Choke valve condition monitoring and prognosis: theory and prototype

Xingheng Liu, 2022

Norwegian University of Science and Technology

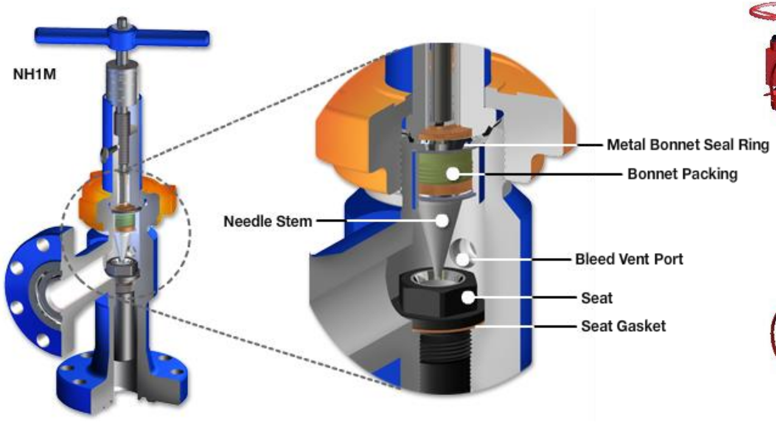
Outline

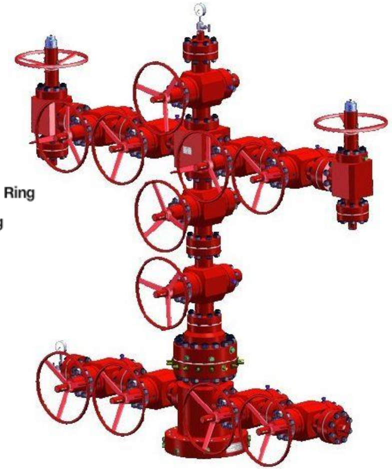
- Problem definition and research objectives (p3-6)
- Data overview (p7-8)
- Methods
  - Two-stage hybrid model (p10-20)
  - Spatio-temporal interpolation (p21-34)
- Summary and comparison (p35-36)
- Toolbox overview (p37-38)

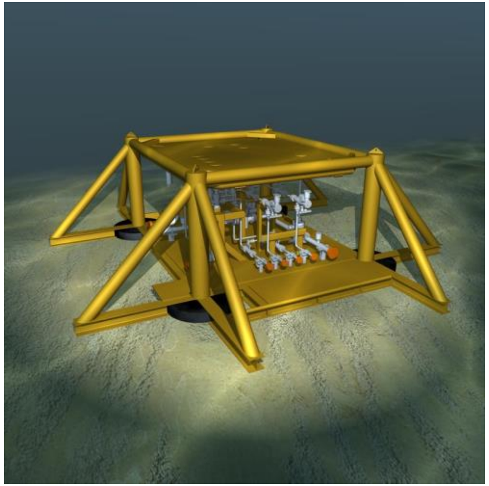
NTNU

Research object: choke valve

- Function: reduce pressure and control flow rate
- Application: production, injection, artificial lift, storage…
- Installation: Xmas tree, manifold, line heater, FPSO…





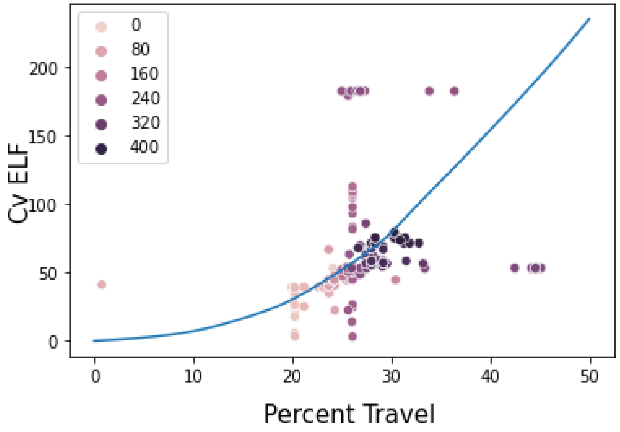


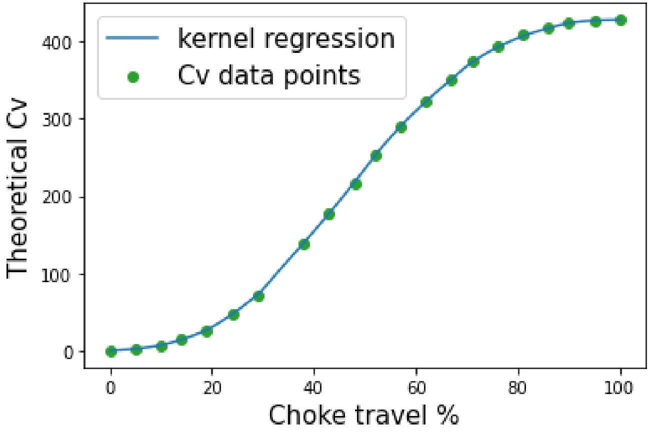
Erosion of choke valves

- Erosive agents:
  - Sand
  - Barite/Calcite
  - Proppants
- Consequences:
  - Body damage
  - Leakage
  - Less controllability
  - …

Erosion monitoring

- Health indicator: flow coefficient (Cv)
- Erosion can cause Cv deviation from its theoretical value
- Evaluate a choke valve's health state (Cv deviation) based on
  - Timestamp
  - Raw and theoretical Cv
  - Percent travel
- Prognosis
  - Degradation trend analysis
  - RUL estimation
- Estimate the Cv deviation for any time, any percent travel, in the past and in the future





NTNU

Problem definition

- Evaluate a choke valve's health state (Cv deviation) based on
  - Timestamp
  - Raw and theoretical Cv
  - Percent travel
- Prognosis
  - Degradation trend analysis
  - RUL estimation
- Estimate the Cv deviation for any time, any percent travel, in the past and in the future

