



Multivariat teknik och processoptimering

Linda Åmand

IVL Svenska Miljöinstitutet

Gruppen för Miljöteknik och modellering



Agenda

- Presentation IVL
- Introduktion multivariat teknik
- Tillämpning av MVA för processoptimering
- Frågor



IVL Svenska Miljöinstitutet AB

- Forskning och konsultverksamhet inom klimat, energi och hållbar tillväxt
- IVL är fristående
- Grundat 1966
- Ca 230 medarbetare
- Avancerade laboratorier
- Kontor i Stockholm, Göteborg, Malmö, Lysekil och Beijing



Temaområden

- Klimat och energi
- Luft och transporter
- Vatten
- Hållbart samhällsbyggande
- **Hållbar produktion**
- Resurseffektiva produkter och avfall

Miljöteknik och modellering: våra verktyg

Statistical
Design of
Experiments

Multivariate
Data Analysis

Automatic
control

Environmental
and Economical
Performance
Indicators

Simulators

Communication
Systems

Multivariat analys

Principal component analysis (PCA)

Partial Least Squares/Projection to Latent Structures (PLS)

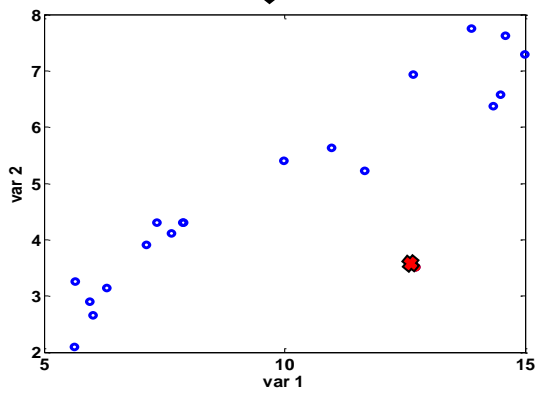
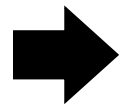
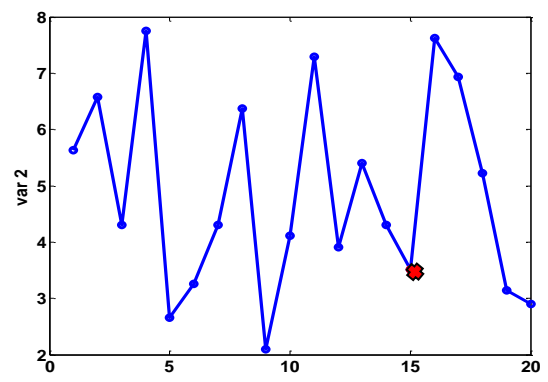
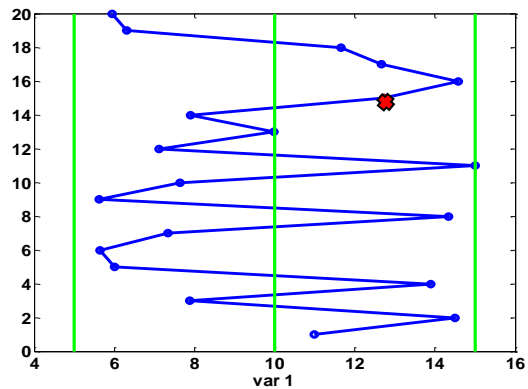
Vad är multivariat analys?

- Verktyg för analys av stora datamängder
 - Grunder i linjär algebra
 - Bredd av metoder
 - Empirisk metod
-
- Inte bara ett hjälpmedel för ingenjörer!

Varför multivariat analys?

- Världen är multivariat
 - En egenskap beror av flera faktorer
- Indirekta observationer
 - Mäta tillgängliga och billiga variabler
 - Förutsäga egenskapen ur mätningar

Samvariation



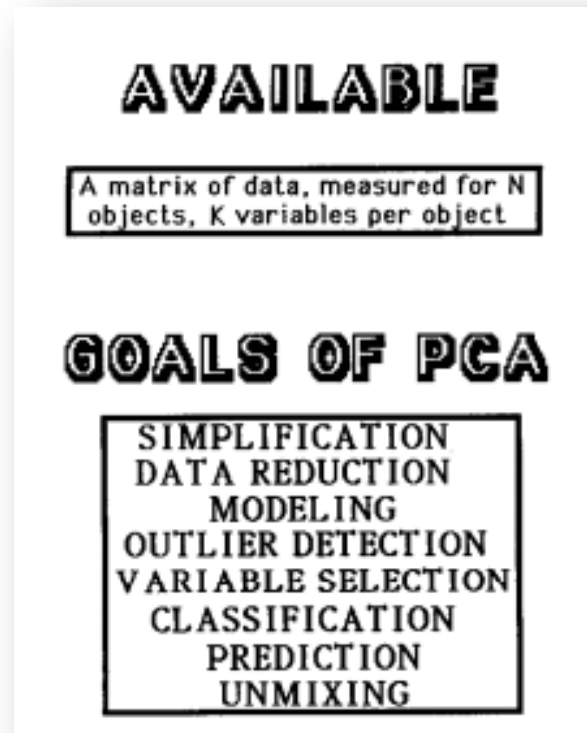
Dolda strukturer i data

- Multivariat data uppvisar ofta stor kovarians
- Det finns ett visst antal fritt varierande faktorer som påverkar data
- Fler mätningar betyder inte alltid mer information

Principalkomponentanalys (PCA)

"One of the most valuable results from applied linear algebra"

PCA enligt S. Wold et al. 1987

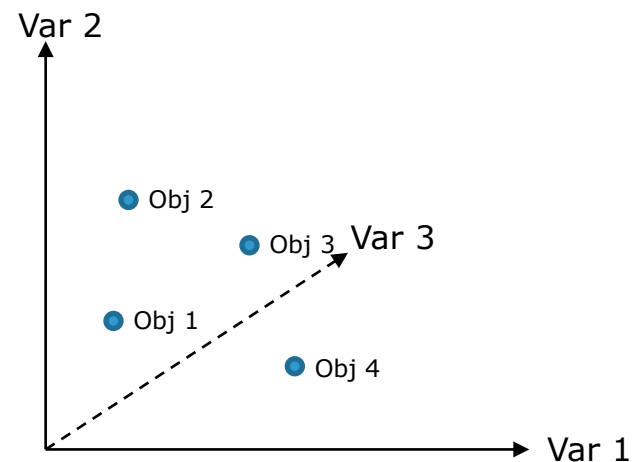


S. Wold et al. 1987

Lite matematik

- Varje objekt kan ses som en vektor i ett vektorrum
- Vektorrummets antal dimensioner bestäms av hur många variabler du har

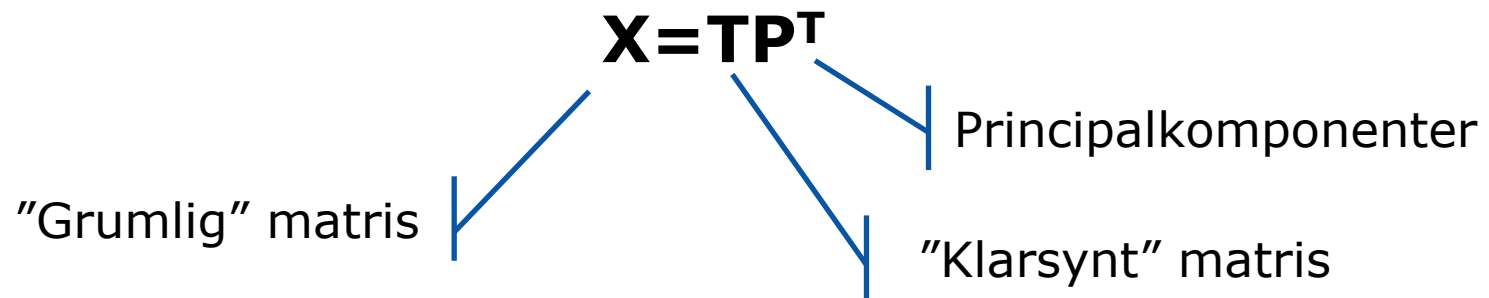
	Var 1	Var 2	Var 3
Obj 1			
Obj 2			
Obj 3			
Obj 4			



PCA ställer frågan:

- Finns det en annan bas, som är en linjär kombination av den gamla basen, som bättre beskriver datasetet?

Grundprincipen för PCA blir därmed ett basbyte:



Vilken är den bästa basen?

- Om mätningarna är bra **antar** vi att riktningen med **störst varians** i vårt vektorrum är **viktigast**
- **Låg redundans** motsvarar **låg kovarians** mellan två variabler
- Maximera variansen, och minimera kovariansen mellan variablerna i din nya matris!

Vilken är den bästa basen?

- Anta att X är matris med variabler och objekt
- Kovariansmatrisen: $\Sigma = E[(X - \mu)(X - \mu)^T]$

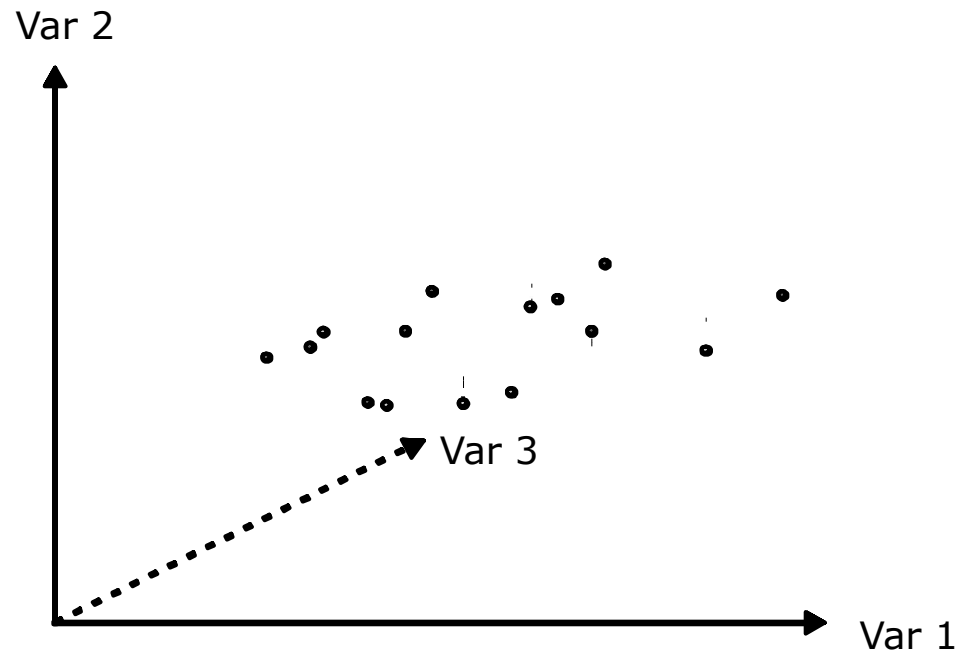
$$\Sigma = \begin{bmatrix} \text{var}(x_1) & \text{cov}(x_1, x_2) & \dots & \text{cov}(x_1, x_m) \\ \text{cov}(x_2, x_1) & \text{var}(x_2) & \dots & \text{cov}(x_2, x_m) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(x_m, x_1) & \text{cov}(x_m, x_2) & \dots & \text{var}(x_m) \end{bmatrix}$$

Vilken är den bästa basen?

- Alltså: Hitta en ortonormal matris P för $T=XP$ sådan att T :s kovariansmatris blir diagonal
- Det visar sig att detta inträffar då P :s kolumner är egenvektorer till kovariansmatrisen till X .

