



# Del 2 – Metoder och tekniker för Datadriven Dammövervakning

IVL @ Energiforsk, 25 januari 2023

Anders Björk, Håkan Fridén, Dag Glebe

# Agenda

- Metoder för signalövervakning
- Metoder för dataanalys
- Metoder för att upptäcka och varna för avvikelser på givare och på damm
- Vår implementation
- Underhåll av övervakningsmodeller
- Kurs om datadriven dammövervakning
- Diskussion och frågor

# Olika metoder behövs för övervakning

- Signalövervakning
  - Syftar till att säkerställa att givarna lämnar korrekta data för att kunna rensa bort rent felaktiga data.
- Tillståndsovervakning
  - Syftar till att upptäcka och varna för avvikelser från givarsignaler som kan indikera avvikelser i dammens tillstånd.

# Metoder för signalövervakning

- Univariat gränsvärdesbaserad validering -> signalbaserade händelser
  - Bortfall
  - Min, max
  - Frusen signal
  - Varians
  - 1:a och 2:a derivata
- Multivariat modellbaserad validering -> modellbaserade händelser
  - Signalmodellering från kringliggande signaler

# Metoder för tillståndsövervakning

- Multivariat modellbaserad övervakning
  - F. n. processlägesövervakning
  - Identifiering av felmoder ligger längre fram i tiden

# Metoder för dataanalys och modellgenerering

# Analysmetoder

- Hydrostatic Season Time (HST) och varianter
  - Pseudolinjär regressionsmodell
  - Fysikalisk relevant för betongdammar, men använd på alla typer av dammar
- Neurala nätverk och andra moderna maskininlärningsmetoder
  - Allmänt olinjära
  - Svåra att tolka
- Projektionsmetoderna PCA och PLS
- Avvikelse-detektion genom SPC / MSPC

# Våra analysmetoder för datadriven dammövervakning

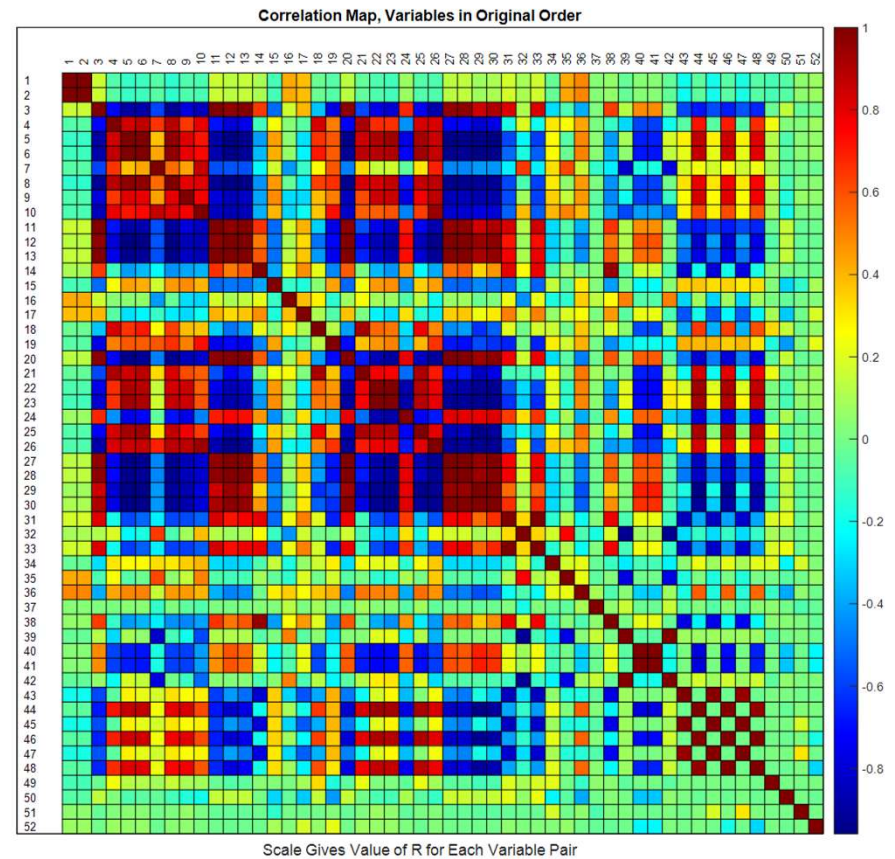
- Trender, korrelation, fasförskjutning
- PCA – för dammtillståndsovervakning
- PLS – för signalvalidering



# Korrelation

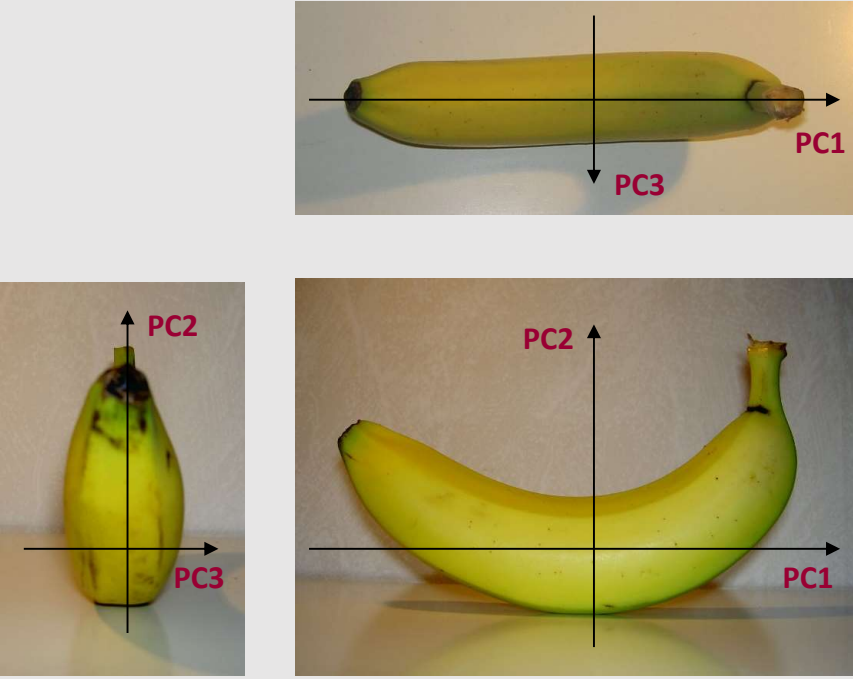
Vilka signaler korrelerar med vilka?

- Positiv korrelation - röd
- Negativ korrelation - blå



# Teckna en banan!

- Hur ritade du?

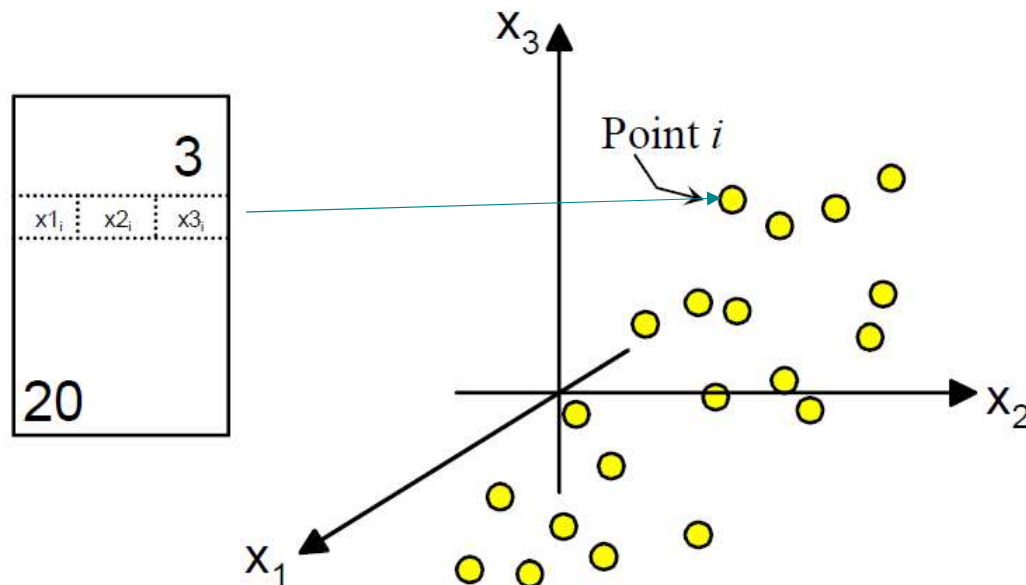


Källa: Fridén Infotech AB

Vi använder oftast intuitivt det perspektiv som ger mest information, = störst variabilitet, = först PC1, sedan PC2