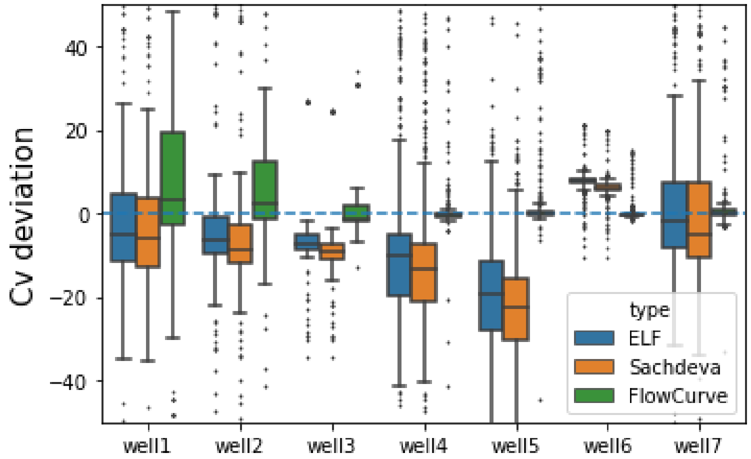
Data overview

- Data collected from 7 chokes of the same type installed at different wells
- Observation time span: 1-2 years
- Effective data length: 200-600
- Variables include:
  - Percent travel (h)
  - Pressure drop (dP)
  - Flow coefficient: ELF, Sachdeva, FlowCurve
  - Date Time

Choke no.	Time span	Nb. data
1	409	284
2	412	287
3	383	260
4	612	597
5	612	597
6	612	597
7	612	597

Choice of the Cv computation method

- ELF, Sachdeva, FlowCurve both are raw Cv
- Computed based on different models from process parameters (pressure drop, flow rate…)
- Which one is more reliable?
- Sachdeva is used in the following.



NTNU

Methods

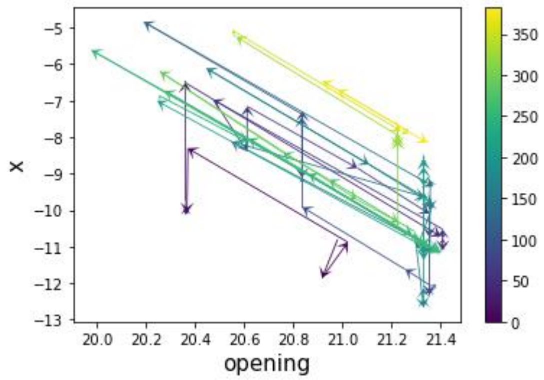
- Two stage hybrid model (TSHM)
- Spatio-temporal interpolation
  - Kriging
  - Inverse distance weighting
  - Trend surface analysis
- Time series analysis
  - ARIMA with exogeneous variable
- Stochastic process
  - Wiener process

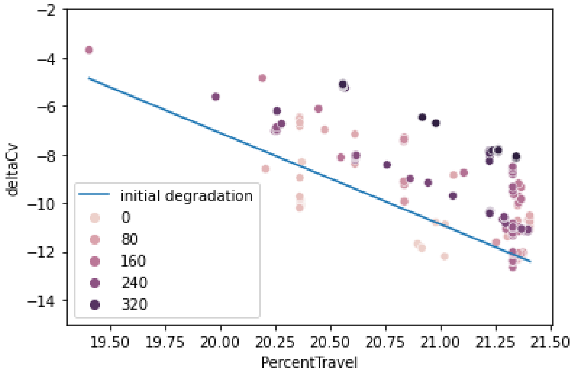
TSHM

- Two stage hybrid model (TSHM)
- Basic assumption: at any time, the Cv deviations at two different percent travel, are linearly dependent
- Stage 1: estimate the initial degradation as a function of percent travel (deterministic)
- Stage 2: modeling of the degradation increments (stochastic)

TSHM: model structure

- Basic assumption: at any time, the Cv deviations at two different percent travel, are linearly dependent
- Notations:
  - H, h: percent travel
  - T, t: time
  - $z_t = Cv^{Sachdeva}(t) - Cv^{Theoretical}(h_t)$ is the raw Cv deviation
- Model structure
- $z_t = \phi(h_t|\theta_\phi) + X_t + g(t, h_t|\theta_g)$
  - $\phi(h|\theta_\phi)$ is the initial degradation, independent of t
  - X_t: "hidden" degradation due to erosion, independent of h, imposes a translation effect on $\phi(h|\theta_\phi)$
  - $g(t, h_t|\theta_g)$: percent travel-dependent degradation growth accounting for the mixed effect between time and percent travel, e.g., the longer a valve operates at a small h, the higher the Cv deviation





TSHM: model identification

- Identification of X_t is done after the determination of ϕ and g!
  - Do not need assumptions on X_t
  - Counter example: assume X_t is a gamma process with mean jump size depending on h and ϕ a polynomial, then the parameters in X_t and ϕ can be deduced simultaneously via state-space model
  - Problem with too many assumptions:
    - Too much subjectivity
    - hard to validate
- Main steps:
  - Determine the structures of $\phi(h_t|\theta_\phi)$ and $g(t, h_t|\theta_g)$
  - Estimate θ_ϕ and θ_g without knowing the distribution of X_t
  - Get X_t
  - Prognosis

Step 1: Determine ϕ and g

- ϕ and g are considered polynomial. Example:
  - $\phi(h) = c_0 + c_1 h + c_2 h^2 + \cdots + c_p h^p$
  - $g(t, h_t) = t * [a_1(h_t - h_{\min}) + a_2(h_t - h_{\min})^2 \ldots a_q(h_t - h_{\min})^q]$
- The choice of model structure, i.e., whether ϕ is a polynomial or a power function, or how the time and percent travel interacts in the function g, is subjective
- Hyper-parameters: p, q can be tuned via grid search cross validation or assigned
- Parameters: $\theta_\phi = c_0 \ldots c_p, \theta_g = a_1 \ldots a_q$ can be estimated once model structure and hyper-parameters are determined

Step 2: Estimate θ_ϕ and θ_g

- Differentiating the raw observations z:

$$\Delta X = \Delta z - \Delta\phi(\theta_\phi) - \Delta g(\theta_g)$$

- Let $\mu = E[X]$, the residuals (with mean 0) are

$$r = \Delta X - \mu = \Delta z - \Delta\phi(\theta_\phi) - \Delta g(\theta_g) - \mu$$

