

PUMPE FLOW TIL DRIFTSSTATUS OG -OVERBLIK

(PUFDO)



PUFDO PUMPE FLOW TIL DRIFTSSTATUS OG -OVERBLIK DANVA VUDP PROJEKTRAPPORT

DATO: 21. august 2019

Projekt ID:

1165.2017

Udgiver:

FORS

Udarbejdet af:

Jesper Ellerbæk Nielsen, AAU
Michael R. Rasmussen, AAU
Søren Højmark Rasmussen, EnviDan
David Getreuer Jensen, EnviDan
Anders Hertz Kristensen, FORS (nu ved EnviDan)

Finansiering:

Vejledningen er finansieret af
VUDP, Vandsektorens Udviklings- og Demonstrationsprogram

Samarbejdspartnere:

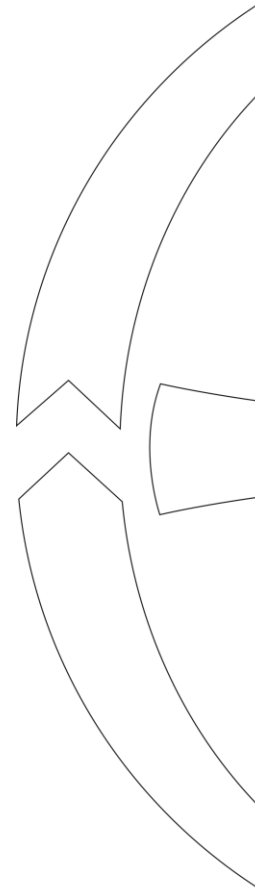
FORS A/S (Hovedansøger)
Aalborg Universitet
EnviDan A/S

Kategori (Spildevand, drikkevand eller klimatilpasning):

Spildevand

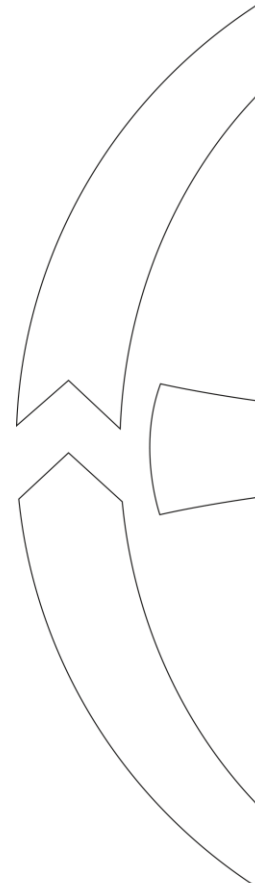
Indholdsfortegnelse

1	Sammenfatning	3
2	English summary	4
3	Introduktion	5
4	Projektets betydning for vandbranchen	6
4.1	Marked og/eller anvendelsesmuligheder	6
4.2	Næste skridt	7
4.3	Formidlingsplan	8
5	Projektet	10
5.1	Formål	10
5.2	Output	11
5.3	Projektresultater	27
5.4	Konklusion	29



1 Sammenfatning

FORS forsyning, EnviDan og Aalborg Universitet har med støtte fra VUDP udviklet en metode til at bruge simple SRO-data til at forudsige vandføring ud fra pumpestationer i oplandet til Holbæk renseanlæg. Den bærende ide har været at bruge Machine Learning metoder til at træne modeller på historiske målinger for at kunne beregne aktuelle vandføring i systemet. Metoden har vist sig meget effektiv til at beregne vandføring fra pumpestationer på basis af pumpernes strømforbrug. Ud over at beregne vandføring fra individuelle pumpestationer kan systemet også bruges til at forudsige tilløbs vandføringen til Holbæk renseanlæg flere timer frem. Dette kan endda gøres, selvom man kun har data fra et færre antal pumpestationer. Projektet har også afsløret, at mængden og måden som forsyningerne gemmer data på, er kritisk for kvaliteten af beregningerne. Det er demonstreret, at man også kan bruge de samme data til at overvåge og vurdere tilstanden af individuelle pumper i en pumpestation. Dette kan have betydning i den daglige drift og planlagt vedligeholdelse. Der er gennem projektperioden testet mange forskellige Machine Learning algoritmer for at finde de metoder, der er bedst egnede til at beregne pumpe vandføring ud af pumpestationerne. Valget faldt på *Long short-term memory* (LSTM) netværket. Denne type netværk bruges i flere *deep learning* applikationer. I dette projekt var dette neurale netværk fordelagtigt, da det både inkluderer momentane målinger af forskellige parametre i pumpestationen, men også viden om hvordan disse parametre har udviklet sig op til beregningstidspunktet. En anden vigtig del af projektet er at vise, at data kan komme hele vejen fra pumpestationen til EnviDans portal, hvor resultater beregnes og vises på en letforståelig måde. Dette muliggør, at man med samme system kan overvåge og analysere mange pumpestationer med det samme system. Overskuelighed og visuel præsentation af data giver operatører og planlæggere et direkte indblik i afløbssystemets funktion.