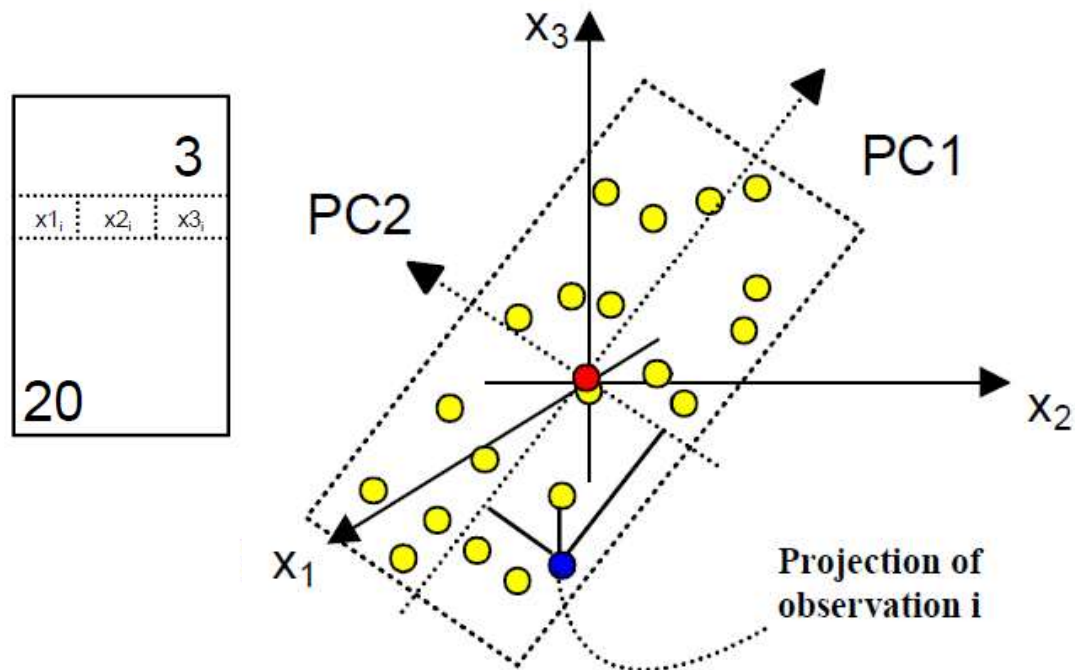
# En himla massa signaler

- En datatabell med  $N$  observationer och  $K$  variabler kan representeras med  $N$  punkter i en  $K$ -dimensionell rymd, variabelrymden! Varje rad,  $i$ , i tabellen blir en punkt i  $K$ -rymden.
- Detta är utgångspunkten för de multivariata projektiionsmetoderna.



# En himla massa signaler

- En datatabell med  $N$  observationer och  $K$  variabler kan representeras med  $N$  punkter i en  $K$ -dimensionell rymd, variabelrymden! Varje rad,  $i$ , i tabellen blir en punkt i  $K$ -rymden.
- Detta är utgångspunkten för de multivariata projektionsmetoderna.



# Multivariat dataanalys

- ➔ PCA och PLS kan användas för tre typer av problem



- ➔ Multivariata projektionsmetoder undviker överanpassning och fungerar med databortfall

# Multivariata projektionsmetoder

- ➔ **Generellt: Data = struktur + brus -> modell + residual**
- ➔ **PCA (Principal Components Analysis)**
  - PCA finner den dominerande strukturen i data. Med en PCA-modell med ett fåtal komponenter beskriver man ofta 80-90% av variationen i data.
- ➔ **PLS (Partial Least Squares Projection to Latent Structures)**
  - PLS finner relationen mellan prediktorer (X) och responser (Y) i ett dataset. Med en PLS-modell kan man från nya X prediktera motsvarande Y..
- ➔ **Modellerna visualiseras ofta med bilder, där man ser hur observationer och variabler hänger samman.**

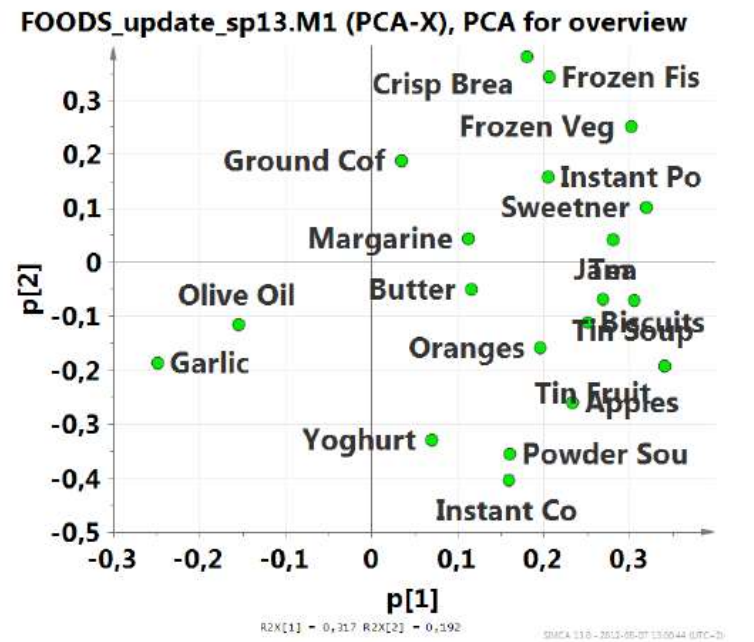
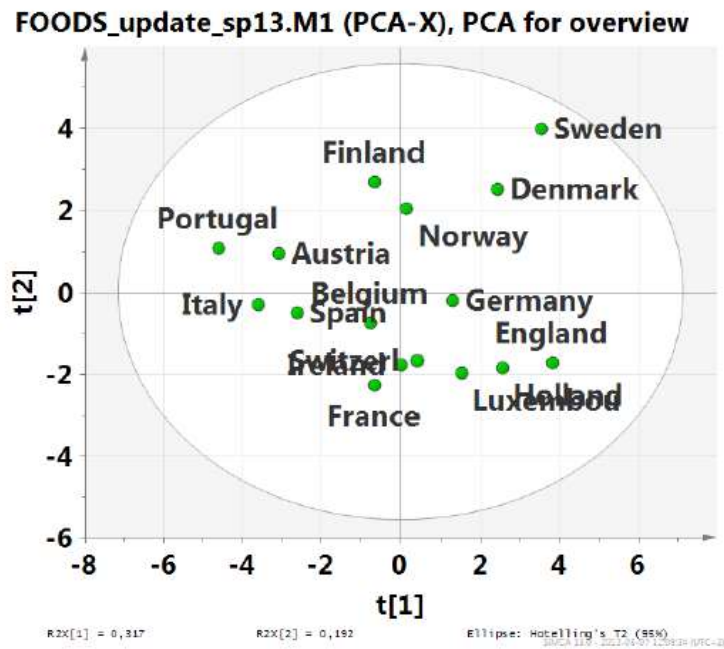
# PCA

- PCA beräknar det hyperplan i A dimensioner (komponenter) som så bäst som möjligt approximerar punkternas läge i variabelrymden.

Primary ID	ONAM	Gr_Coffe	Inst_Coffe	Tea	Sweetner	Biscuits	Pa_Soup	Ti_Soup	In_Potat	Fro_Fish	Fro_Veg	Apples	Oranges	Ti_Fruit	Jam	Garlic	Butter	Margarine	Olive_Oil	Yoghurt	Crisp_Bread
1	Germany	90	49	88	19	57	51	19	21	27	21	81	75	44	71	22	91	85	74	30	26
2	Italy	82	10	60	2	55	41	3	2	4	2	67	71	9	46	80	66	24	94	5	18
3	France	88	42	63	4	76	53	11	23	11	5	87	84	40	45	88	94	47	36	57	3
4	Holland	96	62	98	32	62	67	43	7	14	14	83	89	61	81	15	31	97	13	53	15
5	Belgium	94	38	48	11	74	37	23	9	13	12	76	76	42	57	29	84	80	83	20	5
6	Luxembou	97	61	86	28	79	73	12	7	26	23	85	94	83	20	91	94	94	84	31	24
7	England	27	86	99	22	91	55	76	17	20	24	76	68	89	91	11	95	94	57	11	28
8	Portugal	72	26	77	2	22	34	1	5	20	3	22	51	8	16	89	65	78	92	6	9
9	Austria	55	31	61	15	29	33	1	5	15	11	49	42	14	41	51	51	72	28	13	11
10	Switzerl	73	72	85	25	31	69	10	17	19	15	79	70	46	61	64	82	48	61	48	30
11	Sweden	97	13	93	31		43	43	39	54	45	56	78	53	75	9	68	32	48	2	93
12	Denmark	96	17	92	35	66	32	17	11	51	42	81	72	50	64	11	92	91	30	11	34
13	Norway	92	17	83	13	62	51	4	17	30	15	61	72	34	51	11	63	94	28	2	62
14	Finland	98	12	84	20	64	27	10	8	18	12	50	57	22	37	15	96	94	17		64
15	Spain	70	40	40		62	43	2	14	23	7	59	77	30	38	86	44	51	91	16	13
16	Ireland	30	52	99	11	80	75	18	2	5	3	57	52	46	89	5	97	25	31	3	9

# PCA

- PCA beräknar det hyperplan i A dimensioner (komponenter) som så bäst som möjligt approximerar punkternas läge i variabelrymden.

