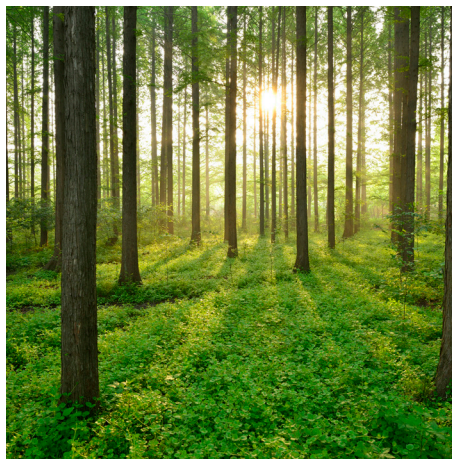
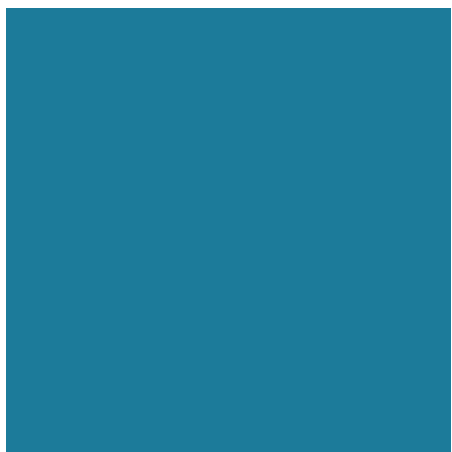


UNI- OCH MULTIVARIATA METODER FÖR UPPTÄCKT AV AVVIKELSER I DAMMAR

RAPPORT 2024:986



DAMMSÄKERHET



UNI- OCH MULTIVARIATA METODER FÖR UPPTÄCKT AV AVVIKELSER I DAMMAR

HÅKAN FRIDÉN, ANTON JACOBSON OCH ANDERS BJÖRK

ISBN 978-91-7673-986-0 | © Energiforsk januari 2024

Energiforsk AB | Telefon: 08-677 25 30 | E-post: kontakt@energiforsk.se | www.energiforsk.se

Förord

Denna rapport behandlar den fortsatta utvecklingen och implementeringen av datadrivna uni- och multivariata metoder för att upptäcka och visualisera både avvikande värden från givare monterade på eller intill vattenkraftsdammar och avvikelser från normalt dammtillstånd.

Rapporten presenterar syfte för arbetet, beskrivning av data, framtagning av metoder för upptäckt av givar- och kommunikationsavvikelser och dess resultat, samt univariat förbehandling av data inför modellering. Vi tillämpar multivariat dataanalys för att både fånga upp avvikande tillstånd hos enskilda givare och för att få en sammanfattning av dammtillståndet.

Rapporten innehåller en beskrivning av använda metoder för implementeringen, och en beskrivning av hur denna implementation gjorts i realtidsmiljö hos dammägarna. Rapporten avslutas med slutsatser.

Rapporten har skrivits av:

Håkan Fridén, Anton Jacobson och Anders Björk på IVL Svenska Miljöinstitutet
Projektledare har varit:

Anders Björk på IVL Svenska Miljöinstitutet

Projektet har haft en referensgrupp bestående av:

Uno Kuoljok (Uniper), Kerim Genel Waldenström (Vattenfall Vattenkraft AB),
Sezar Moustafa Näsval (Fortum), Romanas Wolfsborg (Vattenfall Vattenkraft AB),
Gustav Enmark (Fortum), Andreas Fridh (Statkraft), Anton Lundqvist (Statkraft)
och Rikard Hellgren (Svenska kraftnät)

Fotograf för omslagsbilden:

Uno Kuoljok

Projektet "Optimerade och anpassade datadrivna metoder för dammsäkerhet III"
har finansierats av Energiforsks program för Dammsäkerhet och Stiftelsen
Institutet för Vatten- och Luftvårdsforskning (SIVL).

I denna publikation presenteras resultat och slutsatser från projektet, den publiceras både i Energiforsks rapportserie och i IVL:s rapportserie B.
För innehållet ansvarar IVL. Rapporten har granskats och godkänts i enlighet med IVL:s ledningssystem.

Sammanfattning

Dammövervakning är ett stort och viktigt område för svenska dammägare. Det är en utmaning att kontinuerligt övervaka en dammkonstruktion med dess givare och snabbt kunna reagera på förändringar som visar på en avvikelse i dammen. Rapporten behandlar metoder för datorbaserad upptäckt av givar- och kommunikationsfel samt tillämpning av multivariata metoder för att upptäcka och visualisera avvikande tillstånd för en damm.

Rapporten är en redovisning av det tredje av en rad forskningsprojekt. Projektet har drivits av IVL Svenska Miljöinstitutet AB under 2022–2023. Alla projekt har finansierats av Energiforsk och Stiftelsen Institutet för Vatten- och Luftvårdsforskning (SIVL) och har som mål att implementera övervakning, detektion och visualisering av givarfel samt övergripande tillståndsovervakning av dammar inklusive varning när tillståndet rör sig mot icke normalt område.

Ett syfte för detta projekt har varit att fördjupa kunskapen kring de metoder vi använder genom att komplettera med ytterligare en fallstudie som är av annan karaktär än de tidigare, både gällande dammen och dess övervakningsgivare. Ett andra syfte har varit att öka kunskapen och kompetensen både på bredden och på djupet för branschen. Ett tredje syfte har varit att hitta sätt att öka lättillgängligheten av de resultat som våra metoder ger.

I de tidigare två projekten i serien har vi utvecklat och installerat en generell övervakningsapplikation som körs i realtid hos två av dammägarna, med dataöverföring från och till deras signaldatabaser. Vi har utvecklat metoder för fristående övervakning av enskilda signaler, univariat övervakning samt påbörjat utvecklingen av multivariat övervakning.

I det här projektet har de samlade erfarenheterna använts för att ansluta en tredje damm, fallstudie III, och för att fortsätta utvecklingen av metodiken. En stor del av projektet har också ägnats åt resultatspridning i form av webinarier, en djupare kurs för personal hos kraftbolagen om den använda metodiken och den bakomliggande teorin

Projektserien har nu tre fallstudier, och genom anläggningsägarna har projektet fått tillgång till mer än 5 års historiska data för respektive damm. Dessa data ligger till grund för vidare utveckling och utvärdering av övervakningsmetoderna. Att samla in, bygga upp en gemensam förståelse av och förbehandla dessa data kräver mycket tid och ett gott samarbete mellan dataspecialister och dammägare, en grund som nu är lagd.

Nyckelord

Felövervakning, Tillståndsovervakning, Datadriven, Dammsäkerhet, Feldetektion, Sensordata, Signalövervakning, Processövervakning, Dataanalys, Maskininlärning

Summary

Dam monitoring is a large and important area for Swedish dam owners. It is a challenge to continuously monitor a dam construction with its sensors and be able to quickly react to changes that indicate a deviation in the dam. The report covers methods for computer-based detection of sensor and communication errors and the application of multivariate methods to detect and visualize anomalous states for a dam.

The report is an account of the third of a series of research projects. The project has been run by IVL Swedish Environmental Research Institute AB during 2022–2023. All projects have been funded by Energiforsk and the SIVL foundation and with the aim to implement monitoring, detection, and visualization of sensor errors as well as overall condition monitoring of dams including warnings when the condition is moving towards a non-normal area.

One purposes of this project has been to deepen the knowledge of the methods through an additional case study which has a different character both regarding the dam and its monitoring sensors. A second purpose has been to increase knowledge and competence both in breadth and depth for the involved industry. A third purpose has been to find ways to improve the accessibility of the results that our methods provide.

In the previous two projects in the series, it was reported how we developed and installed a general monitoring application that runs in real time at two of the dam owners with data transfer from and to their signal databases. We have developed methods for independent monitoring of individual signals, univariate monitoring and started the development of multivariate monitoring.

In this project, the accumulated experience has been used to connect a third dam, Case Study III, and to continue the development of the methodology. A large part of the project has also been devoted to dissemination of results in the form of webinars, a deeper course for staff at the power companies on the methodology used and the underlying theory.

The project series now has three case studies and through the plant owners the project has gained access to more than 5 years of historical data for each dam. These data form the basis for further development and evaluation of monitoring methods. Collecting, building a common understanding of and pre-processing this data requires a lot of time and good cooperation between data specialists and dam owners, a foundation that has now been laid.

Innehåll

1	Inledning	7
2	Data	9
2.1	Visualisering av enskilda signaler	10
3	Univariata givar- och kommunikations-avvikelser	12
3.1	Tester	12
3.2	Resultat från fallstudie III (Uniper)	12
3.3	Revidering av gränsvärden för fallstudie I & II	18
4	Multivariata metoder – teori	21
4.1	Principalkomponentanalys, PCA	21
4.2	Partial Least Squares regression, PLS	21
4.3	Avvikelse från modellens giltighetsdomän	22
4.4	Övervakning med PLS-modeller	23
5	Arbetsgång för multivariat modellering	24
6	Förbehandling av data för modellering	25
7	Resultatspridning	28
7.1	Seminarier	28
7.2	Kurs	28
8	Multivariat modellering - resultat	30
8.1	Principalkomponentanalys, PCA	30
8.2	Partial Least Squares regression, PLS	33
9	Modellunderhåll	39
9.1	Utfört modellunderhåll	39
9.2	Fallstudie I (Vattenfall)	39
9.3	Fallstudie II (Statkraft)	46
10	Avvikelsevisualisering - från dammöversikt till detaljnivå	53
11	Implementation	58
12	Slutsatser och rekommendationer	59
13	Referenslista	61

1 Inledning

En damm är ett byggnadsverk vars syfte är att lagra, kontrollera och/eller avleda vatten. Dammar började byggas i Sverige redan under medeltiden då den kraft som kan utvinnas när vatten släpps från dammen utnyttjades för att ersätta muskelkraft vid utförande av tunga sysslor. Vattenkraften utvecklades under 1900-talet då många betongdammar byggdes för att producera elektricitet.

En damm håller inte för evigt: betongen åldras, det sker förändringar i omgivningen, klimatsvängningarna ger högre temperatur- och flödesvariationer, reglerstrategin som från början var ganska statisk blir alltmer volatil då vattenkraften nu också börjat användas för frekvensstabilisering. Det är uppenbart att våra dammar behöver övervakas i ännu högre utsträckning än för 50 år sedan.

Dammätningar behövs för att följa och utvärdera eventuella förändringar, både på kortare och längre sikt, samt för att ge underlag för långsiktig bedömning av anläggningarnas tillstånd och behov av eventuella åtgärder.

Vid övervakning av dammar mäts olika egenskaper vid flera positioner i dammen. Den datamängd som genereras från en damm är stor. Det är svårt att hantera dess kvalitet och kvantitet på ett effektivt sätt. I många fall finns både varnings- och larmgränser definierade för varje enskild givarsignal för att uppmärksamma operatörerna på värden som är högre eller lägre än förväntat vid normalt beteende.

Men vad innebär ett normalt beteende? Det är i många fall är svårt att bedöma vad som kan betraktas som normalt. Ett sätt att delvis hantera detta är att använda sig av historiska data som man litar på och, utöver från erfarenhet, med statistiska metoder etablera vad som är normalt. Då kan man också jämföra hur olika givarsignaler beter sig i förhållande till varandra. Det som inte är normalt kallar vi här en avvikelse.

När en avvikelse väl har upptäckts måste den analyseras. Avvikande mätvärden kan grovt klassificeras i två orsaksgroper.

- Avvikelser som beror på fel i givare eller störningar i signalkedjan.
- Avvikelser som beror på faktiska förändringar i dammens tillstånd, vilka kan bero på inre processer eller yttre påverkan såsom temperatur- eller vattenståndsförändringar. Här antas att alla givarsignaler är korrekta.

Det är problematiskt när mätfel (felaktigt placerade eller trasiga givare) ger upphov till avvikelser, som i sin tur påverkar bedömningen av dammens tillstånd. Idag görs denna bedömning erfarenhetsmässigt, men den ökande mängden mätdata och ändrade reglerförhållanden medför att bedömningen blir allt svårare och även alltmer kritisk. Detta utgör motivationen till att utveckla automatiska metoder för att upptäcka givar- och kommunikationsfel.

I de tidigare två projekten i serien rapporterades (se Jacobson et al., (2021) och Jacobson et al., (2022)) hur vi utvecklat och installerat en generell övervakningsapplikation som körs i realtid hos två dammägare med dataöverföring från och till deras signaldatabaser. Vi har därvid utvecklat metoder för fristående övervakning

av enskilda signaler, univariat övervakning samt påbörjat utvecklingen av multivariat övervakning.

Syftet med detta projekt har varit att

1. utvidga med ytterligare en fallstudie från en damm som har annan karaktär både gällande anläggningen som sådan och dess dammövervakningsutrustning.
2. öka användningen och förståelsen av de utvecklingsarbeten vi gjort. Detta både genom att göra framställningen bredare för att nå olika typer av aktörer som jobbar med dammsäkerhetsfrågor och även på ett djupare plan mot de som ska använda eller förvalta dessa metoder i företagen.
3. förvalta och förbättra de två existerande fallstudierna, vad det gäller signalgränser, modeller, modellgränser samt mjukvara som behöver mindre uppdateringar. Detta omfattar även utveckling av den metodik vi använder.
4. ytterligare visa på nyttan av den metodik och verktyg som utvecklas och använts för förbättrad övervakning av givare för dammövervakning liksom dammarna i sig.

I det här projektet har ännu en anläggning tillkommit. Data från anläggningarna har använts för att vidareutveckla de multivariata modellerna för att övervaka givarsignaler och dammarnas tillstånd. Projektet har även utvecklat och implementerat övervakningsmetoder baserade på dessa modeller för att på ett enklare sätt överblicka den stora mängd givardata som genereras från dammarna. Fallstudierna har genomförts tillsammans med Vattenfall Vattenkraft AB, Statkraft Sverige AB och Uniper. Möten har skett till största del digitalt men med ett kompletterande fysiskt möte i Stockholm den 22 juni 2022.

Rapporten inleds med en redovisning av de data som ligger till grund för arbetet. Därefter beskrivs resultaten från de enkla univariata tester för detektion av givar- och kommunikationsfel som har använts i projektet.

Efter detta följer en något utvidgad och omarbetad teorigenomgång av de multivariata modelleringsmetoderna PCA och PLS som applicerats i projektet. Vi visar hur data förbehandlas, tvättas från konstigheter, inför modelleringen för att de skall vara så rena som möjligt och därmed ge så bra modeller som möjligt. Därefter följer ett kapitel om genomförd resultatspridning i form av webinarier och kurs.

Resultaten från modelleringen av data från fallstudie III (Uniper) används för att illustrera metodernas funktion och prestanda. Sedan förra projektet har ytterligare data erhållits för fallstudie I och II. Dessa har använts till revidering av deras gränsvärden och modeller.

I samband med arbetet med resultatspridning insåg vi att traditionella trendgrafer inte riktigt räcker för visualisering när data nu inte bara består av de loggade givarsignalerna utan även upptäckta avvikelser av olika slag, modellparametrar och prediktioner. Ett kapitel lades därför till kring interaktiv hierarkisk visualisering. Implementeringen i dammägarnas övervakningssystem beskrivs i ett separat kapitel.

2 Data

I detta projekt används data från tre dammar, från olika kraftbolag, som fallstudier. Den första dammen tillhör Vattenfall Vattenkraft AB (fallstudie I) och introducerades i Jacobson et al., (2021), den andra tillhör Statkraft Sverige AB (fallstudie II) vilken introducerades i Jacobson et al., (2022), den tredje tillhör Uniper (fallstudie III).

Då dammarnas utformning skiljer sig åt är dammarna olika instrumenterade, såväl i typ av givare, som antal givare, se Tabell 1.

Tabell 1. Lista på tillgängliga givare i fallstudien

Typ av givare	Storhet	Vattenfall Fallstudie I	Statkraft Fallstudie II	Uniper Fallstudie III
Tryckgivare för nivå	m	33	15	-
Temperatur	°C	31	15	7
Flöde, Läckage	l/s	9	3	-
Flöde, Kraftverk	l/s	-	3	-
Läckage – Turbiditet	FNU	2	-	-
SMHI – Utetemperatur	°C	1	2	2
SMHI – Lufttryck	hPa	1	-	2
Magasinsnivå	m	1	-	-
SMHI - Nederbörd	mm/d	1	-	-
SMHI – Vindhastighet	m/s	1	-	-
SMHI – Vindriktning	deg	1	-	-
Thomsonöverfall	mm	-	2	-
Övre vattenyta	m	-	2	1
Nedre vattenyta	m	-	1	-
Sprickgivare	mm	-	4	13
Temperatur i betong eller monolit	°C	-	4	5
Pendel i damm	mm	-	-	2
Porttryck i berg	mVP (vattenpelare)	-	-	4
Batterispänning	V	-	-	2
Totalt		81	51	33

I fallstudie III introducerades nya typer av givare som ej använts tidigare i metodutvecklingen, till exempel temperaturgivare i monolit samt pendel som mäter dammkroppens rörelse i X- och Y-led. Fallstudie III har ingen givare för utetemperaturen vid dammen. I stället kompletterades data med SMHI:s mätningar av utetemperaturen i regionen där dammen är lokaliserad. För en generell översikt över instrumentering för dammövervakning se Nilsson, (2014).

Data för samtliga fallstudier har exporterats från dammägarnas mjukvara PI som är en långtidsdatabas med verktyg för analys och visualisering (OSIsoft, (2020)).

Fallstudie I och II har uppdaterats med data vid flera tillfällen. Initialt exporterades ett stort historiskt dataset som låg till grund för utvecklingen i Jacobson et al., (2021) samt Jacobson et al. (2022). I samband med dessa uppdateringar reviderades både gränsvärden och modeller. Data från projektets två första fallstudier har ett samplingsintervall på 15 minuter och inkluderar olika tidsserier och presenteras i Tabell 2.

Fallstudie III, som inkluderats i detta projekt (från Uniper), har behandlats med samma univariata metoder som presenterats och använts i Jacobson et al., (2021). Datasetet har analyserats för att erhålla en uppfattning av givarsignalernas egenskaper samt förekomsten av fel. Data för denna fallstudie omfattar 5 års data för 37 givare och med samplingsintervallet en timme.

Tabell 2. De olika fallstudiernas dataset och dess tidsintervall

	Från	Till	Antal observationer	Samplingsintervall
Fallstudie I	2015-01-01	2021-09-06	>250 000	15 min
Fallstudie II	2016-05-06	2021-05-06	>175 000	15 min
Fallstudie III	2015-04-29	2021-06-30	>54 000	60 min

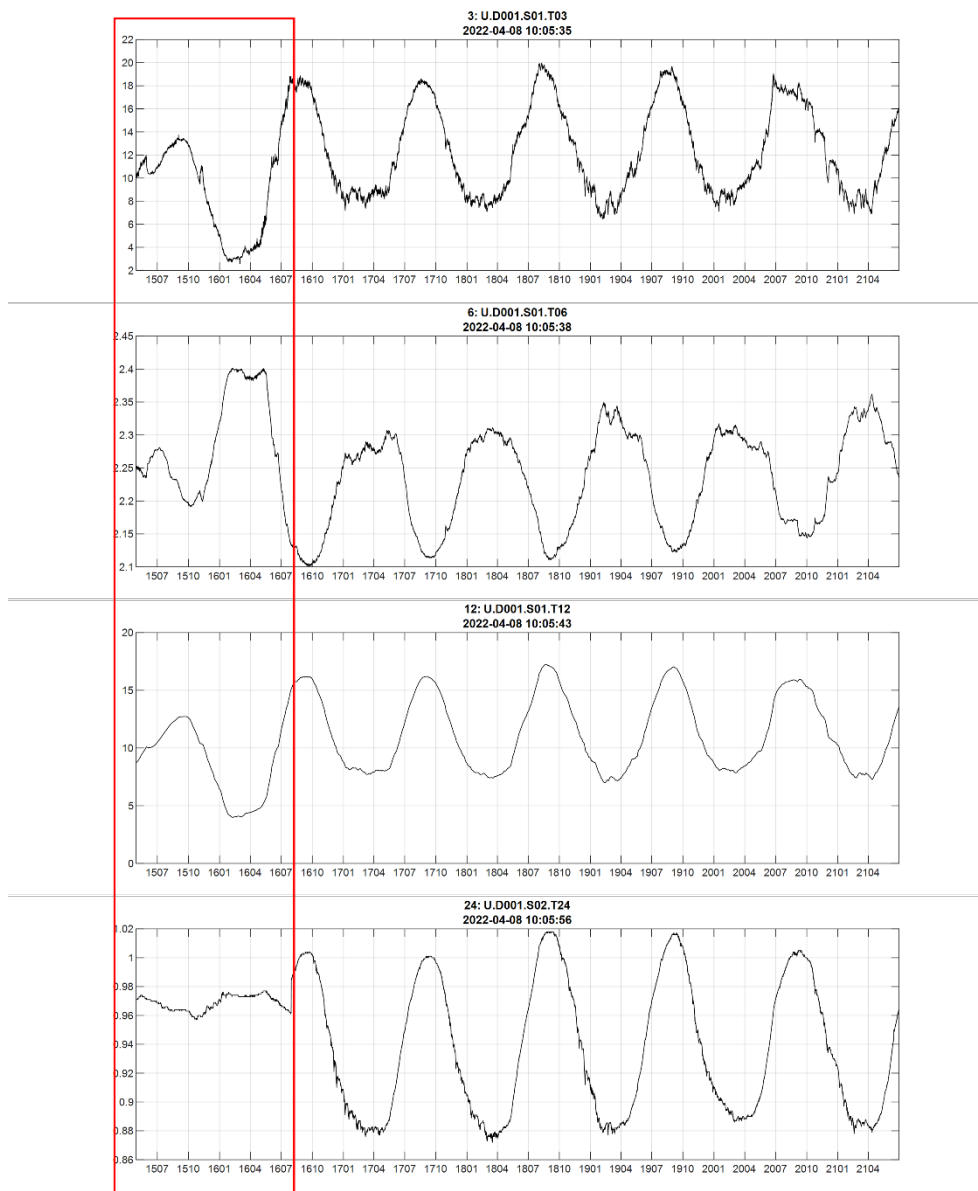
2.1 VISUALISERING AV ENSKILDA SIGNALER

All data behöver analyseras och kontrolleras tillsammans med dammägaren för att ge en förståelse för specifika givare och identifierade avvikelser. Det senare kan kräva att till exempel en dammsäkerhetsingenjör måste samråda med flera aktörer inom sin organisation. Genom att visuellt studera trenderna för samtliga givarsignaler, var för sig, går det att utläsa mycket om signalerna. För detta har vi använt oss av program skrivna i MATLAB, (2021).