2 English summary

FORS utility, EnviDan and Aalborg University, with the support of VUDP, have developed a method for using simple urban drainage system data to predict flow from pumping stations in the catchment to Holbæk wastewater treatment plant. The basic idea is to use Machine Learning methods to train models on historical data to calculate current flow in the system. The method has proven very effective in calculating flow from pumping stations based on the pump's power consumption. In addition to calculating flow from individual pumping stations, the system can also be used to predict the inflow to Holbæk wastewater treatment plant with a lead time of several hours. The project has furthermore revealed that the quality and manner in which utility companies store data is critical to the quality of the calculations. It has been demonstrated that the data also can be used to monitor and assess the condition of individual pumps in a pumping station. This can be of importance in daily operation and scheduled maintenance. Throughout the project, many different Machine Learning algorithms have been tested to find the methods best suited to calculate the flow through pump stations, of which the Long Short-Term Memory (LSTM) was selected. This type of network is used in several deep learning applications. In this project, this neural network was advantageous as it included both instantaneous measurements of different parameters in the pump station, but also knowledge of how parameters have developed up to the time of calculation. Another important part of the project is to show that data can travel from the pump stations to the EnviDan portal, where results are calculated and displayed in an easy-to-understand way to the end user. Transparency and visual presentation of data give operators and urban drainage planners a direct insight into the state and functioning of the drainage system.



3 Introduktion

Det må forventes, at antallet af målepunkter i vores vand-infrastruktur vil eksplodere i de kommende år. Ved anvendelse af IoT og andre teknologier vil omkostningen pr. målepunkt falde drastisk. Det betyder, at der opstår nye muligheder for at forstå og designe vores infrastruktur bedre.

Vandsektoren samler allerede i dag millioner af målinger sammen hver dag. Disse målinger bruges til mange ting: Overvågning, planlægning, analyse og modelberegninger. Det er dog også klart, at potentialet ikke bliver udnyttet fuldt ud. Når man f.eks. måler strømforbruget i en pumpestation, kan det bruges til at opdage, om motoren bruger den forventelige mængde strøm eller om den er defekt. Men det kan også bruges til at beregne f.eks. vandføringen, hvis det kombineres med andre data. Det er den grundlæggende ide, som PUFDO bygger på.

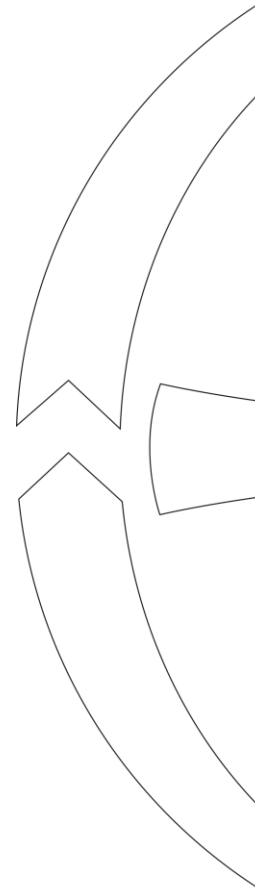
I forbindelse med PUFDO har det ene fokus været rettet mod pumpestationer og de data, man kan finde der. Pumpestationer har været specielt velegnet, da de ofte overvåger flere parametre samme sted.

Når en pumpestation pumper vand videre gennem systemer, sker det ud fra en lokal styring af pumpestationen, som er tilpasset den aktuelle vandføring som løber til stationen. Dette til trods for at tilløbsvandføring ikke direkte måles. Fysikken omkring en pumpestation med pumpeump, som fyldes og tømmes afhængig af indløbsvandføring og pumpernes ydelse, gør det muligt at bestemme pumpestations vandføring, ud fra de observationer, som i dag typisk registreres for en pumpestation - uden at der installeres en egentlig vandføringsmåler til formålet.

På nyere og større pumpestationer er der i dag installeret vandføringsmålere, der måler pumpestationens pumpevandføring, som kan fortælle, hvad vandføringen er forskellige steder i systemet. Men disse pumpestationer udgør dog kun en lille del, og der eksisterer langt flere ældre og mindre pumpestationer, hvor vandføringen ikke måles. Generelt er det kompliceret og meget bekosteligt at få detaljeret viden i realtid om, hvordan vandføringen fordeler sig rundt om i systemet. Etablering af vandføringsmålinger i afløbssystemer, kræver dels store investeringer og dels meget løbende vedligeholdelse for at sikre pålidelige målinger, særligt hvis de foretages på delfyldte gravitationsledninger.

Det andet fokuspunkt PUFDO har haft er at anvende nye metoder til at få maksimal viden ud af de opsamlede data. Det skyldes, at sammen med mange og komplekse data følger også "data overload". Det betyder, at man som menneske ikke kan overskue og vurdere så mange data på én gang. Machine learning metoder er derimod velegnet til at trække sammenhæng og substans ud af store datasæt. En anden fordel er, at de kan implementeres i realtidssystemer, så de kan analysere data umiddelbart efter at de er opsamlet. Det betyder også, at systemerne hele tiden bliver "klogere", des flere data der opsamles.

Rollefordelingen i projektet er, at FORS har stillet SRO-data til rådighed og har været bærende i kravspecifikationen til systemet for at sikre, at resultaterne vil kunne bruges aktivt i den daglige planlægning og drift. Projekt har arbejdet med oplandet til Holbæk Renseanlæg. AAU har stået for at udvikle den bærende algoritme bag det neurale netværk. Herunder nødvendige inputs, netværk design samt træning og validering. EnviDan har stået for operationaliseringen af den udviklede algoritme. Dvs. udstilling af data til FORS, data-flowet, hvor relevant SRO-data fra FORS er synkroniseret til EnviDans portal samt eksekvering af algoritmen.



4 Projektets betydning for vandbranchen

Af de 16.393 (DANVA 2015) pumpestationer i det danske afløbssystem forefindes der vandføringsmålinger i ca. 20-25%. Vandføringsmålinger er i dag essentielle for en række analyser i både planlægning- og driftsfasen af afløbssystemer. Pumpestationer, der ikke fungerer optimalt, kan være en risiko for resten af systemet. Det er derfor essentielt at få mere viden om pumpestationerne og hvordan de påvirker systemet. PUFDO anvender Machine Learning til at estimere vandføringen ud fra logget strømforbrug pr time. En metode, der vil kunne anvendes på de resterende 75-80% pumpestationer, hvor der ikke findes vandføringsmåler.