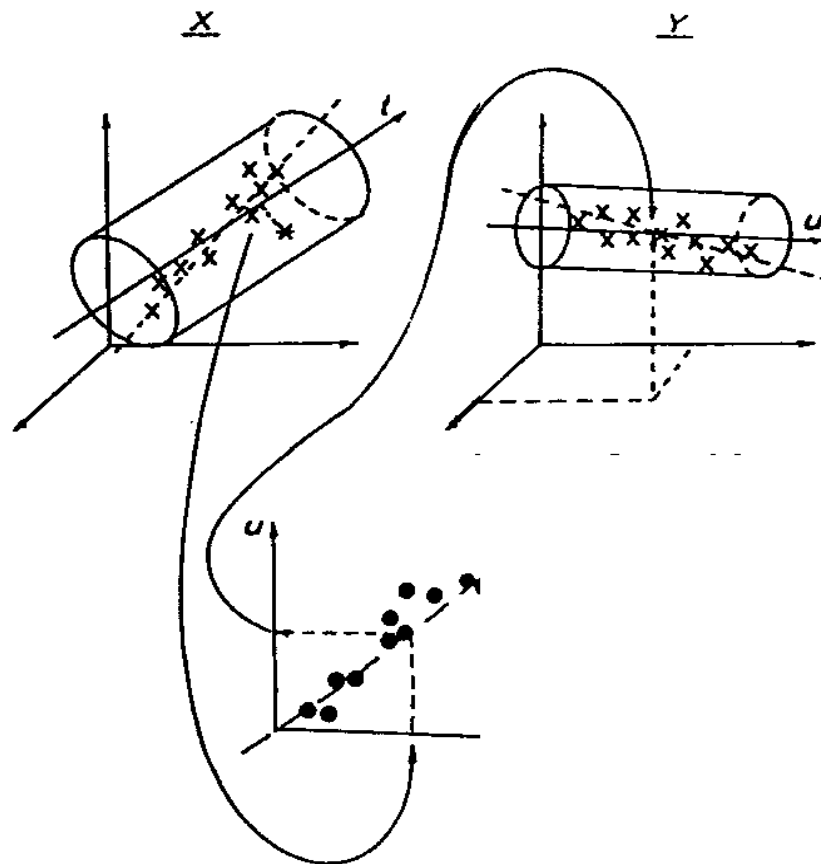




# PLS

Vid PLS representeras varje observation av två punkter, en i X-rymden och en i Y-rymden

PLS beräknar de hyperplan i A dimensioner (komponenter) som så bäst som möjligt optimerar kovariansen mellan observationernas läge i de båda rymderna X och Y.



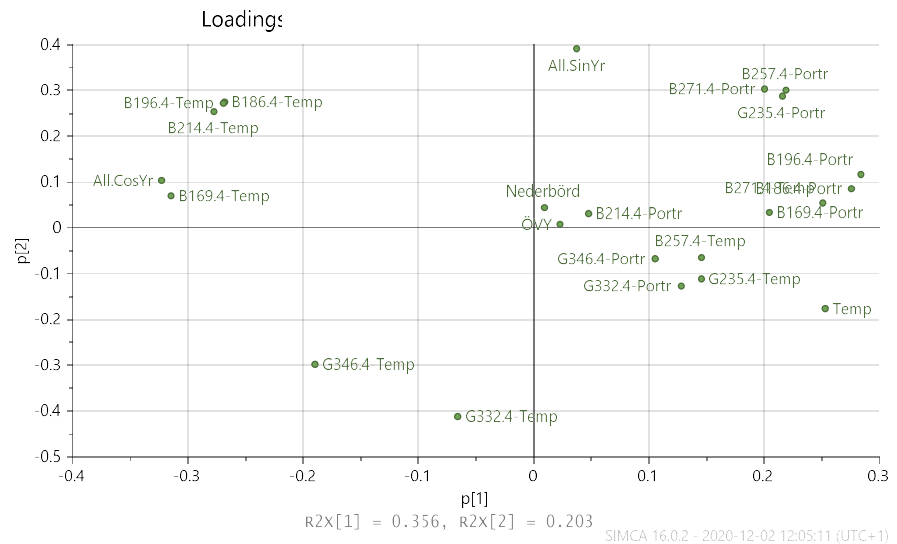
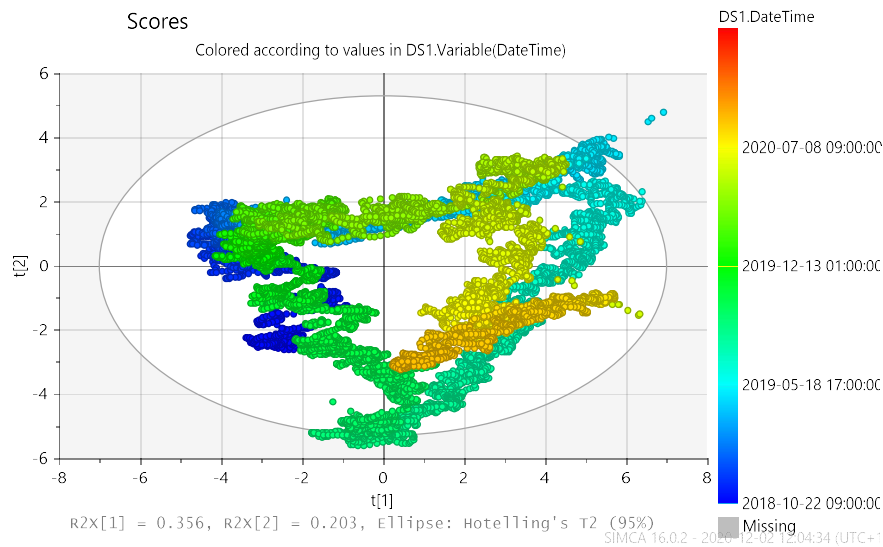
# Demo SIMCA

# PCA kan användas för processövervakning

PCA fångar den dominanta korrelationsstrukturen i data och kan användas för att följa en process och fånga tidiga avvikelser.

Komponent 1 och 2 bildar det plan i vilket data har sin maximala utbredning.

De två första komponenterna beskriver 56% av variationen i våra data

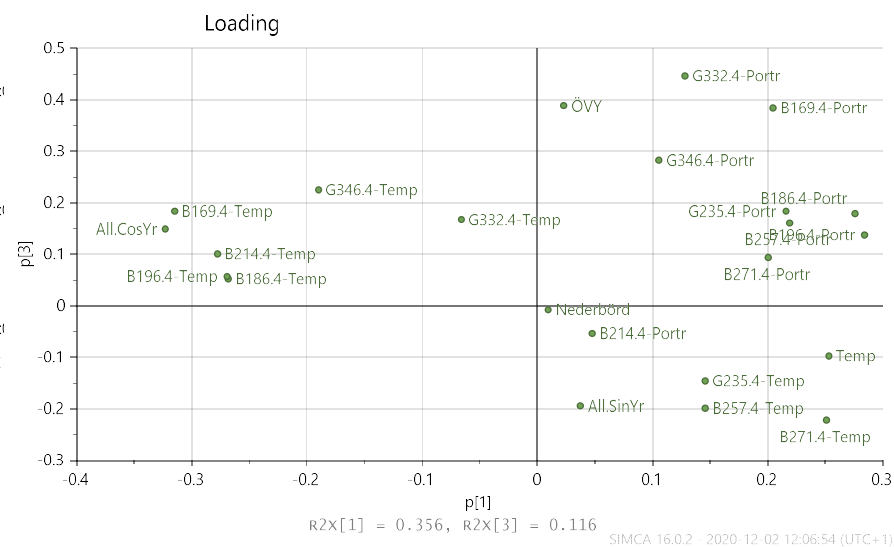
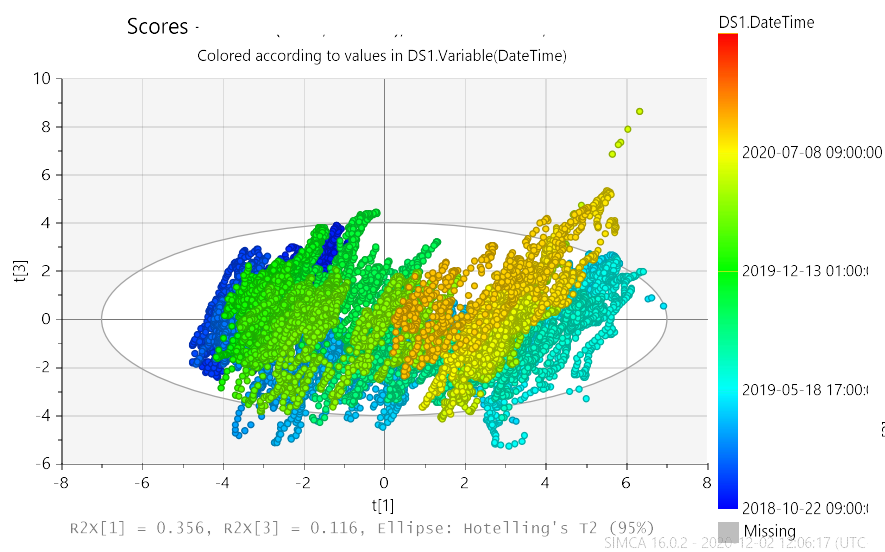


# PCA kan användas för processövervakning

- ➔ Komponent 1 och 3 fångar föregående bild från sidan

Den 3e komponenten beskriver 11% av data

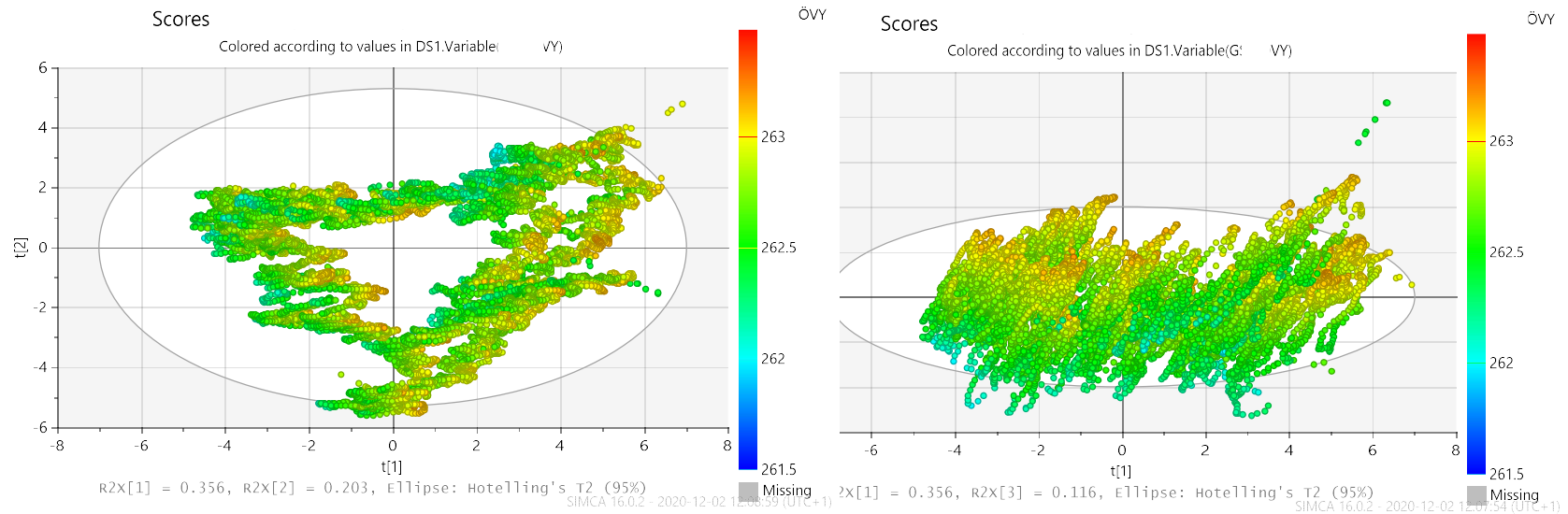
Med 3 komponenter beskrivs 68% av variationen i våra data