Inledningsvis studerades hur samtliga givarsignaler i fallstudie III varierade över tiden. Generellt sett är kvaliteten på det historiska datasetet högt och antalet avvikare är begränsat. Dock har förekomsten av några avvikelser identifierats:

- Sekvenser med beteende som skiljer sig från normaltillståndet
- Spikar (momentant avvikande värde som snabbt återgår till normalläge)
- Givarsignaler med låg upplösning
- En trasig givare
- Nivåskifte i signal

I Figur 1 visualiseras rådata från fyra olika givare från fallstudie III. Givarna är av olika typer och placerade på olika positioner i dammen. Dessa signaler exemplifierar det gemensamt avvikande beteendet som identifierats i början av tidsserien (mellan juli 2015 och augusti 2016), markerat i rött. Signalernas beteende under denna period skiljer sig betydligt från resten av datasetet.



Figur 1. Figuren visualiserar fyra olika givares rådata från fallstudie III, två av dem är temperaturgivare och två är sprickgivare). I X-axeln representeras tiden [ååmm]. Första sekvensen (markerad röd) utgör ett dygt år från juli 2015 till augusti 2016. Dessa signaler exemplifierar det avvikande beteende som identifierats.

3 Univariata givar- och kommunikations- avvikelser

I projektet "Optimerade och anpassade datadrivna metoder för dammsäkerhet I", Jacobson et al., (2021) presenterades ett antal tester för att identifiera avvikande mätvärden som ju kan bero på både givarfel och på faktiskt avvikande beteende. Nedan beskrivs testerna kortfattat samt resultat från alla fallstudierna. Fallstudie III är introducerad i det här projektet, de andra två visar de reviderade resultaten.

3.1 TESTER

Detta delkapitel beskriver de test som applicerats på data från fallstudie III.

- **Min- och maxgränser**

En enkel och effektiv metod för att undersöka om den nya observationen befinner sig inom givarens fysikaliskt förväntade min- och maxvärden. Framtagandet av lämpliga min- och maxgränser baseras på en visuell inspektion av alla givares data var för sig. Det mänskliga ögat är mycket bra på att tolka data genom plottade tidsserier. Givarsignalens mätvärden kan bedömas vara normala eller onormala och min- och maxgränser används för att urskilja dessa två grupper så gott det går. Denna process kräver manuellt arbete. Diskussioner pågår rörande möjligheter att automatisera denna gränssättning.

- **Frusen signal**

Den här metoden finner om en givare har rapporterat samma värde under en längre period. Detta fel kan orsakas av att en givare är trasig, vilket bör upptäckas. Metoden lämpar sig inte för övervakning av signaler som under perioder kan rapportera samma värde över tid, som till exempel flöden med styrning där luckan kan vara stängd en längre period, eller signaler med låg amplitudupplösning.

- **Differensgränser**

Genom att jämföra differensen mellan successiva mätvärden kan snabba förändringar mellan observationerna upptäckas. Om skillnaden är ovanligt stor mellan närliggande observationer kan detta tyda på att det rapporterade värdet är avvikande. Metodiken används för att upptäcka stora hopp eller spikar i data.

3.2 RESULTAT FRÅN FALLSTUDIE III (UNIPER)

Här sammanfattas våra observationer vid användning av testerna som beskrivs i avsnitt 3.1 på data från fallstudie III.

I varje figur visas två grafer: Den övre visar historiska rådata från fallstudien, i något fall beskuren i y-led för att ge en bättre överblick av hur givarsignalen varierar under normalt beteende. Den nedre visar binära signaler som visar om

signalen, utifrån varje testvillkor, befinner sig i ett normalt läge, eller om testet har uppmärksammat en avvikelse. Gränsvärdena för signalerna är individuellt framtagna baserade på historiska data. Dessa värden kan justeras om bedömningen görs att funktionerna reagerar på beteenden som bör anses vara normala.

Sju olika tester finns representerade i figurena. En offset har adderats till vardera av testsignalerna för att underlätta tydlig presentation.

Följande tester är representerade i figurena nedan, och är numrerade enligt följande i samtliga figurer:

- I. Mingränsvärde
- II. Maxgränsvärde
- III. Givardrift (ej implementerat)
- IV. Frusen signal
- V. Glidande varians (ej implementerat)
- VI. Andra ordningens differens
- VII. Första ordningens differens

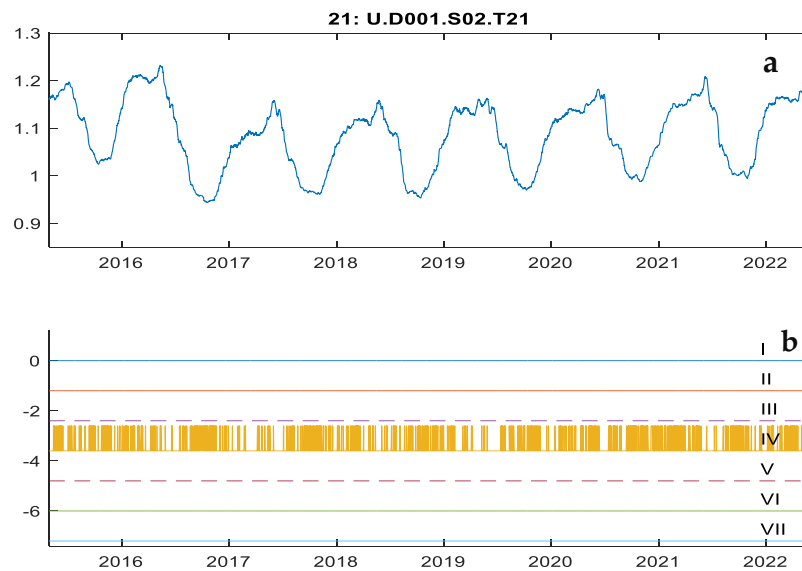
Två av testen är ej implementerade för denna fallstudie. Testet för givardrift exkluderades då slutsatsen i Jacobson et al., (2021) var att detta test är mycket svårt att utveckla univariat. Testet för glidande varians utvecklades och implementerades för fallstudie I då ett specifikt beteende identifierades i den fallstudien. Detta beteende har ej identifierats i fallstudie III vilket reducerar detta tests nytta.

Testet för frusen signal har i tidigare fallstudier begränsats till ett dygns varighet. Initialt implementerades även denna begränsning för fallstudie III. Sampling-sintervall i datasetet för fallstudie III skiljer sig från tidigare fallstudier. Detta har medfört att begränsningen som sätts på ett antal observationer varit tvunget att anpassas till frekvensen som används i fallstudie III. Gränsen för fallstudie I och II är satt till 96 observationer ($96 \text{ obs} * 15 \text{ min} = 1 \text{ dygn}$), medan den för fallstudie III är satt till 24 ($24 \text{ obs} * 1 \text{ h} = 1 \text{ dygn}$).

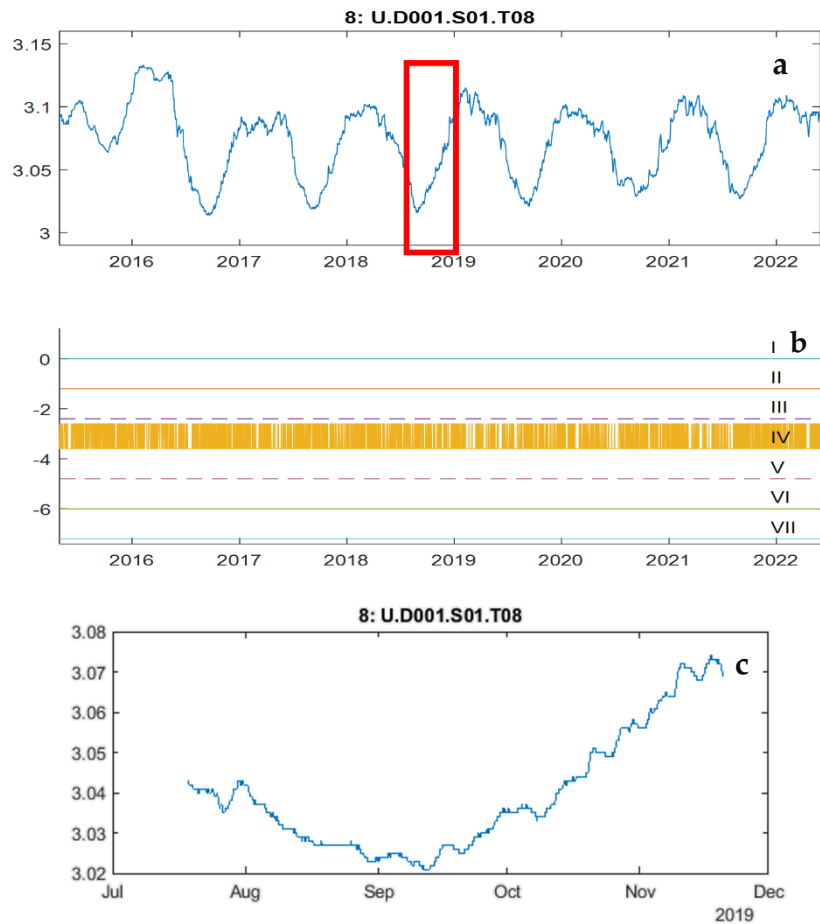
Vid en första utvärdering med detta test visade det sig att testen markerade flertalet givare som mäter sprickor med mycket hög frekvens. Detta visualiseras i Figur 2 och Figur 3.

Efter en noggrannare undersökning av dessa givarsignaler har det visat sig att upplösningen av signalen är relativt låg i förhållande till den årliga variabiliteten. Upplösningen i signalerna är liten, vilket orsakar ett steg-beteende i signalen, vilket går att se i Figur 3c. Detta betyder att mätvärdet trunkerats till samma värde under en längre tid.

Då beteendet är väntat, givet vetenskapen om bristande upplösning, skall inte frusen signal markeras vid detta beteende. För att motverka detta har justeringar av larmgränser för frusen signal genomförts. För de signaler då detta är vanligt förekommande har testet för frusen signal kopplats ur, och för övriga signaler har gränsen ökat till 2 dygn.



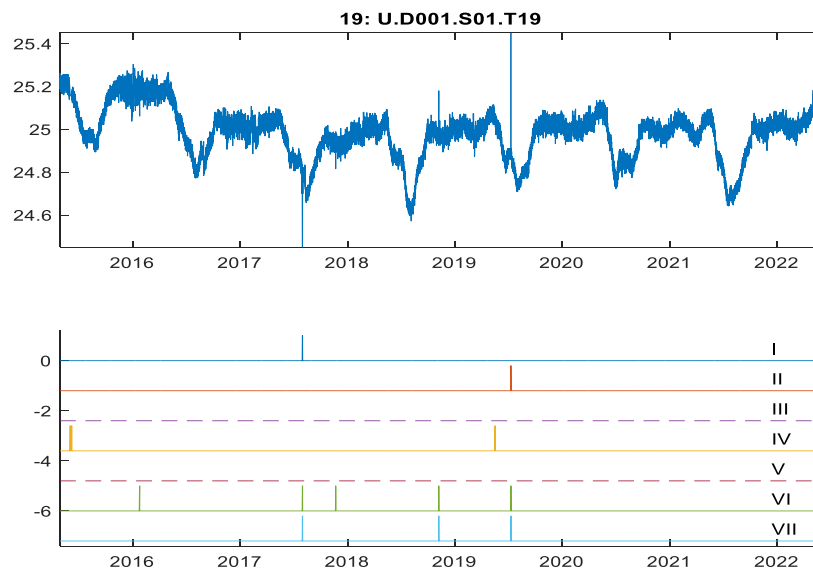
Figur 2. Figuren visar en sprickgivar signal från fallstudie III, med tillhörande övervakningssignaler. Frusen signal rapporteras med hög frekvens över hela tidsserien.



Figur 3. Figuren visar en sprickgivarignal från fallstudie III med tillhörande övervakningssignaler. Frusen signal upptäcks med hög frekvens över hela tidsserien. Figur 3c visar en mindre del av tidsserien som är markerad med röd rektangel i Figur 3a.

Som nämnts var data från fallstudie III vid första anblick generellt sett bra och antalet exempel på oönskade beteenden är begränsat. Dock finns ett fåtal exempel som presenteras nedan.

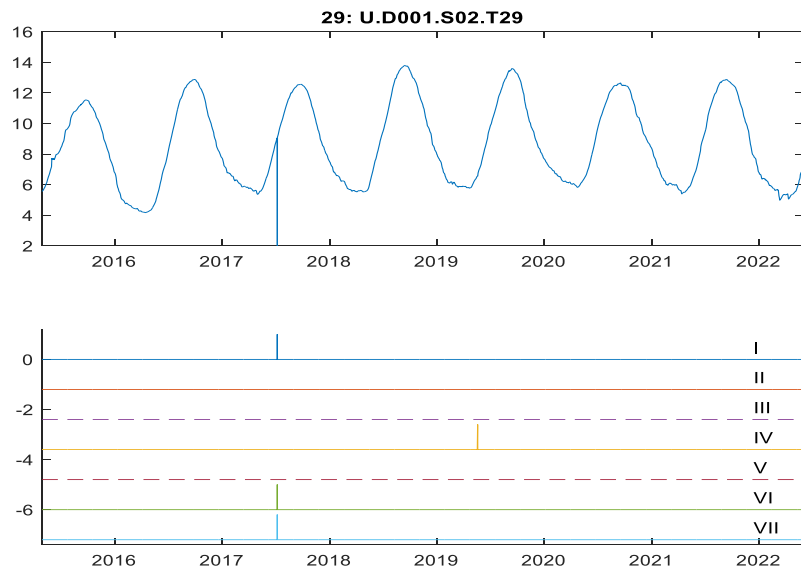
I Figur 4 visas en givarsignal som mäter dammens rörelse genom en X/Y-pendel vilken hänger fritt inuti dammkroppen. Denna signal behandlar rörelsen i Y-led, det vill säga rörelsen längs dammkroppen. Signalen innehåller ett antal spikar. Dessa upptäcks och markeras av min- och maxtesterna bara då avvikelserna är stora medan differenstagarna upptäcker och markerar spikarna även då de värdena befinner sig inom signalens normala intervall (inom min- och maxvärdet).



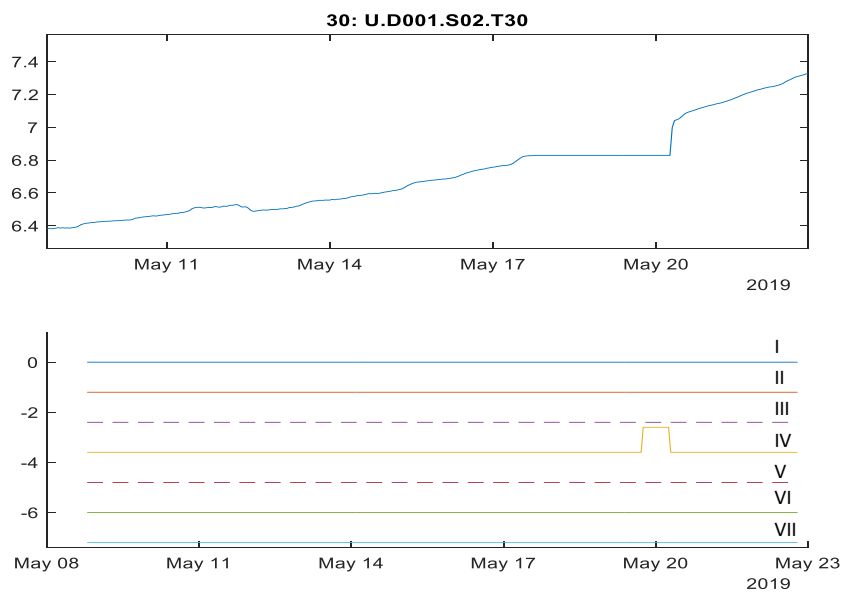
Figur 4. Figuren visar en givarsignal från en pendel som indikerar dammens rörelse i riktning längs med dammkroppen, med tillhörande univariata övervakningssignaler. Givarsignalen innehåller ett antal spikar. Ej implementerade metoder visas med streckade linjer.

I tidserien finns det två tillfällen som är av intresse att analysera vidare. Det första tillfället inträffade tidigt den 7 juli 2017. Då visade samtliga temperaturgivare som mäter vattentemperatur en stor negativ spik vid samma tillfälle. Spiken är en observation lång och sker simultant. Det andra tillfället ses i maj 2019 när flertalet givare simultant rapporterar konstant värde över 2.5 dygn. I Figur 5 kan båda dessa tillfällen studeras.

Tillfällena representeras här av en temperaturgivare som mäter vattentemperatur i fallstudie III. I Figur 5 har tidsaxeln (Y-axeln) begränsats kraftigt till tillfället med den rapporterade frusna signalen. I figuren går att utläsa att den pågående ökande signalen frusit under ca 2,5 dygn och sedan återhämtat sig med ett språng. Detta är genomgående för de signaler som delar denna händelse med givaren visualiserade i figurerna.



Figur 5. Figuren visar en givarsignal som mäter vattentemperatur i fallstudie III. Denna signal inkluderar de båda tillfällena av händelser som är gemensamma för flera givare samtidigt. Under 2017 observeras en stor negativ spik, samt under 2019 markerar frusen signal att givaren rapporterar ett konstant värde i mer än 2 dygn.



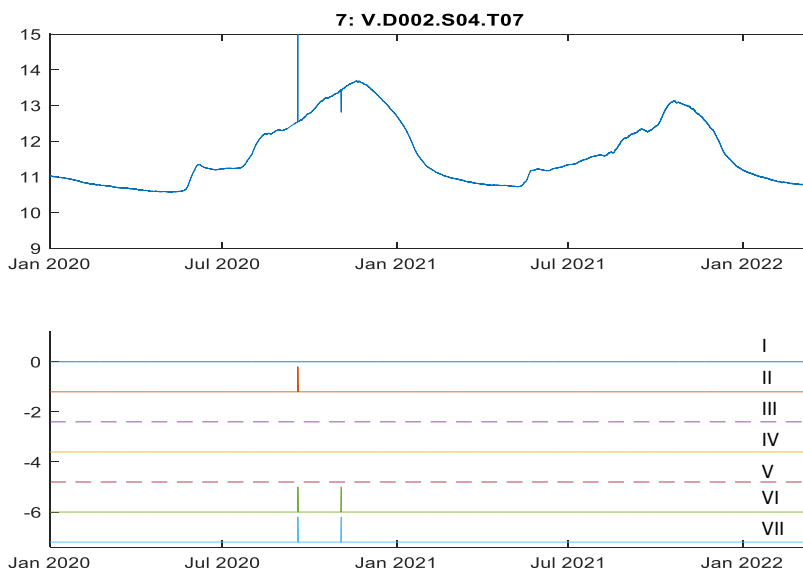
Figur 6. Både den övre och nedre grafen är en inzoomning av Figur 5 under perioden med frusen givarsignal (8:e maj till 23:e maj).