Ud over at PUFDO har vist gode estimater af vandføringen for de enkle pumpestationer, samt potentialet i detektering af anomalier vha. gentræning, har PUFDO vist mulighed for at prædiktere tilløbet til renseanlægget, udelukkende på baggrund af SRO-data fra pumper i oplandet. Således er det muligt at forudsige vandføringen til renseanlægget 1-3 timer frem i tiden.

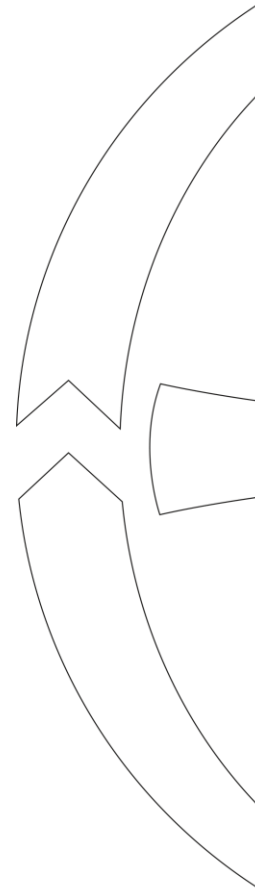
PUFDO giver hermed mulighed for at kende vandføringen mange steder i afløbssystemet - alle de steder hvor der er pumpestationer. Samtidig er der mulighed for at bruge samme data til forudsigelse af tilløbet til fx renseanlægget. Dette giver bedre systemoverblik samt bedre lokalisering af uvedkommende vand, bedre grundlag for optimering af driftsrutiner og asset management.

PUFDO-projektet bygger på de SRO-data, der i dag er tilgængelig. PUFDO har vist et anvendelsespotentiale af de loggede data, men PUFDO har samtidigt givet anledning til at overveje, hvad der logges og gemmes af data, samt med hvilken tidsopløsning og hvor længe det opbevares. For at få fuld udnyttelse af PUFDO-konceptet kan det evt. være nødvendigt at justere på dette. PUFDO-projektet har indikeret, at der er et potentiale for yderligere driftsinformation og anomali detektering i de højt-opløselige-loggerdata, hvilket kunne være interessant at undersøge nærmere i et kommende opfølgingsprojekt til PUFDO.

4.1 Marked og/eller anvendelsesmuligheder

PUFDO henvender sig til alle spildevandforsyninger, som har pumpestationer. PUFDO anvender EnviSync, der giver mulighed for at hente SRO-data fra forsyningerne. EnviSync kan i dag kommunikere med stort set alle danske forsyningers SRO-systemer, således det er muligt at hente data. På baggrund af disse data vil det være muligt at træne PUFDO-algoritmen op til beregning af vandføring fra pumpestationerne, hvorved alle danske spildevandforsyninger er potentielle kunder. EnviDan har hele Skandinavien som markedsområde og der er dermed potentiale for, at PUFDO-konceptet kan udbredes til hele Skandinavien eller mere.

PUFDO-konceptet giver effektiviseringsmuligheder i form af driftsoptimeringen og bedre datagrundlag til asset management. PUFDO giver anvendelse af foreliggende logget SRO-data, samtidig forædles data ved, at de omregnes til vandføring. Omregning til vandføring giver mulighed for at opnå systemoverblik og vandføring i systemet. Disse vandføringsdata vil yderligere kunne bruges til bedre datagrundlag for planlægning af afløbssystemer og til analyser for uvedkommende vand. Når vandføringen er kendt mange steder i afløbssystemet, er der mulighed for at kunne lokalisere hvilke områder, der er særlig plaget af uved-



kommende vand, ved at se på uventet stor tilvækst i vandføring mellem to pumpestationer. Med lange kontinuerlige tidsserier er der mulighed for hydrografanalyser til opdeling af vandføring på forskellige afstrømningskomponenter og herigennem kvantificere uvedkommende vand.

Samling af SRO-data fra flere pumpestationer på et sted giver mulighed for at benytte data på tværs af pumpestationerne. Fx har PUFDO vist at anvendelse af SRO-data fra en række pumpestationer kan bruges til at forudsige indløbet, til renseanlægget. Ligeledes tænkes det muligt at validere vandføringsestimater fra en pumpestation ved at lave estimering ud fra andre pumpestationers data. En anden anvendelsesmulighed vil være at anvende PUFDO til udfyldning af huller i logget data og dermed opnå kontinuerlige tidsserier. EnviDans portal indeholder data fra flere kunder og på denne måde er det muligt at lave big data lignende analyser på tværs af de forskellige kunder. Fx til at kunne lave dynamiske alarmniveauer for normaldrift, baseret på datamining analyser af anonymiserede data fra flere kunder. Denne sammenstilling af data fra mange pumpestationer fra flere kunder, kunne være et emne for en kommende opfølgende PUFDO-projekt.