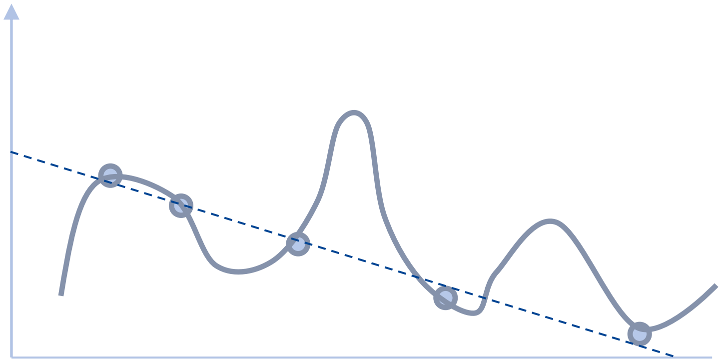
- To infer θ_ϕ and θ_g, we can minimize a loss function associated with r when μ is known

$$L(r_{1:n}) = \left(\sum_{i=1}^{n} |r_i|^s\right)^{\frac{1}{s}}$$

- Two options when μ is unknown:
  - Option 1: estimate μ together with θ_ϕ and θ_g (1st estimator)
  - Option 2: ignore μ and minimize a loss function of ΔX (2nd estimator)
- Robust estimation: use the 2nd estimator with $s = 1$ (L1 loss, sum of absolute residuals)

Step 2: hyper-parameter tuning

- The more complex the polynomial, the smaller the loss, but also the more difficult it is to ascribe physical meaning to it.
- Use grid search CV to determine the best p and q
- Schema:
  - Split the data into train and test sets (no shuffling)
  - Construct the searching space: $P = [1,2,3,\dots], Q = [1,2,3,\dots]$
  - For p, q in P, Q:
    - Use training set to estimate parameters θ_ϕ^{train} and θ_g^{train}
    - Compute the loss function on the test set with θ_ϕ^{train} and θ_g^{train}
    - Obtain the average loss over all partitions of train/test split
  - Best p and q: minimize the average loss

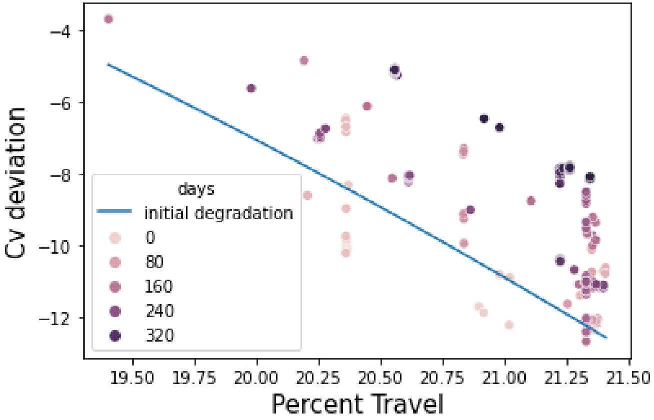


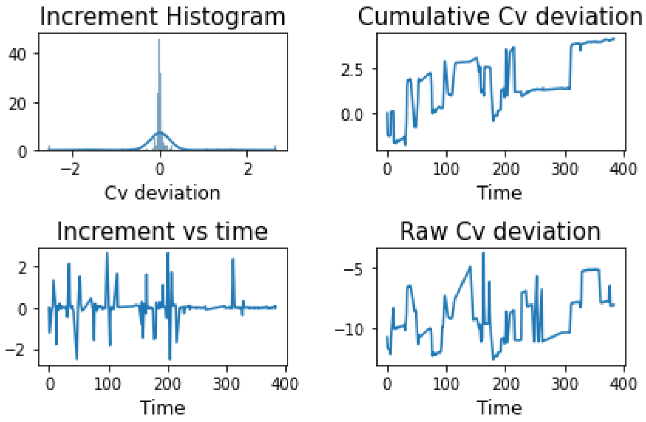
...train...				Test set	...train...				
1	2	3	4	5	6	7	8	9	10

NTNU

THSM: applied to valve 3

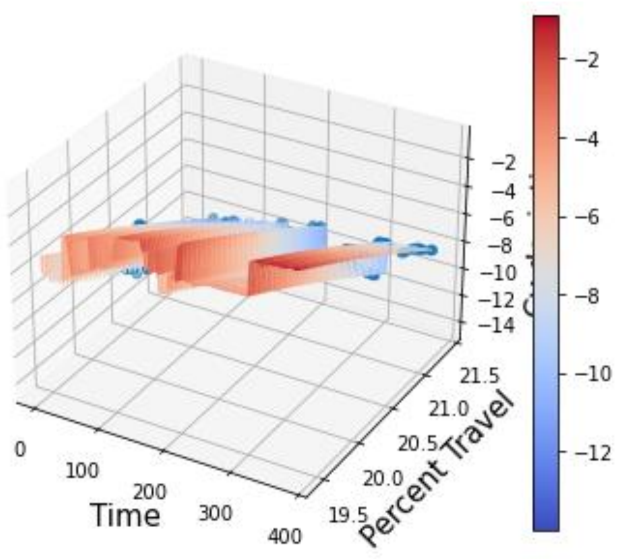
- ϕ and g are considered polynomial.
- (p, q) are tuned as (2,0): translation effect on the initial degradation ϕ
- $\phi(h) = 4.18h - 0.20h^2 - 12.67$
- Hidden degradation X_t: Excess kurtosis and heavy tailed
- Follows a student t distribution (Cramér–von Mises, p-value=0.18)