3.3 REVIDERING AV GRÄNSVÄRDEN FÖR FALLSTUDIE I & II

Inom tidigare projekt (Jacobson et al., (2021) och Jacobson et al., (2022)) har implementation av de univariata övervakningsfunktionerna gjorts för fallstudie I och II. Testerna implementerades vintern 2020 respektive hösten 2021 och har varit i drift sedan dess.

Att sätta gränser för dessa tester är en iterativ process som har genomförts ett par gånger under våren 2022 då nya data har tillkommit. Revisionen har inneburit att analysera testernas utslag och identifiera felaktigt beteende. Gränsvärdena har justerats för både fallstudie I och II.

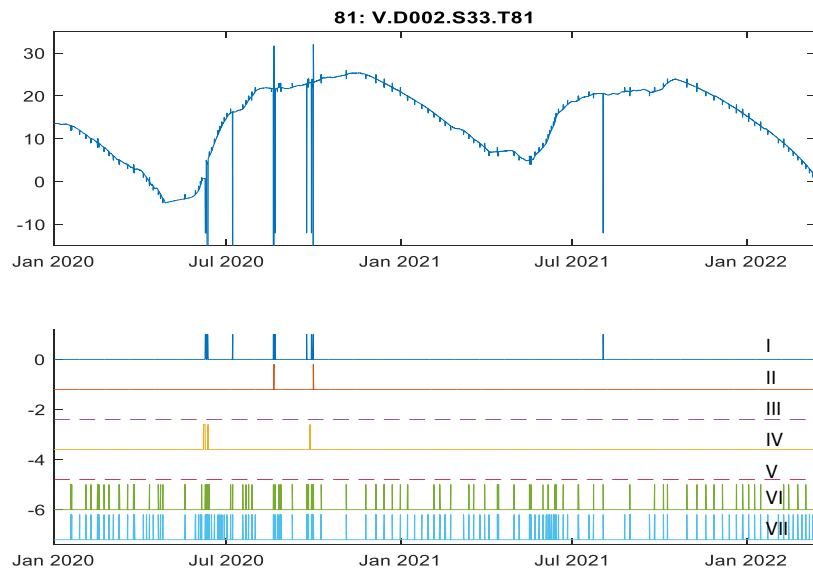
I Figur 7 och Figur 8 visar givarsignaler från fallstudie I medan Figur 9 och Figur 10 har givarsignaler från fallstudie II. Signalerna är utvalda då det förekommer uppenbara mät- eller kommunikationsfel som markerats av övervakningstesterna.



Figur 7. Figuren visar en nivågivare från fallstudie I. Tidssekvensen är begränsad till den data som adderats sedan första implementation hos fallstudie I.

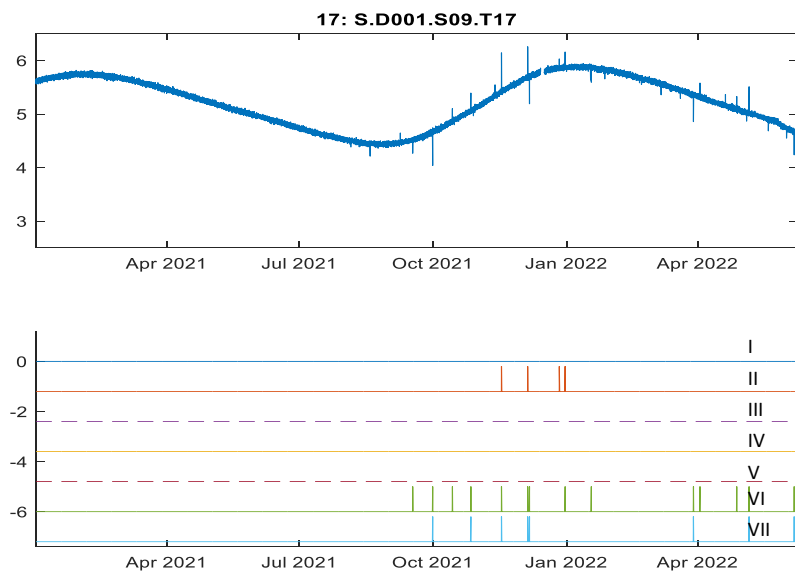
I Figur 7 visas en givarsignal för nivå i vattenståndsrör. Under hösten 2020 förekommer två spikar i data, vilka markerats av övervakningen av signalens differenser.

Figur 8 visar magasinsnivå för fallstudie I. Ett fel uppstod för denna givarsignal under 2020, vilket fortfarande kvarstår. Var gång mätvärdet korsar en exakt meter, alltså när mätvärdets decimaler är 0, rapporteras även mätvärden ± 1 meter. Felet upptäcktes under 2021 och har kommunicerats till dammägaren. Vid utvärdering av övervakningsmetoderna upptäcks dessa felaktiga mätvärden av testen av signalens differenserr.

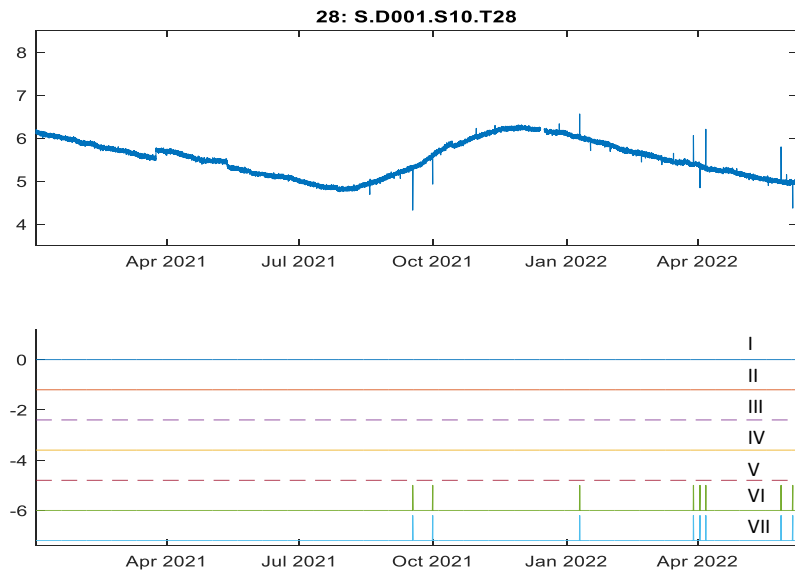


Figur 8. I denna figur visas en givarsignal som mäter magasinssnivå i fallstudie I. Tidssekvensen är begränsad till de data som adderats sedan första implementation hos fallstudie I.

Figur 9 och Figur 10 visar två vattentemperaturgivare från fallstudie II. Båda dessa givarsignaler uppvisar ett antal spikar under senaste året. Dessa felaktigheter markeras av testen för differenser, vilket är ett önskat resultat.



Figur 9. Figuren visualiserar en temperaturgivare som indikerar vattentemperatur i fallstudie II. Tidssekvensen är begränsad till de data som adderats sedan första implementation hos fallstudie II.



Figur 10. Här visualiseras en temperaturgivare som mäter vattentemperatur i fallstudie II. Tidssekvensen är begränsad till de data som adderats sedan första implementation hos fallstudie II.

4 Multivariata metoder – teori

Den teori som presenteras här är en modifiering och utvidgning av den teori som presenterades i kapitel 7 i Jacobson et al., (2022).

Metoderna PCA och PLS är multivariata projektionsmetoder. De modellerar korrelation i data och kan inte indikera kausalitet även om man ibland kan vilja tro de. För att finna kausala samband måste man göra kontrollerade försök, företrädesvis med statistisk försöksplanering.

4.1 PRINCIPALKOMPONENTANALYS, PCA

PCA kan sägas vara modern till all multivariat dataanalys. PCA är ett sätt att matematiskt beskriva de mest dominanta korrelationsstrukturerna i ett dataset. Ett dataset kan sägas bestå av N observationer, var och en med K variabler. Varje observation kan representeras som en punkt i ett K -dimensionellt koordinatsystem, vilket resulterar i att en tidsserie kommer utgöra en punktsvärm i detta koordinatsystem. En PCA-modell definierar ett annat koordinatsystem, ett underrum, som har mindre antal dimensioner än det ursprungliga koordinatsystemet. För att åstadkomma detta, och samtidigt bibehålla så mycket information som möjligt beträffande punkternas position, så definieras axlarna i koordinatsystemet för PCA-modellen baserat på i vilka riktningar som utbredningen och variabiliteten är som störst i punktsvärmen/datasetet.

- Den första axeln, principalkomponenten, i detta koordinatsystem fångar den största utbredningen eller variabiliteten i punktsvärmen/datasetet.
- Den andra axeln, principalkomponenten, fångar den näst största utbredningsriktningen vinkelrätt mot den första principalkomponenten.

Koordinatsystemet kan utökas med fler dimensioner, vinkelräta mot tidigare, upp till maximalt samma antal som det finns variabler (K) i datasetet. Men då varje komponent beskriver den största kvarvarande utbredningsriktningen brukar en till tre komponenter räcka för att få en god visualiserbar överblick av en datamängd. Mycket förenklat kan PCA sägas fånga den dominanta strukturen i två komplementära (1–3)-dimensionella bilder. I dessa bilder går det att observera hur data fördelar sig, både observationsvis (scores) och variabelvis (loadings). Det går även att studera grupperingar av observationer och variabler, såväl som avvikare och trender.

Läs mer om PCA i Varmuza & Filzmoser, (2008), Johansson et al., (2020) och Principal Component Analysis, (2019).

4.2 PARTIAL LEAST SQUARES REGRESSION, PLS

PLS är en regressionsmetod som kan ses som en utvidgning av PCA till att omfatta data i två rymder, X och Y , där K variabler representerar X -rymden medan M variabler, en eller flera, representerar Y -rymden. Varje observation representeras alltså med två punkter: en punkt i X -rymden och en punkt i Y -rymden.

PLS identifierar matematiska modeller som maximerar kovariansen mellan observationernas lägen i X och Y. Identifieringen av dessa modeller sker stegvis för att nå en så hög prediktionsförmåga som möjligt. För varje steg ökar antalet komponenter i X- och Y-rymderna. Antal komponenter i rymderna är av stor vikt då för många kan orsaka överanpassning och därmed försämrade prediktionsförmåga. För att minimera denna risk används inre korsvalidering där delar av data iterativt utesluts och predikteras.

Vid utveckling av en PLS-modell finns olika godhetstal, som till exempel Q^2 . Detta tal uttrycker modellens prediktionsförmåga (Eriksson et al., (2003)). Under den iterativa processen jämförs Q^2 mellan modellerna för att säkerställa att modellens prediktionsförmåga ökar. Andra godhetstal kopplat till en modell är R^2X , R^2Y och VIP.

R^2X beskriver hur stor del av X-variationen som är involverad i modellen och R^2Y hur stor del av Y-variationen som är modellerad. En PLS-modell kan uttryckas med en regressionslikning,

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + \epsilon$$

Koefficienterna, b_x kan användas för att exkludera givarsignaler som ger ett litet eller inget bidrag till prediktionen.

För PLS-modellerna kan också Variable Importance of Projection (VIP) (Chong & Jun, (2005)) beräknas. VIP är ett annat mått på hur viktig varje variabel i X är i modellen. Måttet handlar dels om hur viktiga variablerna är för att beskriva X, dels hur väl de korrelerar med Y.

För mer information mer om PLS se Eriksson et al., (2003), Varmuza & Filzmoser, (2008) och Johansson et al., (2020).

4.3 AVVIKELSE FRÅN MODELLENS GILTIGHETSDOMÄN

En multivariat modell definieras som ett hyperplan i processrymden. Hyperplanet har så många dimensioner/komponenter (A) som anses vara rimliga för att ge modellen en bra kvalitet. Varje observation i den ursprungliga K-dimensionella rymden kan relateras till, projiceras på, modellens hyperplan. Den observerade punkten, i K-rymden, ligger sällan i hyperplanet utan oftast något under eller över detta plan. Vinkelräta distansen från observationen till hyperplanet kallas Distance to Model (DModX). DModX beskriver alltså hur nära observationen är till det hyperplan som modellen utgör. Om detta avstånd ökar över tiden betyder det att observationerna rör sig bort från modellens giltighetsdomän, medan en liten distans betyder att observationen ligger i det område som modellen kan hantera. Även DModX kan övervakas genom konfidensintervall. Detta intervall kan beskrivas som mellanrummet mellan två skivor som ligger parallellt med modellens hyperplan och begränsar den vinkelräta distansen till planet. Gränsvärdet kallas Critical Distance to Model (DModXCrit). Observationer som har högre DModX än DModXCrit anses som avvikare.

Om observationen projiceras ned på hyperplanet (DModX sätts till noll) kan distansen mellan observationens position i hyperplanet och modellens mitt (origo)

anges med Hotelling's T2 (HT2). Positionen för en observation i modellens hyperplan ges av koordinater i planet (scores), ett värde per komponent (dimension). Måttet HT2 väger samman avståndet till modellens mittpunkt, för respektive komponent tillvärdet, HT_i^2 . Precis som för DModX kan ett gränsvärde för HT2 beräknas, kallat Critical HT2 (HT2Crit). Observationer som har högre HT2 än HT2Crit anses som avvikare.

Beräkningen av konfidensintervall utgår från sannolikheten att ett värde skall ligga inom detta intervall. I tidigare projektet i serien använde vi 99% intervall vilket rent statistiskt gav 1 falsk avvikelse på 100 värden. Därför har intervallet utsträckts till 99,9%, vilket ger ett rimligt antal falska avvikelser. Nackdelen med stora konfidensintervall är att man riskerar missa sanna avvikare, så en avvägning måste göras.

En mer ingående förklaring av DModX och HT2 finns att läsa i Eriksson et al., (2003).

4.4 ÖVERVAKNING MED PLS-MODELLER

Övervakningen med en PLS-modell kan ske på många sätt. Inom projektet har vi framför allt fokuserat på modellens residual, alltså skillnaden mellan observerat och predikerat värde. Vid framtagandet av modellen beräknas standardavvikelsen (std) för residualen över hela det historiska datasetet. Standardavvikelsen används sedan för att beräkna ett statistiskt 2-sidigt 99,9% konfidensintervall (CIYpred):

$$CIY_{pred} = \pm t(p = 0.9995) * std$$

Då många givarsignaler övervakas ökar risken för falska avvikelser. Eftersom korta avvikelser i residualen är av mindre intresse så skulle olika sätt kunna testas för att implementera en tröghet i övervakningen.

En kumulativ summering (CuSum) av residualen med dödbandsfilter har implementerats. Dödbandsfiltret är parametriserat i implementationen med varje signals standardavvikelse, std, och faktorerna cuSumDeadbandFactor (0.5) och cuSumActionlimitFactor (100). Det går att justera faktorerna efter önskat beteende (Montgomery, (2013)). I de fall som ligger till grund för denna rapport är dödbandet definierat enligt följande:

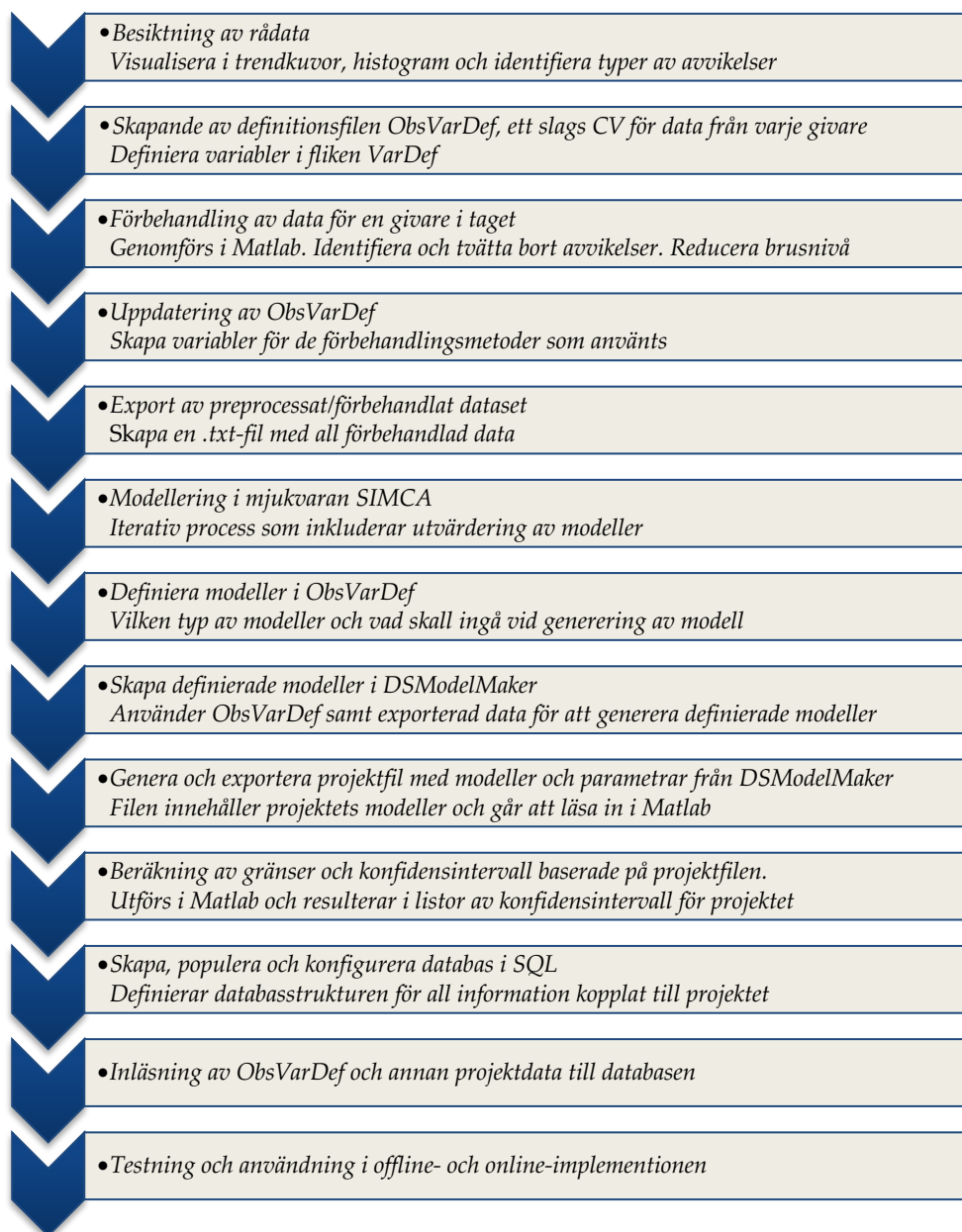
$$dödband = std / cuSumDeadbandFactor.$$

Ett gränsvärde är också implementerat för övervakningen av CuSum:

$$gränsvärde_{CuSum} = \pm std * cuSumActionlimitFactor.$$

5 Arbetsgång för multivariat modellering

I Jacobson et al., (2022) låg fallstudie I och II till grund för metodutveckling av multivariata modeller anpassade för data från en vattenkraftsdamm. Arbetsgången visas i **Fel! Hittar inte referensskälla.** nedan. Dessa metoder här implementeras för fallstudie III.



Figur 11. Arbetsgång vid modelleringsarbetet

6 Förbehandling av data för modellering

För att utveckla och implementera multivariata modeller för dammövervakning behöver data förbehandlas innan modellerna identifieras. Vid förbehandlingen av data ersätts identifierade avvikande värden från de univariata testerna med bortfall så att dessa värden inte skall påverka modellen och dess prediktionsförmåga. Data behöver alltså tvättas från onormala tillstånd eftersom modellerna skall baseras på korrekta data och användas för att indikera onormala tillstånd. Efter tvätten decimeras data, om de har kortare samplingsintervall, till samplingsintervallet 1 timme med intervallvis medelvärdesbildning, detta för att hålla antal observationer i datasetet på rimlig nivå, dvs 5 års data innebär cirka 44 000 observationer. Proceduren för tvättningen beskrivs i Jacobson et al., (2022).