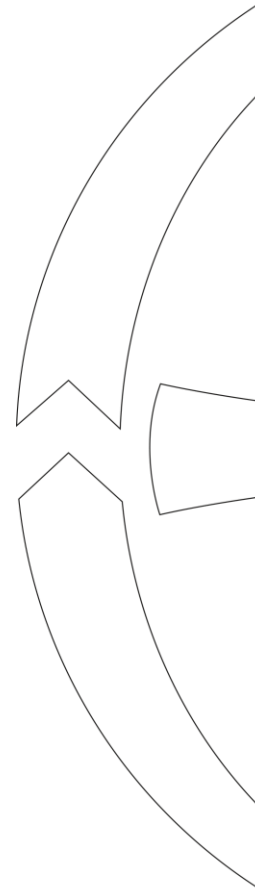
PUFDO-projektet har vist, at det er muligt at beregne vandføring på baggrund af SRO-data med meget god overensstemmelse mellem model og målinger, samt at det er muligt at forudsige vandføring i centrale punkter i afløbssystemet baseret på målinger andre steder i afløbssystemet. PUFDO har vist potentialet for datatjek, anomalidetektion og driftsstand, der kan føre til automatisk kvalitetstjek og styring af afløbssystemer. PUFDO har potentiale til at levere datagrundlag til behovstilpasset drift og optimeret asset management.

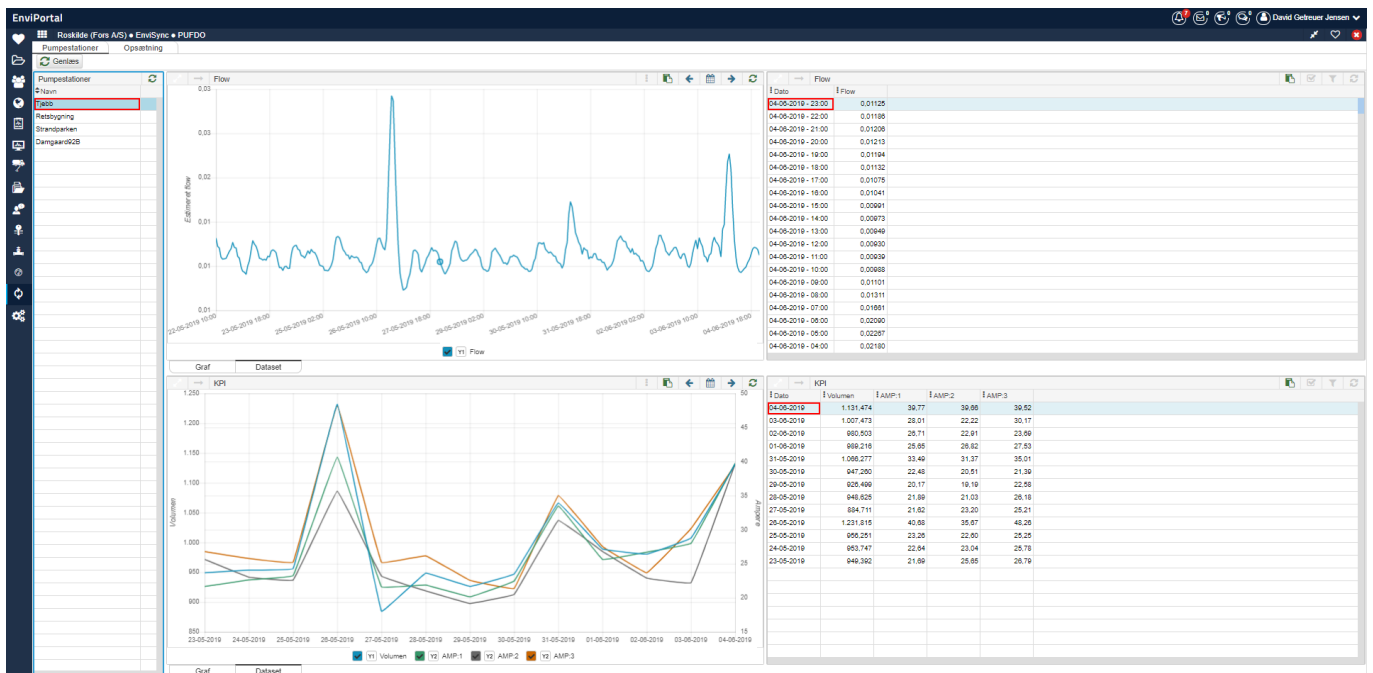
Logning af strømforbrug bliver mere og mere almindelig, og tilmed ved brug af online målinger. Således vil det være muligt at få information om strømforbrug fra samtlige af forsyningernes pumpestationer. Dette vil give datagrundlag for at kunne implementere PUFDO eller dele af det for at lave samanalyzer af strøm eller vandføring til at kunne detektere unormal drift og analyser for uvedkommende vand.

4.2 Næste skridt

Der arbejdes på at finde flere interesserede forsyninger til at være med til et kommende opfølgende udviklingsprojekt eller lignende med PUFDO-konceptet. Alle de forsyninger som vi indtil nu har været i dialog med har alle udvist interesse for, hvad PUFDO har opnået og perspektiverne i et kommende projekt.

PUFDO-konceptet passer godt ind i den generelle digitaliseringsstrategi der pågår i branchen. EnviDans portal giver mulighed for at processere PUFDO vandføring og driftsoverblik. PUFDO-projektet har vist, at konceptet virker og EnviDans portal vil løbende blive udviklet til bedre at understøtte PUFDO og PUFDO lignende processer. På nedenstående billede vises PUFDO implementeret i EnviDans portal. Her udstilles data som estimeret vandføring på pumpestationer, strømforbrug og volumen af pumpet vand.