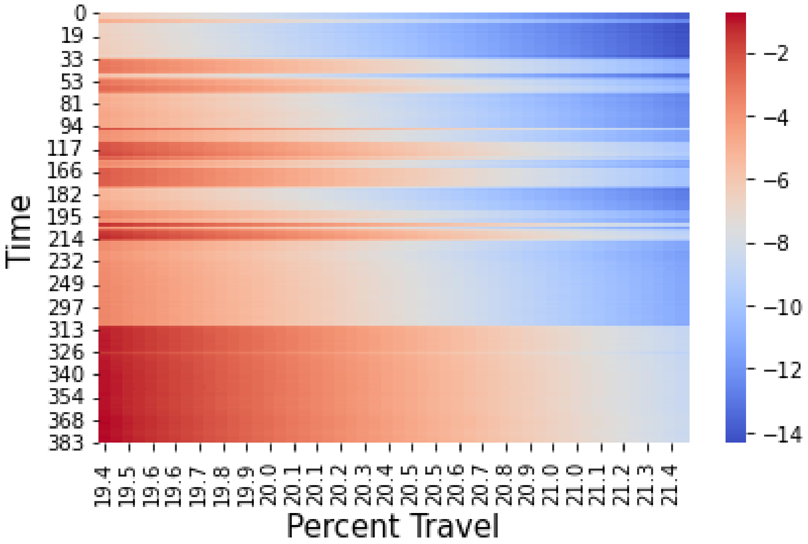




THSM: applied to valve 3

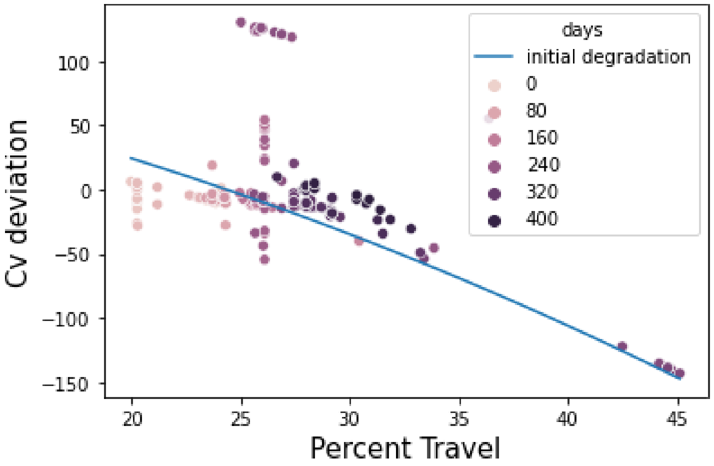
- Interpolation: not smooth in the t axis

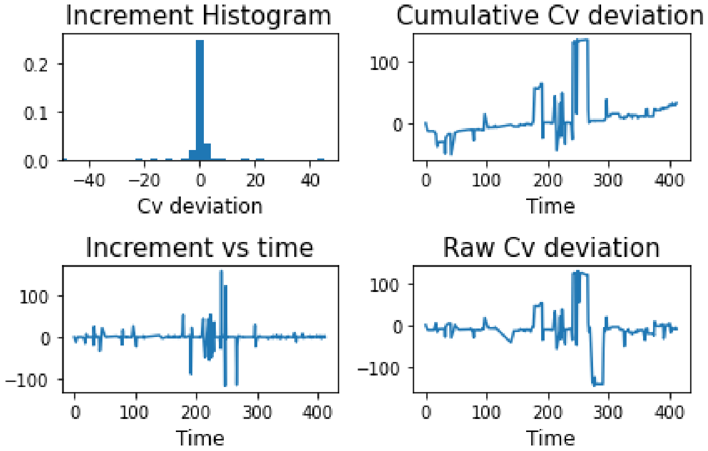




THSM: applied to valve 2

- ϕ and g are considered polynomial.
- (p, q) are tuned as (2,0): translation effect on the initial degradation ϕ
- $\phi(h) = -2.83h - 0.06h^2 + 105.24$
- Hidden degradation X_t: Excess kurtosis and heavy tailed