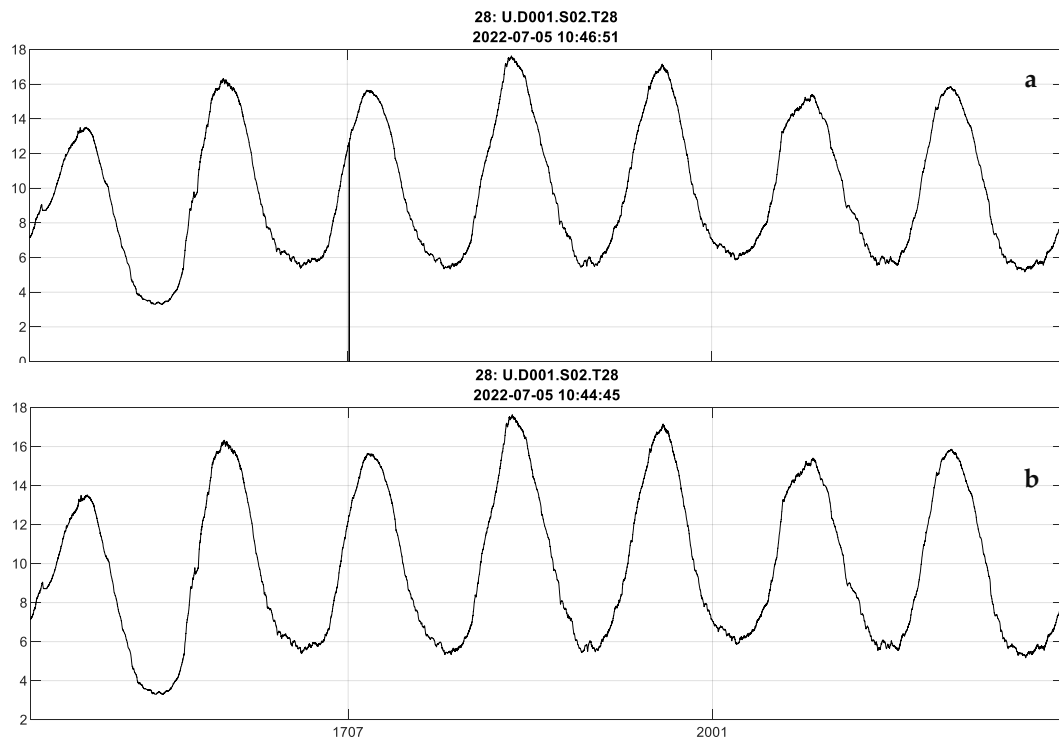
Antal fel som upptäckts i den historiska datamängden för fallstudie III är förhållandevis få till antalet. De fel som upptäckts är spikar, där en eller flera mätvärden i följd rapporterats utanför den årliga variationen och bryter mönstret i signalen, se nedan. Ett antal givare uppvisar även en hög brusnivå, vilket gör det svårt att utläsa mer information om den naturliga variationen. Ett vanligt förekommande fel är givarsignalens låga upplösning. Detta fel kommer ha en viss påverkan på den multivariata modelleringen, men påverkan är begränsad. Ingen förbehandlingsmetod har implementerats för att justera data med dålig upplösning. Detta skulle medföra att förbehandlingen justerade data till mer specifika mätvärden, för vilket evidens saknas.

Nedan visas resultat av förbehandling av ett antal givarsignaler från fallstudien. Varje figur innehåller två grafer,

- a. Rådata
- b. Data efter förbehandling

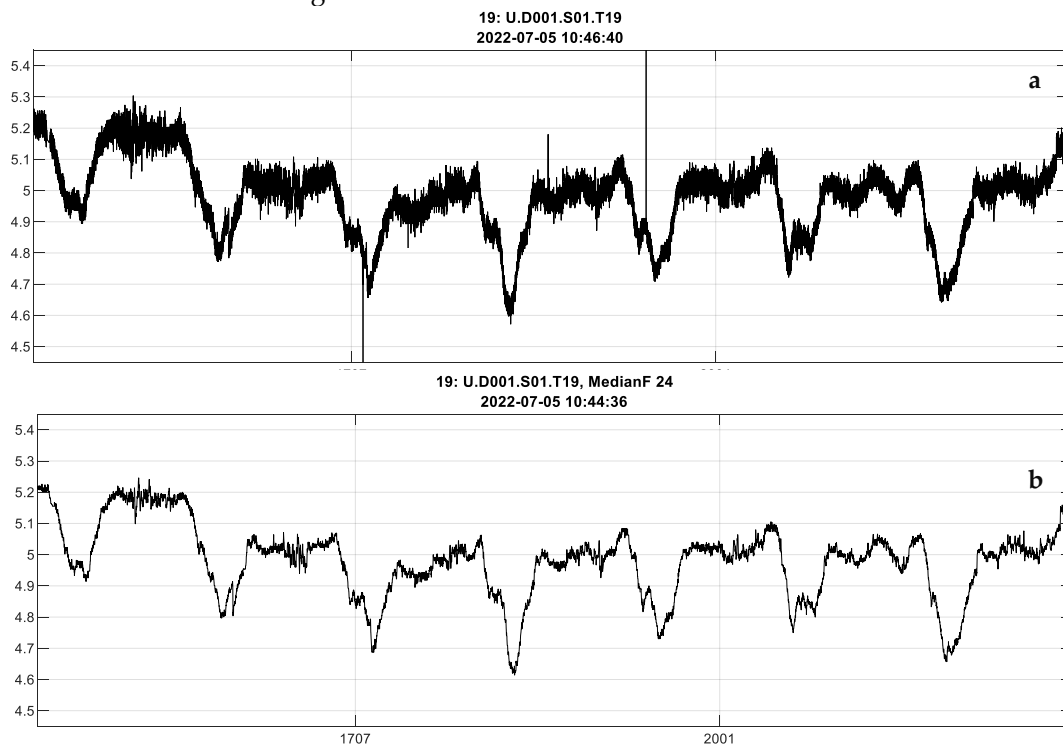
Den nedersta grafen i varje figur visar alltså resultatet efter genomförd förbehandling enligt ovan.

Figur 12 visar en givarsignal som mäter temperatur från fallstudie III. Signalen innehåller en tydlig spik i juli 2017. Denna har rensats bort i förbehandlingsresultat.



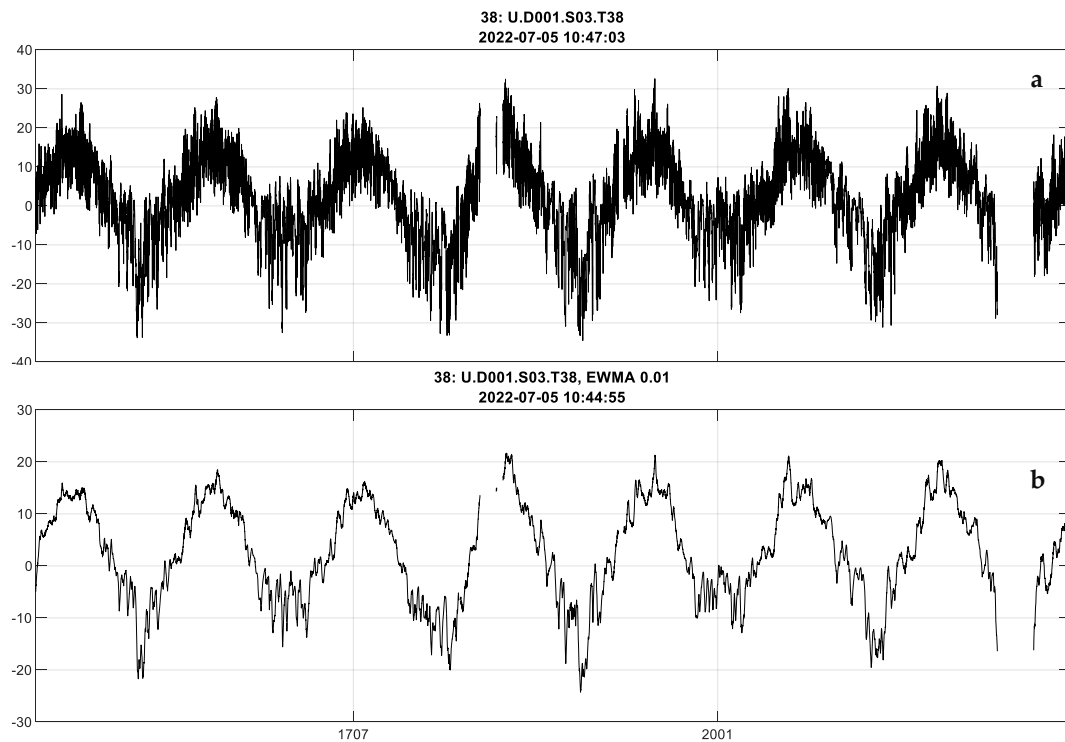
Figur 12. Figuren visar data från en temperaturgivare från fallstudie III där X-axeln beskriver tid [ååmm].

Figur 13 visar samma data som i Figur 4, data från en pendelgivare i fallstudie III. Signalen innehåller ett antal mindre spikar såväl som är något brusig och svår att avläsa. Genom tvättning av data genom differenser och filtrering ses det förbehandlade resultatet i Figur 13b.



Figur 13. Figuren visar data från en pendel från fallstudie III där X-axeln beskriver tid [ååmm]. Pendel mäter rörelse i dammkroppen längs dammens utbredning.

Figur 14a visualiserar SMHI:s data för temperaturen i regionen vilket används som utetemperatur för fallstudie III. Rådatat är relativt brusigt och svårt att avläsa. Signalen filtreras i förbehandlingen och resultatet visas i Figur 14b.



Figur 14. Här visualiserar SMHI data för utetemperatur vilket används som utetemperatur för fallstudie III. X-axeln beskriver tid [ååmm].

7 Resultatspridning

Den resultatspridning som skett har bestått av seminarier och kurser under rubrikerna nedan. Utöver detta har även konferensbidrag presenterats Fridén et al., (2023).

7.1 SEMINARIER

Ett webinarium i två entimmes delar genomfördes den 25/1 2023 med ett 10-tal åhörare på plats i Energiforsks lokaler i Stockholm, och cirka 70 via webben. Den första delen var mer översiktlig och den andra mer detaljerad.

Den första delen – "Introduktion till datadriven dammövervakning" hade följande rubriker:

Introduktion

- Varför behövs datadriven dammövervakning och vilken nytta kan den göra?
- Vad omfattar datadriven dammövervakning på en övergripande nivå
- Kort om metoder vi använt
- Praktiska resultat från implementering av datadriven dammövervakning
- Visualisering för händelser i ett ögonkast
- Visioner
- Kurs om datadriven dammövervakning

Den andra delen – "Metoder och tekniker för datadriven dammövervakning" hade följande rubriker:

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

7.2 KURS

En kurs "Datadriven dammövervakning" utvecklades under hösten 2022/våren 2023 och hölls den 9/3 och den 30/3 2023 i Energiforsks lokaler i Stockholm med cirka 8 deltagare. Inför kursstart fick deltagarna några förberedelseuppgifter för att mjuka upp de statistiska tankebanorna. Under första kursdagen presenterades en hemuppgift att arbeta med till andra kursdagen. Ett avstämningstillfälle hölls via webben den 20/3 inför den andra kursdagen. Kursen avslutades den 4/5 med en diskussion via webben.

Första kursdagen hade följande rubriker:

- Kursintroduktion
- Dataförbehandling
- Teori om multivariat maskininlärning
- Demo mjukvaran SIMCA

- Eget arbete med förberedda övningsuppgifter
- Genomgång av hemuppgiften

Andra kursdagen hade följande rubriker:

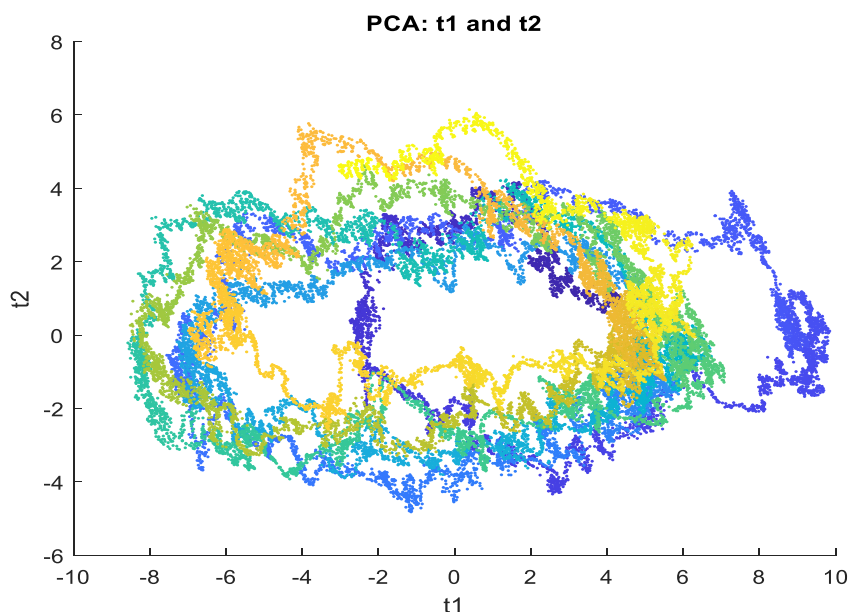
- Introduktion till dag 2
- Genomgång av hemuppgiften
- Teorirepetition
- Eget arbete med preparerade data
- Modellbaserad avvikelседetektion
- Mjukvarulösning
- Demo online

Kursen lovordades av deltagarna som ansåg den angelägen och att den borde nå många fler i branschen.

8 Multivariat modellering - resultat

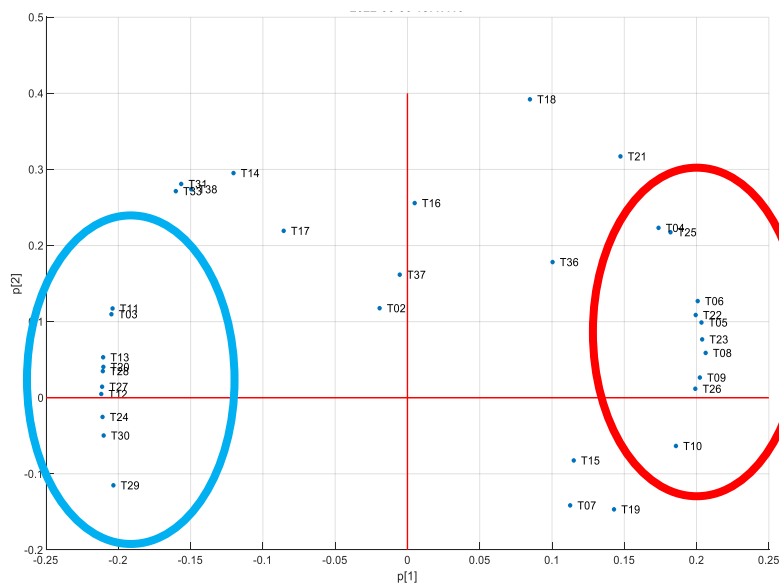
8.1 PRINCIPALKOMPONENTANALYS, PCA

För att få en överblick av dammens beteende för fallstudie III har PCA-modellering använts. I Jacobson et al., (2022) beskrevs att de flesta givare i en dammkropp uppvisar ett cykliskt beteende med ett års frekvens. Detta beroende på att den årliga vädervariationen har stor påverkan på dammens givare. Detta medför att dammdatas dominanta beteende beror av årets cykliska väder och således är starkt korrelerad med tiden på året. I Figur 15 visualisera en scoreplot, dammobservationernas dominanta variabilitet över tiden. X-axeln utgörs av analysens första principalkomponent (t1) och Y-axeln den andra principalkomponenten (t2). De data som visualiseras är 5 års observationer från fallstudie III, från 29 april 2015 till 30 juni 2021. Figuren visar alltså dammens dominanta datastruktur i två dimensioner. Det går att utläsa fem ovaler i figuren, varav fyra ligger relativt nära varandra och en oval (blå) ser ut att vara förskjuten i t1. Varje oval representerar ett år av data och den årliga cykeln är orsaken till det ovala beteendet. Ovan har påvisats att första årets data skiljer sig från följande år (första årets data har därför tagits bort vid utvecklandet av modellerna) och det är det första året som här representeras med blå färg. Med hjälp av denna scoreplot kan vi urskilja att dammens datanormalläge förändrats efter första årets data.



Figur 15. Figuren visualiserar en scoreplot över t1 och t2 för fallstudie III. Data består av 5 års data där varje observation representeras med en prick. Axlarna representerar t1 och t2, alltså de två första principalkomponenterna i PCAn.

Scoreplotten, ovan, visar hur datas observationer relaterar till varandra. På analogt sätt visar loadingplotten hur variablerna i data relaterar till varandra. Variabler som ligger nära är positivt korrelerade, de som ligger diametralt motsatta är negativt korrelerade. Variabler som ligger långt från origo är bättre modellerade än de som ligger nära, se Figur 16.

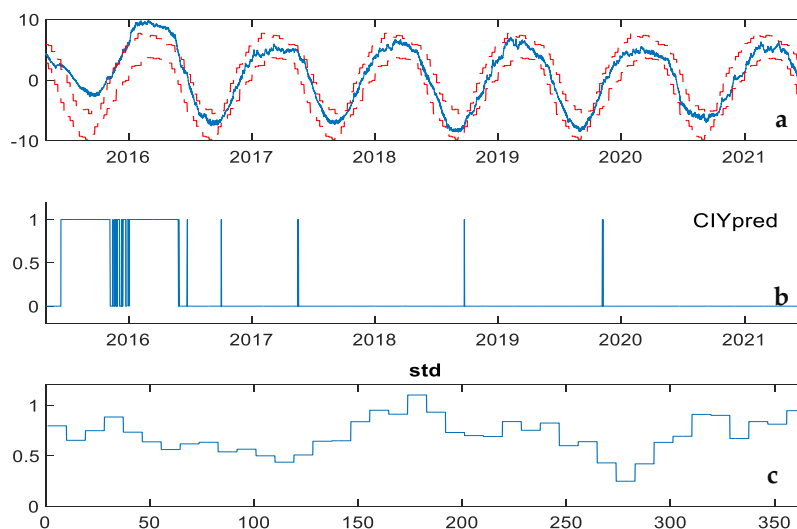


Figur 16. Figuren visar loadingplotten för PCA-modellen för fallstudie III. Till vänster (blå) samlas i huvudsakligen temperaturgivare, medan till höger (röd) har sprickgivare grupperats. Temperatursignalerna är alla nära korrelerade, så också spricksignalerna. Därtill är temperatur negativt korrelerat med sprickor, dvs hög temperatur korrelerar med små sprickor och tvärtom.

För att förtydliga denna förskjutning i t1 plottas t1 och t2 separat i Figur 17 och Figur 18. Figuren baseras på samma data men är plottade som tidserier från 29 april 2015 till 30 juni 2021. I figuren visas tre olika trender.

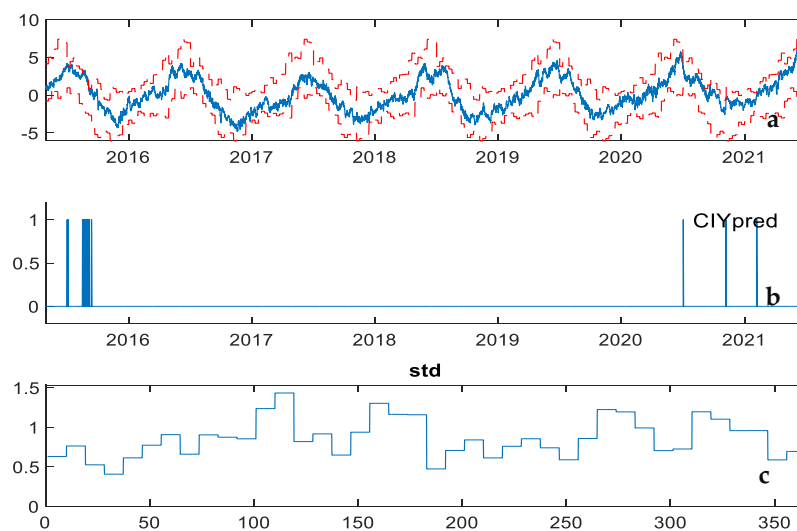
I översta figuren (a) visas t1 (blå) och ett beräknat konfidensintervall (streckad röd) baserat på metoden presenterad i Jacobson et al., (2022). Den mittersta figuren visualiserar larmfunktioner som aktiveras när mätvärdet går utanför konfidensintervallet. Den blå översta linjen är den faktiska larmsignalen som har en viss tidsfördröjning innan den aktiveras (observationen måste rapporteras utanför konfidensintervallet under en viss tid innan larmet aktiveras). Den nedersta trenden (c) visar den beräknade standardavvikelsen som ligger som grund till konfidensintervallet. X-axeln skiljer sig i denna trend då standardavvikelsen beräknas per år. X-axel är dag på året och går till 365.

I Figur 17 går det att se den förskjutning i t1 som diskuterades kring Figur 15. T1 är tydligt högre under första året än vid övrig tid i tidserien, Man kan även se detta från rådata för utvalda givare i Figur 1. Vi ser även att larmfunktionen som utvecklades i Jacobson et al., (2022) skulle fångat detta beteende.



Figur 17. Här visas t_1 från PCA baserad på samtliga givarsignaler i fallstudie III. Figuren består av tre delbilder. a) t_1 (blå), konfidensintervall (streckad röd), tid på x-axel. b) larmfunktioner som aktiveras beroende på om observerad t_1 befinner sig utanför konfidensintervallet. c) en trend över standardavvikelsen för olika tidsavsnitt över året.

I Figur 18 visar motsvarande figur för t_2 . Som anats redan vid studerandet av Figur 15 ser vi inget tydligt avvikande beteende i t_2 , vilket bekräftas i Figur 18.