# PCA kan användas för processövervakning

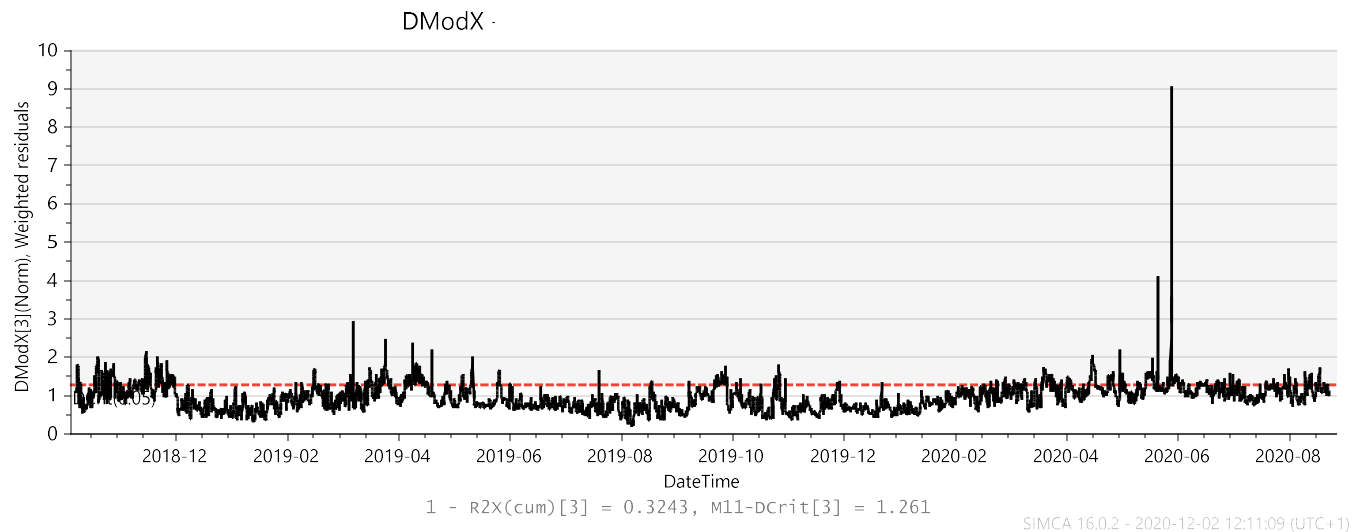
- ➔ Vad är det som varierar nerifrån vänster – uppåt höger?

Här är data färgade efter ÖVY.



## PCA, även det som inte modelleras kan visualiseras

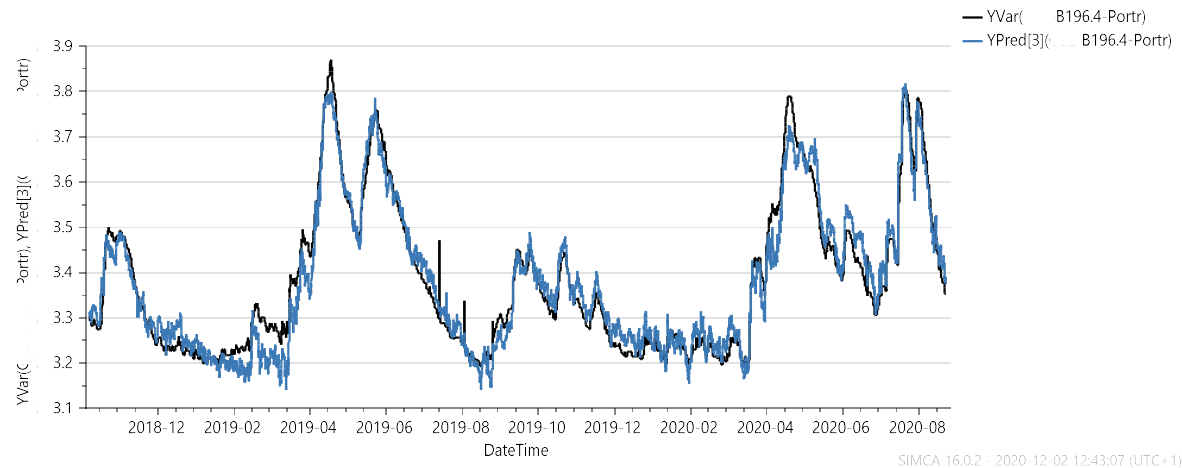
- Det som inte syns i PC1-3 är det som inte är modellerat, ligger i hyperplanet = det som återstår = residualen.
- Den rödstreckade linjen, DCrit, är den 95%-iga konfidensgränsen. Ovanför den kan vi hitta avvikare.



## PLS – en prediktionsmetod i k-rymden

### ➔ PLS för signalvalidering

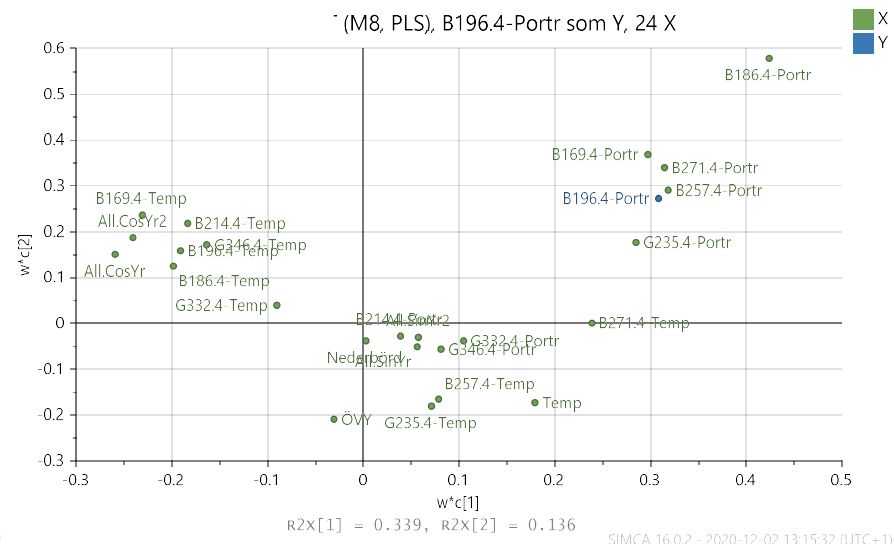
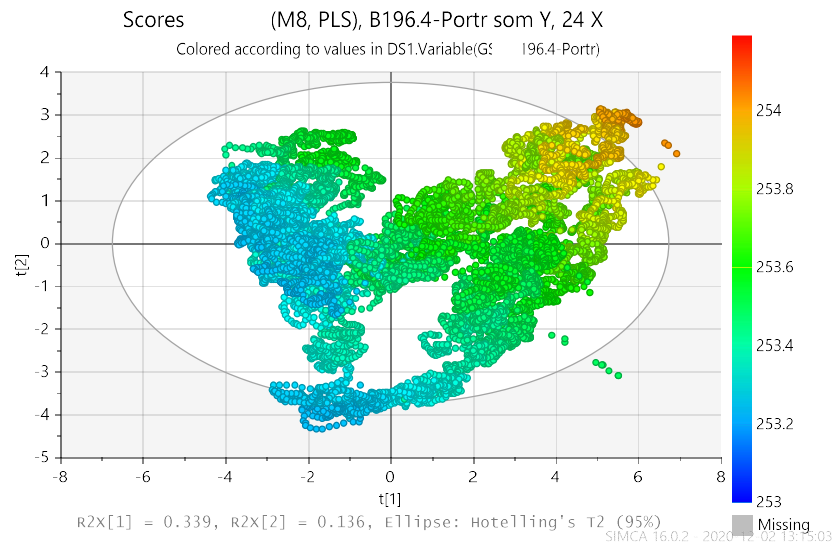
- Med PLS kan man låta valda omgivande signaler prediktera en övervakad signal (vad tycker de andra givarna om givarvärdet?)
- Här predikteras B196.4Portr från övriga 24 signalerna



# PLS – en prediktionsmetod i k-rymden

## → PLS för signalvalidering

- Med PLS kan man låta valda omgivande signaler prediktera en övervakad signal (vad tycker de andra givarna om givarvärdet?)
- Här predikteras Portr från övriga 24 signalerna