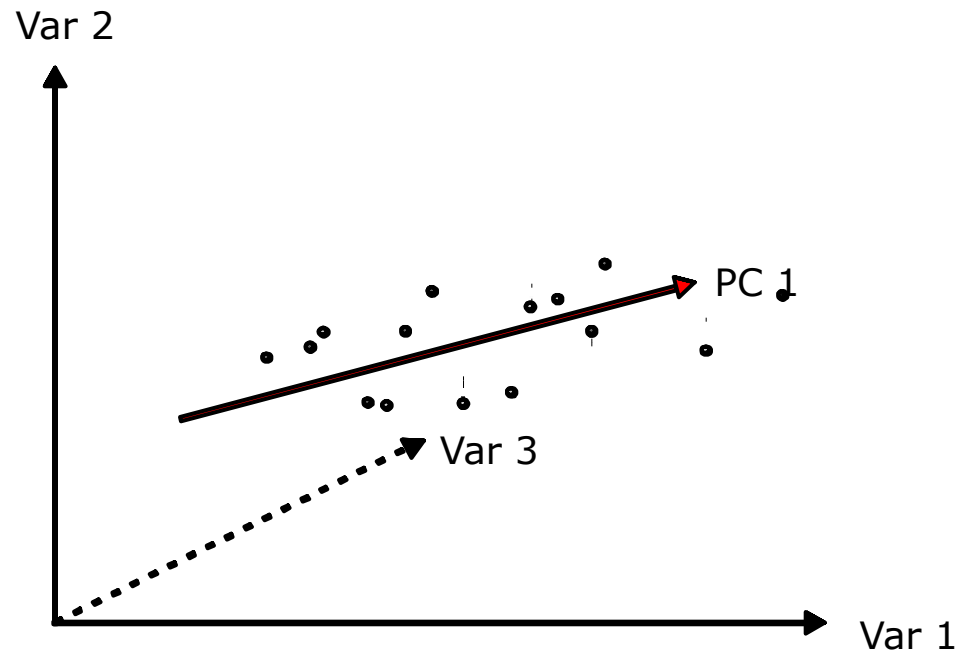
- Två metoder:
 - Kovariansmetoden
 - Singulärvärdesuppdelning (SVU)

Ursprunglig bas



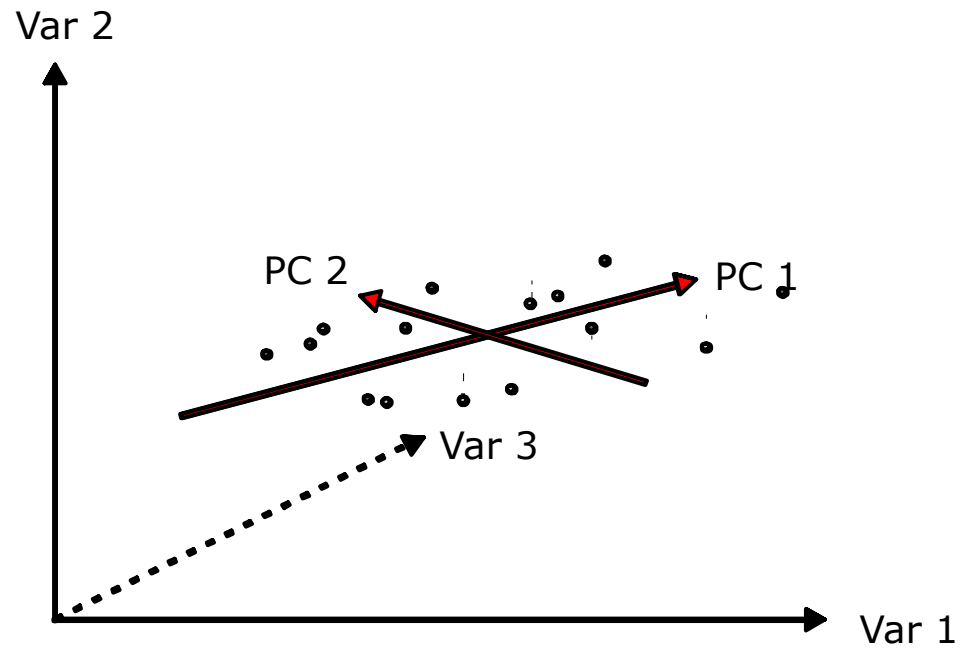
Steg 1

- PC 1 i riktning av största variansen



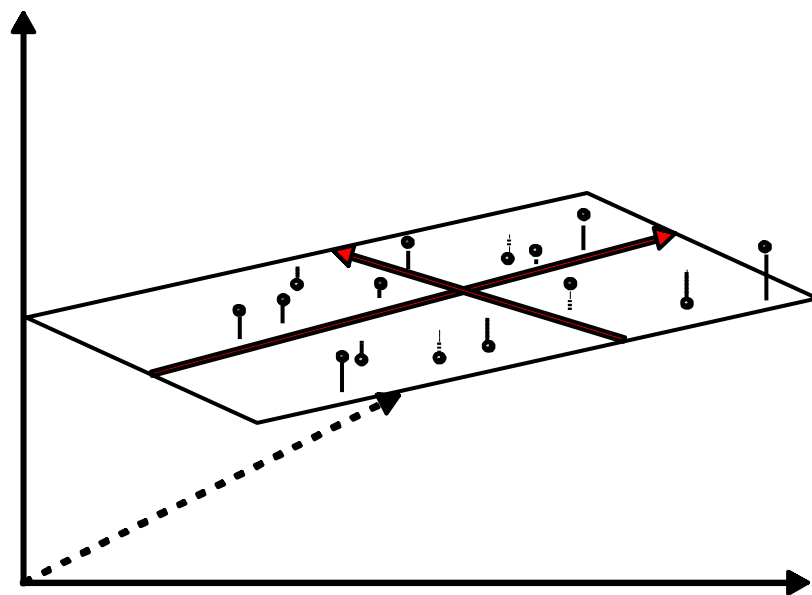
Steg 2

- PC 2 beskriver resterande variansen så bra som möjligt, och är vinkelrät mot PC 1



Steg 3

- Observationerna projiceras på nya basen



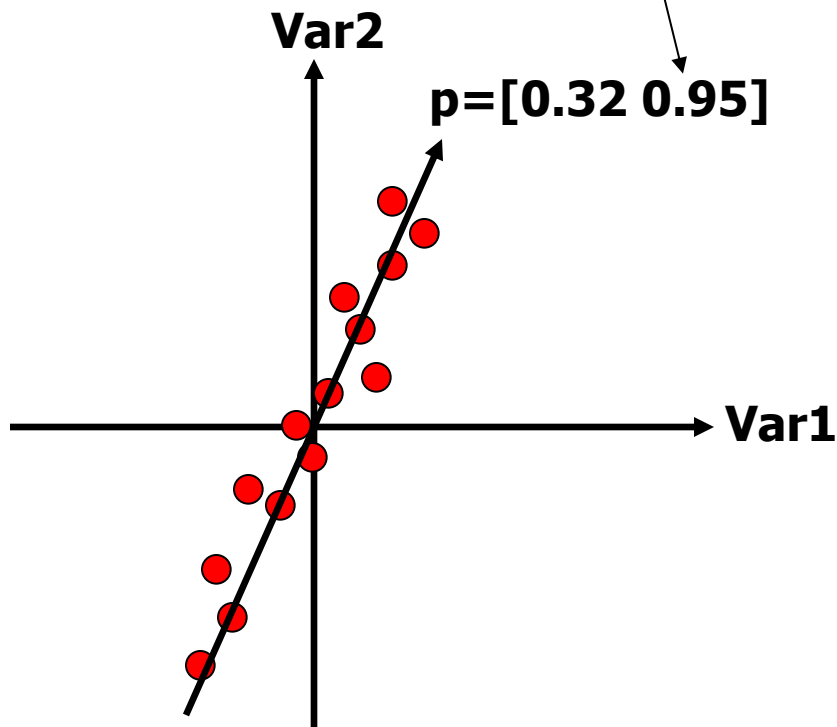
Det finurliga är att...

- ...vi tittar på färre dimensioner än innan
- ...de första principalkomponenterna innehåller den mest intressanta informationen
- ...vi får en grafisk tolkning med loadings och scores

Loadings (p)