PUFDO-projektet forventes yderligere udviklet og fokus for et kommende projekt kunne være:

- Automatisk kvalitetstjek, samt syntetiske data til udfyldning af huller i logget data
- Detaljeret systemoverblik ved at estimere vandføring mv. og Data Mining af data med høj tidsopløsning
- Prædiktering af vandføring i nøgle lokaliteter, fx til styring
- Estimering af uvedkommende vand på baggrund af estimeret vandføring

4.3 Formidlingsplan

PUFDO er blevet præsenteret på flere DANVA-arrangementer. Nærværende slutrapport samler alle erfaringer der er opnået gennem PUFDO, dette er fra datahøst til resultater og hvordan datalogning kan ændres til at opnå endnu bedre mulighed for anvendelse i drifts- overvågningen og vandføringsestimater. Herunder listes alt udført formidling, samt planlagt kommende formidling.

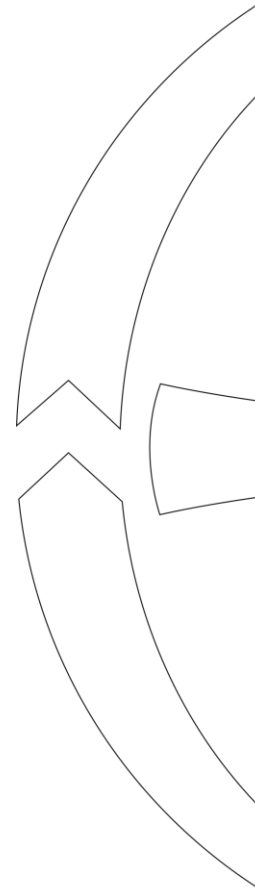
Projekter er blevet formidlet ved:

- Præsentation på Dansk Vand Konference 2018, november, Aarhus. Anders Hertz Kristensen, FORS og Søren Højmark Rasmussen, EnviDan.
- Vandforum, 2019, April, Skanderborg, Michael R. Rasmussen, Aalborg Universitet.
- DANVA Årsmøde, Maj, 2018, Kolding, Anders Hertz Kristensen, FORS
- Projekt case beskrivelse på EnviDans hjemmeside under Digitalisering.
- VUDP Slutrapport, hvilken ligeledes anvendes til opsamling af erfaringer fra projektet.
- Udkast til opfølgende projekt hvor flere forsyninger har været indbudt til at deltage.

Plan for fremtidig formidling:

- Artikel i Dansk Vand
- Poster præsentation på IWA-konferencen i København 2020
- Opfølgende projekt gerne med flere forsyninger involveret
- PUFDO på EnviDans portal forsætter med at køre så længe, at der hentes SRO data fra FORS

EnviDan vil arbejde videre med PUFDO konceptet, som vil indgå i den generelle udvikling af EnviDans portalprodukter rette mod forsyningskunder.



5 Projektet

5.1 Formål

Projektets har det overordnede mål at skabe merværdi af de mange data fra pumpestationer der i dag opsamles hos danske forsyninger som led i den daglige drift og vedligeholdelse. Når sensorer installeres i en pumpestation, sker det typisk med et bestemt formål for øje. Dette kan for eksempel være at monitorere et vandstands niveau i en pumpeump, for derved at have et styresignal til pumpestationens pumper. Eller opsamling af pumpernes strømforbrug og driftsfrekvens, for at have realtids overvågning af pumpernes drift. Observationerne opsamles typisk lokalt og sendes herfra til centrale servere og gemmes i behandlet form i databaser. Men i og med at de individuelle observationspunkter er installeret og tænkt til specifikke og individuelle anvendelser, ligger der et uudnyttet potentiale i at kombinerer forskellige typer af sensordata til virtuelle software sensorer, som kan estimere kvantiteter, der ikke direkte måles af de enkelte sensorer.

Grundideen bag PUFDO er således, at det for enhver pumpestation er muligt, på en omkostningseffektiv måde at etablere virtuelle softwaresensorer, som på baggrund af pumpe-data kan estimere vandføringen pålideligt og præcist. Det direkte udbytte er større viden om systemets aktuelle funktion og tilstand.

I projektet er afprøvet, hvordan forskellige Machine Learning metoder og datadrevne modeller, kan anvendes til at omsætte data fra pumpestationer til vandføringsestimater i afløbssystemet. I projektet implementeres de virtuelle software sensorer i EnviDan's data-portal.

Afprøvningen er foretaget med primær fokus på:

- Beregning af vandføring fra pumpestationer på baggrund af stationens primære data
- Overvågning af pumpestationers driftstilstand
- Estimering og forudsigelse af tilløbsvandføringen til Holbæk renseanlæg

5.2 Output

Beregning af vandføring fra pumpestationer på baggrund af stationens primære data