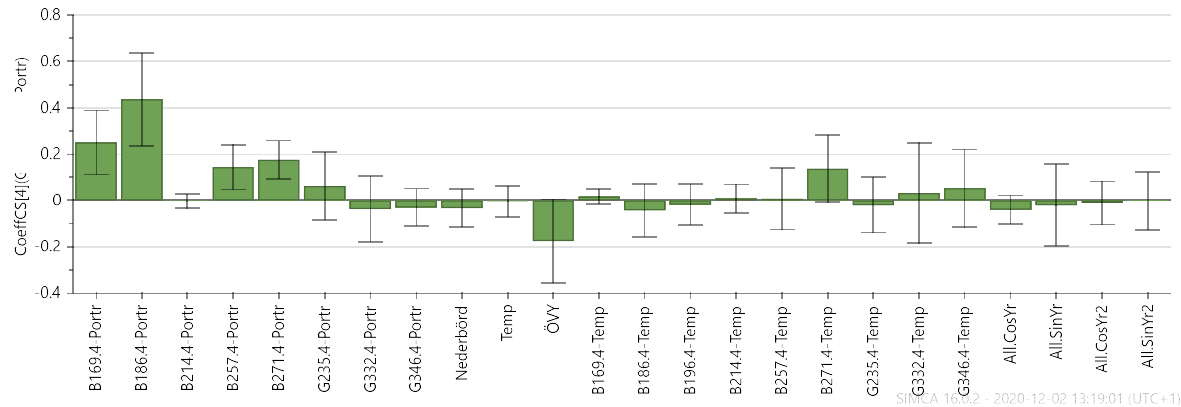




## PLS – en prediktionsmetod i k-rymden

### ➔ PLS för signalvalidering

- Med PLS kan man låta valda omgivande signaler prediktera en övervakad signal (vad tycker de andra givarna om givarvärdet?)
- Här predikteras B196.4-Portr från övriga 24 signalerna

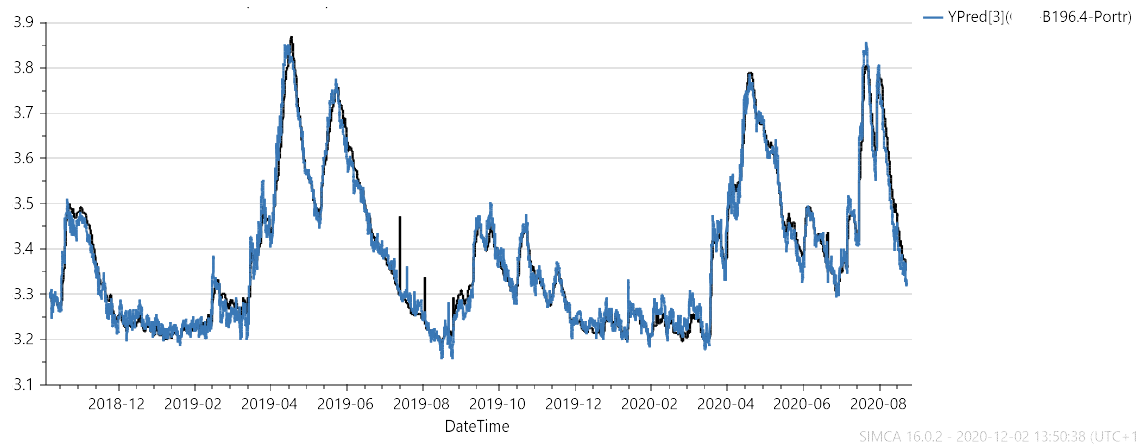


SLIPKA 16.22 - 2020-12-02 13:19:01 (UTC+1)

## PLS – en prediktionsmetod i K-rymden

### ➔ PLS för signalvalidering

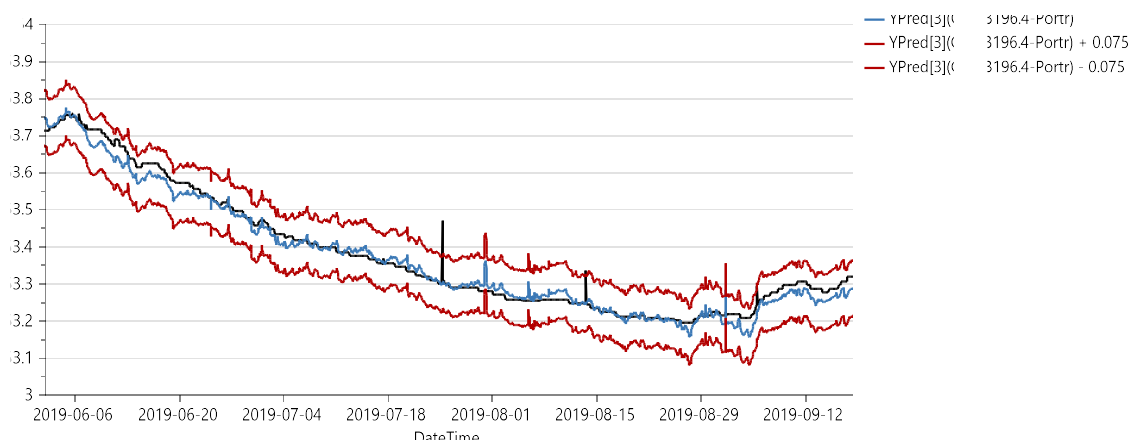
- Här har vi fått en bättre modell geom att bara ta med de 6 viktigaste X-variablerna
- Med PLS kan man övervaka en signal genom att prediktera den från omgivande signaler och vakta över differensen.



## PLS – en prediktionsmetod i K-rymden

### ➔ PLS för signalvalidering

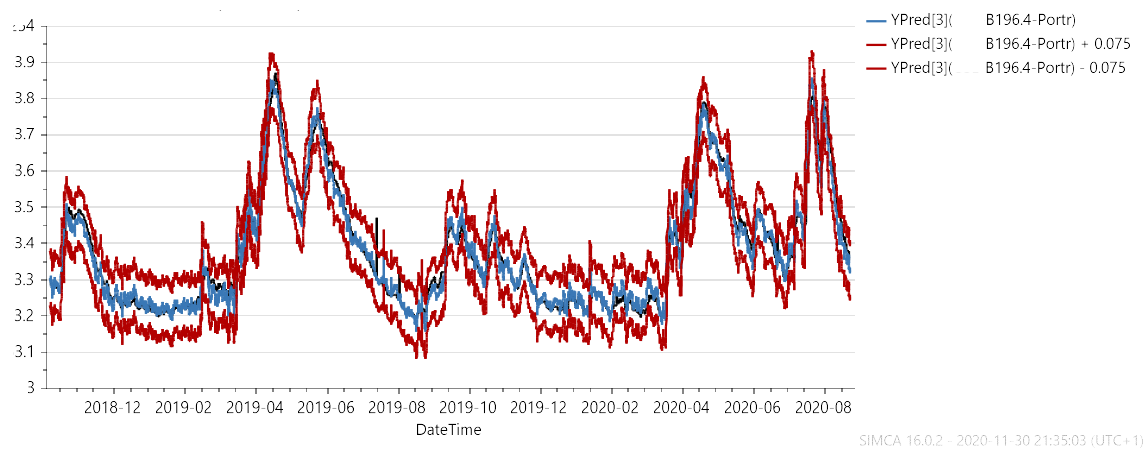
- Här har vi fått en bättre modell genom att bara ta med de 6 viktigaste X-variablerna
- Med PLS kan man övervaka en signal genom att prediktera den från omgivande signaler och vakta över differensen.



## PLS – en prediktionsmetod i K-rymden

### ➔ PLS för signalvalidering

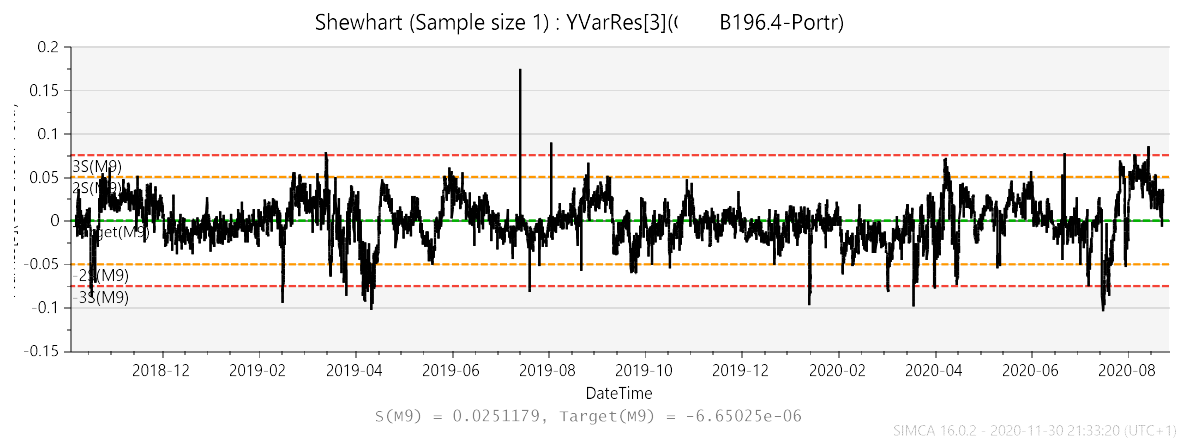
- Här har vi fått en bättre modell geom att bara ta med de 6 viktigaste X-variablerna
- Med PLS kan man övervaka en signal genom att prediktera den från omgivande signaler och vakta över differensen.