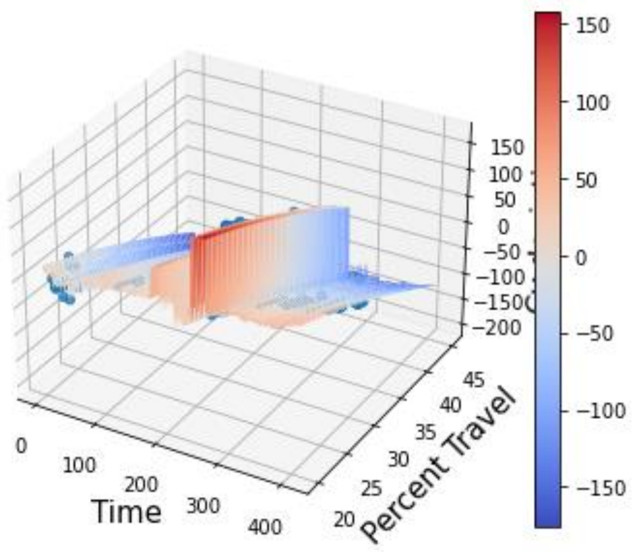


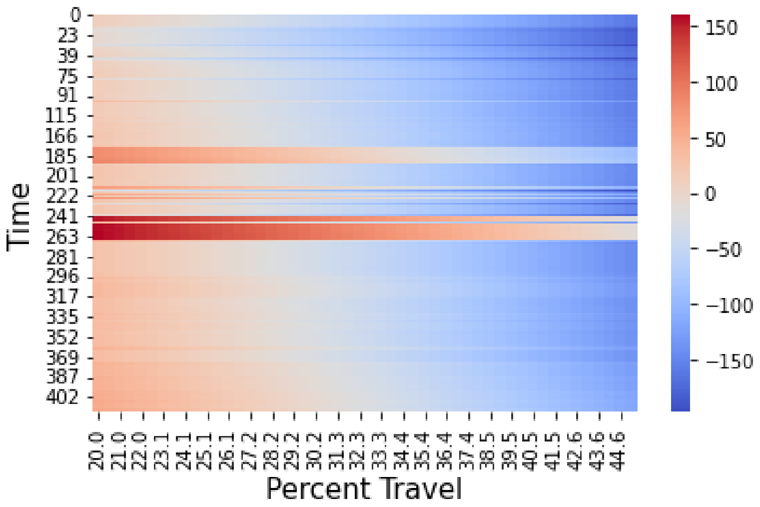


NTNU

THSM: applied to valve 2

- Interpolation: not smooth in the t axis



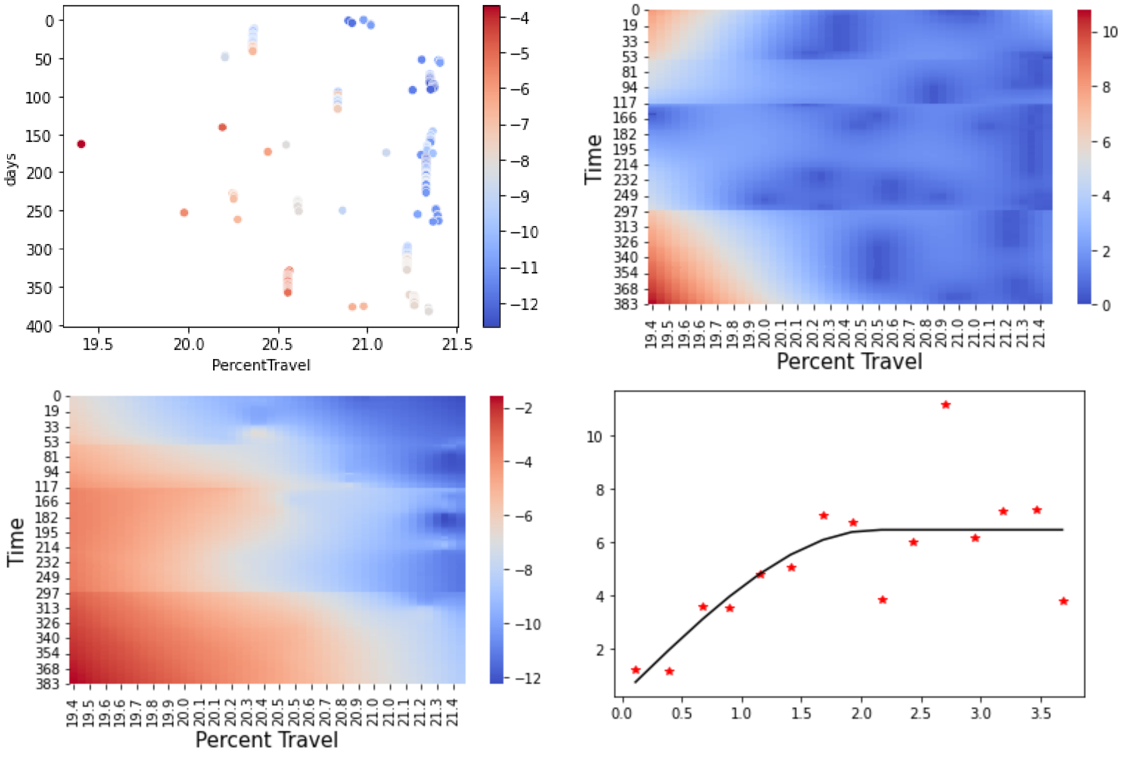


THSM: pros and cons

- Features:
  - Percent travel and degradation: treated separately
  - Fast
  - Intuitive and easily understandable
- Drawbacks
  - Linear relation in the H axis: too strong an assumption
  - Data sparsity at certain locations: increased uncertainty
  - Sampling bias due to changes in the percent travel
    - Valve 3: $19 < h < 22$
    - Valve 2: no data collected between 35 and 40
    - Valve 5: h increases from 25 to 60, data unbalance
  - Non-smooth interpolation surface
  - Unable to derive the standard deviation for the interpolated points
  - Degradation trend may be masked by noise/large jumps

Kriging applied to valve 3

- Hyper-parameter tuning:
  - Anisotropy scaling factor: 0.002 (time span: 380--0.76 days)
  - Variogram: spherical (sill=6.21, nugget=0.25, effective range =2.12)



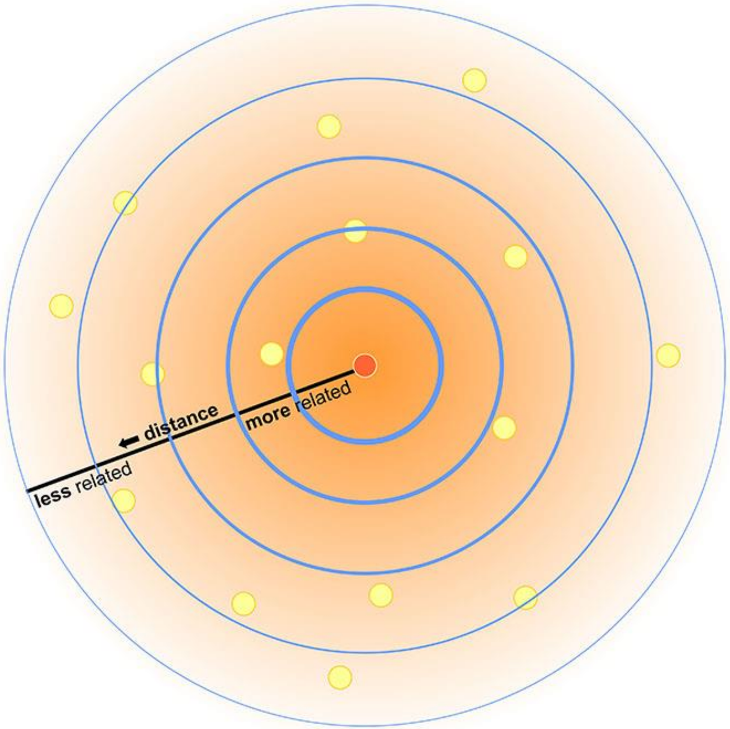
NTNU

Spatio-temporal interpolation

- Observed Cv deviation: $z_1(h_1, t_1), z_2(h_2, t_2) \ldots z_n(h_n, t_n)$
- Degradation evaluation for the past: find $z(h, t)$ for $t < t_n, h \in H$
- Prognosis: estimate $z(h, t)$ for $t > t_n, h \in H$
- The percent travel H forms the 1D spatial dimension
- The time T forms the 1D time dimension

First Law of Geography

- “Everything is related to everything else, but near things are more related than distant things.”-- Waldo R. Tobler.



https://www.e-education.psu.edu/maps/l2_p2.html

Inverse distance weighting

- The assigned values to unknown points are calculated with a weighted average of the values available at the known points.

$$z(x, y) = \sum_i w_i \, z_i \, , w_i = \frac{\left(\frac{1}{d_i}\right)^p}{\sum_k \left(\frac{1}{d_k}\right)^p},$$

- z_i: values at (x_i, y_i)
- w_i: weights
- d_i: Euclidean distances between (x_i, y_i) and (x, y)
- p: exponent that controls the weighting of z_i on z
  - small p tends to yield estimated values as averages of z_i in the neighborhood
  - large p tends to give larger weights to the nearest points and increasingly down weights points farther away.

