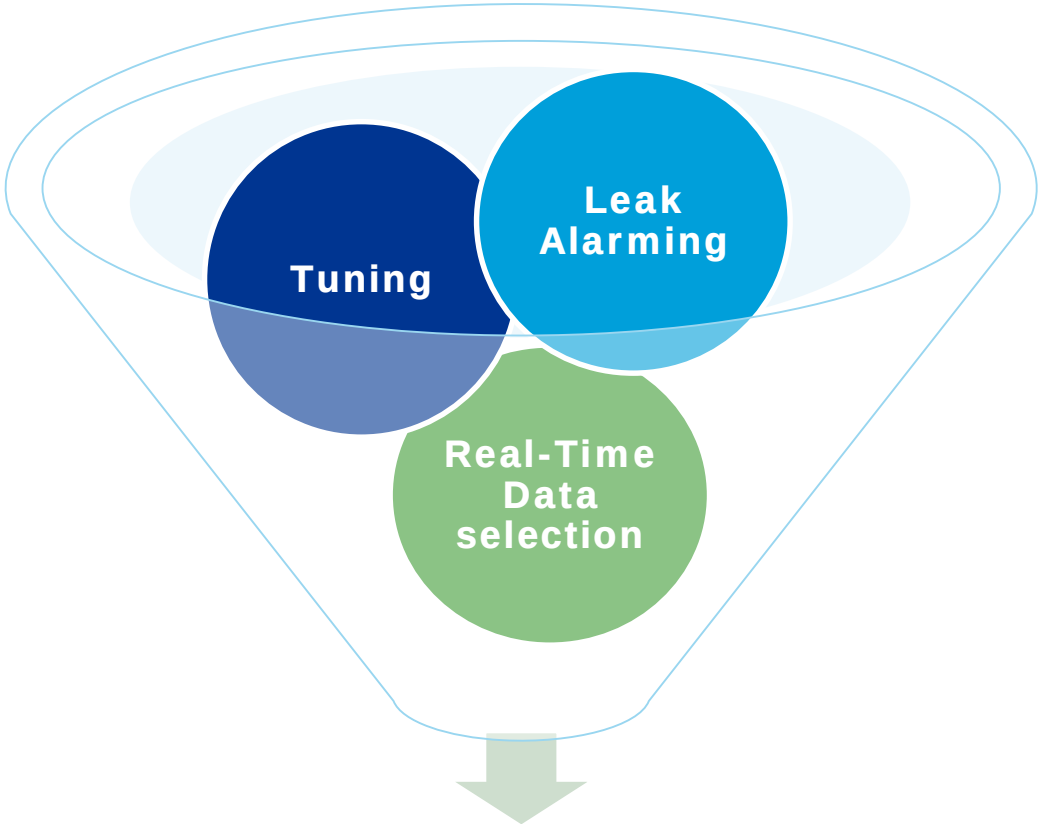
- The work is ongoing:
  - Has already proven useful in the quick analysis of new data sets (most for data quality)
  - Rolled out and made available to our project team for use
  - Scanning through the charts offers quick high-level analysis
  - Not a failure, but not yet at the end goal

- Next Steps:
  - Continue to use as part of project work
  - Obtain feedback from other users
  - Continue refining the analysis

 DNV GL © 2019

Wrap up

- Next Steps:
  - Continue to use as part of project work
  - Obtain feedback from other users
  - Continue refining the analysis
  - Use it in other model configuration and project implementation tasks:
    - Real-Time Data Selection
    - Automate tuning process
    - Leak alarming



Alarm Analysis

Special thanks to others at DNV GL: David Stobb, Sophie Massolas, Sanjay Yadav

Follow DNV GL – Digital Solutions on social media!

LinkedIn

- www.linkedin.com/company/dnvgl-software

Facebook

- www.facebook.com/dnvglsoftware

YouTube

- www.youtube.com/user/dnvsoftware

Twitter

- @DNVGL_S

Vimeo

- www.vimeo.com/search?q=DNV+GL+Software

DNV GL Software blog

- blogs.dnvgl.com/software

Thank you!

www.dnvgl.com

SAFER, SMARTER, GREENER

The trademarks DNV GL®, DNV®, the Horizon Graphic and Det Norske Veritas®
are the properties of companies in the Det Norske Veritas group. All rights reserved.