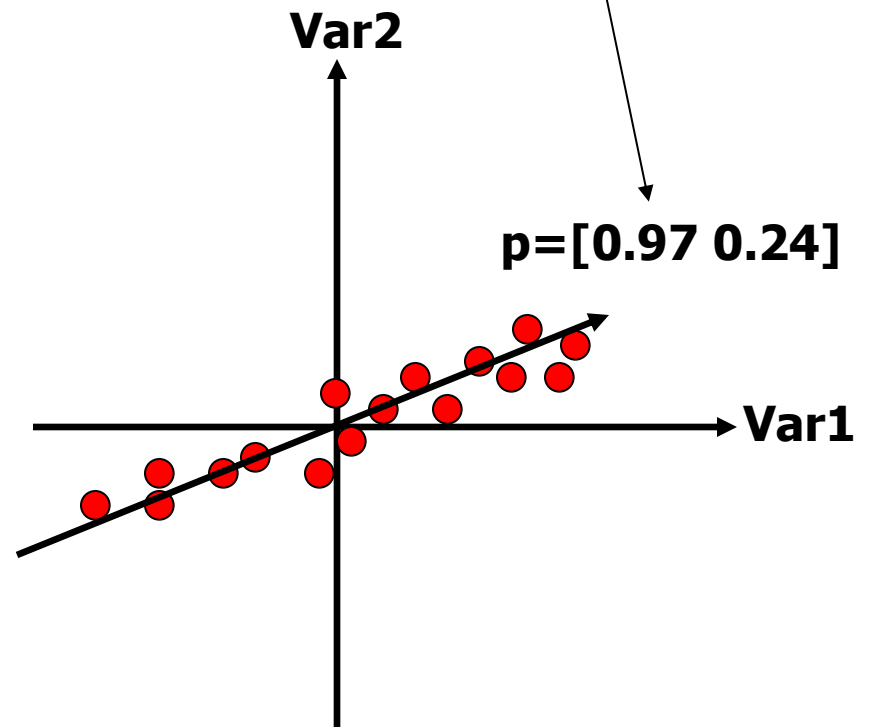
Var2 varierar mest

Loading för var2 är högre



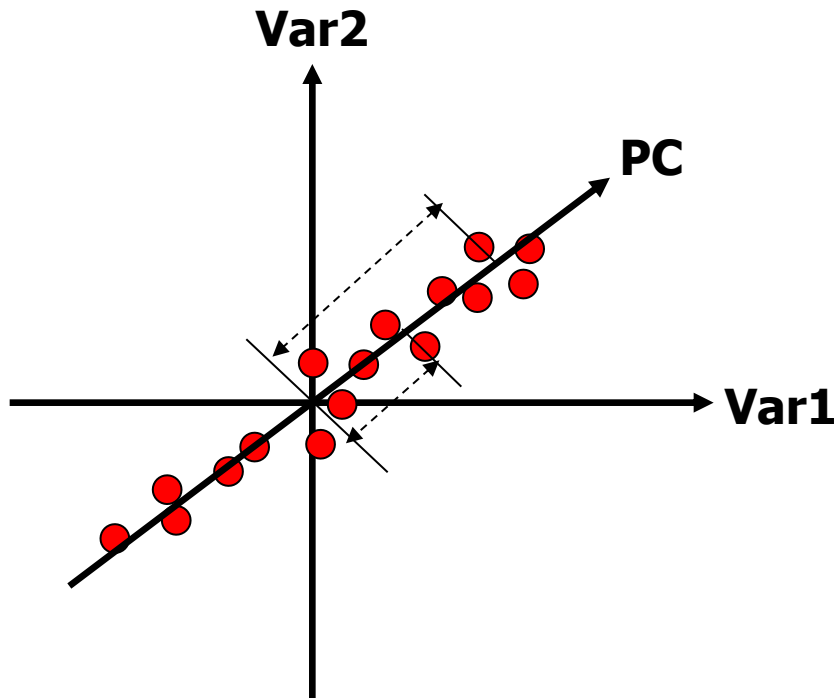
Var1 varierar mest

Loading för var1 är högre



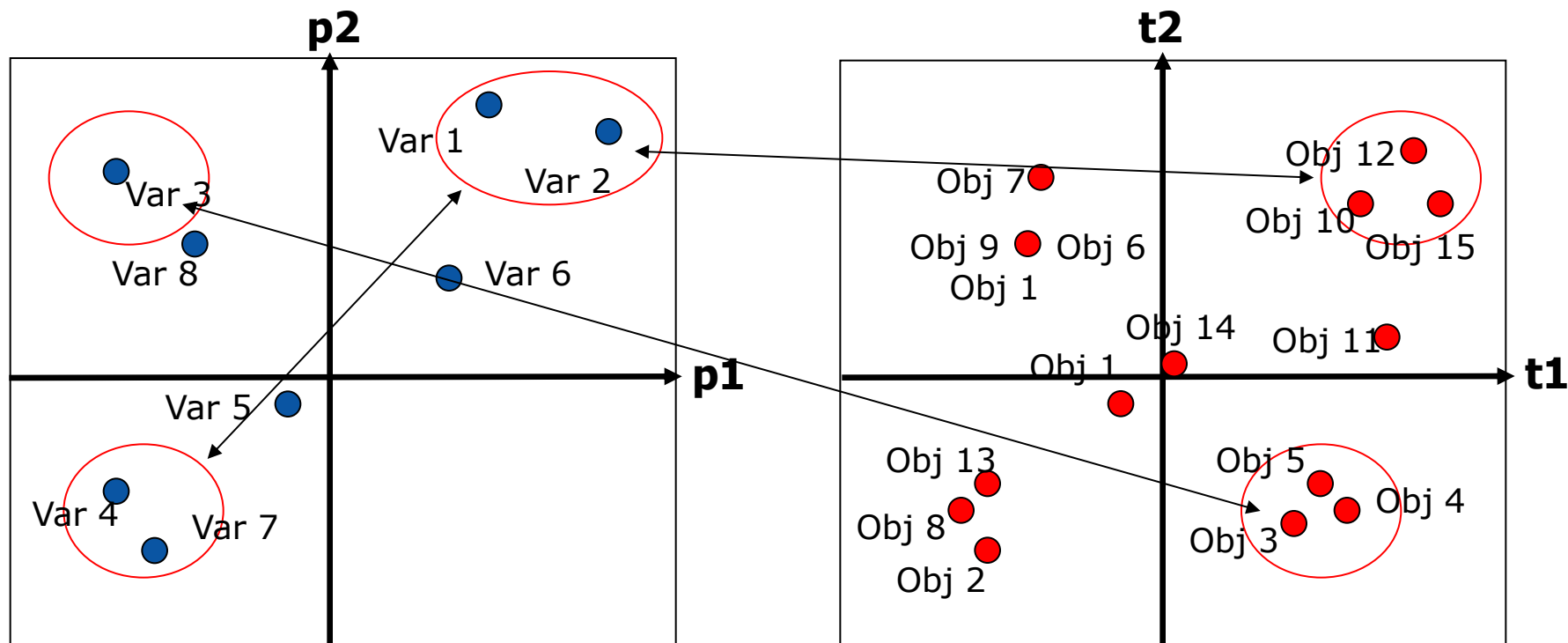
Scores (t)

- Beskriver objekten
- Koordinater i det nya koordinatsystemet



Projicera data på våra nya axlar. Längden av projektionerna kallas **scores**.

Tolkning av PCA



PCA i matriser

- X kan beskrivas som

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$$

The diagram illustrates the matrix equation $\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$. The term \mathbf{X} is the data matrix. \mathbf{T} is labeled as 'Scores'. \mathbf{P}^T is labeled as 'Loadings'. \mathbf{E} is labeled as 'Residualen'. Lines connect the labels to their respective matrices in the equation.

Att arbeta med PCA

- Använd bra kalibreringsdata
- Förbehandla alltid data
- Outliers måste bort!

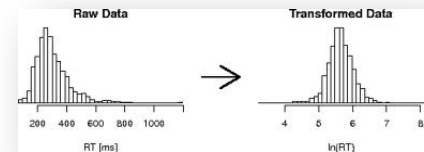
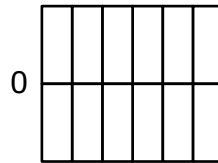
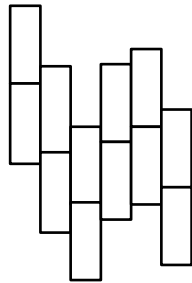
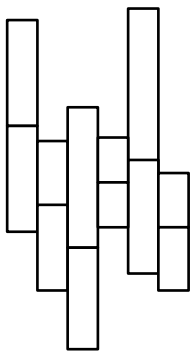
Krav på prover kalibreringsdata

- Tillräckligt många
- "Tillräckligt olika"
- Representativa

Förbehandling av data

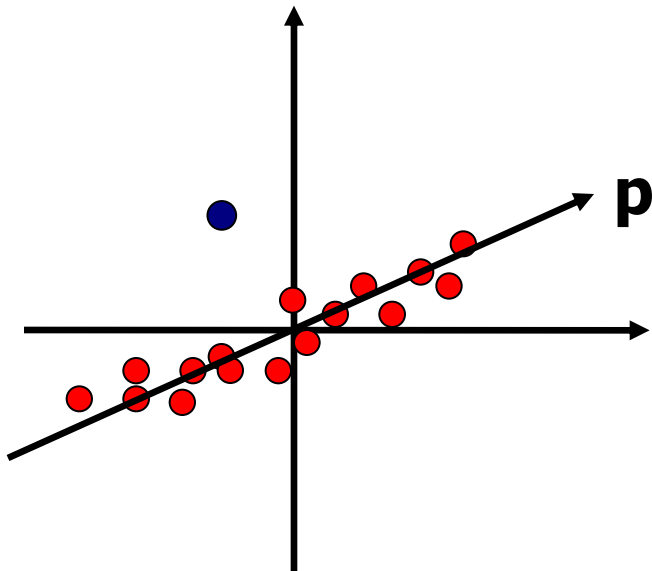
- Auto-skalning
 - Dividera variabler med dess standardavvikelse
 - Ger alla variabler samma inverkan på modellen
- Centrerings
 - Subtrahera medelvärde av varje variabel
 - Behåller skillnader mellan objekt
- Viktning
 - Vikta ner oinformativa variabler
- Transformerings
 - Exempelvis $A = \ln(T)$
 - Bygger oftast på förkunskaper

Förbehandling av data

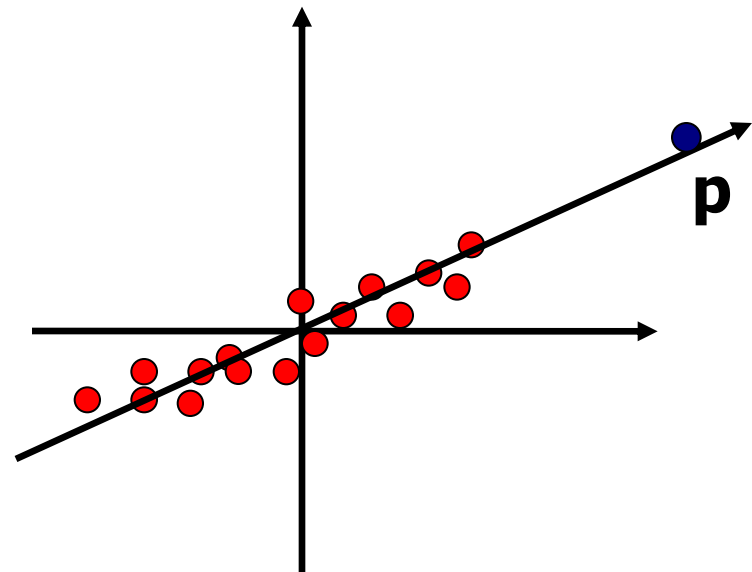


Outliers i PCA, två exempel

Hög residual



Högt leverage



Sammanfattning PCA

- Gör ett basbyte för att maximera variansen och minimera kovariansen i dina nya basvektorer (variabler)
- Metod för tolkning och klassificering av data
 - Antalet dimensioner som behövs för att beskriva data reduceras
 - Data blir mer överskådlig och tolkningsbar
 - Grafisk tolkning

Att tänka på om PCA

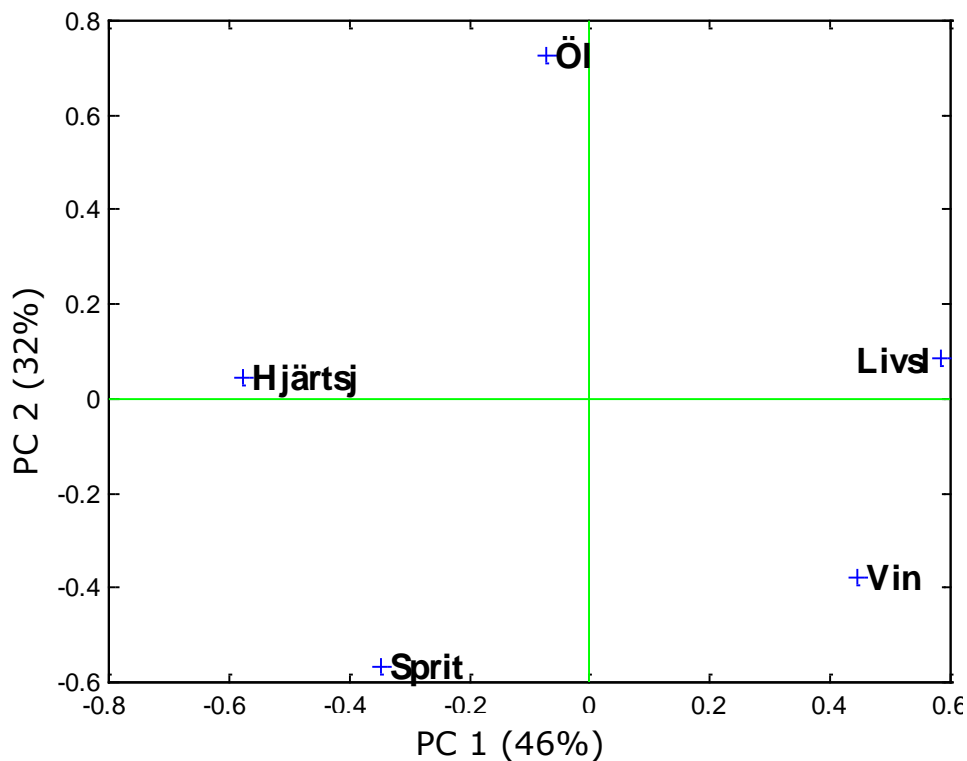
- PCA antar linjäritet i data
- PCA antar att stor varians beskriver viktig dynamik
- Medelvärde och varians är tillräckliga för att beskriva data

Exempel: Data på dryckesvanor och hälsa

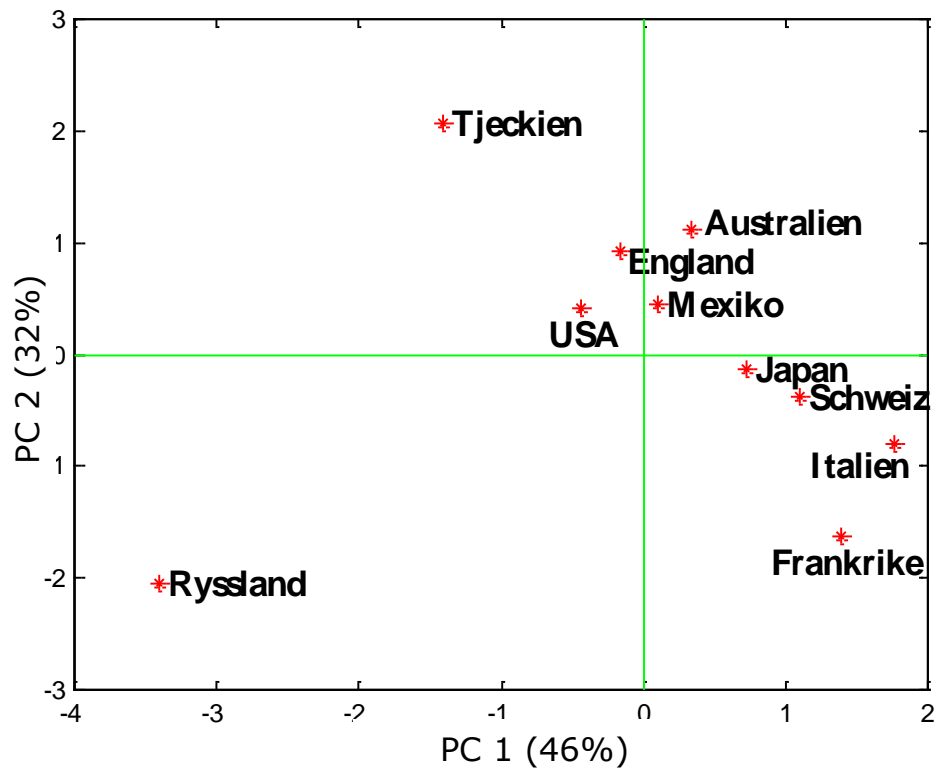
	Vin (liter/år)	Öl (liter/år)	Sprit (liter/år)	Förväntad livslängd (år)	Hjärtsjukdom (antal/100 000, år)
Frankrike	63.5	40.1	2.5	78	61.1
Italien	58	25.1	0.9	78	94.1
Schweiz	46	65	1.7	78	106.4
Australien	15.7	102.1	1.2	78	173
England	12.2	100	1.5	77	199.7
USA	8.9	87.8	2	76	176
Ryssland	2.7	17.1	3.8	69	373.6
Tjeckien	1.7	140	1	73	283.7
Japan	1	55	2.1	79	34.7
Mexiko	0.2	50.4	0.8	73	36.4