## PLS – en prediktionsmetod i K-rymden

### ➔ PLS för signalvalidering

- Här har vi fått en bättre modell geom att bara ta med de 6 viktigaste X-variablerna
- Med PLS kan man övervaka en signal genom att prediktera den från omgivande signaler och vakta över differensen.



## Kvalitetsmått för modelleringsgrad – R2 & Q2

- ➔ Modellpassning: Goodness-of-fit, R2,

$$R^2 = 1 - \frac{ReSS}{TotSS}$$

där ReSS är "Residual Sum of Squares" och TotSS är "Total Sum of Squares"

–  $R^2 \leq 0$  indikerar att modellen inte förklarar någon variation i data

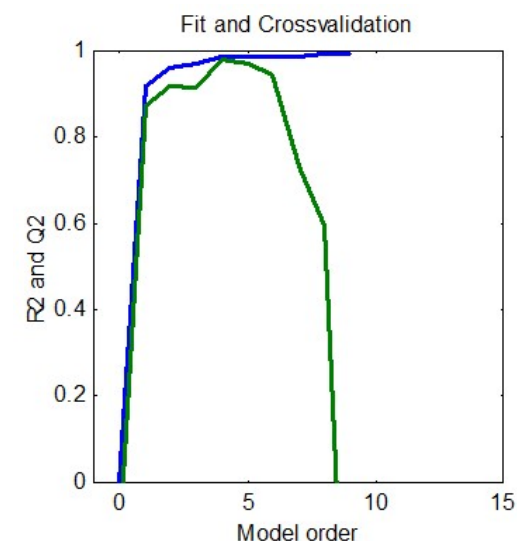
–  $R^2 = 1$  indikerar att modellen förklarar all variation inklusive brus

- ➔ Modellens prediktiva kvalitet: Goodness-of-prediction, Q2 (korsvaliderad prediktiv goodness-of-fit)

$$Q^2 = 1 - \frac{PrESS}{TotSS}$$

där PrESS är summan av kvadrerade värden av prediktionsfel

- ➔ Driver man R2 för högt minskar Q2, modellen inkluderar bruset. Vi säger då att modellen är överanpassad.

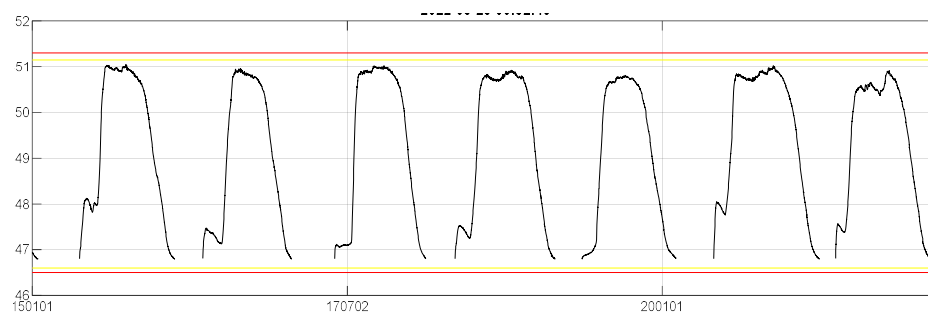
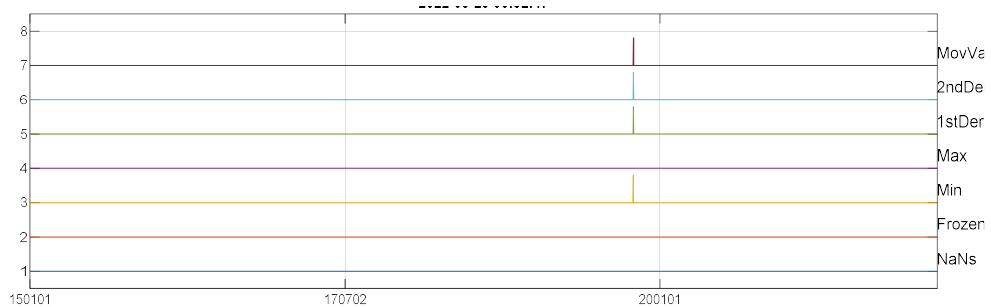
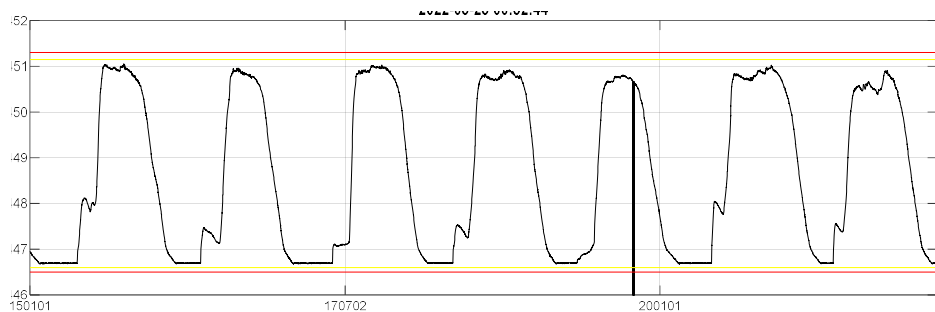


# Sammanfattning

- Datadrivna metoder är redo att implementeras
- Bra förarbete är A och O vid implementation av datadrivna metoder
- Metoder bör kunna förstås och analyseras



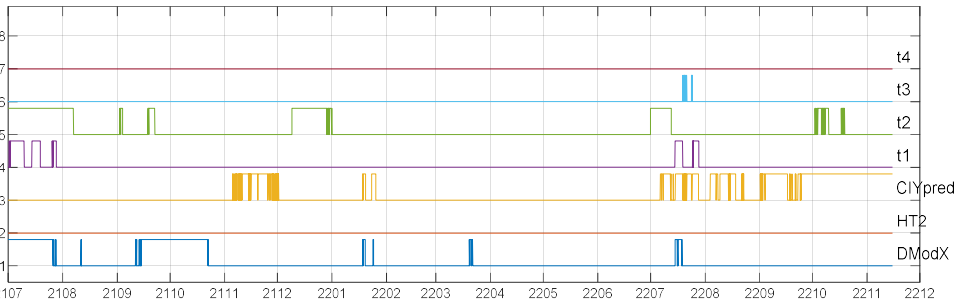
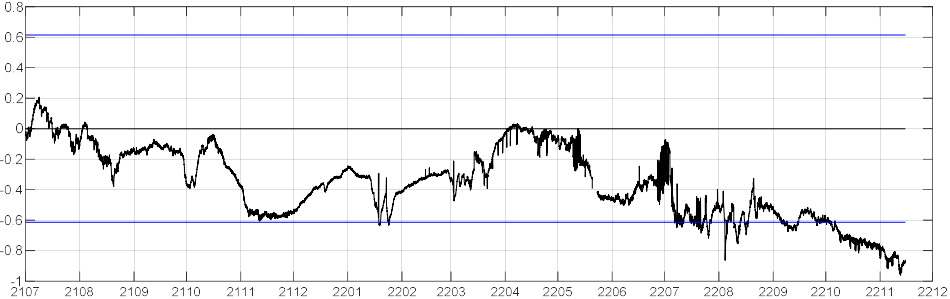
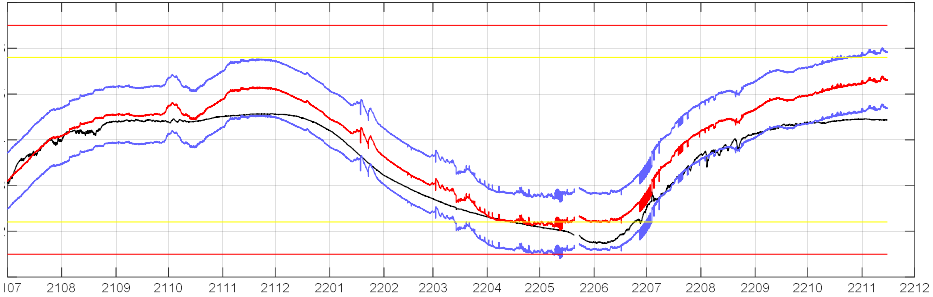
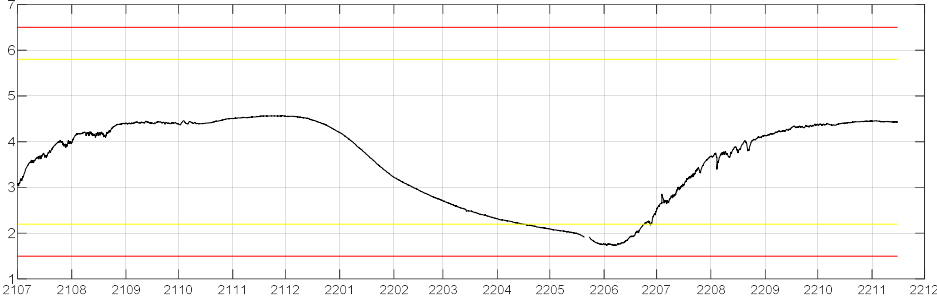
# Signalvalideringen (gränsvärdesbaserad)