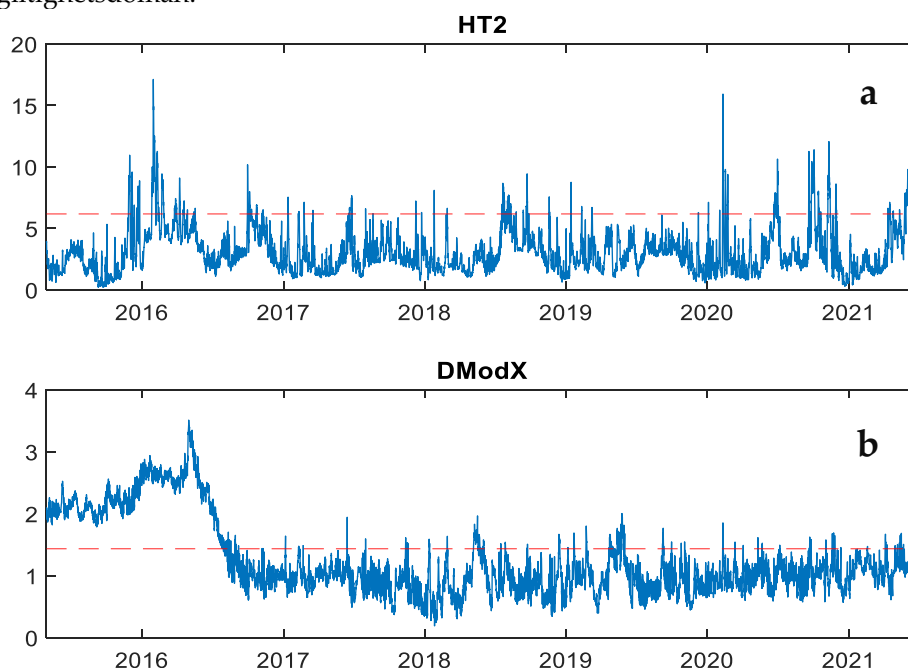
Figur 18. Figuren visar t_2 från PCA för samtliga givarsignaler i fallstudie III. Figuren består av tre plottar. a) t_2 (blå), konfidensintervall (streckad röd), tid på x-axel. b) larmfunktioner som aktiveras beroende på om observerad t_2 befinner sig utanför konfidensintervallet. c) en trend över standardavvikelsen för olika tidssekvenser över ett år.

I Figur 19 visualiseras distanserna till modellens giltighetsdomän, Hotelling's T^2 och DModX. HT^2 ser ut att ha ett ganska likartat beteende över hela tidsperioden.

Detta är något förvånande då första året uppvisat ett avvikande beteende i t_1 . Detta förväntades ha ett genomslag även på Hotelling's T^2 , vilket inte är tydligt i Figur 19a.

I Figur 19b visas ett tydligt avvikande första året, vilket tidigare identifierats vid analys av enskilda givare. Att DModX avviker tyder på att processen befinner sig

långt från det hyperplan som modellen utgör och därav utanför modellens giltighetsdomän.



Figur 19. Figuren visualiserar tidsserier för Hotelling's T2 och Distance-to-Model (DModX). Det två streckade röda linjerna representerar Hotelling's T^2_{Crit} och DModXCrit.

8.2 PARTIAL LEAST SQUARES REGRESSION, PLS

För fallstudie III har en annan princip vid framtagandet av PLS-modeller använts. Då det kunnat fastslås att kvaliteten på data generellt sett är god i jämförelse med tidigare fallstudier har ett stort antal modeller utvecklats, denna gång utan att försöka maximera modelleringsförmågen genom variabelselektion. En signal i taget har valts för att predikteras (utgöra Y-rymden) och resten av dammens signaler har fått utgöra X-rymden, för samtliga PLS-modeller. I och med detta har det varit enkelt att utveckla många modeller för fallstudie III och det finns en modell för varje givare i dammen. I Tabell 3 listas samtliga modeller för fallstudie III.

Modellens prediktionsförmåga (Q^2) är mycket hög för större delen av modellerna, vilket är bra. Även om modellerna innehåller många variabler i X, ingen variabelselektion, lyckas ändå prediktionen av Y mycket väl. En nyckel till detta är att kvaliteten på data är bra. I Tabell 2 specificeras att datasetet från fallstudie III består av över 54 000 observationer.

För modellerna i Tabell 3 används endast 43 824 observationer. Detta beror på att det första årets observationer inte använts vid modellerandet. Figur 19b visar DModX för PCA-modellen för hela dammen och där avviker det första året, som exkluderats vid byggandet av modellerna, från det övriga datasetet.

Då X-rymden för PLS-modellerna också består av många givare med det angreppssätt som används för fallstudie III kommer detta beteende troligtvis avspeglas i DModX för PLS-modellernas första år.

Tabell 3. Tabellen listar alla modeller, och dess godhetstal, för fallstudie III.

Modell- namn	Typ av modell	Antal X	Antal Y	Antal observa- tioner (N)	R ² X	R ² Y	Q ²	Namn på Y vari- abeln
Mdl01	PCA	34		43 824	0,727			
Mdl02	PLS	43	1	43 824	0,853	0,604	0,475	U.D001.S01.T02
Mdl03	PLS	43	1	43 824	0,815	0,989	0,988	U.D001.S01.T03
Mdl04	PLS	43	1	43 824	0,722	0,969	0,961	U.D001.S01.T04
Mdl05	PLS	43	1	43 824	0,773	0,989	0,98	U.D001.S01.T05
Mdl06	PLS	43	1	43 824	0,757	0,988	0,979	U.D001.S01.T06
Mdl07	PLS	43	1	43 824	0,759	0,902	0,789	U.D001.S01.T07
Mdl08	PLS	43	1	43 824	0,836	0,977	0,968	U.D001.S01.T08
Mdl09	PLS	43	1	43 824	0,759	0,95	0,925	U.D001.S01.T09
Mdl10	PLS	43	1	43 824	0,76	0,979	0,958	U.D001.S01.T10
Mdl11	PLS	43	1	43 824	0,795	0,991	0,99	U.D001.S01.T11
Mdl12	PLS	43	1	43 824	0,722	0,994	0,993	U.D001.S01.T12
Mdl13	PLS	43	1	43 824	0,557	0,988	0,987	U.D001.S01.T13
Mdl14	PLS	43	1	43 824	0,556	0,784	0,769	U.D001.S01.T14
Mdl15	PLS	43	1	43 824	0,759	0,392	0,359	U.D001.S01.T15
Mdl16	PLS	43	1	43 824	0,784	0,572	0,304	U.D001.S01.T16
Mdl17	PLS	43	1	43 824	0,787	0,544	0,4	U.D001.S01.T17
Mdl18	PLS	43	1	43 824	0,787	0,949	0,937	U.D001.S01.T18
Mdl19	PLS	43	1	43 824	0,772	0,78	0,74	U.D001.S01.T19
Mdl20	PLS	43	1	43 824	0,783	0,985	0,982	U.D001.S02.T20
Mdl21	PLS	43	1	43 824	0,557	0,987	0,982	U.D001.S02.T21
Mdl22	PLS	43	1	43 824	0,76	0,978	0,963	U.D001.S02.T22
Mdl23	PLS	43	1	43 824	0,753	0,987	0,977	U.D001.S02.T23
Mdl24	PLS	43	1	43 824	0,755	0,989	0,988	U.D001.S02.T24
Mdl25	PLS	43	1	43 824	0,557	0,97	0,961	U.D001.S02.T25
Mdl26	PLS	43	1	43 824	0,761	0,979	0,964	U.D001.S02.T26
Mdl27	PLS	43	1	43 824	0,76	0,99	0,988	U.D001.S02.T27
Mdl28	PLS	43	1	43 824	0,557	0,988	0,986	U.D001.S02.T28
Mdl29	PLS	43	1	43 824	0,557	0,989	0,988	U.D001.S02.T29
Mdl30	PLS	43	1	43 824	0,722	0,985	0,984	U.D001.S02.T30
Mdl31	PLS	43	1	43 824	0,556	0,983	0,98	U.D001.S02.T31
Mdl32	PLS	43	1	43 824	0,773	0,983	0,98	U.D001.S02.T33
Mdl33	PLS	43	1	43 824	0,774	0,743	0,686	U.D001.S02.T36
Mdl34	PLS	43	1	43 824	0,785	0,132	-0,05	U.D001.S03.T37
Mdl35	PLS	43	1	43 824	0,184	0,923	0,901	U.D001.S03.T38

Nedan visas ett urval av modellerna listade i Tabell 3. De modeller som presenteras har valts då deras prediktionsförmåga är antingen mycket hög eller låg. Varje modell representeras med tre bilder. Den första bilden (a) visar

- Svart - Observerat värde, efter förbehandling
- Röd – Predikterat värde från modellen
- Blå – Konfidensintervallen inom vilka observerade värdet bör befinna sig

Den andra bilden (b) visar modellens residual, alltså skillnaden mellan observerat och predikterat värde. Även denna bild har två blå linjer som markerar samma konfidensintervall som tidigare. Men då residualen rör sig runt noll och konfidensintervallet är konstant över tiden representeras det som två horisontella linjer.

Den sista bilden (c) visar de övervakningsfunktioner som är implementerade. Precis som för univariata metoderna representeras de som binära signaler över hela tidsserien med en offset i Y-led för att förtydliga varje övervakningssignals beteende. De övervakningsfunktioner som finns med i denna figur är

- Övervakning av principalkomponenter, var för sig (tx)
- Negativ kumulativ summering (LowerCuSum)
- Positiv kumulativ summering (HighCuSum)
- Residual, predikterat Y-värde – observerat Y-värde
- HT2 och DModX

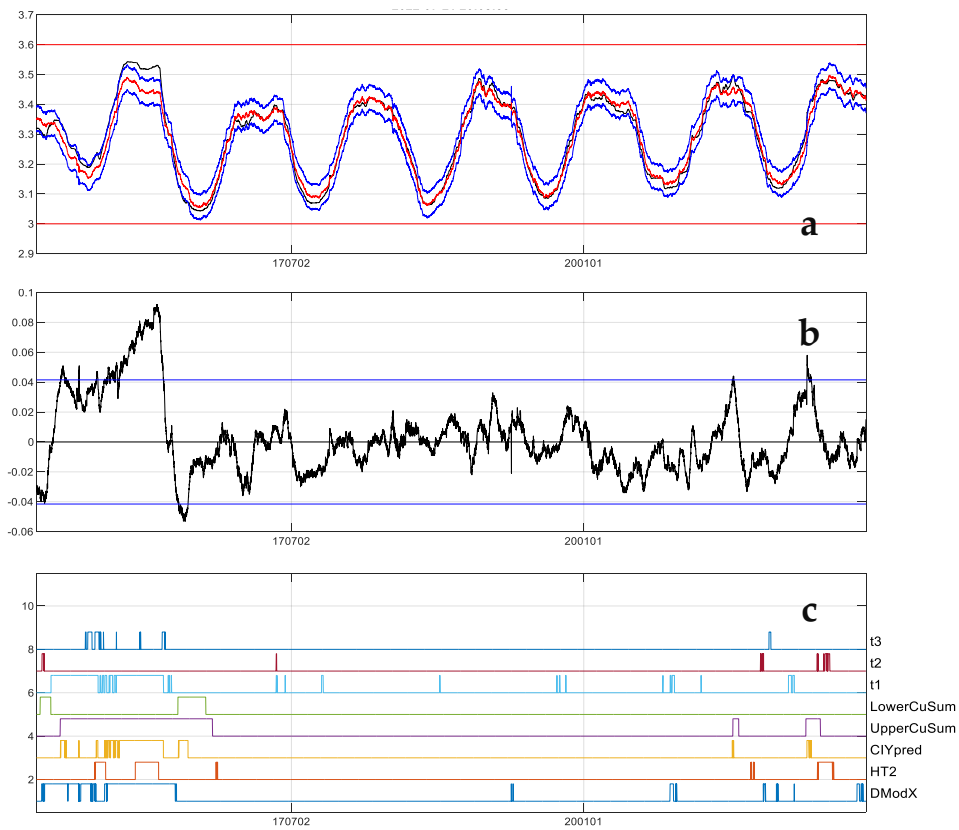
I Figur 20 visualiseras Mdl06, vilken predikterar en sprickgivare i dammen.

För denna modell är prediktionsförmågan mycket hög ($Q^2 = 0,979$) vilket medför att residualen, skillnaden mellan prediktion och observation, är liten.

Detta resulterar också i att konfidensintervallet som beräknas baserat på statistik blir mycket snävt kring det predikterade värdet. Under det första året, vilket tidigare identifierats som avvikande, går att se att residualen är betydligt större än för övriga delar av tidserien. Detta går att se både i Figur 20a & Figur 20b.

I Figur 20a befinner sig det observerade värdet till stor del över konfidensintervallet. I Figur 20b syns avvikarna tydligt när residualen plottas med konfidensintervallet.

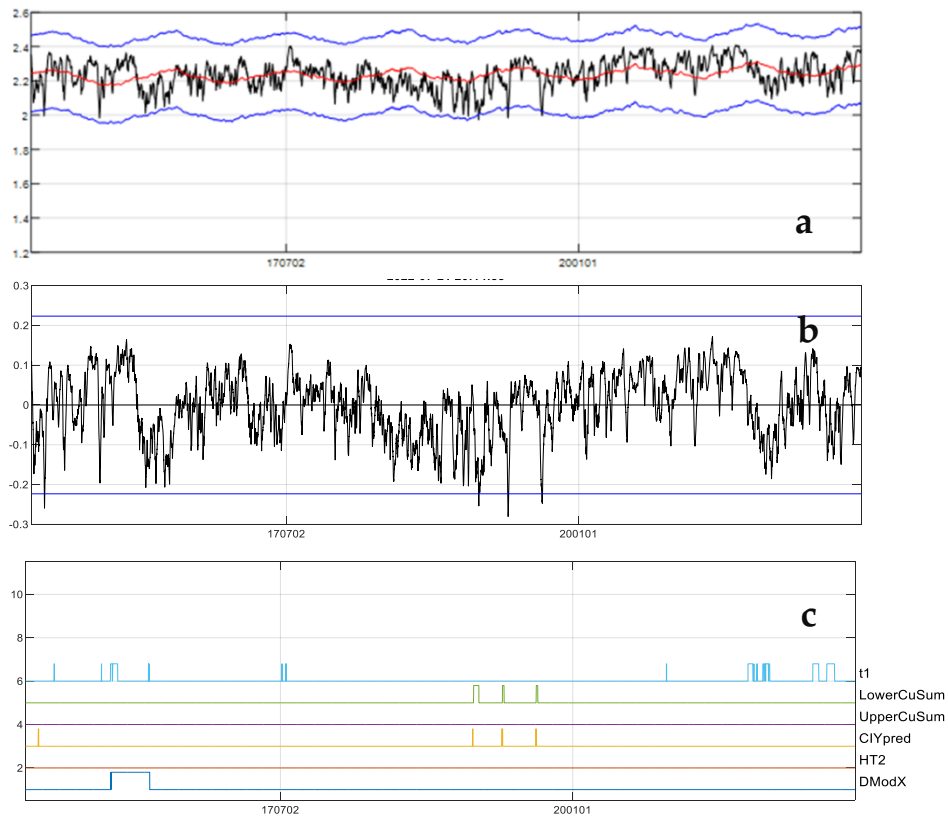
Det avvikande beteendet speglas också i Figur 20c där intensiteten i funktionernas markeringar är höga. Notera även att DModX går hög under en stor del av denna period vilket förväntades baserat på resonemanget ovan.



Figur 20. Figuren visar Mdl06 vilket predikterar signalen U.D001.S01.T06 vilket är en sprickgivare.

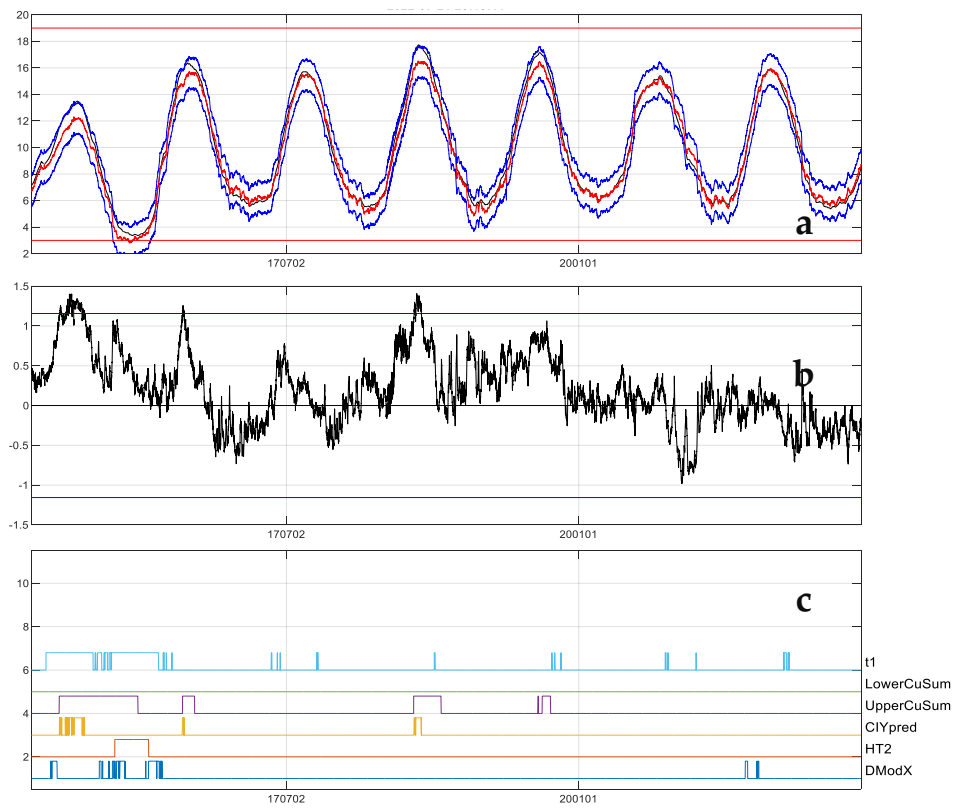
Figur 21 visar data från modell Mdl34. Prediktionsförmågan för denna modell är negativ ($Q^2 = -0,05$), vilket betyder att modellen är värdelös. Modellen syftar till att prediktera ÖVY, vilken inte korrelerar alls med variabiliteten i de närvarande X-variablerna. Därmed kan man konstatera att ÖVY är en isolerad parameter, som lever sitt eget liv, till skillnad från de andra dammsignalerna vilka i hög grad har andra samkorrelerande signaler. Detta är intressant eftersom vi för en annan damm konstaterat samvariation med porttrycksgivare och här kan få en indikation om det låg något i detta. I Figur 21a ses tydligt att prediktionen inte följer observationens mindre rörelser, samt att den predikterade årliga variabiliteten inte alls korrelerar med de observerade värdena.

Det går att övervaka en givare med en modell med låg prediktionsförmåga på samma sätt som med en modell med hög Q^2 . Men då gränserna för övervakningen är statistiskt baserade kommer dessa att bli mer generösa, vilket medför att det blir svårt att upptäcka om signalen faktiskt avviker från sitt normala beteende.



Figur 21. Figuren visualiserar Mdl34 vilket predikterar signalen U.D001.S01.T37 vilket är givaren för ÖVY (övre vattenytan).

I Figur 22 visar data från Mdl27 vilken predikterar en temperaturgivare som mäter vattentemperatur. Denna modell har en mycket hög prediktionsförmåga ($Q^2 = 0,988$) och det kan ses i Figur 22a att prediktionen följer observationen väl. Även här är andelen observationer med högt DModX mer frekvent under det första året av data. Annars ser övervakningen ut att fungera bra med undantag för övervakningen av t1, vilken avviker från konfidensintervallet i slutet av flera av åren.



Figur 22. Denna figur visar Mdl27 vilken predikterar signalen U.D001.S01.T27 som är en temperaturgivare som mäter vattentemperatur.

9 Modellunderhåll

Vid skapandet av en modell används historiska data som träningsmaterial för modellen. Dessa historiska data visar hur dammens normala tillstånd sett ut historiskt och antas kunna användas för att beskriva dammens tillstånd även fram i tiden. Modellen förutsätter alltså att dammens grundförutsättningar förblir desamma.

Om drastiska förändringar sker i dammen, som till exempel förändringar i hur kraftverket drivs eller vid en omorganisation av givarövervakningen, kommer dammens datagrundtillstånd troligtvis förändras. Detta medför att modellen inte längre är giltig. En annan orsak till detta skulle kunna vara vissa år med betydligt högre medeltemperatur än vad som förekommit historiskt. Detta är något som kan påverkas av jordens naturligt varierande klimat såväl som klimatförändringar.

Detta betyder att tillförlitligheten för modeller som tagits fram naturligt kommer att minska med tiden och detta kan uttryckas som att modeller har en halveringstid. Hur lång en sådan halveringstid skulle vara är omöjligt att säga då många (icke definierade) faktorer spelar in på modellens giltighet.