Spatio-temporal IDW

- IDW adapted to interpolate spatio-temporal data, with time as an additional dimension

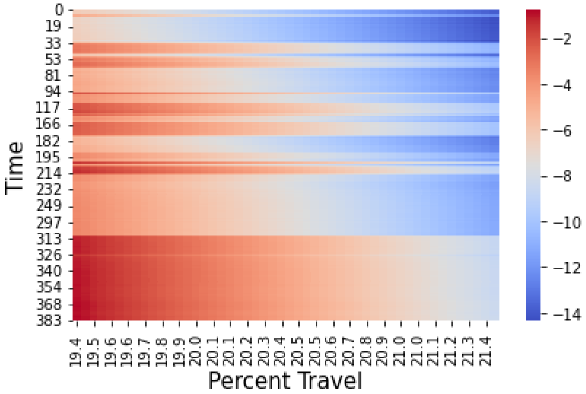
$$z(h,t) = \sum_i w_i \, z_i \, , w_i = \frac{\left(\frac{1}{d_i}\right)^p}{\sum_k \left(\frac{1}{d_k}\right)^p},$$

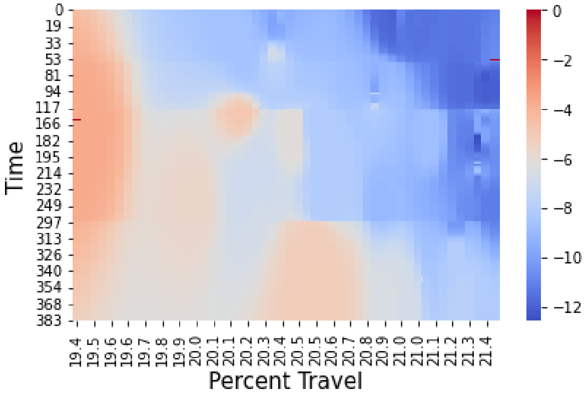
$$d_i = \sqrt{(h_i - h)^2 + c^2 (t_i - t)^2}$$

- z_i: values at (h_i, t_i)
- w_i: weights
- c: scaling factor
- Model structure is determined by c and p, which can be tuned by grid search CV

IDW applied to valve 3

- Hyper-parameter tuning: $c = 0.002, p = 5$
- Left: THSM, right: IDW
- IDW is a deterministic method and cannot provide a measure of uncertainty.





NTNU

Spatial correlation and variogram

- The theoretical variogram $2\,\gamma(\boldsymbol{s}_1, \boldsymbol{s}_2)$ is a function describing the degree of spatial dependence of a spatial random field or stochastic process, Z(s)

$$2\,\gamma(\boldsymbol{s}_1, \boldsymbol{s}_2) = Var[Z(\boldsymbol{s}_1) - Z(\boldsymbol{s}_2)]$$

- If the spatial random field has constant mean:

$$2\,\gamma(\boldsymbol{s}_1, \boldsymbol{s}_2) = E\left[\left(Z(\boldsymbol{s}_1) - Z(\boldsymbol{s}_2)\right)^2\right]$$

- If the covariance function of a stationary process exists it is related to variogram by:

$$2\,\gamma(\boldsymbol{s}_1, \boldsymbol{s}_2) = C(\boldsymbol{s}_1, \boldsymbol{s}_1) + C(\boldsymbol{s}_2, \boldsymbol{s}_2) - 2C(\boldsymbol{s}_1, \boldsymbol{s}_2)$$

- where

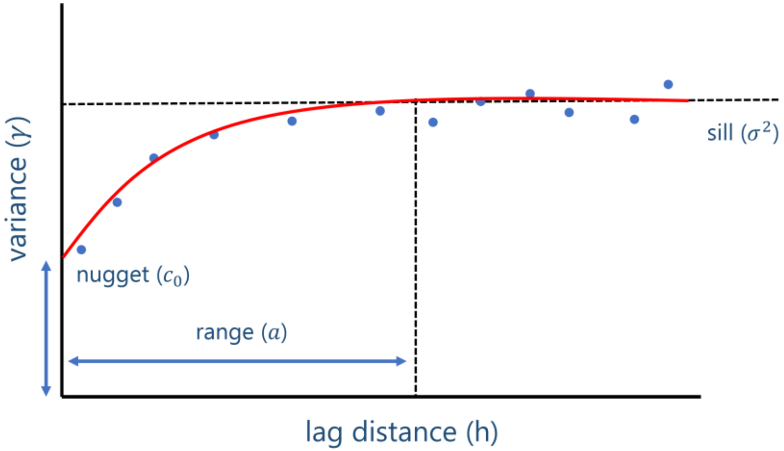
$$C(x, y) = Cov\big(Z(x), Z(y)\big)$$

- Variogram is a measure of dissimilarity over a distance. It shows how two data points are correlated from a spatial perspective, and provides useful insights when trying to estimate the value of an unknown location using collected sample data from other locations.

NTNU

Variogram explained

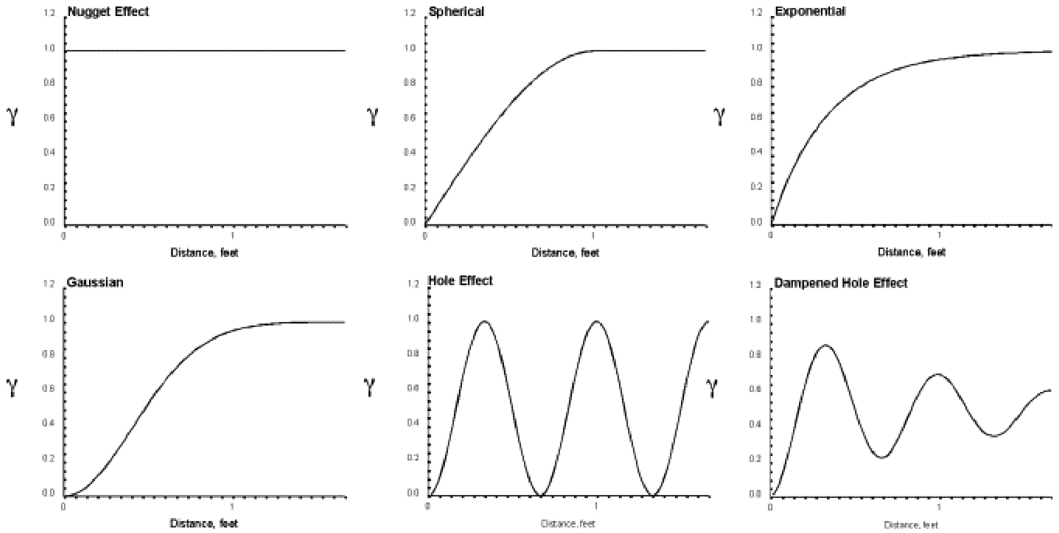
- Sill: the variance in which spatial data pairs lose correlation.
  - As the distance between two data points increases, it will be less likely that those two data points are related to one another.
- Nugget: the nonzero intercept of the variogram.
  - It is an overall estimate of error caused by measurement inaccuracy and environmental variability occurring at fine enough scales to be unresolved by the sampling interval.
- Range: a distance in which the spatial variability reaches the sill.
  - It is the distance beyond which observations are no longer correlated



https://aegis4048.github.io/spatial-simulation-1-basics-of-variograms

Variogram models

- The empirical variograms are approximated by theoretical models to ensure validity (e.g., conditionally negative definite function), which will then be used for kriging.
- Common bounded models:
  - Spherical
  - Exponential
  - Gaussian
- Unbounded models:
  - Linear
  - Power