# Modellövervakningen (en modell i taget)

17: V.GR010TH.Temp.valuePP  
2023-01-20 11:25:57



# Vår implementation

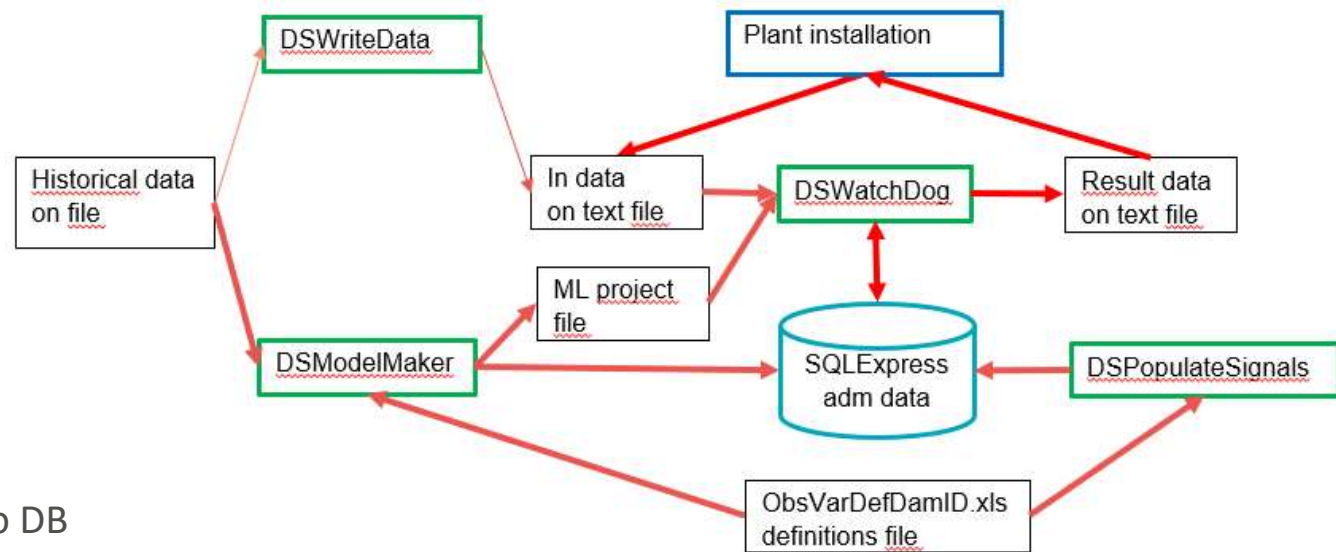
# Vårt tillvägagångssätt

- Utveckling i nära samarbete med användare
  - Domänkunskap ovärderlig
- Iterativ utveckling - erfarenhetsåterföring
- Val av dataanalysmetod
  - PCA och PLS är enkla att förstå och tolka
- Förbehandling / "datatvätt"
  - Validering, filtrering, statistisk karaktärisering, korsanalyser,...
- Utveckling på historiska data - provdrift på realtidsdata



# Vår implementation

- Virtuellt dator
  - Enkel drift och backup
- Datakommunikation
  - Textfiler in och ut med tidsstämpel
- En infofil i Excel per damm
  - Signallista med gränsvärden
  - Modellista med konfigurering
- Program i C#
  - DSPopulateSignals – data transfer to DB
  - DSModelMaker – modellgenerering
  - DSWriteData – offlineövervakning
  - DSWatchDog – realtidsövervakning



# Från modelldefinitioner till vakthund

- Definiera gränsvärden och modeller i ObsVarDef.xlsx
- DSPopulateSignals populerar gränsvärdestabellen Signals i databasen från ObsVarDef
- DSModelMaker använder preprocessade data och ObsVarDef, skapar projekt med alla modellerna samt uppdaterar tabellen Models i databasen
- DSWriteData skriver historiska data till fil eller OPC (för offline-tester)
- DSWatchDog exekverar nya data, först gränsvärdesbaserad validering per signal och sedan med alla aktiva modeller. Utdata skrivs till fil och/eller OPC beroende på inställningarna.

➡ *DSWatchdog är en generell och flyttbar mjukvara*

# ObsVarDef - signalgränsvärden

I ObsVarDefDnnn.xlsx definieras gränsvärdena för respektive signal

Varlx	TagName	EngUnit	Range Min	Range Max	PlotMin	PlotMax	Median Minus	Median Plus	MinLimit	MaxLimit	MinWarn	MaxWarn	MaxC onst	EWMA Lambd a	Moving Varianc eLimit	FirstDer Limit	Second DerLimi t	IVLMin Lim	Monit ored	Description
1	T01	möh	96	102	96	102			96.5	101.3	96.6	101.15	96		0.05	1.3	2	96.8	1	
2	T02	degC	1	7	1	7	12	12	1.5	6.5	2.2	5.8	96		0.05	0.8	1.5		1	Misstänkt trasig
3	T03	möh	102	107	102	107			102.75	106.5	103	106.2	96		0.05	2	3	103.3	1	
4	T04	degC	1	7	1	7			1.5	6.5	2.2	5.8	96	0.01	1	3.5	5		1	
5	T05	möh	105.5	110	105.5	110			105.75	109.75	106	109.5	96		10	0.04	0.07	106.4	1	
6	T06	degC	1	8.5	1	8.5			1.7	8	2.2	7.5	96	0.01	2	2	3		1	
7	T07	möh	109	115	109	115	6	6	110	114	110.25	113.8	96		0.05	0.05	0.2	110	1	
8	T08	degC	2	5	2	5			2.2	4.8	2.7	4.3	96		2	0.7	1		1	20200827 - Bekräftad trasig
9	T09	möh	106.5	112	106.5	112	6	6	106.75	111.5	107	111.3	96		0.5	0.04	0.07	106	1	
10	T10	degC	4	13	4	13			5	12.5	5.5	12	96	0.01	0.5	2	3		1	
11	T11	möh	106.5	111	106.5	111	6	6	107	110.75	107.1	110.5	96		0.5	0.04	0.2	106	1	
12	T12	degC	-2	10	-2	10			-1	9.5	-0.5	9	96	0.01	0.5	0.4	0.7		1	20200827 - Bekräftad trasig
13	T13	möh	107	111	107	111			107.25	110.85	107.4	110.7	96		0.5	0.02	0.04	107.2	1	
14	T14	degC	1.5	6.5	1.5	6.5			1.9	6.3	2.4	5.8	96	0.01	0.5	1	2		1	
15	T15	möh	70	72.5	70	72.5			70.25	72.35	70.5	72.2	96		0.5	0.7	1	70.7	1	
16	T16	degC	-0.5	5.5	-0.5	5.5			0	5	1	4.5	96	0.01	0.5	2	3		1	
17	T17	möh	38	41	38	41			38.25	40.6	38.5	40.4	96		0.5	1.3	2	38	1	

# ObsVarDef - modelldefinitioner

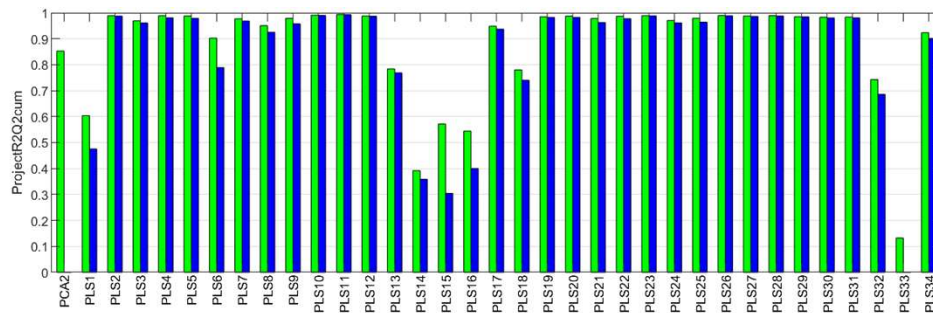
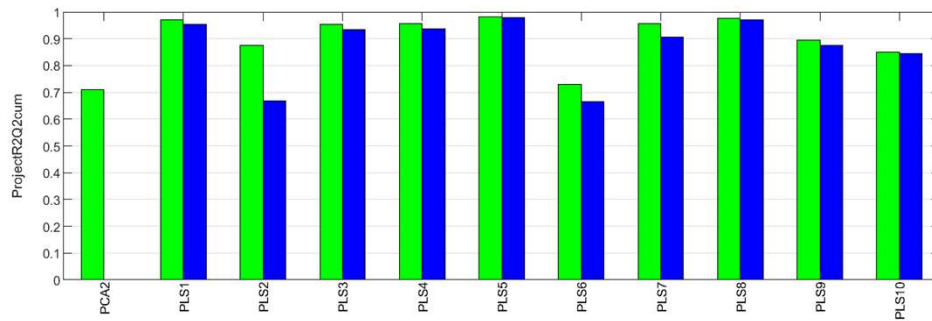
I ObsVarDefDnnn.xlsx definieras vilka variabler som ingår som X och Y i respektive modell

Varlx	TagName	Mdl 01	Mdl 02	Mdl 03	Mdl 04	Mdl 05	Mdl 06	Mdl 07	Mdl 08	Mdl 09	Mdl 10
1	T01	X				X	X		X	X	
2	T02	X		X	X			X	X	X	
3	T03	X					X		X	X	
4	T04	X	Y	X	X			X	X	X	Y
5	T05	X				X	X			X	
6	T06	X	X		X	X		X		X	
7	T07	X				X			X	X	
8	T08										
9	T09	X				X			X	X	
10	T10	X	X			X		X	X	X	
11	T11	X				X			X	X	
12	T12										
13	T13	X				X			X	X	
14	T14	X	X	X	X	X			X	X	
15	T15	X								X	
16	T16	X	X	X		X		X	X	X	
17	T17	X							X	X	
18	T18	X		X		X		X		X	