Grafisk tolkning från scores och loadings

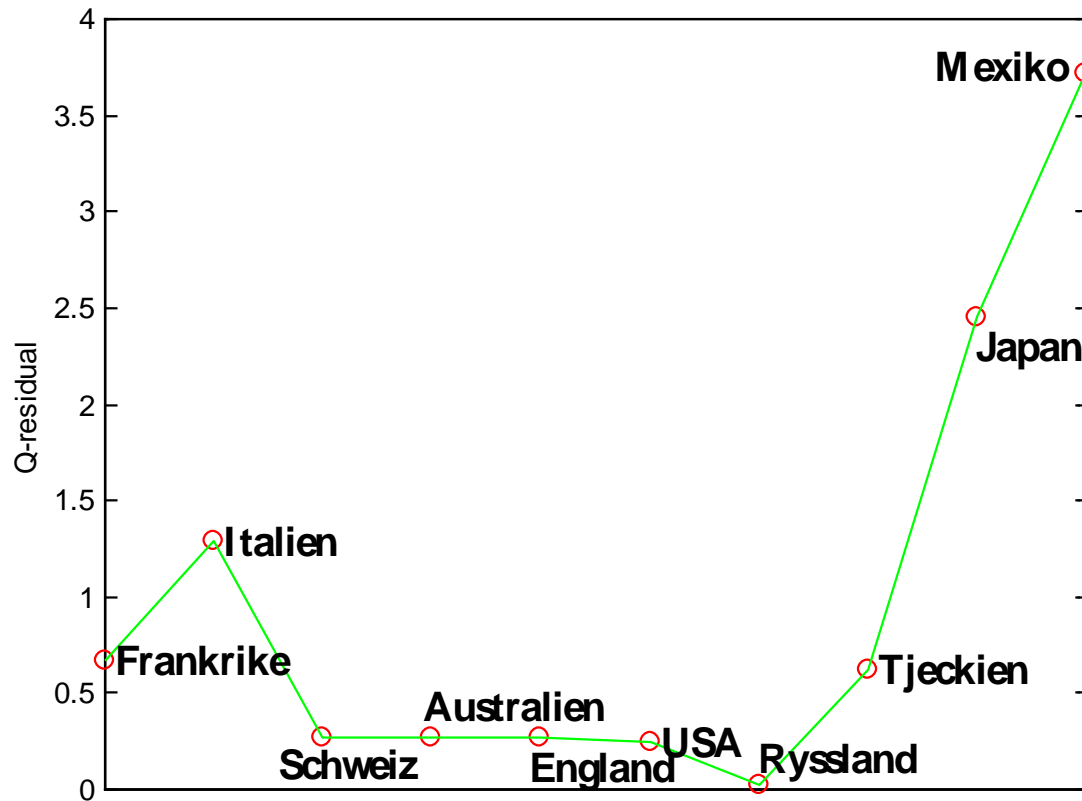
Loadings - relation mellan variabler



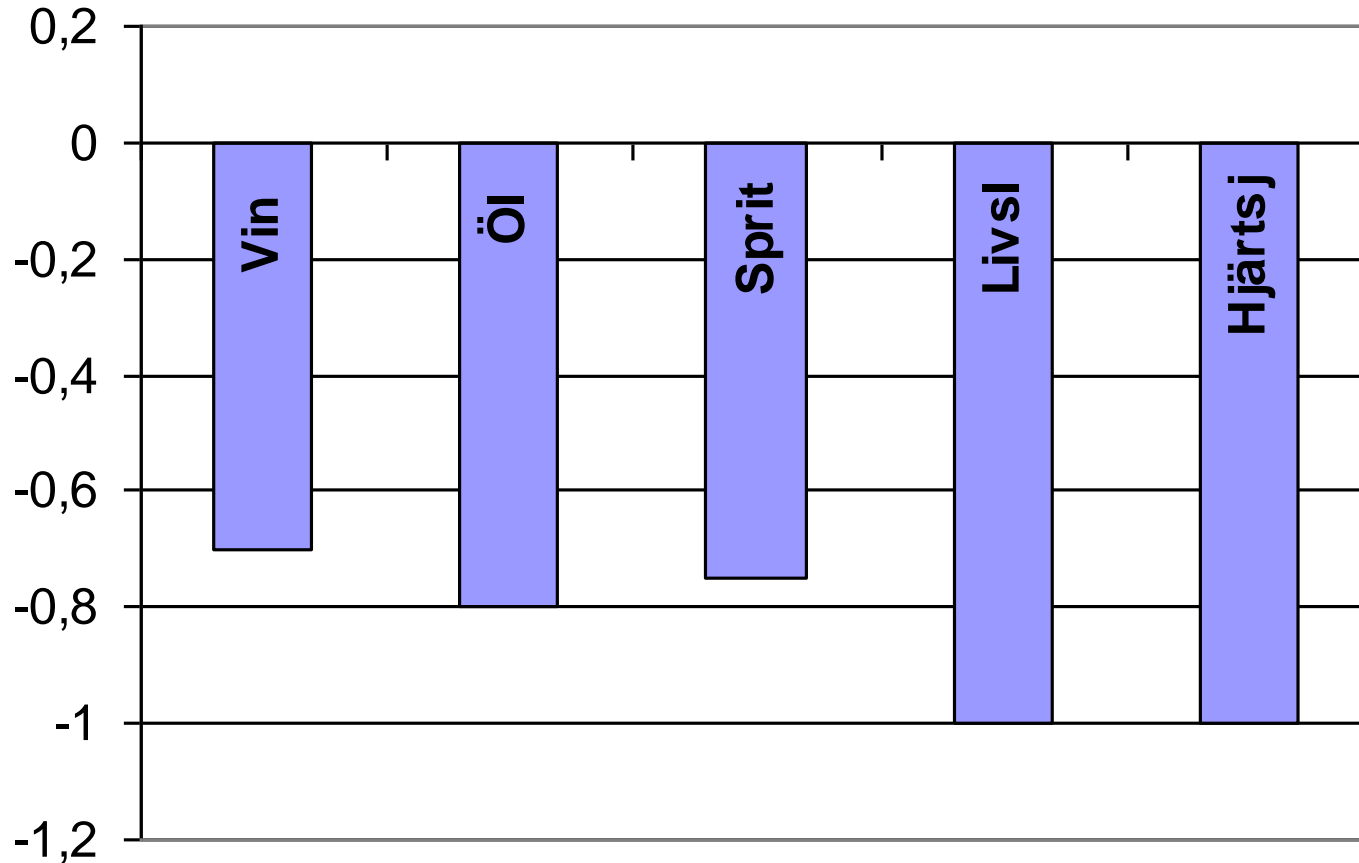
Scores - relation mellan objekt



Residual - Avstånd till modellen



Avvikelser för Mexiko

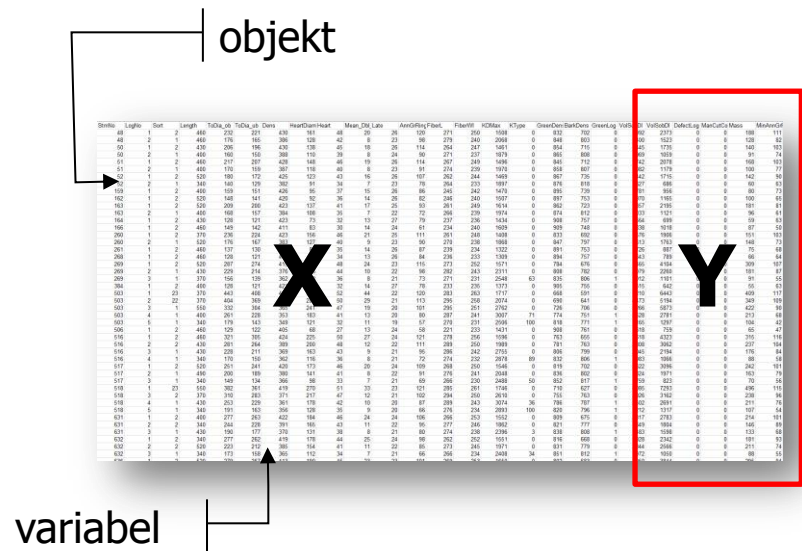


Multivariat kalibrering

- Kan ge möjlighet att ersätta dyra/långsamma mätningar med billigare/snabbare
- Utnyttja all tillgänglig information
 - Moderna instrument genererar mycket data
- Möjliggör detektion av "outliers", modelldiagnostik

Organisation av data

- Som i PCA men med utsignaler/målvariabler (Y)



Olika regressionsmetoder

- MLR (Multiple Linear Regression)
- PCR (Principal Component Regression)
- PLS (Partial Least Squares)



Olika regressionsmetoder

- MLR (Multiple Linear Regression)
- PCR (Principal Component Regression)
- PLS (Partial Least Squares)

**Klarar inte av
kovarians i X**

**Löser problem
med kovarians i X**

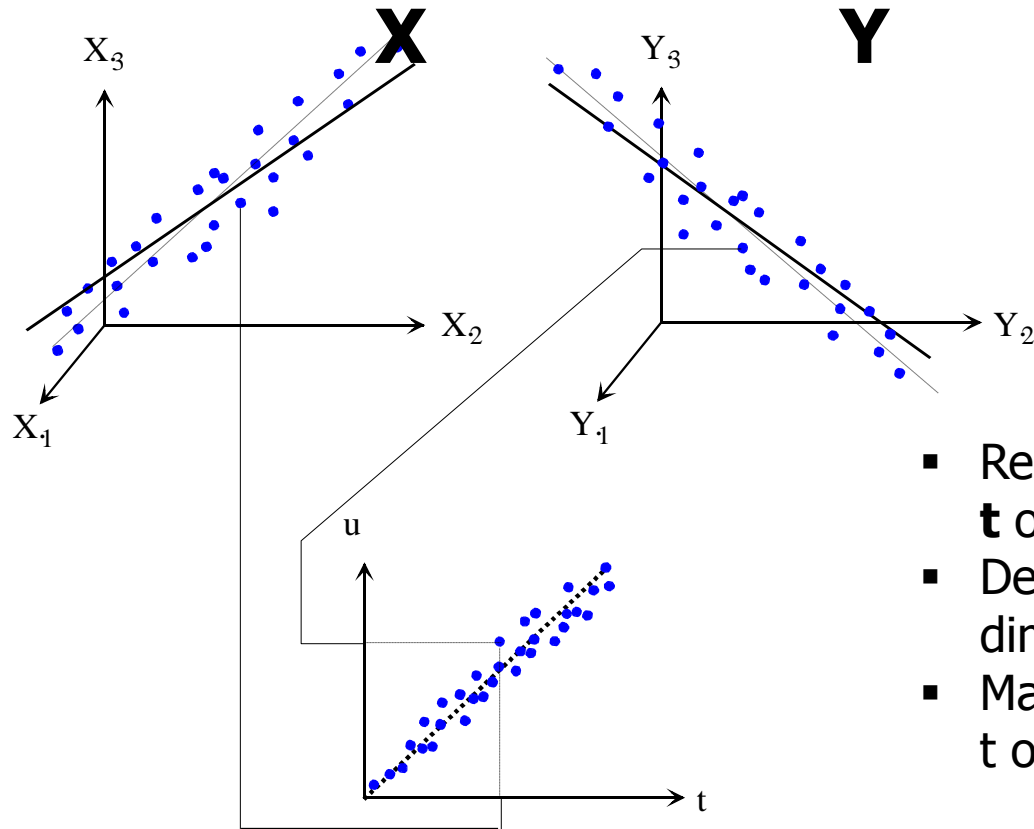
**Tar hänsyn till både
X och Y från början**

Inslag av
PCA

PLS en svensk idé

- Introducerat av svenska statistikern Herman Wold som fortsatt utveckla metoden med sin son, Svante Wold
- Svante Wold föredrar: *Projection to latent structures (PLS)*

PLS - grafiskt



- Regression sker mellan **t** och **u**
- Detta sker för varje ny dimension
- Maximera kovariansen mellan **t** och **u**

PLS liknar PCA

- Kräver bra kalibreringsdata
 - Förbehandling av data
 - Borttagning av outliers
 - Loading och scores
-
- MEN: Modellvalidering viktig!