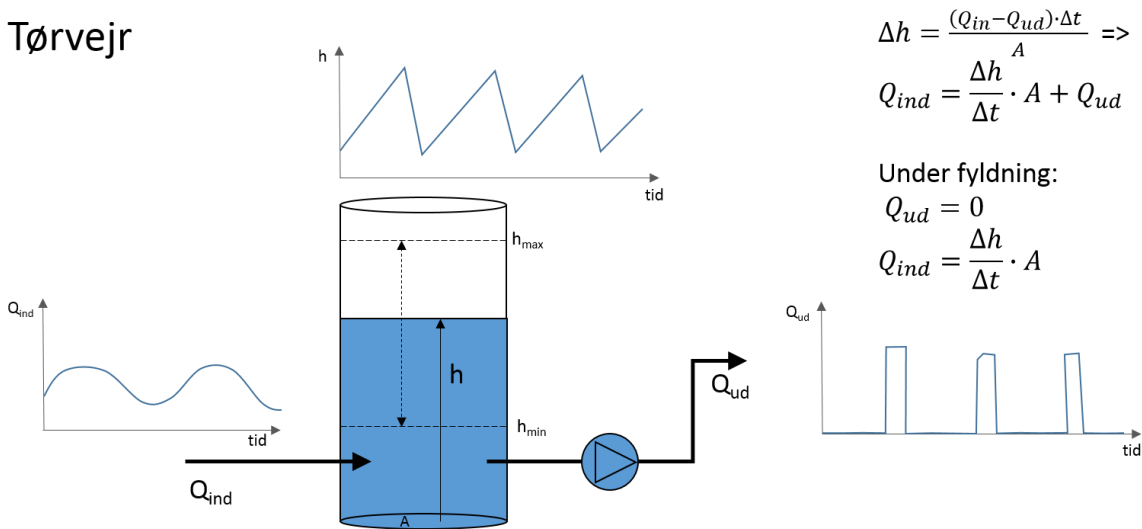
Beregning af en pumpes vandføring, kan i princippet beregnes hydraulisk ud fra pumpens energiforbrug og kendskab til pumpens ydeevne, vandets løftehøjde og systemets energitab. I praksis er det dog usandsynligt at resultatet af en sådan fysisk model vil give præcise estimater for vandføringen. Pumpens ydeevne afhænger i stor grad af pumpens tilstand, da slitage på pumpens pumpehjul og elektriske komponenter påvirker pumpens virkningsgrad. En fuldstændig og korrekt beskrivelse af pumpesystemets hydrauliske energitab, vil herudover kræve en detaljeret gennemgang af hele tryksystemet – fra hvor pumpen suger vand ud af pumpeumpen og igennem trykledningen til dets udløb. Der eksisterer semiempiriske metoder til bestemmelse af såvel enkelt og ledningstab, men for at sikre den nødvendige præcision vil det kræve et detaljeringniveau af information, som 1) sjældent vil være tilgængeligt og 2) skal opdateres løbende i takt med at systemet slides under drift. (Kallesøe og Knudsen, 2016)

Dette ændrer dog ikke på, at der eksisterer en fysisk sammenhæng mellem pumpens energiforbrug og pumpens vandføringsydelse – om end sammenhængen er ukendt og tidsafhængigt. Machine Learning og data drevne modeller har deres styrke i, at de kan anvendes til at identificere sammenhænge i data. Uden en forud defineret fysisk model, identificerer data drevne modeller de fysisk betingede sammenhænge, som eksisterer i dataene gennem læring på baggrund af træningsdata. Herefter anvender modellen de identificerede sammenhænge til at forudsige/estimere vandføringen på baggrund af nye data, som modellen ikke er trænet på.

En grundlæggende forudsætning for at etablere sammenhængen mellem pumpestationens primære data og stations vandføring, er at vandføringen kendes i træningsperioden. Dette er naturligvis et paradoks, for hvis vandføringen allerede kendes for stationen, er behovet for at estimere dette ikke umiddelbart relevant. Den praktiske udfordring er også, at for pumpestationer uden vandføringsmålinger eksisterer de nødvendige træningsdata umiddelbart ikke.

Principielt vil det være muligt at generere de nødvendige træningsdata ved hjælp af midlertidige målekamper med vandføringsmålere. Men da dette er bekosteligt og tidskrævende, vil denne tilgang reelt begrænse metodens anvendelighed og udbredelse. Derfor er det i projektet i stedet valgt at løse denne udfordring ved at anvende højopløselige niveau-målinger i pumpestationens pumpeump. Princippet bag er skitseret på figur 5.1 og figur 5.5, som illustrerer indløbsvandføringen, pumpevandføringen og vandstandsvariationen i pumpeumpen i en tørvejrssituation og under regn.

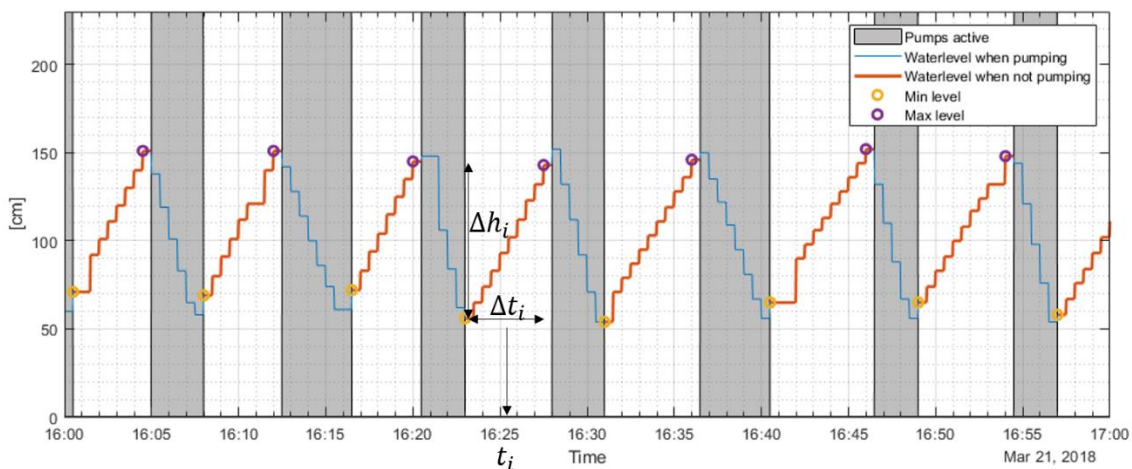
Tørvejr



Figur 5.1: Principskitse for variation i indløbsvandføring, pumpevandføring og vandstandsvariation i en tørvejrssituation.