I och med modellernas egenövervakning av deras giltighet, DModX och HT², går det faktiskt att avläsa när en modell bedömer sig själv som icke giltig i förhållande till inkommande data. Med tiden kommer frekvensen när dessa överskrider sina kritiska gränser öka, vilket tyder på att tillståndet i dammen förändrats gentemot det när modellen skapades.

För att modellerna skall uppfylla sitt syfte behöver de multivariata modellerna således underhållas. Vid de tillfällen då modellerna nått en gräns av icke giltighet som överskrider en viss nivå då behöver modellerna beräknas om på ett uppdaterat historiskt dataset. Detta kan inkludera de historiska data som användes vid den ursprungliga utvecklingen. Ett annat alternativ, som är fördelaktigt om tillståndet förändrats baserat på faktiskt förändringar i dammen eller övervakningen, är att låta det historiska datasetet i stället starta efter att denna förändring skett.

9.1 UTFÖRT MODELLUNDERHÅLL

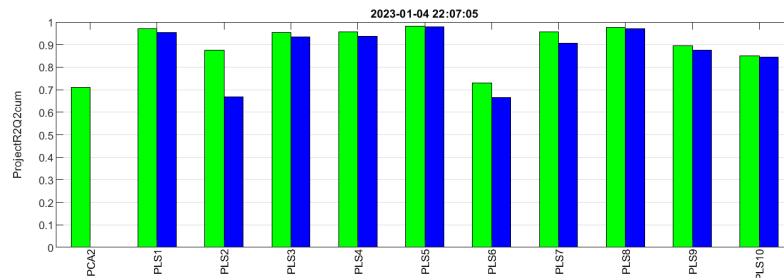
Sedan modellerna för fallstudie I och II identifierades har nya data tillkommit och gränsvärden och modellerna har reviderats.

9.2 FALLSTUDIE I (VATTENFALL)

Gränser och modeller är initialt baserade på data från perioden 2015-01-01 – 2021-09-07. Vilka signaler som då ingick i PLS-modellerna är grundade i diskussioner med Vattenfall om olika avrinningsområden för dammen.

Vattenfallsmodellerna är 11, varav en är en översikts-PCA och de övriga är PLS-modeller för utvalda enskilda signaler.

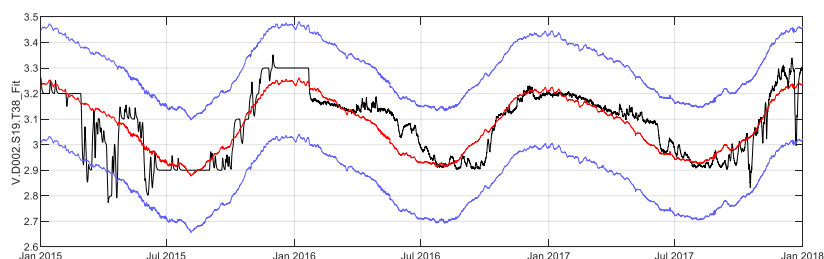
Den senaste modelluppdateringen påbörjades i april 2022 och avslutades i september 2022. De senaste gränsvärdesjusteringarna gjordes i september. Modelleringsresultaten framgår översiktligt i figuren nedan.



Figur 23. Anpassningsgrad (grön) och prediktionsförmåga (blå) för de senaste VF-modellerna.

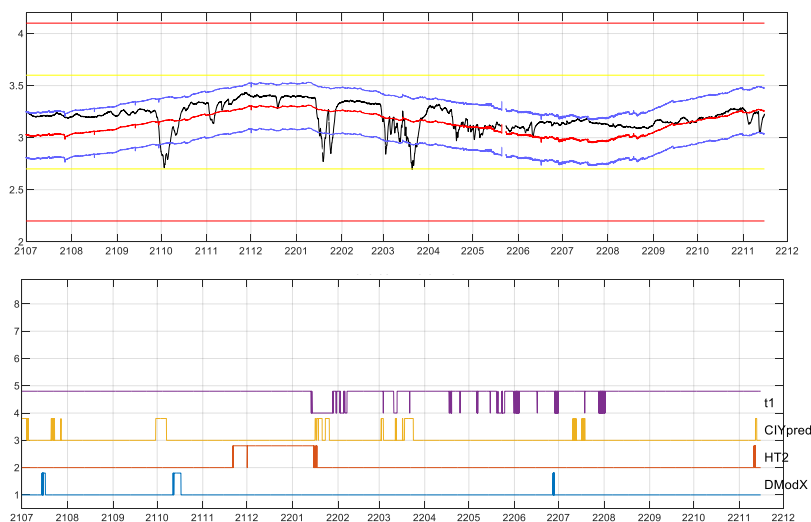
I Figur 23 kan vi se att sex av modellerna är mycket bra PLS (1, 3–5 och 7–8), två bra PLS (9 och 10) och två halvbä PLS (2 och 6).

Som exempel på modelleringsresultat visas modeller för temperatur, nivå och flöde. En modell, PLS6 (temperatur), är bra och flera modeller, såsom PLS1 (temperatur), PLS8 (flöde) samt PLS7 (nivå), är mycket bra.



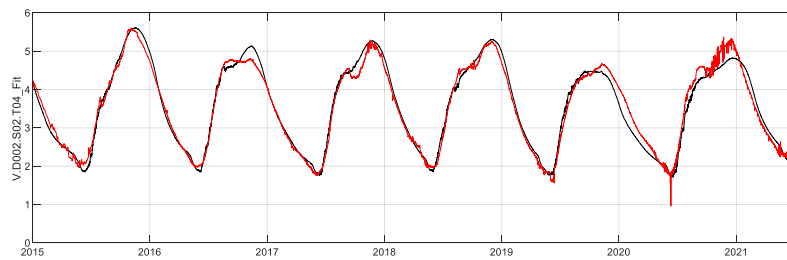
Figur 24 Predikterad (röd), från modellen PLS6, observerad signal (svart) och konfidensintervall (blå) för V.D002.S19.T38, en temperatur. Modellen har en prediktionsförmåga Q2 på 0.67 en bra modell. Formatet på x-axeln är ÅÅÅÅ.

Vi ser i Figur 24 att den observerade signalen och predikterade överensstämmer bättre mot den senare halvan av kalibreringsperioden. Man kan se att den senare halvan av den observerade signalen har ett beteende som är mer troligt för en temperaturmätning. Detta då signalen i den första halvan innehåller hopp och är periodvis frusen.



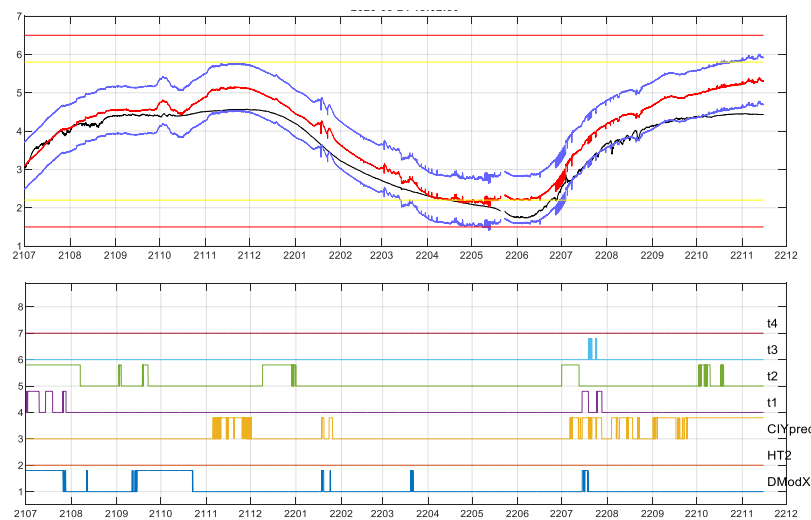
Figur 25. Övre figuren visar predikterad (röd) för PLS6, observerad signal för V.D002.S19.T38 tillsammans med konfidensintervall (blå) för nya data (2107–2211). Den nedre figuren visar modellhändelserna. Formatet på x-axeln är ÅÅMM.

Vi ser i Figur 25 övre grafen att modellen verkar fortfarande vara användbar trots att den tränades på data mellan 201501-201801. Men det är på gränsen eftersom t1 ligger som detekterad händelse under huvuddelen av tiden i nedre grafen av Figur 25. Noterbart är dock att vi har få händelser i DModX, HT2 och CIYpred.



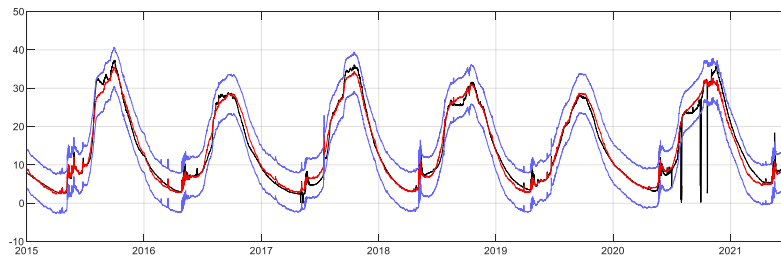
Figur 26. Predikterad (röd), från modellen PLS1, observerad signal (svart) för V.D002.S02.T04, en temperatur. Modellen har en prediktionsförmåga Q^2 på 0.95, en mycket bra modell. Formatet på x-axeln är ÅÅÅÅ.

I Figur 26 ser vi en mycket god överensstämmelse mellan predikterad (röd) och observerad signal (svart).



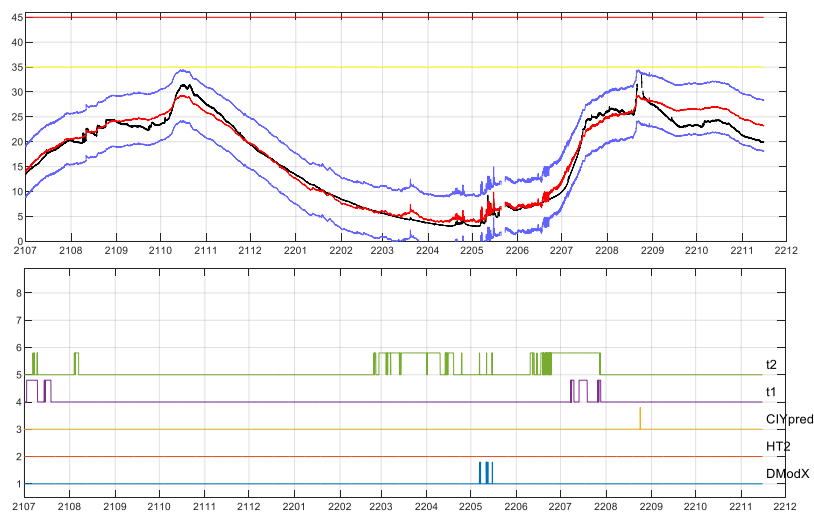
Figur 27. Övre figuren visar predikterad (röd) för PLS6, observerad signal (svart) för V.D002.S02.T04 tillsammans med konfidensintervall (blå) för nya data (2107–2211). Den nedre figuren visar modellhändelserna. Formatet på x-axeln är ÅÅMM.

I övre delen av Figur 27 syns att den observerade signalen underskrider det nedre konfidensintervallet i tiden mellan 2111–2112 samt 2207–2210. Detta fångas även i Figur 27 nedre delen av CIYpred. Där kan vi även se modellhändelser för DModX för perioden 2107–2111.



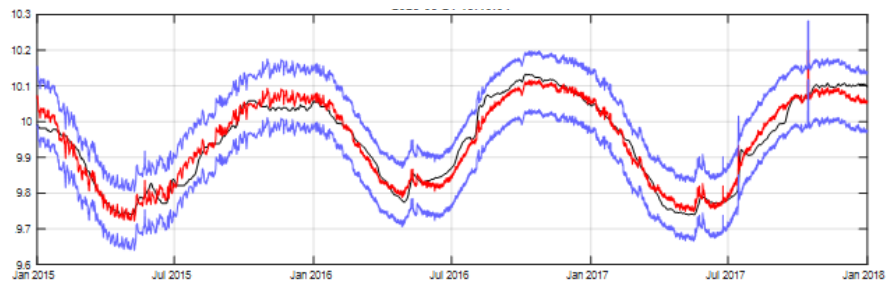
Figur 28. Predikterad (röd), från modellen PLS8, observerad signal (svart) och konfidensintervall (blå) för V.D002.S34.T70, en flödesgivare. Modellen har en prediktionsförmåga Q2 på 0.97 en mycket bra modell.

Även här ses i Figur 28 vi en mycket god överensstämmelse mellan observerad och predikterad signal.



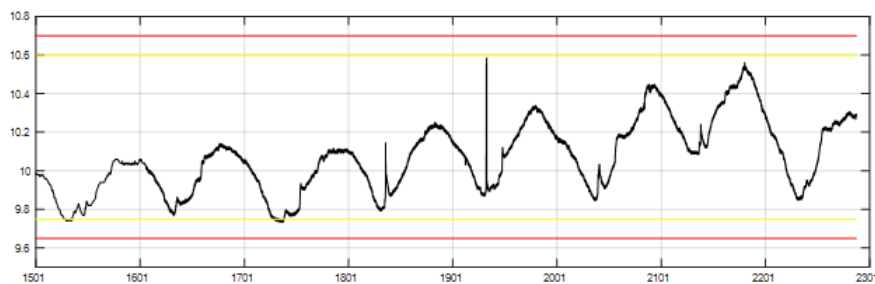
Figur 29. Övre figuren visar predikterad (röd) för PLS6, observerad signal (svart) för V.D002.S02.T04 tillsammans med konfidensintervall (blå) för nya data (2107–2211). Den nedre figuren visar modellhändelserna.

I nedre delen av Figur 29 indikeras få signalhändelser för dessa nya data.



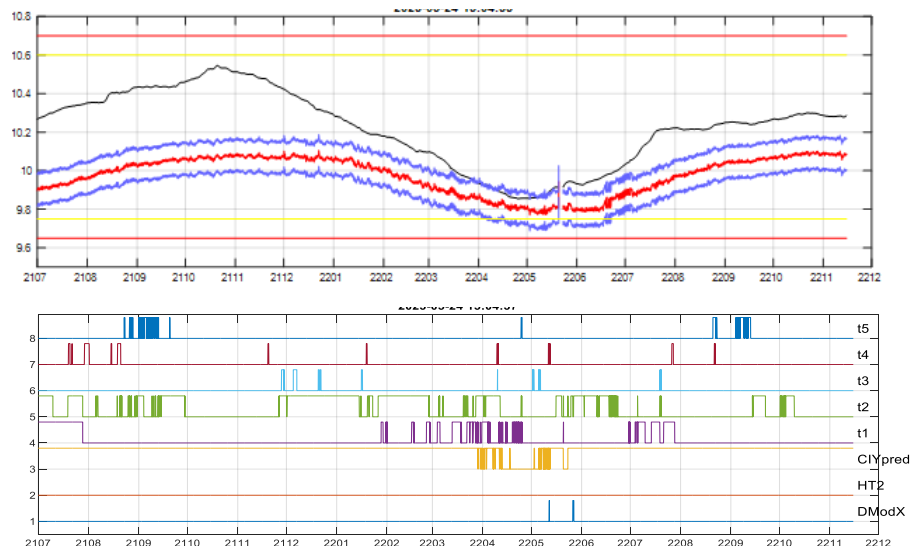
Figur 30. Predikterad (röd), från modellen PLS7, observerad signal (svart) och konfidensintervall (blå) för V.D002.S18.T35, en . Modellen har en prediktionsförmåga Q^2 på 0.91 en mycket bra modell.

Det ser i Figur 30 som förväntat ut att den predikterade signalen följer den observerade signalen mycket väl. Bedömningen av perioden mellan 201806 och 202106 är att givaren var felaktig mellan dessa datum (Figur 32 övre delen). Därför uteslöts data under denna period vid kalibrering.



Figur 31. Observerad signalvärde för V.D002.S18.T35 för perioden 1501-2211.

Vi kan se i Figur 31 att medelvärde (1501–1801) för V.D002.S18.T35 ungefär varit 9.69, vilket kan jämföras med (1801–2211) då medel varit cirka 10.01.

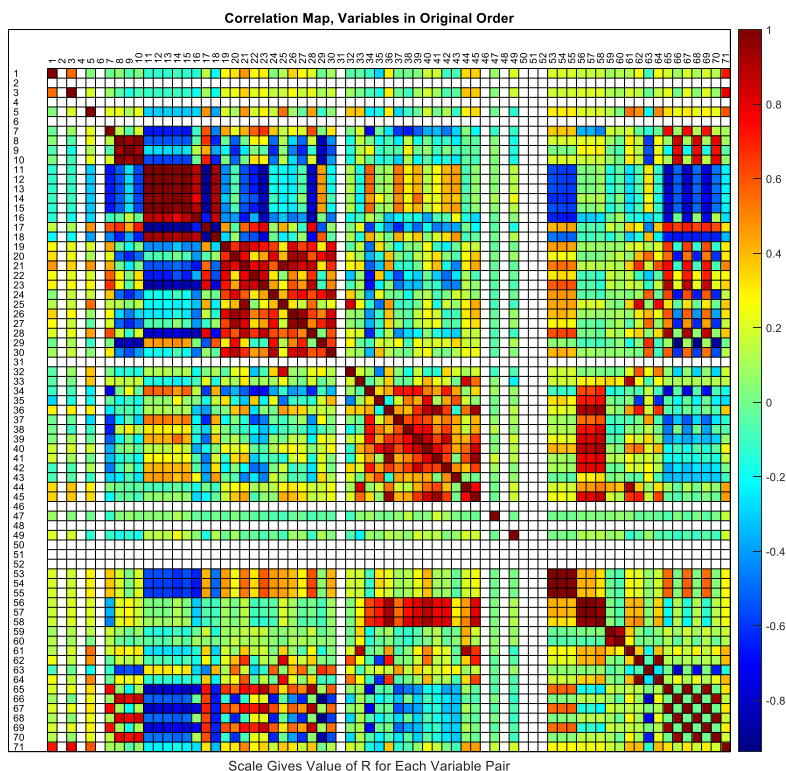


Figur 32 Övre figuren visar predikterad (röd) för PLS7, observerad signal (svart) för V.D002.S18.T35 tillsammans med konfidensintervall (blå) för nya data (2107–2211). Den nedre figuren visar modellhändelserna.

Vi kan i den övre grafen i Figur 32 se att observerade trenden (svart) ligger ovanför konfidsintervallen (blå) och avviker kraftigt från predikterad. I den nedre grafen ser vi att signalhändelsen för CIYpred är aktiverad i stort sett hela tiden från 2107–2211, undantaget mellan 2204 och 2205. Detta är en klar indikation på att man måste se över givaren och även revidera modellen.

9.3 FALLSTUDIE II (STATKRAFT)

Statkraftsdata kompletterades med väderdata och modellerna reviderades på förslag från Statkraft i december 2022. För uppdatering av modellerna hos Statkraft bad vi dem välja vilka Y- respektive X-variabler som skulle ingå i modellerna. De valde ut 14 Y-variabler att modellera. Som hjälp för modellkonfigureringen tog vi fram en korrelationsmatris som visar hur datasetets 71 variabler korrelerar, Figur 33.

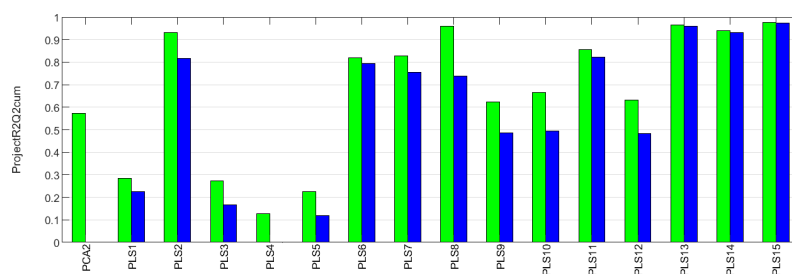


Figur 33. Parvisa korrelationen, R, mellan datasetets 71 signaler. Positiv korrelation är röd och negativ är blå. Skalan till höger visar hela R-skalan.

Statkraftsmodellerna är 16 till antalet, varav en är en översikts-PCA och de övriga är PLS-modeller för de utvalda signalerna, varav det för en signal gjordes två PLS-modeller.