Modellvalidering

- Validering är viktigt!
 - Hur många PC/PLS-komponenter är bäst?
 - Fungerar modellen verkligen?
 - Hur bra blir prediktioner från modellen?

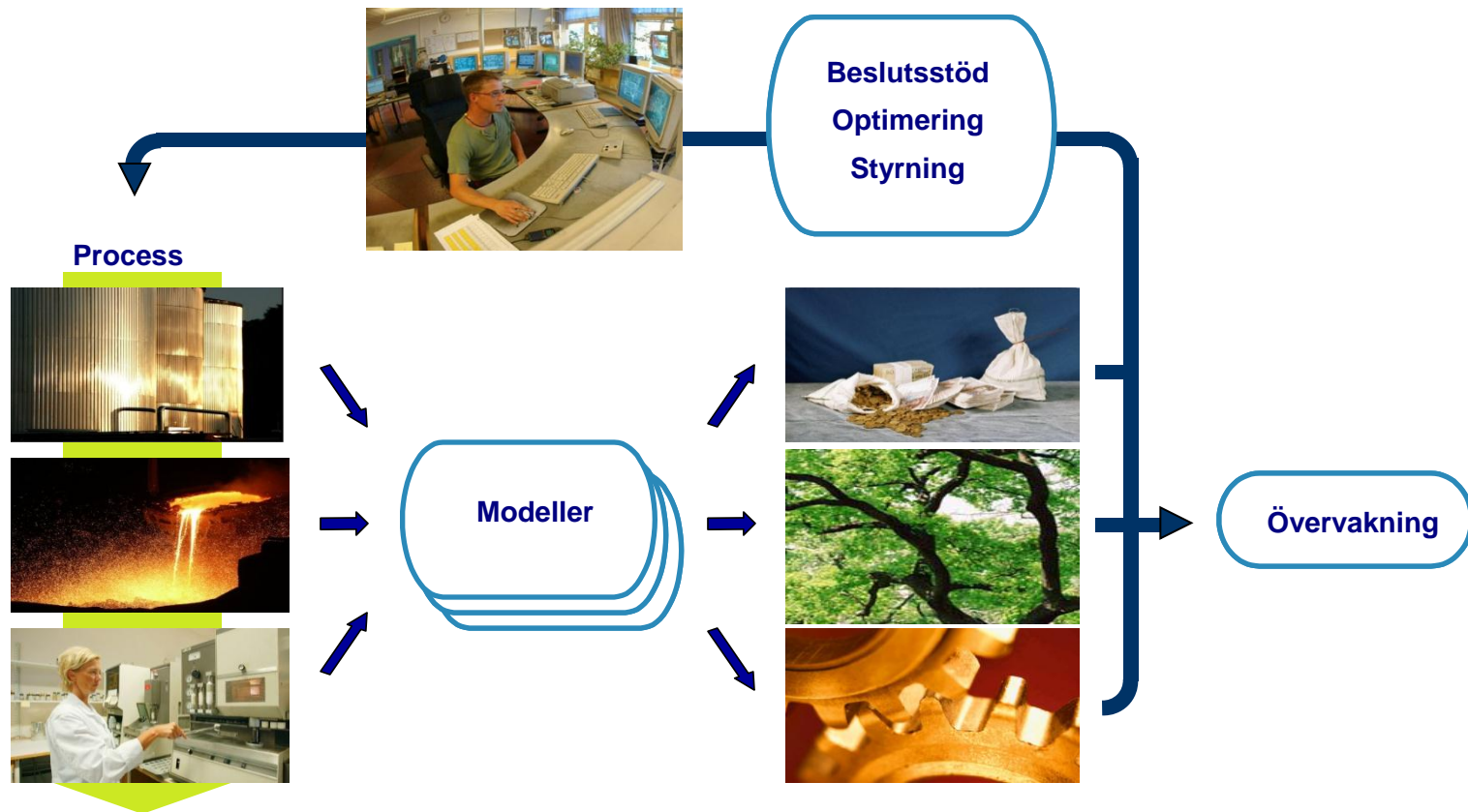
- Tre vanliga metoder
 - Intern validering
 - Korsvalidering
 - Testsets-validering

Modellvalidering

- Undvik över- och underanpassningar
 - Underanpassning: Modellen täcker inte in alla intressanta fenomen
 - Överanpassning: Onödigt mycket brus modelleras

Tillämpning av MVA för processoptimering

Vår princip för processoptimering



MVA för processindustrin

- PCA-modellering
 - lära sig mer om processen
 - optimering
 - övervakning
- PLS-modellering
 - virtuella sensorer (mjukvarusensorer)
 - lära sig mer om processen

Extrahera mer information ur redan tillgängliga data!

Exempel från industrin

- Industriell avloppsvattenrening
 - ökad processförståelse
- Stålintusti
 - Klassificering
- Petrokemisk process
 - virtuell sensor
 - övervakning i realtid
- Vattenverk
 - PLS-modell för automatisk dosering

Exempel I: Industriell avloppsvattenrening

- Biologisk rening
- Data: 3 månader, 1 h upplösning

Inkommande vatten

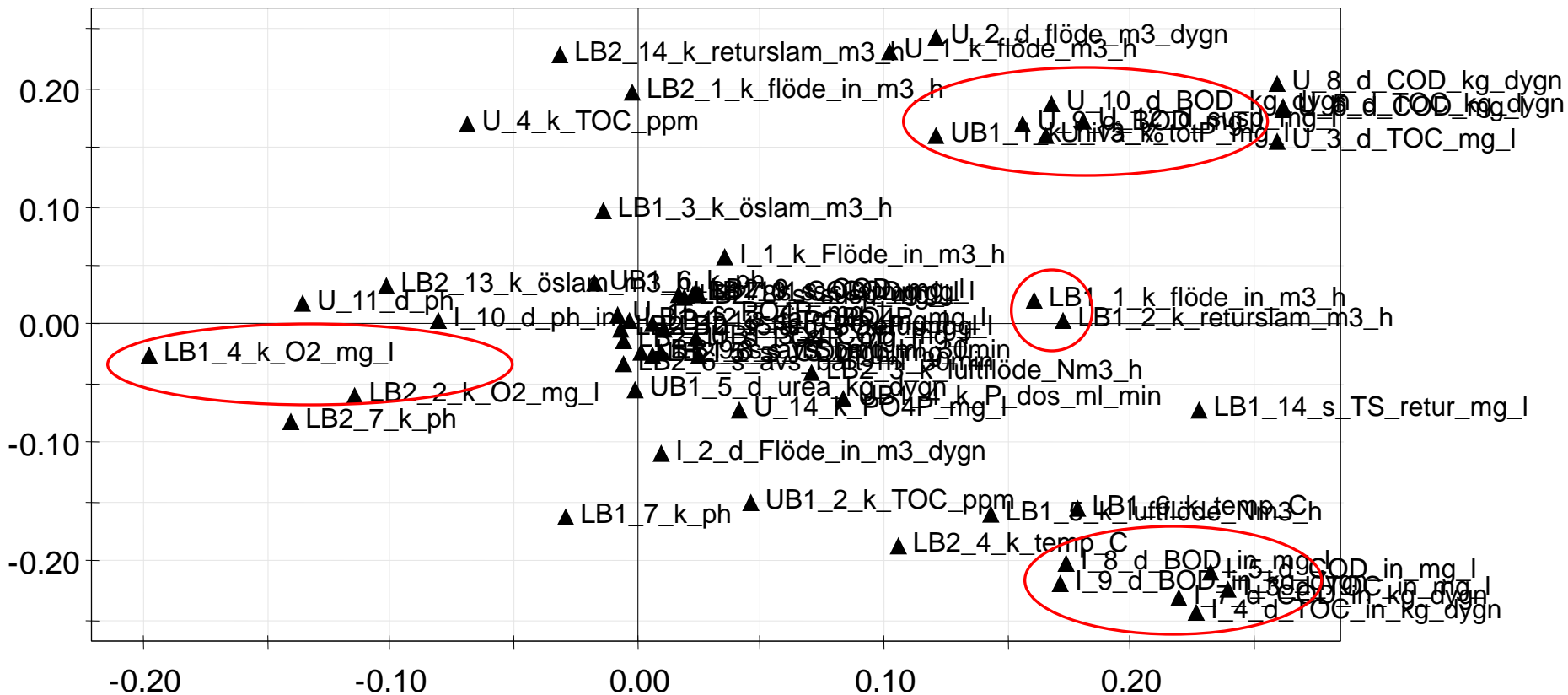
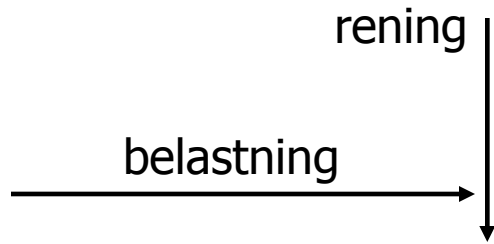
Utvämning