Clayton V. Deutsch, in Encyclopedia of Physical Science and Technology (Third Edition), 2003

Spatio-temporal covariance function

- Since H forms the 1D spatial dimension and T forms the 1D time dimension, a more "correct" model would be spatio-temporal covariance function
- The random field $Z(h, t)$ can be described by

$$2\,\gamma(h; t) = Var[Z(h_0 + h; t_0 + t) - Z(h_0;\ t_0)]$$

- Or equivalently by its covariance function

$$C(h_1, h_2; t_1, t_2) = \gamma(h_1; t_1) + \gamma(h_1; t_1) - \gamma(h_1 - h_2; t_1 - t_2)$$

- However, due to data sparsity (only one measurement per day), the empirical spatio-temporal variogram cannot be estimated.
- Time is treated as the 2nd spatial dimension, with an anisotropy scaling factor.

Trend and Anisotropy

- Anisotropy:
  - An isotropic phenomenon is a process that is not directionally dependent. In spatial studies, this process is considered to evolve similarly in all the directions in space.
  - On the contrary, anisotropy refers to a process that varies differently according to the direction of interest. Often, it denotes a characteristic of a random process that shows higher autocorrelation in one direction than another.
  - Time and percent travel: different autocorrelation
- Trend:
  - Cv deviation increases in percent travel
  - Cv deviation increases in time
  - Can be described by a deterministic function
  - Incorporate external variables

Ordinary kriging

- Assumption: data are intrinsically stationary

$$Z(\boldsymbol{s}) = \mu + \epsilon(\boldsymbol{s})$$

- Interpolation at the target position s_0:

$$Z^*(\boldsymbol{s_0}) = \sum_{i=1}^{n} w_i \, Z(\boldsymbol{s_i})$$

- The estimator is unbiased ($E[Z^*(\boldsymbol{s_0}) - Z(\boldsymbol{s_0})] = 0$), and the weights w_i are determined by minimizing the estimation variance

$$\sigma_E^2 = E\left[\left(Z^*(\boldsymbol{s_0}) - Z(\boldsymbol{s_0})\right)^2\right]$$

- Subjected to

$$\sum_{i=1}^{n} w_i = 1$$

Universal kriging

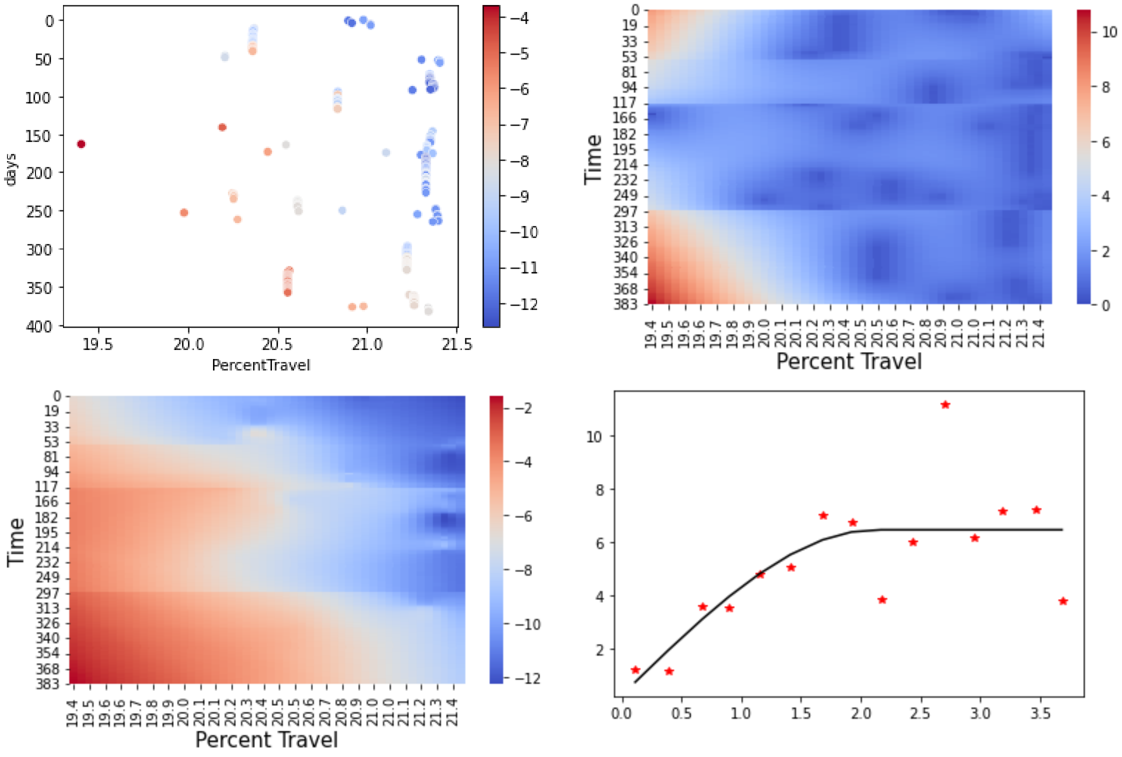
- Kriging with unknown mean or external drift

$$Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$$

- The trend can be linear, quadratic…
- Universal kriging: trend modeled as a function of the coordinates
- Kriging with external drift: trend defined via auxiliary variables

Kriging applied to valve 3

- Hyper-parameter tuning:
  - Anisotropy scaling factor: 0.002 (time span: 380--0.76 days)
  - Variogram: spherical (sill=6.21, nugget=0.25, effective range =2.12)



NTNU

Compare model performances by rolling forecasting

- Train/test split:
  - Fit a model Data[1:k]
  - Use the fitted model to predict data[k+1:k+m], where m is the prediction length
  - Let k move forward
- Example: valve 3.
  - Total data length: 210
  - Prediction start at $k_0 = 100$, prediction length $m = 5$.
  - Overall performance is averaged over 105 predictions
- Metrics:
  - mean absolute error (MAE),
  - root mean squared error (RMSE),
  - median absolute relative error (MARE),
  - R-squared (R2)