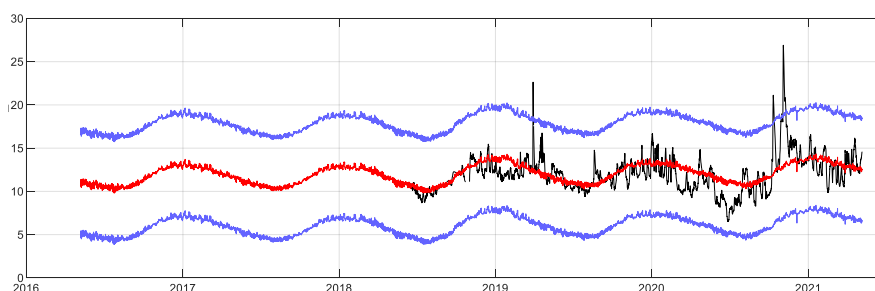
Den senaste modelluppdateringen gjordes 2022-12-22 och de senaste gränsvärdesjusteringarna gjordes 2022-12-27. Resultaten diskuterades sedan med Statkraft i januari 2023. Modelleringsresultaten framgår översiktligt i figuren nedan.

Modellerna har tränats/kalibrerats på data från perioden 201604 till 202106 (Format ååmmdd). Utvärdering har gjorts med nya data (202107 – 202210) tillsammans med en kortare del av kalibreringsdata (202101 – 202106).



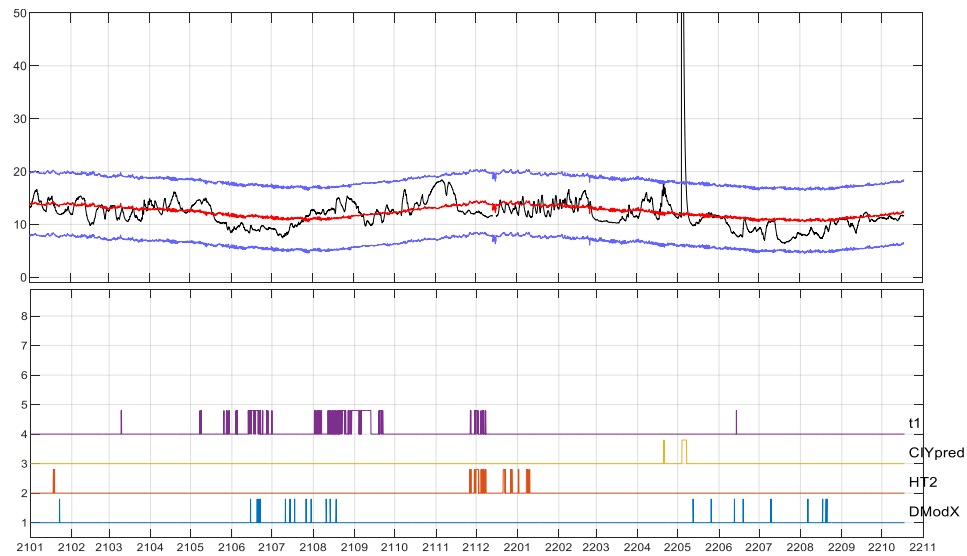
Figur 34. Anpassningsgrad (grön) och prediktionsförmåga (blå) för de senaste Statkraftsmodellerna.

Tre av modellerna är mycket bra (PLS13–15), fyra bra (PLS2, 6–8), tre mindre bra (PLS9, 10, 12) och fyra undermåliga (PLS 1, 3–5), se Figur 34 Vi har valt ut modellerna PLS1-5, 8 och 10, vilka diskuteras i denna ordning i texten som följer.



Figur 35. Predikerad (röd) från modellen PLS1 och observerad signal (svart) med konfidensintervall (blå) för S.D001.S27.T53 som är ett Thomsonöverfall med beräknat flöde i l/min. Modellen har en prediktionsförmåga Q2 på 0.26, något som innebär att modellen anses som en undermålig modell.

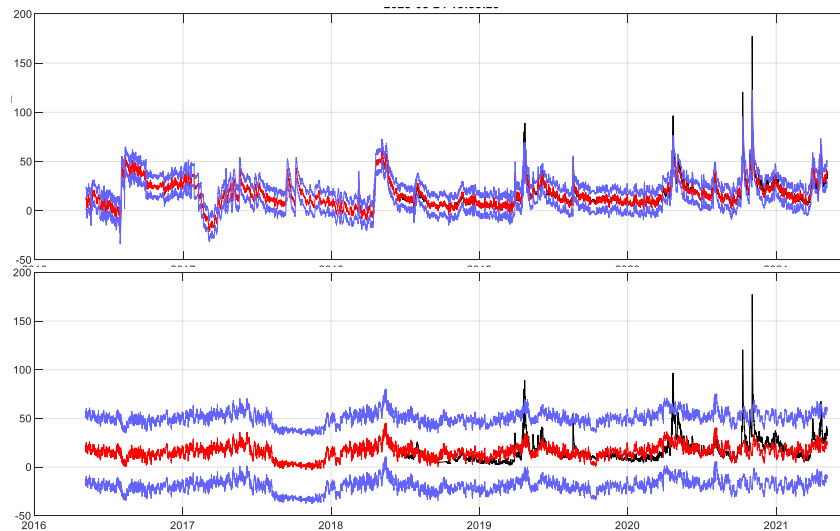
Från Figur 35 visas flödes hastigheten från ett Thomsonöverfall med beräknat flöde i l/min. Notera att vi inte har några data för givaren förrän efter 1804 (ååmm). Modellen har bara 5 X-variabler: två grundvattenrör med tillhörande temperaturer samt ÖVY. Dessa X är inte tillräckliga för att modellera Y-variabeln. De breda konfidensintervallen är en konsekvens av modellens dåliga prediktionsförmåga, vilket diskuterades i slutet av 8.2.



Figur 36. Övre grafen visar predikterad (röd) för PLS1 observerad signal för S.D001.S27.T53 (svart) tillsammans med konfidensintervall (blå) för nya data (2107–2210) föregånget av sista biten av kalibreringsdata (2101–2106). Den nedre figuren visar modellhändelserna. Formatet på x-axeln är ÅÅMM.

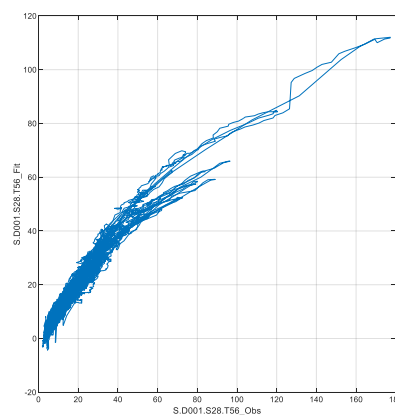
Även nya data bekräftar att de X vi byggt modellen med inte fångar variationen för flödet över Thomsonöverfallet. Men ändå kan man i övre grafen i Figur 36 se att den observerade signalen ligger utanför predikterat konfidensintervall med en stor topp under 2205. I nedre grafen i Figur 36 syns att CIYpred fångade händelsen med toppen i övre grafen.

Modellerna PLS2 och 3 modellerar båda ett annat flöde än PLS1. Modellerna har i stort sett olika X-variabler. PLS2 är bra medan PLS3 är undermålig, Q^2 på 0.82 resp. 0.17. PLS2 har 23 x-variabler och PLS3 har 7 x-variabler. Det saknas data för flödes hastigheten före cirka 2018-06.



Figur 37. Predikerad (röd) från modellen PLS2 resp. PLS3 och observerad signal (svart) tillsammans med konfidensintervall (blå) för S.D001.S28.T56 som är ett Thomsonöverfall med beräknat flöde. Modellerna PLS2 resp. PLS3 har en prediktionsförmåga Q^2 på 0.82 (bra) resp. 0.17 (undermålig).

I både övre och nedre grafen i Figur 37 predikterar modellerna samma givare S28.T56 ett Thomson-överfall med beräknat flöde i l/min.



Figur 38. Predikterade från modell PLS2 mot observerade värden för Thomsonöverfallet S.D001.S28.T56.

I Figur 38 visas predikterade mot observerade värden för den bästa modellen av de två, PLS2. Här ser vi att sambandet är olinjärt. Sådana modeller kan ofta förbättras genom att använda transformerade värden av Y som $\sqrt{Y}$. Här kan man vinna på att transformera Y genom att använda kvadratroten ur Y. Ett annat alternativ är att använda mätningen i mm för Thomsonöverfallet i stället för beräkningen till l/min. När detta gjordes ökade Q^2 till 0.9.

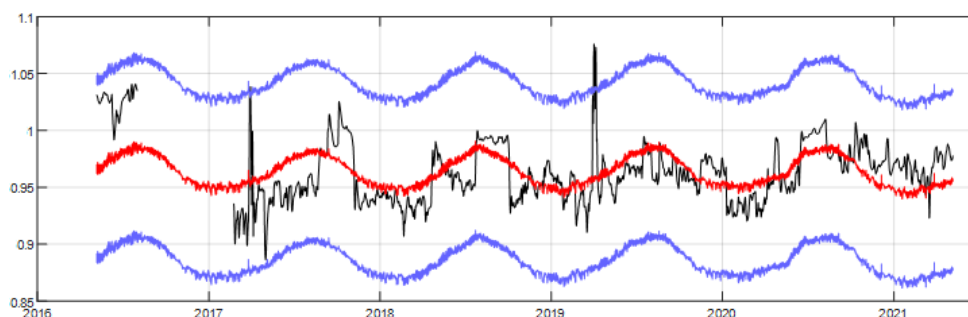
Den dåliga modellen har troligen orsakats av det stora signalsprånget i okt 2018, som missats under modelleringsarbetet



Figur 39. Predikterad (röd) från modellen PLS4 och observerad signal (svart) för S.D001.S09.T32 som är ett grundvattenrör. Modellen har en prediktionsförmåga Q2 på -0.134 vilket kan ses som tecken på en undermålig modell.

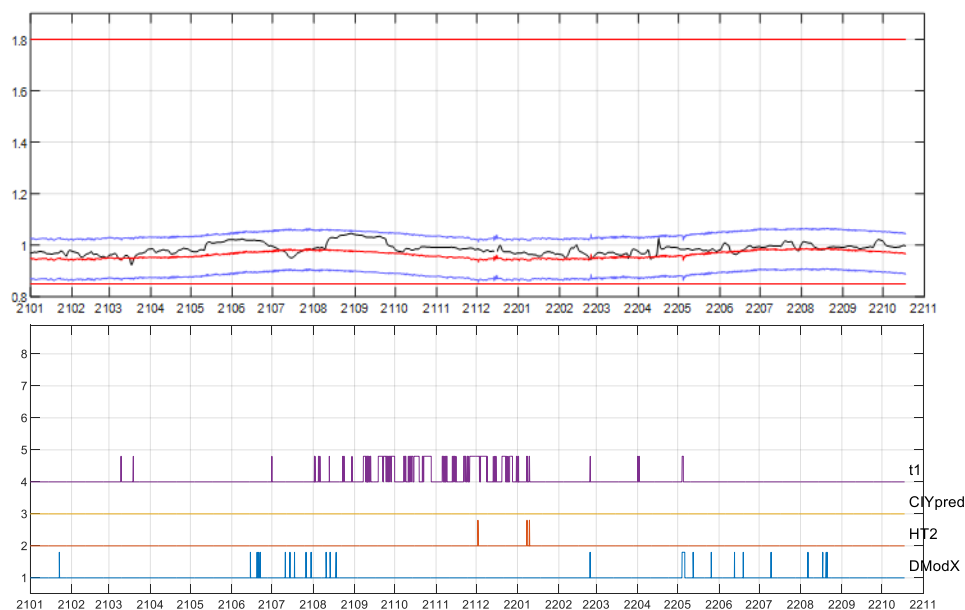
Modellen PLS4, Figur 39, påvisar ett problem som förekommer då och då när man gör en manuell mätning och justering av givaren. I figuren gör den observerade signalen (svart) ett vertikalt språng uppåt. Det vore rimligt att retroaktivt justera signalen bakåt i tiden från kalibreringstidpunkten. Detta skulle då bli den nya signalen som man använder för att bygga kommande uppdaterade modeller. Vi har tagit med detta som ett exempel på hur en del datasystemen fungerar och att det finns förbättringspotential.

Modellen PLS5 (nedan) tas med som på ett exempel där man kunde vinna på att inkludera fler X variabler.



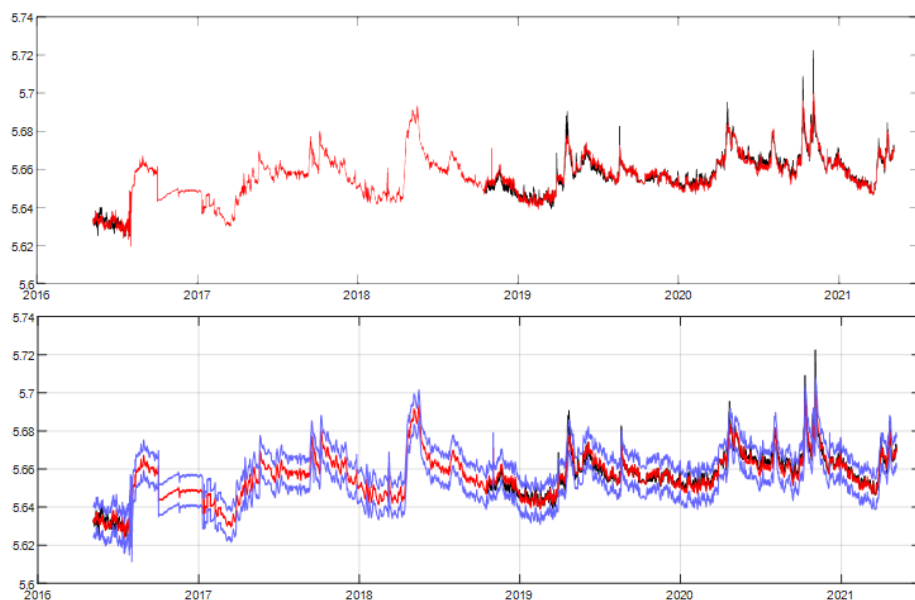
Figur 40. Predikterad (röd), från modellen PLS5, Observerad signal (svart) och konfidensintervall (blå) för S.D001.S10.T43, ett grundvattenrör. Modellen har en prediktionsförmåga Q2 på 0.13 en undermålig modell.

Figur 40 visar resultaten från modellen PLS5 som har X-variabler: två grundvattenrör inkl. dess temperaturgivare, ett Thomson-överfall, ÖVY, NVY samt temperatur och nederbörd från närmaste väderstation så totalt 9 x-variabler. Här kan vi misstänka att de signaler som inkluderats inte korrelerar tillräckligt med Y för att man ska kunna bygga en bra modell. En annan förklaring är att givaren är felaktig.



Figur 41. Övre grafen visar predikterad (röd) för PLS5, observerad signal för S.D001.S10.T43 tillsammans med konfidensintervall (blå) för nya data (2107–2210) och en bit av kalibreringsdata (2101–2106). Den undre grafen visar modellhändelserna. Formatet på x-axeln är ÅÅMM.

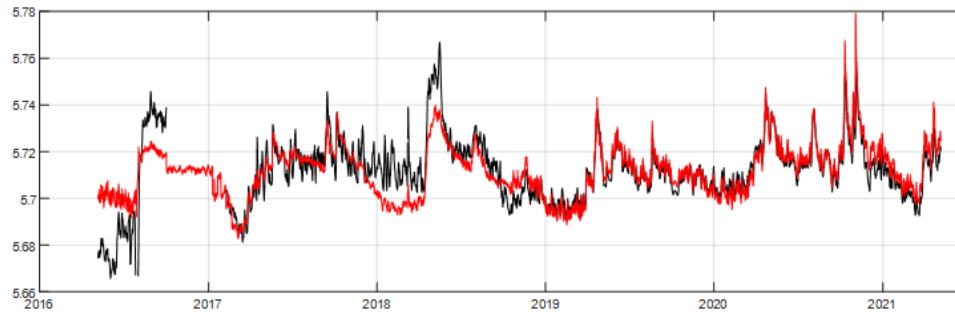
Från övre grafen i Figur 41 ser vi även att predikterad ger en mycket lugnare signal än den uppmätta och dessutom ligger Predikterad i huvudsak under det sensorn visar. Det ger ytterligare stöd att fler sensor behöver ingå i modellen för att vi ska få en önskad god prediktionsförmåga och där med även kunna fånga modellhändelser med god kvalitet, se undre grafen i Figur 41.



Figur 42. Övre grafen visar predikterad (röd) från modellen PLS8 och observerad signal (svart) för S.D001.S15.T36, ett grundvattenrör. Den nedre visar i tillägg till detta även konfidensintervallen (blå). Modellen har en prediktionsförmåga Q2 på 0.74 vilket kan ses som en bra modell.

I den övre Figur 42 framgår, liksom för flera av modellerna att det är signalbortfall under lång tid (ingen svart linje). Vi ser bra modellanpassning både före och efter

signalbortfallet. Med en så bra modell kan vi anta att modellprediktionerna under bortfallet är rimliga. Detta illustrerar möjligheten att använda modellprediktioner i brist på observerade signaler som en så kallad virtuell redundant givare.



Figur 43. Övre grafen visar prediktionerna (röd) från modellen PLS10 och observationerna (svart) för S.D001.S20.T41, ett grundvattenrör. Prediktionsförmågan Q2 för modellen ligger på 0.50, en halvbra modell.

I Figur 43 kan man se att modellen överensstämmer mycket bra från 2019 vilket ger antyder att modellen fungerar bra ändå.

10 Avvikelsevisualisering - från dammöversikt till detaljnivå

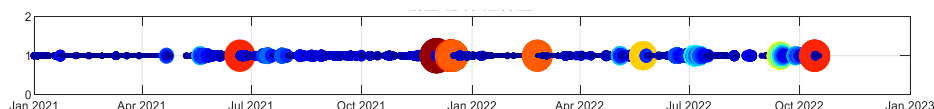
Under projektets gång insåg vi att traditionella trendgrafer inte riktigt räcker. Användaren behöver såväl enkelt kunna få en översikt över inträffade händelser som kunna gräva sig ner i enskilda signaler och modeller. Vi utarbetade ett hierarkiskt presentationssystem med 5 nivåer, med alla dammhändelserna i toppen.

Händelserna på en nivå visas punktvis som funktion av tiden. Punkternas storlek och färg indikerar relativa antalet registrerade händelser vid varje tillfälle.

Tabell 4. Händelsevisualiseringen är hierarkiskt ordnad, från översta dammnivån, ner till enskilda signaler och modeller.

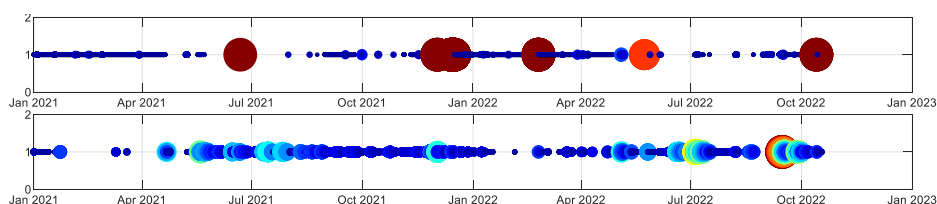
1 Toppnivån	Summan av alla signal- och modellhändelser
2	Separata signal- eller modellhändelser
3	Händelser per signal eller modell
4	Trender och separerade händelser per signal eller modell
5 Mest detaljerad	Enskild signal eller modellparametrar

På den högsta nivån, nivå 1, summeras alla avvikelser från en damm, se Figur 44.



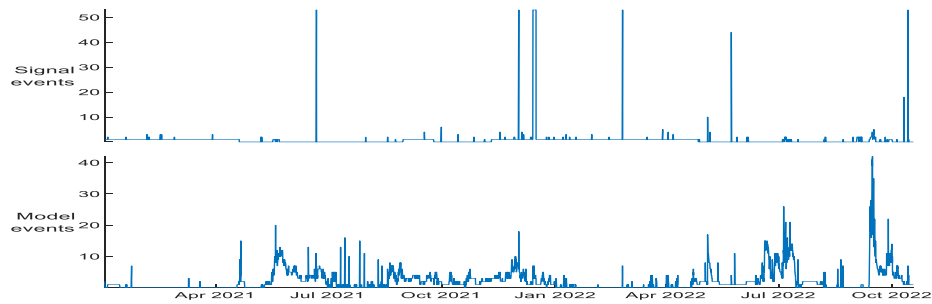
Figur 44. Översta nivåns statusgraf visar summan av signal- och modellhändelser

Grafen kan delas upp, med ett musklick, så att signal- och modellhändelser visas separerade, nivå 2. Se Figur 45



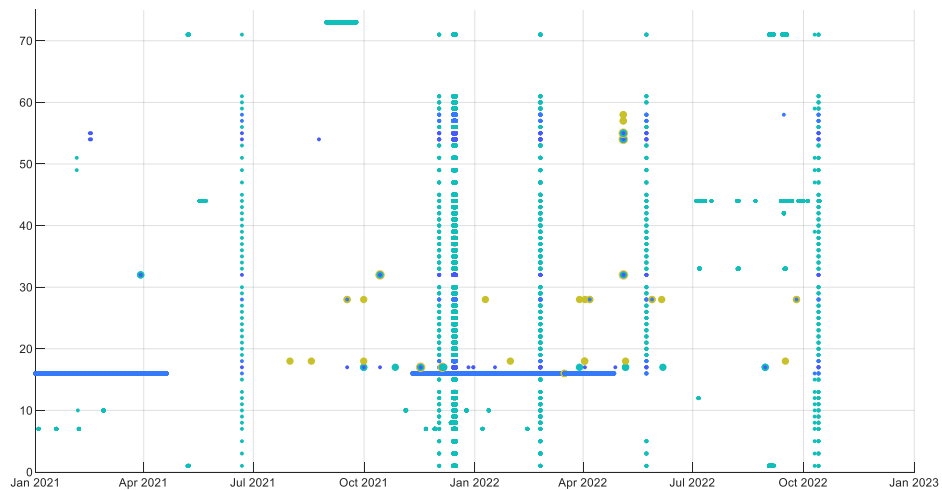
Figur 45. Separerade signal- och modellhändelser.