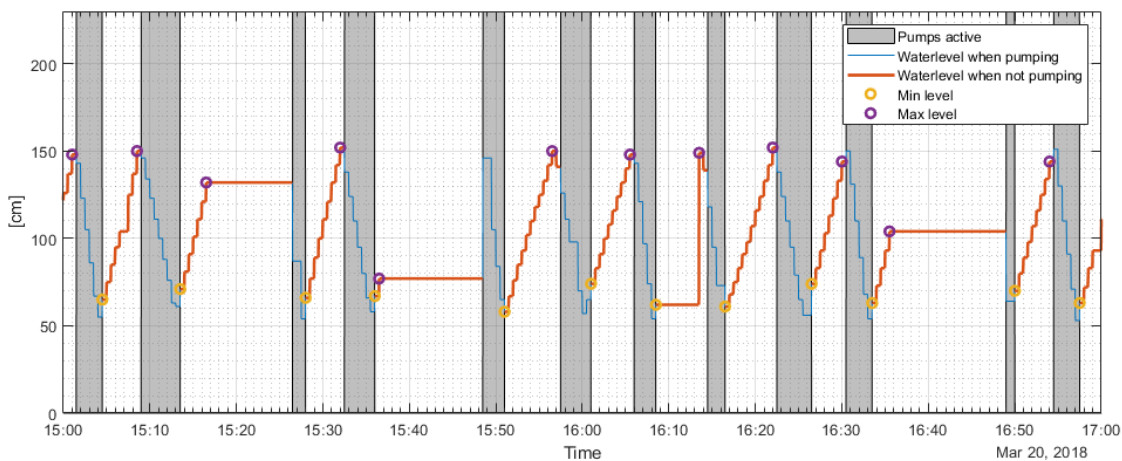
Pumpesumpen i en pumpestation har til formål at udjævne forskellen mellem indløbsvandføring og pumpernes vandføring, således at pumperne ikke konstant er i drift. Når pumperne ikke kører, fyldes sumpen indtil at maksimumniveauet (h_{\max}) er nået. Her starter en eller eventuelt flere pumper, som tømmer sumpen. Hastigheden hvormed sumpen fyldes er udelukkende afhængig af indløbsvandføringen, mens tømmetiden afhænger både af indløbsvandføringen og pumpens vandføring, som vist på figur 5.1.

Under hver fyldning er det således muligt at estimere tilløbsvandføringen mens pumperne er inaktive. Typisk har en pumpestation 5 til 15 starter i timen, hvilket betyder, at det er muligt at bestemme timevandføringen ud fra det samme antal estimater.



Figur 5.2: Data fra Tjebberup pumpestation fra den 21. marts 2018. Grå angiver at en pumpe er aktiv, rød kurve viser niveauet i sumpen under fyldning og blå under tømning.

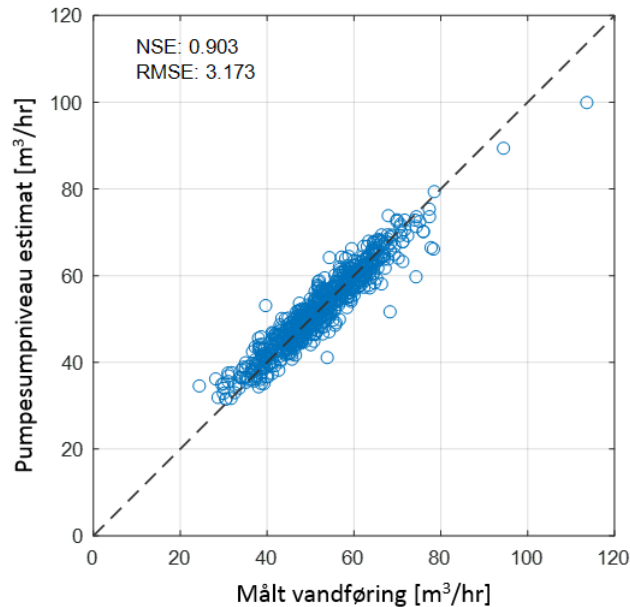
Til formålet blev der i projektet udviklet en algoritme, som automatisk genererer den nødvendige træningsvandføring på baggrund af variationen i pumpesumps niveau. Algoritmen er baseret på højopløselige vandstandsvariationer i sumpen, samt højopløselige data for pumpernes strømforbrug. Sidstnævnte anvendes udelukkende til at identificere perioderne hvor pumperne er aktive, som illustreret på figur 5.2. Figur 5.2 viser en periode med 'pæne' data, mens figur 5.3 viser en periode med udfald som algoritmen også er designet til at håndtere.



Figur 5.3: Data fra Tjebberup pumpestation fra den 20. marts 2018. Data indeholder tydelige fejl. Grå angiver at en pumpe er aktiv, rød kurve viser niveauet i sumpen under fyldning og blå under tømning.