Table 4: Case study 1: forecasting errors

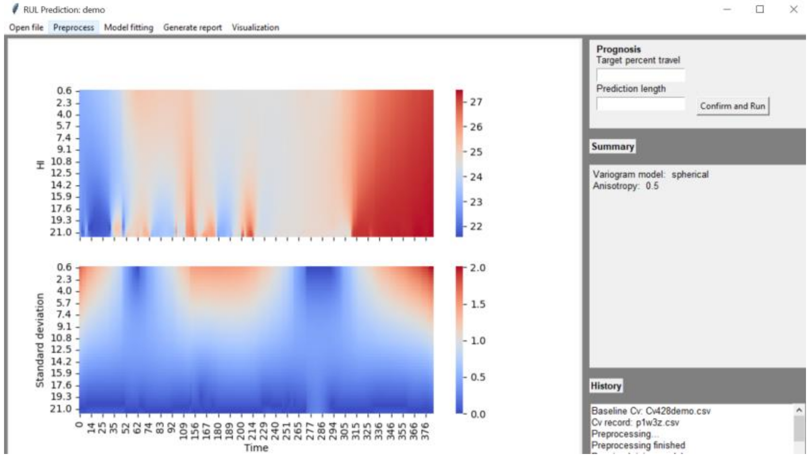
	TSHM				Kriging			
	MAE	RMSE	MARE	R^2	MAE	RMSE	MARE	R^2
1	0.16	0.50	0.003	0.94	0.23	0.63	0.004	0.91
2	0.26	0.69	0.005	0.88	0.40	0.87	0.005	0.82
3	0.29	0.72	0.007	0.87	0.48	0.96	0.008	0.78
4	0.35	0.80	0.009	0.84	0.58	1.07	0.010	0.71
5	0.41	0.90	0.012	0.79	0.69	1.20	0.012	0.63
	Inverse Distance Weighting				Trend surface analysis			
1	0.24	0.62	0.003	0.91	1.35	1.60	0.13	0.40
2	0.40	0.83	0.006	0.83	1.43	1.67	0.14	0.33
3	0.47	0.92	0.011	0.80	1.48	1.71	0.14	0.29
4	0.57	1.03	0.015	0.74	1.51	1.75	0.15	0.24
5	0.67	1.13	0.024	0.67	1.56	1.79	0.16	0.18
	ARIMAX				Wiener process			
1	1.18	1.42	0.13	0.52	0.50	1.18	0.004	0.67
2	1.42	1.67	0.15	0.33	0.80	1.52	0.008	0.44
3	1.55	1.80	0.16	0.21	0.96	1.69	0.015	0.31
4	1.63	1.89	0.17	0.11	1.02	1.67	0.025	0.30
5	1.70	1.97	0.18	0.00	1.19	1.81	0.045	0.16

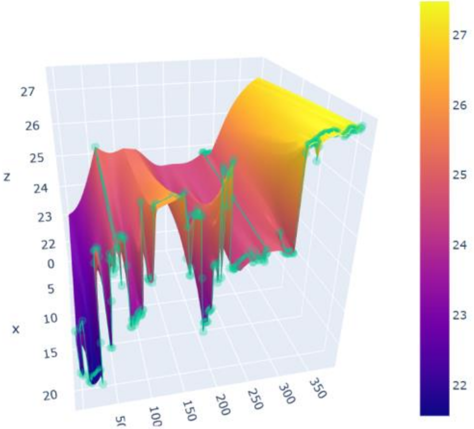
Summary of the two methods

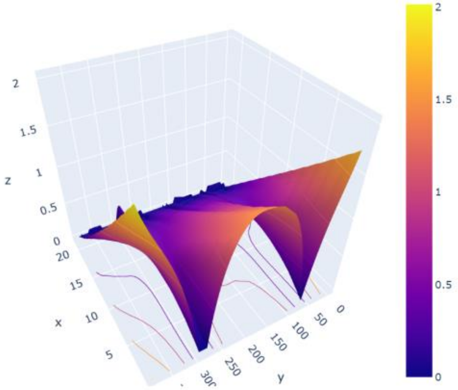
	Two stage hybrid model	Kriging
Basic assumption	Linear dependence in percent travel	First law of geostatistics
Additional assumptions	Structure of ϕ (initial degradation) and g (need to specify how the operation time and percent travel interacts)	Second order stationarity of the random field Structure of the trend function
Hyper parameters	Polynomial orders of ϕ and g	Variogram type, anisotropy scaling factor
Type	Deterministic ϕ and g Stochastic in forecasting	Stochastic
Uncertainty	Only in forecasting	Over the random field
Running time	Fast	Slow (for hyper-parameter tuning if the data is large)

Toolbox

- Functionalities
  - Cv and Cv deviation overview
  - Degradation trend estimation
  - Uncertainty quantification
  - 3d interactive visualization
- Programming language: Python
- Methods:
  - Two stage hybrid model
  - Kriging
- Working mode: offline. Data should be manually imported.
- Full documentation will come by the end of September.







Toolbox: input file

- Theoretical Cv
  - Percent travel as the first column, from 0 to 100
  - Corresponding Choke Cv as the second column
- Observed Cv (ELF, Sachdeva or Flowcurve)
  - Time, percent travel and Cv as the first, second and third columns

Choke travel (% or °)	Choke Cv
100	428
95	427.41
90	425
86	417
81	408
76	393
71	375

Time (Choke 1 Percent Trave	Choke 1 Percent Travel -	Choke 1 Flow Coefficient - Sachdeva (IFM.WellChkPerf.PDM.Day)-avg
2020-06-14T22:00:00Z	0	0
2020-06-17T22:00:00Z	20.98048615	23.26839193
2020-06-18T22:00:00Z	20.89753161	22.0464681
2020-06-21T22:00:00Z	20.91711355	21.95760091
2020-06-23T22:00:00Z	21.0205885	22.0242513
2020-06-24T22:00:00Z	21.02232761	23.35725911
2020-06-28T22:00:00Z	20.3707406	23.37947591
2020-06-29T22:00:00Z	20.36745929	21.60213216
2020-06-30T22:00:00Z	20.36589345	21.57991536
2020-07-01T22:00:00Z	20.36501546	21.66878255
2020-07-02T22:00:00Z	20.36238757	21.53548177