Biologisk
rening

Sedimentering

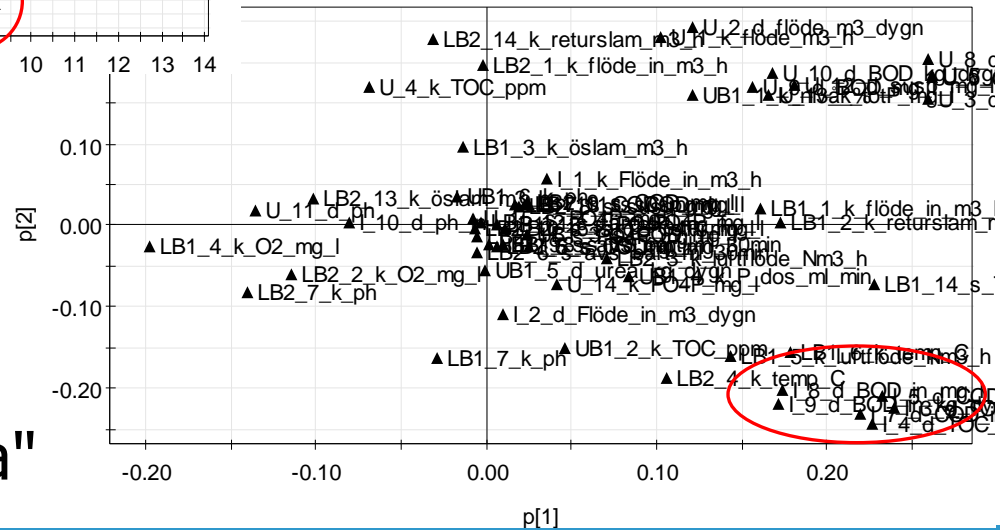
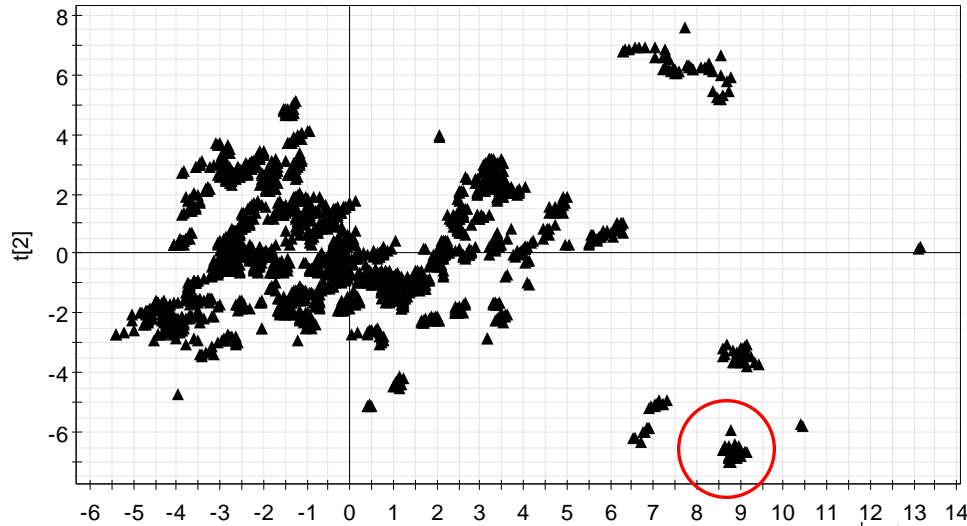
Efterbehandling

Loadings



Jämför scores med loading

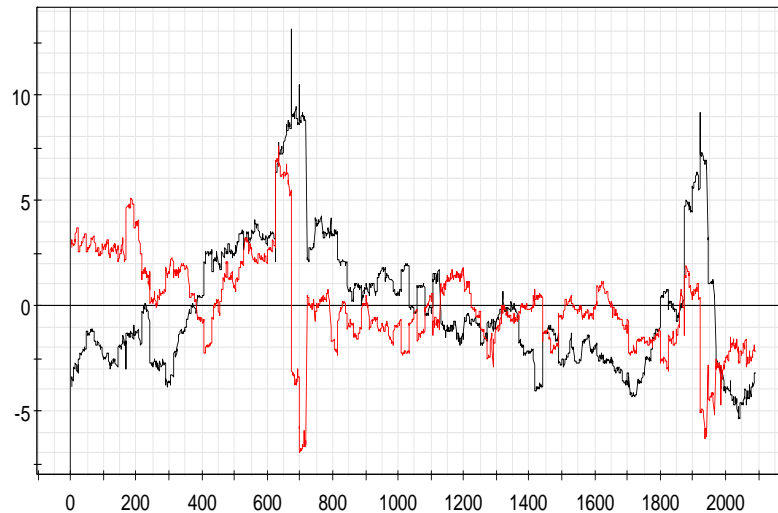
Scores – "processkarta"



Loadings – "variabelkarta"

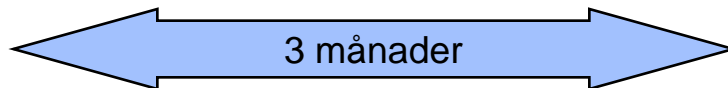
Tillståndsvariabler

Mycket tydligare än trender av de 60 enskilda signalerna



Svart: PC1 belastning

Röd: PC2 rening



Exempel II: Valsning



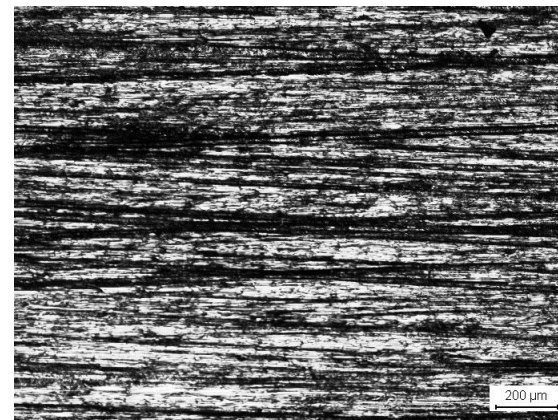
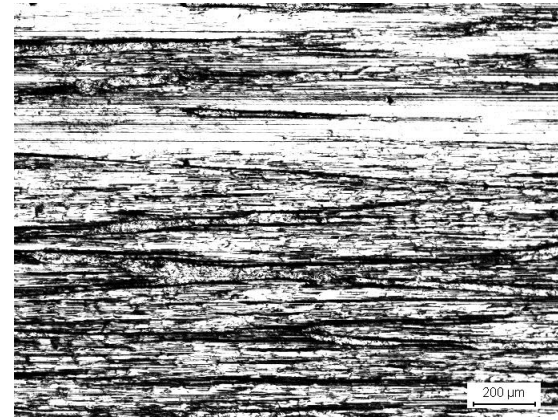
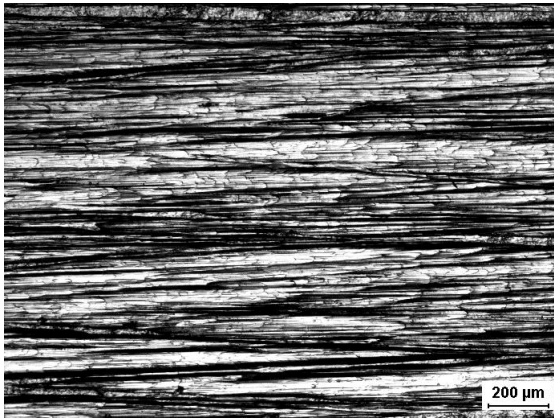
- Valsning: Plastisk bearbetning för att minska materialets tjocklek
- IVL har utvecklat tekniker för att separera och recirkulera valsolja och vatten från valsning och metallbearbetning
- IVL har utprovat en lösning av vatten och kemikalier (villkorad emulsion) som ersättning för valsoljaemulsionen
- Denna lösning används idag av SAPA, som därmed kan arbeta med en helt oljefri lösning, vilket har stora arbetsmiljö-, miljö- och energifördelar

Enkel multivariat analys av bilder

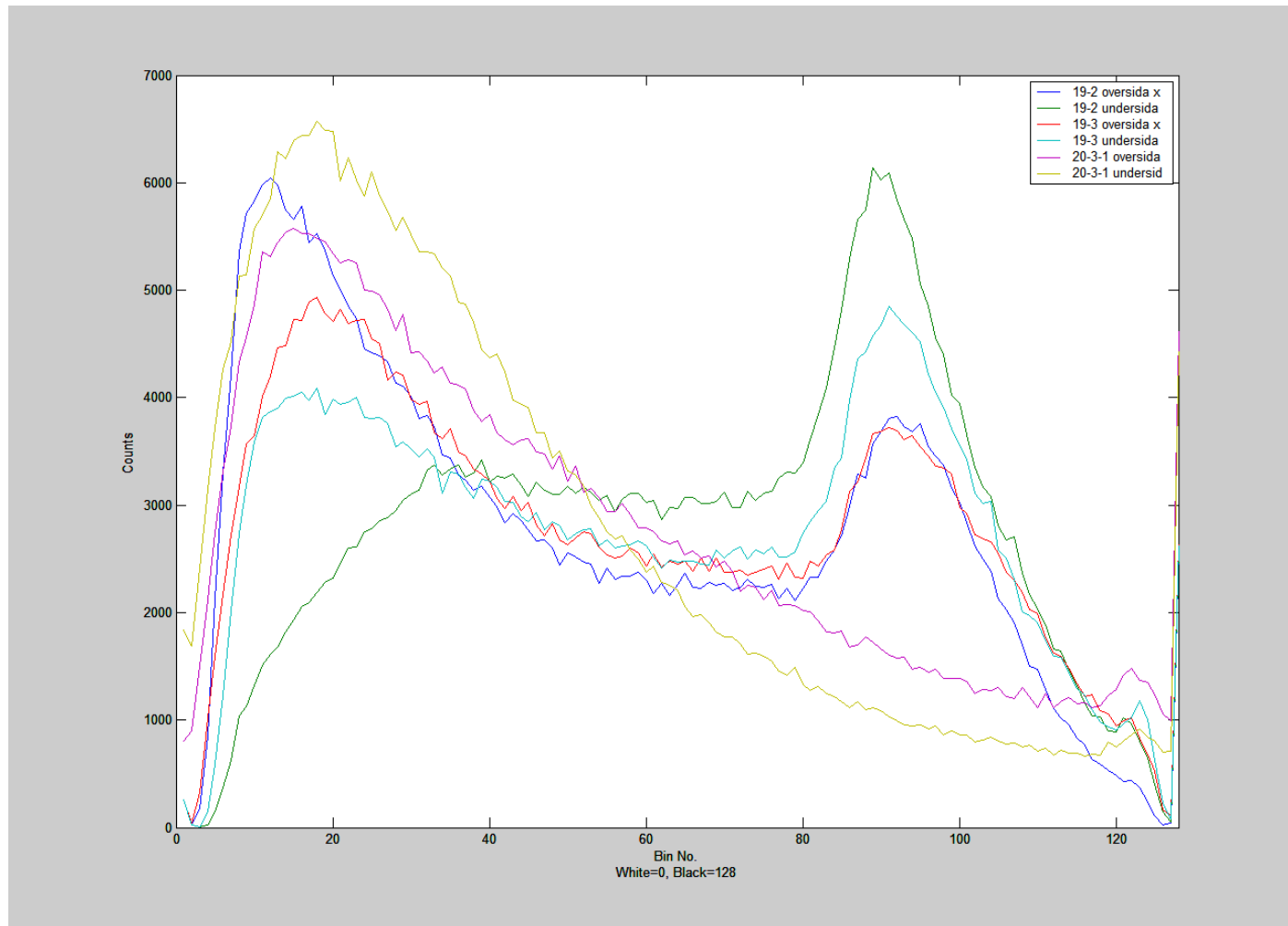
- Vid valsning undersöker man ytorna med hjälp av ljusmikroskop som tar svartvita bilder
- Van mikroskopist granskar ytorna för klassificering
- Vår idé: gör histogram över gråskalan och gör sedan PCA
- Om processdata är inläst samtidigt kan vi färga med hjälp av dessa variabler

Bilderna kommer från Pilotvalsningar inom projektet Optilub där SAPA valsade aluminium hos Swerea MEFOS i Luleå