# DSModelMaker - modellbyggaren

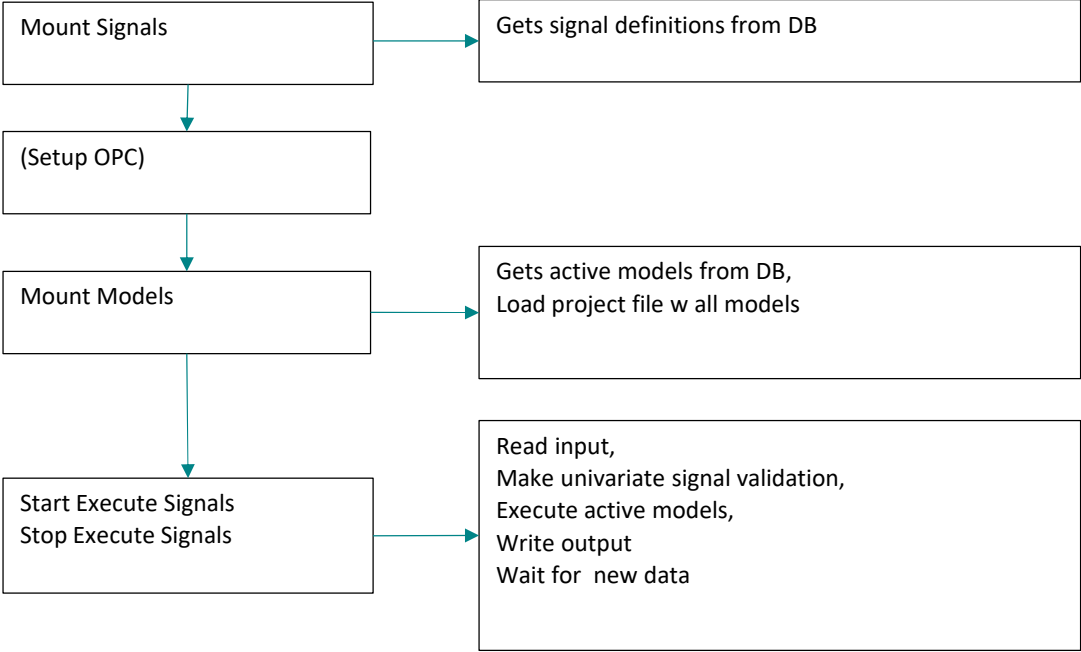
Varje PCA-modell redovisas med R2 efter 3 komponenter

Varje PLS-modell med R2/Q2 efter antal korsvaliderade komponenter





# Realtidsövervakningen – vakthunden / DSWatchDog



# Indatafil

- Första raden är DateTime
- Sedan följer rader med TagNamn; Värde

```
1 IN_211217_1000.csv *
2021-12-17 10:00:00
S01.T01;49.2481
S01.T02;4.741591
S02.T03;54.8999
S02.T04;4.421692
S03.T05;57.3397
S03.T06;4.879759
S04.T07;61.5193
S04.T08;-6.575829
S05.T09;59.7352
S05.T10;8.334814
S06.T11;58.5661
S06.T12;2.61655
S07.T13;59.8361
S07.T14;4.99017
S08.T15;21.3984
```

# Utdatafil

- Första raden är DateTime, antal överskridna signalGränser och modellGränser
- Sedan följer rader med signalNamn; statusVärde
- Därefter, för varje model: {statusVärde, modellVärden inkl predikterat signalVärde}

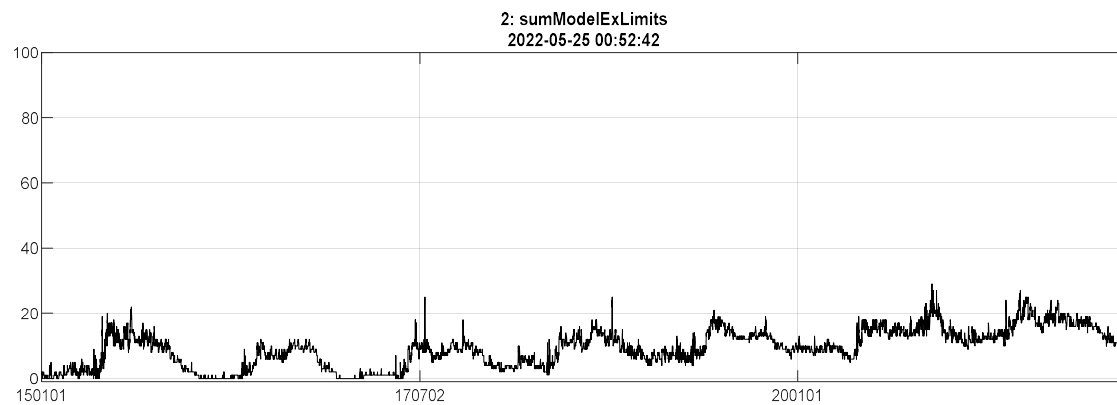
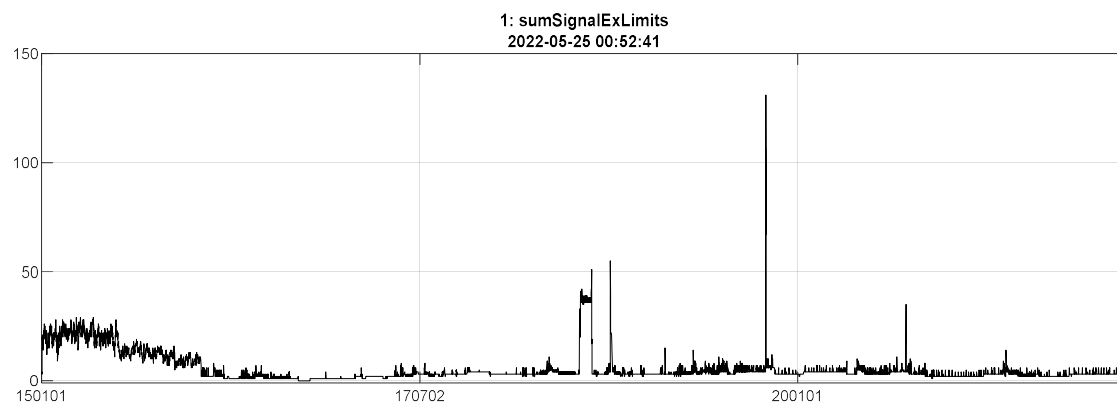
1 Dnnn\_201208\_134500.txt

```
2020-12-08 13:45:00;sumSignalExLimits;4;sumModelExLimits;48  
S01.T01.status;0  
S01.T02.status;0  
S02.T03.status;0  
S02.T04.status;0  
S03.T05.status;0  
S03.T06.status;0  
S04.T07.status;0  
S04.T08.status;2  
S05.T09.status;0  
S05.T10.status;0  
S06.T11.status;0
```

```
M010.status;97  
M010.DModXpred;2.342  
M010.t1pred;-0.689  
M010.t2pred;-0.575  
M010.HT2pred;0.252  
M010.T04.pred;3.207  
M010.T04.CuSumH;0.000  
M010.T04.CuSumL;0.000  
M011.status;32  
M011.DModXpred;1.324  
M011.t1pred;-0.345  
M011.HT2pred;0.031  
M011.T70.pred;12.657
```

# Utdata

- Sammanfattning: antal överskridna händelser för signaler och modeller



# Underhåll av övervakningsmodeller

# Underhåll av övervakningsmodeller

- Vi har sett att övervakningsmodellerna behöver revideras periodvis, initialt årsvis allt eftersom nya data tillkommer.  
Möjligen kan behovet minska ju mer träningsdata som ingått vid modelleringen.
- För att klara detta underhåll behövs
  - Kunskap på om hur de datadrivna metoderna fungera.
  - Kunskap om aktuella dammar och mätsystemet för övervakning.
  - Budgeterade ekonomiska resurser
- Vi tror att de bästa resultaten erhålls om dammsäkerhetsingenjörer och de som arbetar med mätteknik för dammsäkerhet utbildas i datadrivna metoder.

# Kurs om datadriven dammövervakning

# Kurs om datadriven dammövervakning

Datum	Moment
23-03-01	Material för inläsning tillsammans med mindre datorövning i Excel.
23-03-09	Fysisk kursdag 1 Introduktion till verktyg för att genomföra datorövningar, följt av datorövning. Teori och användandet av PCA och PLS.
23-03-10 - 23-03-27	En datorövning att genomföra på egen hand med PLS, på data från en damm.
23-03-20	Kort avstämning via Teams i grupp kring datorövningen.
23-03-27 – 23-03-29	Avrapportering datorövningsarbetet. Granskning av andras rapporter.
23-03-30	Fysisk dag 2 Genomgång av metoder för att detektera avvikelser med hjälp av PLS modeller. Diskussion om datorövningen och de erfarenheter man fått, en kort datorövning om avvikelседetektion samt summering av kursen och behov kursdeltagarna identifierat.
23-04-18	Kort uppföljningsmöte via Teams



Slutsatser

# Nyttor med datadrivna metoder

- Minskar personberoende övervakning
- Kan ge snabböversikt av en eller flera dammar
  - Men ändå mindre risk att man bortser från små händelser
- Förädlad information genom ML/AI-metoderna
- Möjlighet till mer intelligenta varningar



# Utvecklingen inom dammövervakning har bara börjat

- Intåget av AI i dammövervakningsvärlden har börjat
- Datakvalitet är viktigt: skräp in -> skräp ut! ("GIGO")
- Utökat datorstöd ger nya möjligheter
- Nya visualiseringsmetoder förenklar arbetet
- Många lärdomar att bygga vidare på från projekten
  - Iterativ utveckling, tumregler och metodik
- Fortbildning är ett måste
- Vi har bara börjat dataresan!

Diskussion och frågor