I stället för punktrepresentationen kan man förstås visa händelserna som traditionella trender vilket kan ses i Figur 46

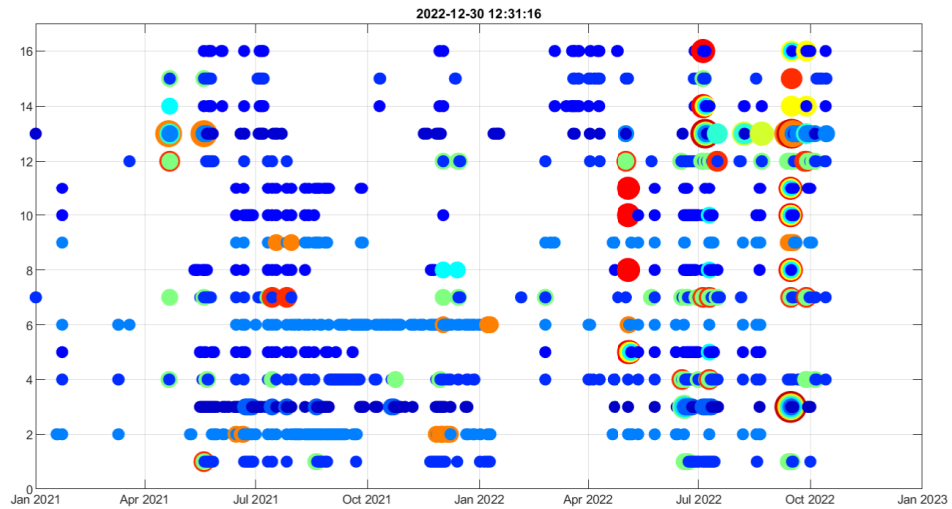


Figur 46. Summa signal- och modellhändelser som vanliga tidsserier

På nästa nivå, nivå 3, visas signal- eller modellhändelser uppdelade med en linje per signal eller modell, Figur 47 och Figur 48.

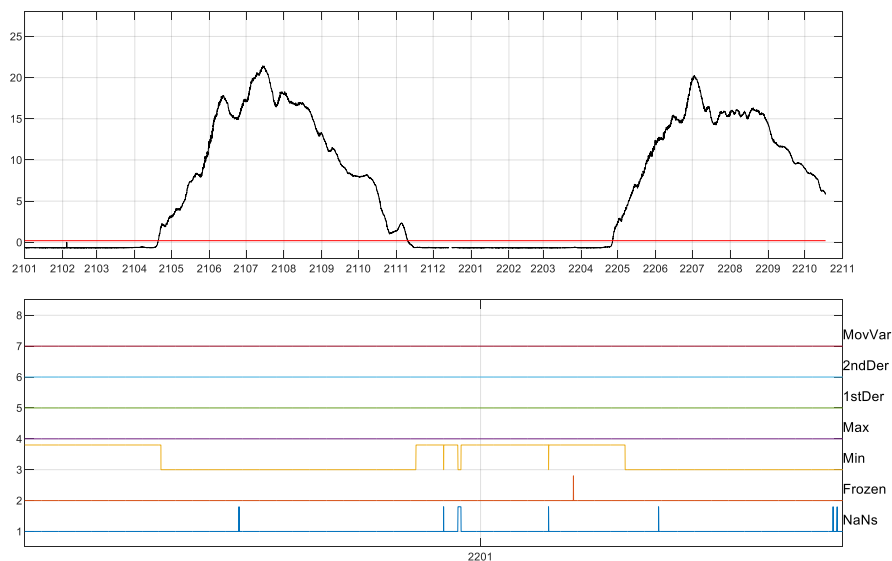


Figur 47. Summa händelser visas här per signal, här 74 signaler. Man kan se vertikala stråk av händelser för nästan alla signalerna vid vissa tidpunkter. Signal 17 har två längre perioder av händelser.



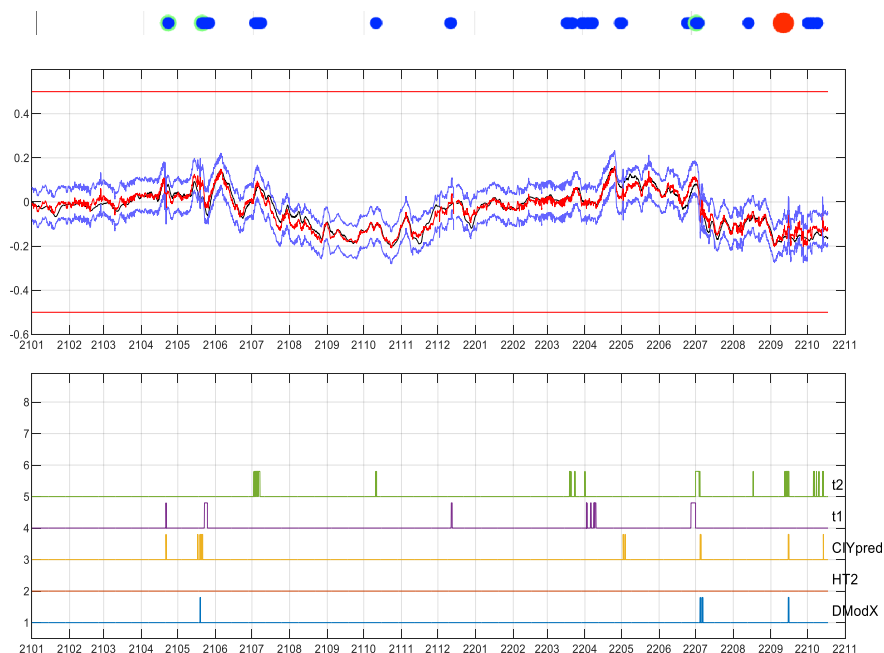
Figur 48. Summa händelser visas här per modell, här 16 modeller från anläggning 2. Flera av händelserna under senare halvan av år 2022 kan förklaras av att anläggningen engagerades i nätfrekvensstyrning

Från nivå 3 kan man välja att fokusera på en enskild signal eller modell, nivå 4. Då visas signalen eller modellen som en tidsserie samt alla händelsetyper som separata trender, se Figur 49.



Figur 49. Trend och händelser för signal 17, en dammtemperatur. I trenden är det tydligt att signalen har legat under sitt undre gränsvärde i två långa perioder. Orsaken identifierades som ett kalibreringsfel. Bland signalhändelserna kan man se indikationer på bortfall och en period av frusen signal. Bortfallsindikationerna kan spåras tillbaka till de vertikala strecken i Figur 31 som indikerar mer omfattande signalbortfallstillfällen. Tiden i graferna ges av YYMM.

På nivå 4 finner vi också trender och händelser för enskilda modeller, här modell 15, Figur 50.



Figur 50. Trender för observerad (svart) och predikerad (röd) signal plus konfidensintervall (blå) för modell 15, övre delen, samt enskilda modellhändelser, nedre delen. Knappt skönjbart i övre delen men i trenden för CIYpred ser vi att konfidensgränsen överskreds vid några tillfällen.

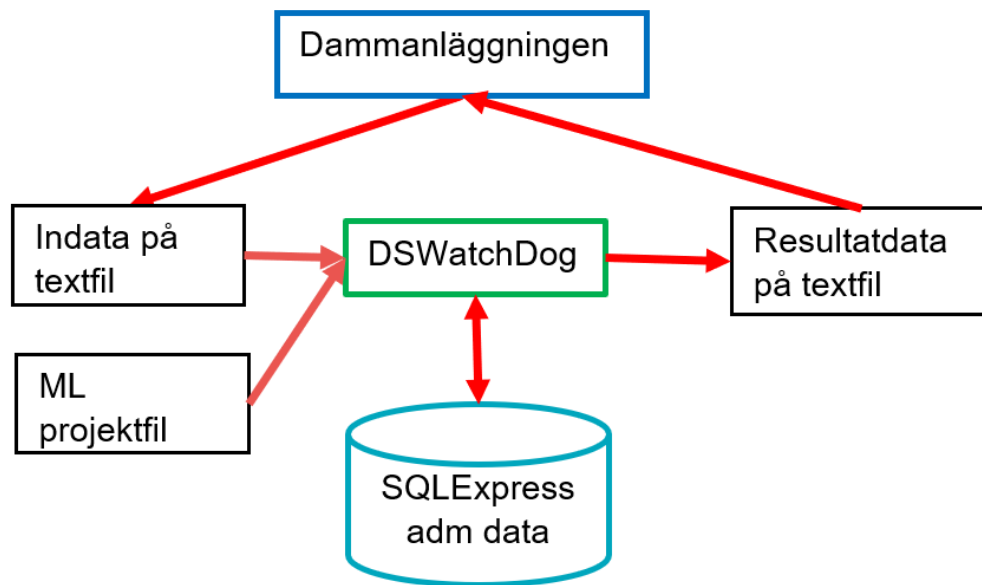
Med den här tekniken kan man utvidga det hierarkiska perspektivet med ännu ett, övre, lager som omfattar en hel dammsträcka. Med ett musklick borde man kunna expandera grafen till en graf per damm.

Andra visualiseringsidéer som kom under projektet är ökad interaktivitet, 3D-grafer och användning av geobaserade AR-verktyg. Bra visualisering är en förutsättning för ökad användbarhet.

11 Implementation

Utvecklingsversionen av mjukvaran är beskriven i Jacobson et al., (2022). Här ingår program för modellbyggande, databashantering och realtidsexekvering.

Från utvecklingsversionen är realtidsdelen, Figur 51, utbruten och installerad hos respektive dammägare på en virtuell dator med fjärrskrivbordsåtkomst, en per installation. Det är samma mjukvara som installerats för alla tre installationerna, det enda som skiljer är den specifika konfigureringsfilen.



Figur 51. Blockschema över den installerade realtidsmjukvaran hos dammägarna.

Under våren 2022 upptäcktes en bugg i mjukvaran som orsakade att filtreringen av signalerna ej fungerade som förväntat. Denna bugg korrigerades och ny version av mjukvaran genererades. Under maj 2022 installerades mjukvaran för fallstudie III, därvid uppdaterades också mjukvaran för fallstudie I och II.

Fram till februari 2023 kördes programmet DSWatchDog av en inloggad lokal användare med administratörsrättigheter via Fjärrskrivbordet. Då bestämde en av dammägarorganisationerna att inga sådana program, av säkerhetsskäl, fick vara aktiva över natten. Därför skapade vi ett serviceprogram, DSWDService som administreras och körs av den lokale användaren System på den virtuella datorn. DSWDService startas automatiskt då den virtuella datorn startas och startar i sin tur övervakningsprogrammet DSWatchDog.

12 Slutsatser och rekommendationer

Den övervakningslösning som tagits fram i denna serie av projekt är generell och endast mindre anpassningar behövs för att införa lösningen för ytterligare dammar. Lösningen är inte begränsad till dammar utan passar alla kontinuerliga processer såsom exempelvis annan kraftproduktion. Vi har fångat upp sensorer med avvikande beteenden under arbetet, något som har diskuterats med dammsäkerhetssakkunniga för respektive fallstudie och referensgruppen under arbetet med metoderna.

Acceptansen och därmed användningen att den framtagna övervakningslösningen beror mycket på visualiserings- och interaktionsmöjligheterna. Goda modeller med goda gränsvärden samt väl injusterade gränsvärden för olika test för en signal i taget är inte tillräckligt. Man bör kunna peka, klicka och zooma sig fram i data i både tid och rum, i 1-D, 2-D och 3-D. Detta kräver mer än vad som är tillgängligt i de flesta av dagens automationssystem.

Vi skapade en hierarkisk prototyp för visualiseringen. Där går man från en överblicksnivå för alla detekterade avvikelsehändelser för att på nästa nivå dela upp dem i signal- och modellhändelser. Där se man hur händelserna för de olika givarna respektive modellerna är fördelade över tiden. På nästa nivå kan man se vilka specifika signal- eller modellhändelser för en signal som går till och från över tiden.

Vi anser att mer utvecklingsarbete återstår kring visualisering för fallstudierna. Visualiseringen hos den enda fallstudie som har implementerat detta är i huvudsak statistik och omfattar endast signaler och signalhändelser per signal. Vi tror att det är först då dammsäkerhetsansvariga, underhållspersonal eller anställda på en driftcentral själva kan titta närmare från överblicksnivå till specifika sensorer i realtid och även kan zooma in mer på de relevanta tidsperioderna som man kommer uppleva den fulla nyttan. Det är då de själva sannolikt känner ett större engagemang för detta och även internt kan visa nyttan av verktygen.

En viktig slutsats som dragits flera gånger är vikten av hög datakvalitet. I och med arbetet med fallstudie III har vi nu sett exempel på hur data med hög kvalitet underlättar arbetet och förkortar tiden vid förbehandling av signaler och design av modeller. För att data skall komma till nytta i framtiden behöver data valideras och den goda kvaliteten säkerställas redan vid insamlingen.

En PLS-modell med hög prediktionsförmåga har ett snävare konfidensintervall, jämfört med en modell med lägre prediktionsförmåga. Båda modellerna går att övervaka med vår metodik, men modellen med snävare konfidensintervall kommer vara snabbare på att upptäcka felaktigheter.

Dammars tillstånd kan förändras över tiden. Detta kan till exempel bero på att det sker förändringar i hur kraftverket drivs eller att medeltemperaturen kring dammen förändras. Modellerna som skapas förutsätter att tillståndet är detsamma vid användningen som under perioden som använts för att generera dem. Detta betyder att modeller behöver underhåll och tränas om med tiden. Parametrarna DModX och Hotelling's T^2 bedömer modellens giltighet och kan användas på indikatorer på att modeller behöver underhåll.

Vi rekommenderar starkt att varje givare ska i datalagringsammanhang få ytterligare en tagg/fält för vilken man bör göra justeringar bakåt i tiden då det utförts en kalibrering. En sådan tagg/fält skulle man även kunna använda då en givare är trasig och man av olika anledningar inte kan byta ut den så snabbt man önskar och ersätter dess signal med modellprediktioner. Vi föreslår att RIDAS - Energi-företagens riktlinjer för dammsäkerhet bör överväga att ta med detta i sina rekommendationer.

Arbete med övervakning baserad på maskininlärningsmetoder innebär förstås behov av kunskap inom området inom varje dammägerorganisation. Kursen som framtagits inom projektet är avsedd att ge tillräcklig bakgrund för fortsatt kunskapsinhämtning på egen hand för dem som blir ansvariga för övervakningslösningens framtida drift.

Projektserien har nu tre fallstudier och genom anläggningsägarna har projektet fått tillgång till mer än 5 års historiska data för respektive damm. Dessa data ligger till grund för vidare utveckling och utvärdering av övervakningsmetoderna. Att samla in, bygga upp en gemensam förståelse av och förbehandla dessa data kräver mycket tid och ett gott samarbete mellan dataspecialister och dammägare, en grund som nu är lagd.

13 Referenslista

- Bulyaculov, R. R., Schogoleva, K. P., Yakovlev, I. N., & Roskostov, R. A. (2017). Modelling and analysis of the median filter algorithm of suppression of impulse noise. Rapport presenterad på 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), Saint Petersburg, Russia.
- Chong, I.-G., & Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and intelligent laboratory systems*, 78(1-2), 103-112.
- Ekonomifakta. (2020). Sveriges totala elproduktion fördelat på energislag, 2019, TWh. Hämtad från https://www.ekonomifakta.se/Fakta/Energi/Energibalans-i-Sverige/Elproduktion/?gclid=Cj0KCQiA7qP9BRCLARIsABDaZziz1w25ccMa5Uqt-Zf_Ij0AJPXpjMpNoJrzTqa0focf9QzZ0zGBX9IaAm4gEALw_wcB&graph=/12218/all/all/
- Eriksson, L. B., T, Johansson, E. T., J, & Vikström, C. (2003). *Multi- and Megavariate Data Analysis. Principles and Applications: Umetrics*.
- Jacobson, A., Fridén, H., & Björk, A. (2022). Metoder för datadriven dammövervakning (Rapport 2022:855). Hämtad från <https://energiforsk.se/media/31252/metoder-for-datadriven-dammovervakning-energiforskrapport-2022-855.pdf>
- Jacobson, A., Torbjörn, J., Håkan, F., & Björk, A. (2021). Detektion av givar- och kommunikationsfel vid dammövervakning (Rapport 2021:735). Hämtad från <https://energiforsk.se/media/29553/detektion-av-givar-och-kommunikationsfel-vid-dammovervakning-energiforskrapport-2021-735.pdf>
- Johansson, T., Fridén, H., Sundberg, M., Samuelsson, O., & Björk, A. (2020). Datadrivna metoder för att detektera avvikande mätvärden inom dammsäkerhet (Rapport 2020:640). Hämtad från <https://energiforsk.se/media/27536/datadrivna-metoder-for-att-detektera-avvikande-matvarden-inom-dammsakerhet-energiforskrapport-2020-640.pdf>
- MATLAB. (2021). Hämtad från <https://sv.wikipedia.org/wiki/Matlab>
- Miljöbalken (1998:808). (1998). Stockholm: Regeringskansliet.
- Montgomery, D. C. (2013). *Introduction to statistical quality control* (Vol. 7): John Wiley & Sons.
- OSIsoft. (2020). Hämtad från <https://en.wikipedia.org/wiki/OSIsoft>
- Principal Component Analysis. (2019). Hämtad från https://en.wikipedia.org/wiki/Principal_component_analysis

- Regeringskansliet. (2015, 2015-04-02). Dammsäkerhet, Tydliga regler och effektiv tillsyn. Hämtad från <https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2012/06/sou-201246/>
- RIDAS - Energiföretagens riktlinjer för dammsäkerhet. (2019). (J. Bladh Ed.). Stockholm: Energiföretagen Sverige.
- Varmuza, K., & Filzmoser, P. (2008). Introduction to Multivariate Statistical Analysis in Chemometrics (pp. 1-328). Hämtad från <http://kcl.ebib.com/patron/FullRecord.aspx?p=427098>
- H. Fridén, A. Jacobson, D. Glebe and A. Björk, Development and application of data-driven surveillance of sensors and dams, Symposium "Management for Safe Dams" - 91st Annual ICOLD Meeting – Gothenburg 13-14 June 2023, p 1594-1603, Gothenburg, Sweden

UNI- OCH MULTIVARIATA METODER FÖR UPPTÄKT AVVIKELSER I DAMMAR

Dammövervakning är ett stort och viktigt område för svenska dammägare. Det är en utmaning att kontinuerligt övervaka en dammkonstruktion med dess givare och snabbt kunna reagera på förändringar som visar på en avvikelse i dammen.

Vi har tidigare rapporterat hur vi utvecklat och installerat en generell övervakningsapplikation som körs i realtid hos två av dammägarna, med dataöverföring från och till deras signaldatabaser. Utveckling har skett av metoder för fristående övervakning av enskilda signaler, univariat övervakning, samtidigt som utvecklingen av multivariat övervakning har påbörjats.

I det här projektet har de samlade erfarenheterna använts för att ansluta en tredje damm, fallstudie III, och för att fortsätta utvecklingen av metodiken. Genom anläggningsägarna har projektet fått tillgång till mer än 5 års historiska data för respektive damm. Dessa ligger till grund för vidare utveckling och utvärdering av övervakningsmetoderna. Att samla in, bygga upp en gemensam förståelse av samt förbehandla data kräver mycket tid och ett gott samarbete mellan dataspecialister och dammägare, en grund som nu är lagd.

Resultatspridning har skett i form av seminarier och en djupare kurs för personal hos kraftbolagen.

Ett nytt steg i energiforskningen

Forskningsföretaget Energiforsk initierar, samordnar och bedriver forskning och analys inom energiområdet samt sprider kunskap för att bidra till ett robust och hållbart energisystem. Energiforsk är ett politiskt neutralt och icke vinstutdelande aktiebolag som ägs av branschorganisationerna Energiföretagen Sverige och Energigas Sverige, det statliga affärsverket Svenska kraftnät, samt gas- och energiföretaget Nordion Energi. Läs mer på energiforsk.se.