Enkel multivariat analys av bilder



Histogram över gråskalan





Exempel III: On-line prediktion med mjukvarusensor

Nynas AB

Tillämpning på Nynäs Refining, Göteborg



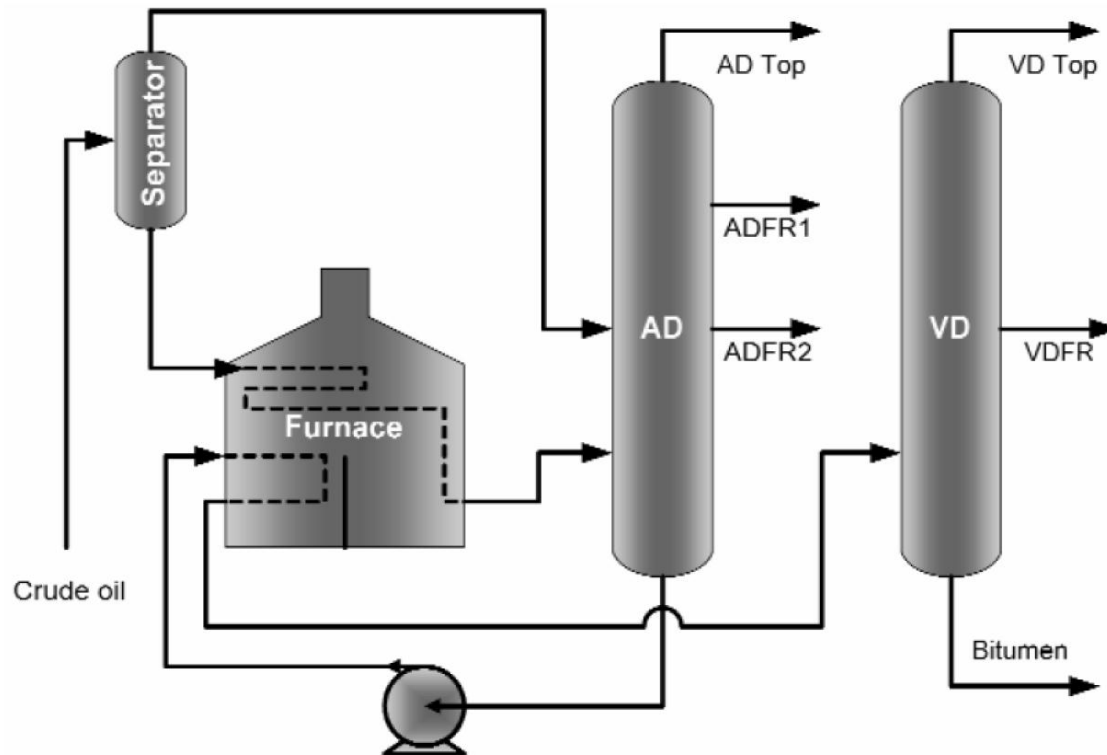
Identifiering av

- Hittills okända störningar
- Potentiell energieffektivisering
- Sätt att nå ökat utbyte av värdefull produkt i kombination med energibesparing

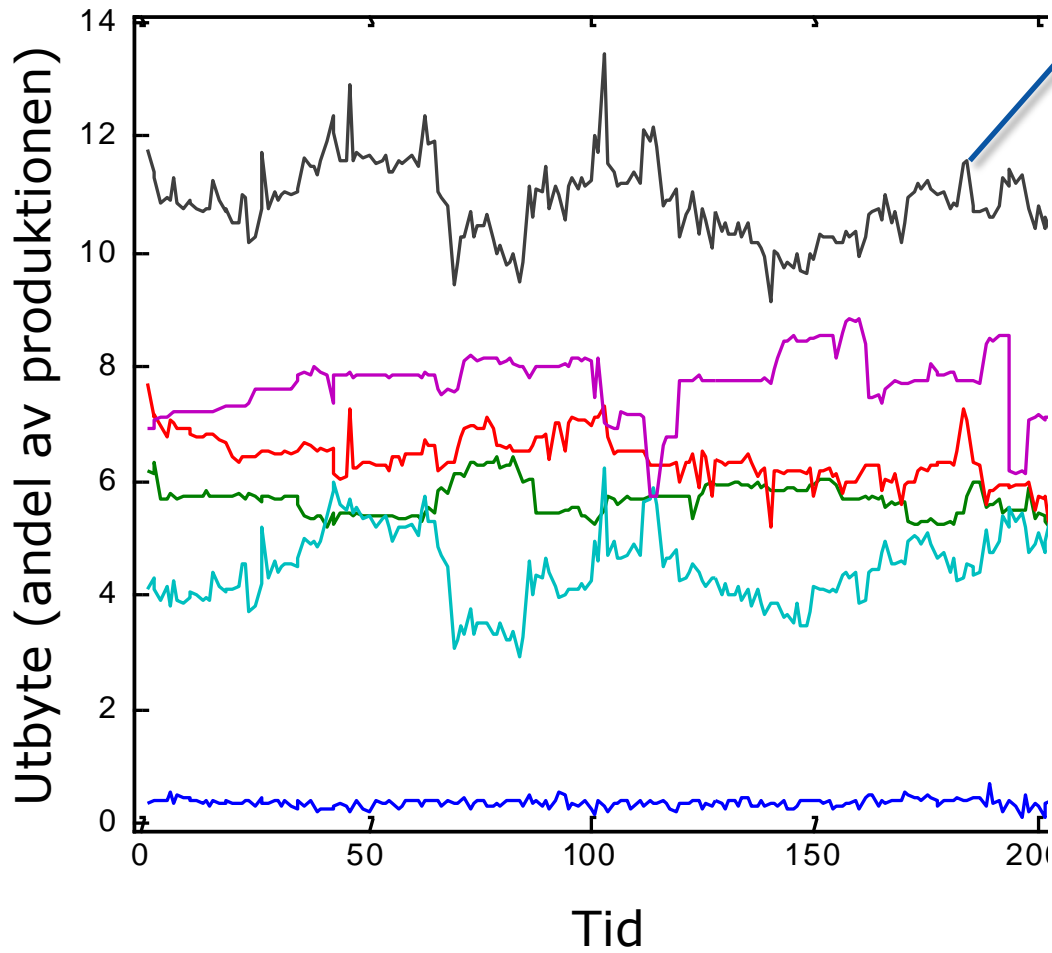
”Vi sitter på en guldgruva!”

Lars Hidfors, produktionschef på Nynäs Refining AB

Processen i Nynäs

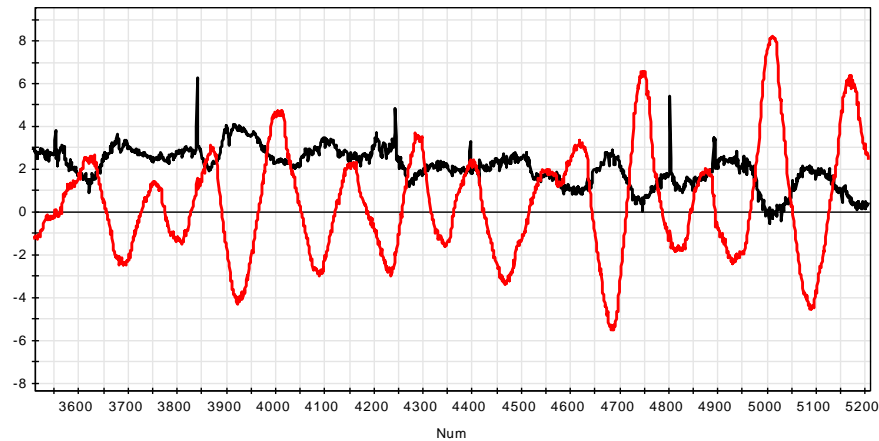


Maximera vinsten



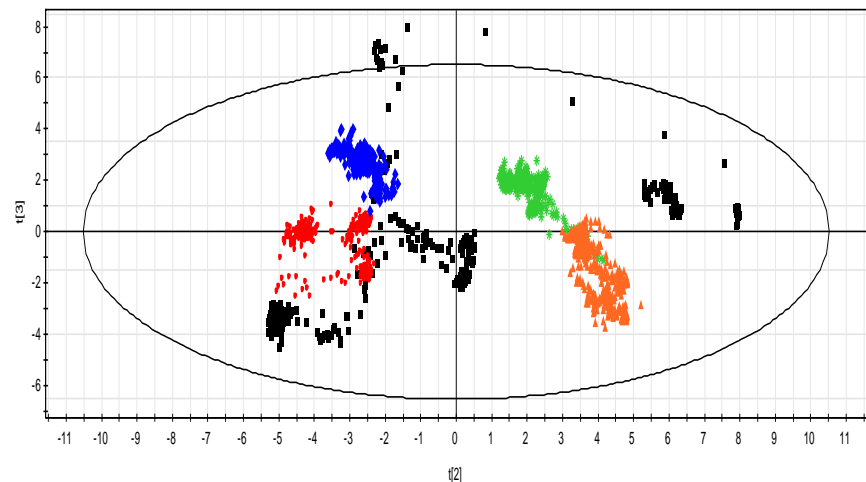
Identifiering av störningar i AD tornet

- Snabba oscillationer i toppen av en av kolonnerna
 - Onödig energikonsumtion
 - Påverkar produktegenskaper
- Identifiering av möjliga orsaker



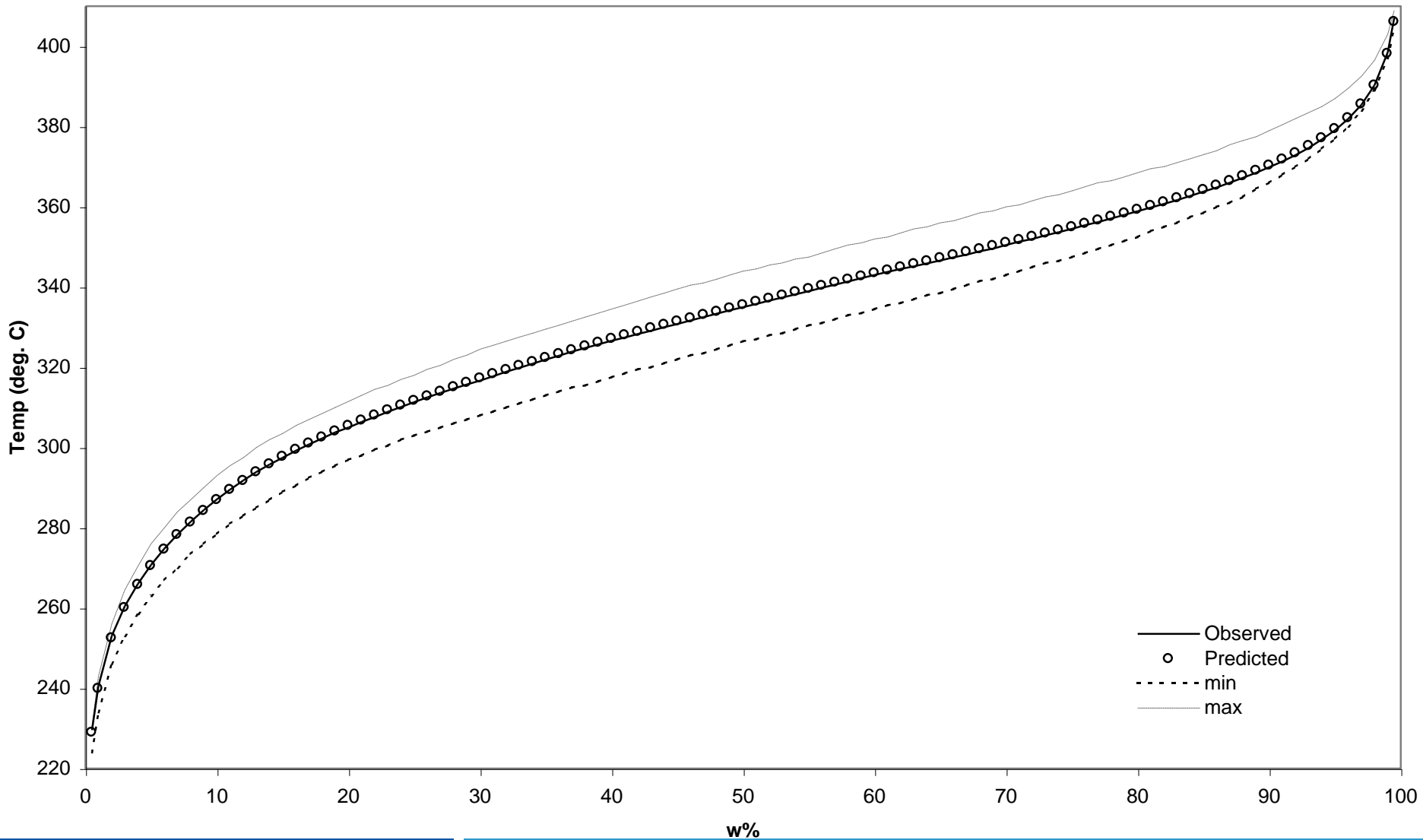
Variationskarta VD-kolonnen

- Olika körperioder har olika körsätt. Visualiseras!
- Mer uniform processtyrning önskvärd
- Tolkning av skillnader

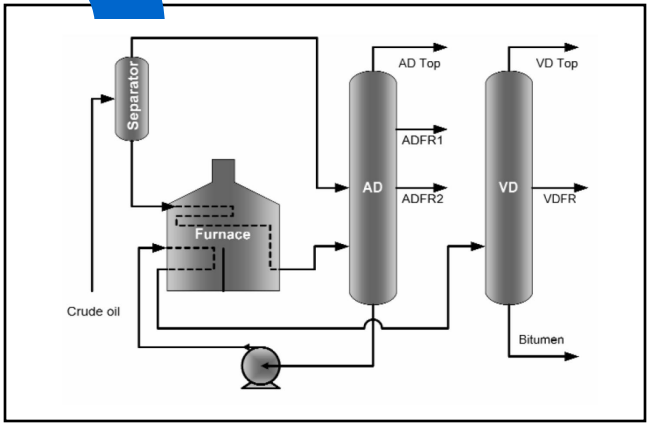
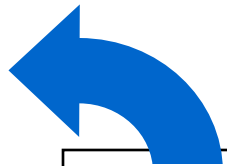
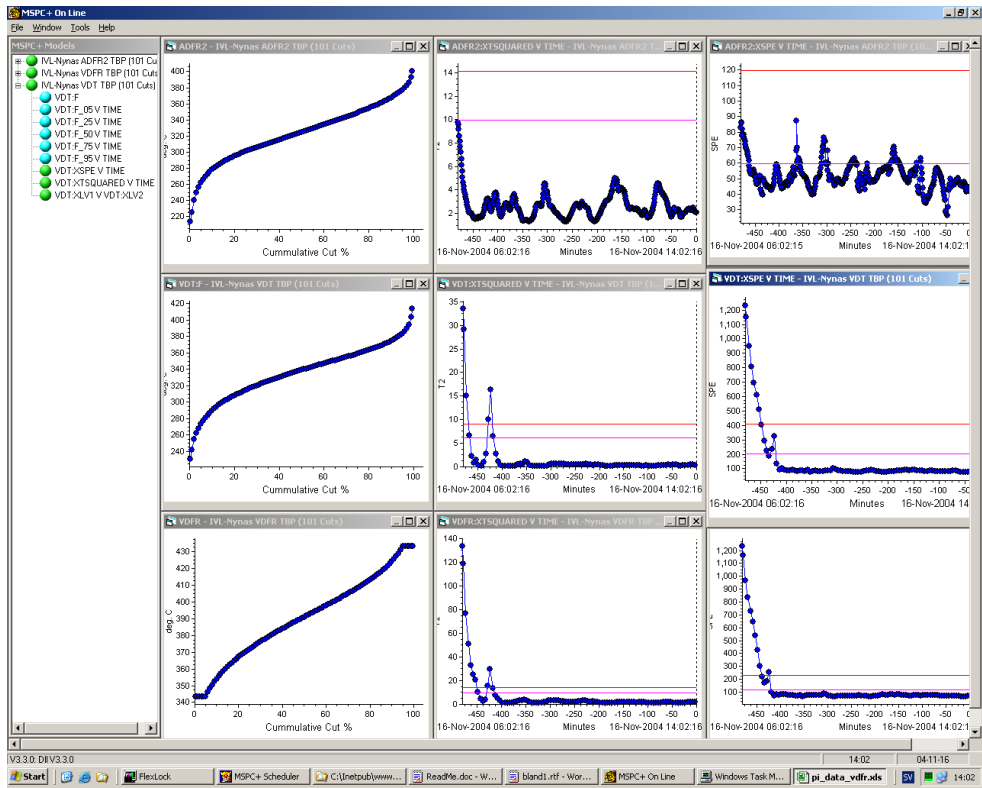


Mjukvarusensor för produkttegenskaper

- Produkten karakteriseras av en kokpunktskurva (TBP=true boiling point)
 - många variabler påverkar
 - info om andra fysikaliska egenskaper behövs
- Ursprunglig metod
 - Analys var 8:e timme
 - Upp till 5 timmars fördröjning
- Efter projektet
 - Online-prediktion i realtid
 - Helt nya möjligheter att styra processen



On-line: TBP prediktion och diagnostik



Resultat från projektet

Teori innan projektet:

- Ökat produktutbyte 5%
- Energibesparing 5%
- Ökad vinst 4 MSEK/år

Verkliga resultat:

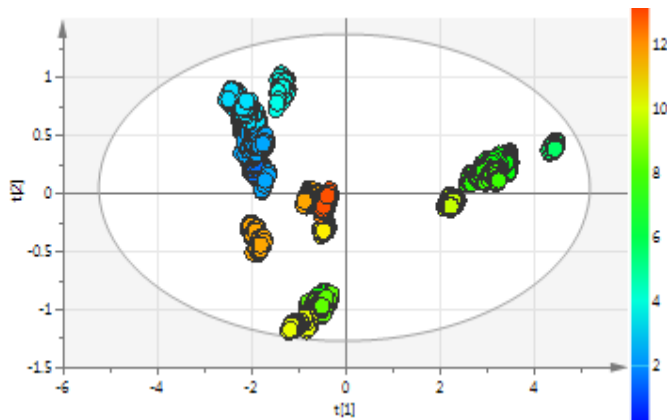
- Ökat produktutbyte 14% av värdefullaste produkten
- Energibesparing
- Ökad vinst med över 20 MSEK/år

Exempel IV: Dosmodell för Norrvatten

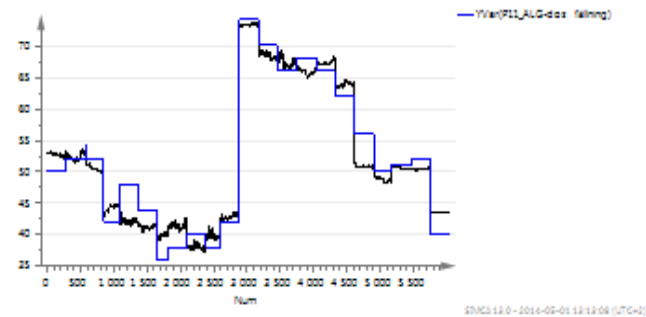
- **Syfte:** Att gå från manuell till automatisk dosering av fällningskemikalier
- **Vinster:** att kunna reagera snabbare på förändringar av råvattnet och därmed få stabilare kvalitet på dricksvattnet
- **Vill också:** Undvika ytterligare hårdvara, implementera i ABB 800XA

Resultat förstudie

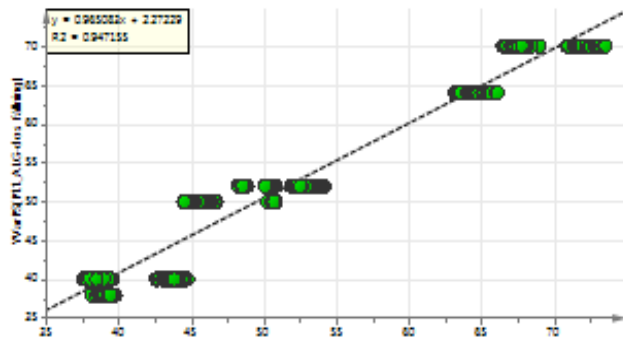
Scores



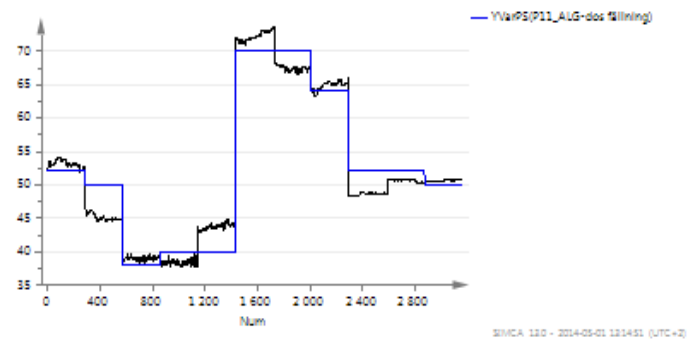
Kalibrering



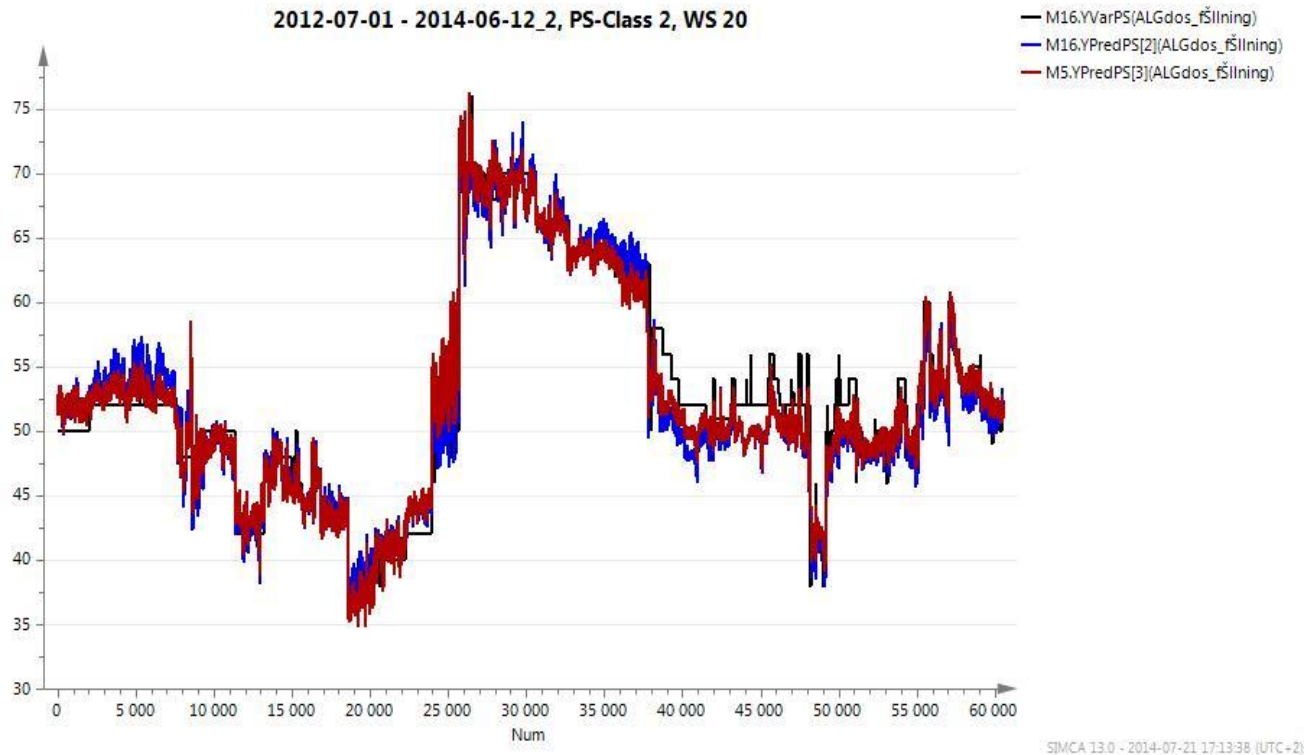
Modell vs manuell dosering



Externvalidering



Validering juli 2012 till maj 2014



Fortsättning

- Implementering i styrsystemet i början av hösten 2014
- Validering över längre tidsperiod på historiska data
- Validera efter driftsättning innan den slås på skarpt

- Framtida möjligheter
 - Optimera dosering genom försöksplan för doseringen beroende på inkommande kvalitet och utgående kvalitet

För att sammanfatta

- Multivariat dataanalys är
 - ett verktyg för att få ut mer information ur data
 - grafisk tolkning
- Tillämpningar
 - Övervakning
 - Virtuella sensorer
 - Optimering
 - Skrivbordet eller operatörsrummet

Ökad kunskap om processen!

Tänk på...

- MVA är inte bra för helt oberoende data
- Inte nödvändigtvis kausala samband
- Variationen i indata bestämmer vilken kunskap som kan erhållas från modellen



Mer info

- www.ivl.se, linda.amand@ivl.se

Några programvaror för MVA

- MATLAB PLS-toolbox, Eigenvector
- SIMCA, Umetrics



Tack för uppmärksamheten!

FRÅGOR?