Som vist på figur 5.3, indeholder de højopløselige data logningsfejl i perioder. Grunden hertil er ukendt, men der er i udviklingen fokuseret på at træningsvandføringen ikke påvirkes af disse fejl. Dette er implementeret ved hjælp af en statistisk analyse af de enkelte vandføringsestimater for hver sumpfyldning inden integrering til timevandføring, som er den endelige træningsvandføring

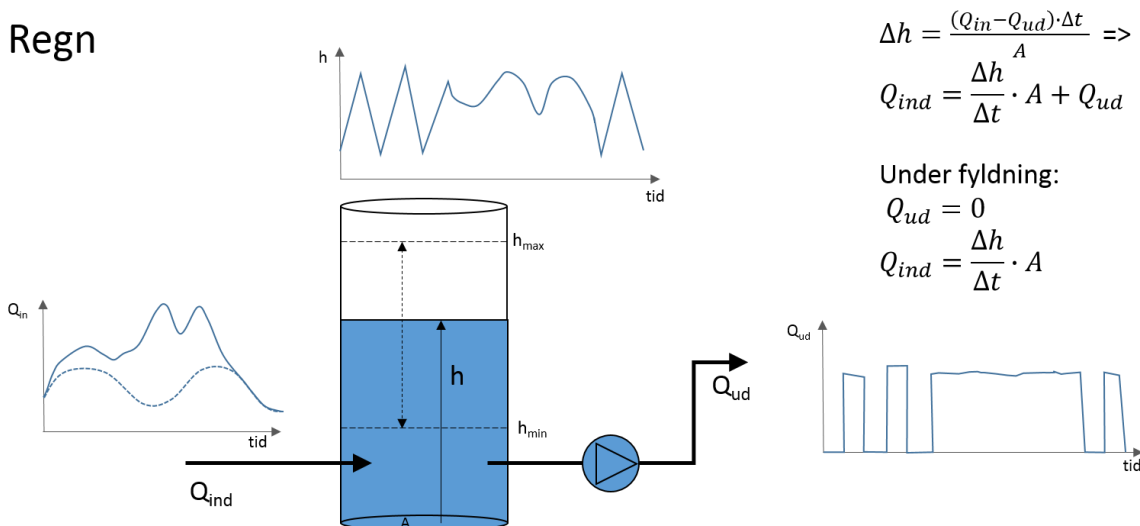
I metodeudviklingen anvendes data fra Tjebberup pumpestation, som er en forholdsvis stor station med 3 pumper i altdrift og et pumpeareal på 4,908 m². Grunden til dette valg er, at denne station også er udstyret med en vandføringsmåler, som måler pumpernes vandføring. At der er en vandføringsmåler installeret, gør det muligt at validere den sumpniveau-baserede træningsvandføring. Figur 5.4 viser en sammenligning mellem pumpestationens vandføringsmåling og træningsvandføringen baseret på vandstandsvariationen i pumpearealen. Det ses at vandføring estimeret ud fra vandstandsvariationen i pumpearealen giver en træningsvandføring, som er meget lig den målte fra stationen. Her skal det dog nævnes, at den målte vandføring også må forventes at være behæftet med en vis måleusikkerhed.



Figur 5.4: Sammenligning af timevandføring målt med vandføringsmåler og estimat baseret på pumpesumpvariation. NSE: Nash Sutcliffe performance. RMSE: Root Mean Square Error

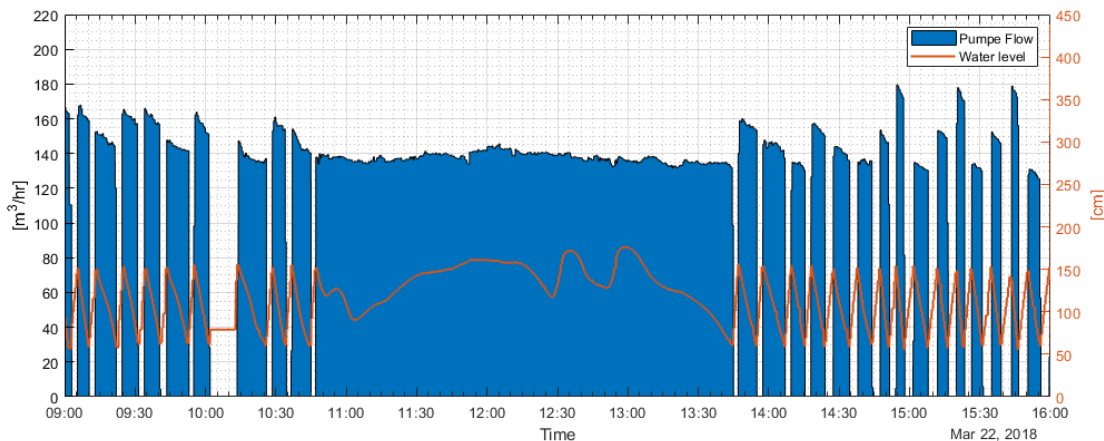
Vandføringen baseret på niveauvariation i pumpesumpen kan kun bestemmes for perioder, hvor sumper fyldes og pumperne er inaktive. I perioder med høje vandføringer f.eks. under regn er en eller flere pumper i drift i længere tid ad gangen. Dette er illustreret på principskitzen på figur 5.5

Regn



Figur 5.5: Principskitse for variation i indløbsvandføring, pumpevandføring og vandstandsvariation i en situation med høj vandføring f.eks. under regn.

Er indløbsvandføringen så høj at sumpen ikke tømmes, er det i denne periode ikke muligt at bestemme en vandføring ud fra variationen i pumpeumpen. Derfor kan træningsvandføringen ikke i sig selv anvendes som det endelige estimat for pumpestations timevandføring. Som eksempel på en sådan periode er der i figur 5.6 vist data fra Tjebberup pumpestation.



Figur 5.6: Eksempel på periode, hvor variationen i pumpeumpen ikke kan anvendes til at bestemme vandføringen. Pumpeumpen fyldes og tømmes fra kl. 9:00 til kl. 10:45 og igen 13:45 til 16:00, mens der i perioden fra 10:45 til 13:45 pumpes konstant uden at sumpen tømmes. Desuden ses et dataudfald fra ca. kl. 10:00 til 10:15

Som figur 5.6 viser, er der en periode på ca. 5,5 time, hvor tilløbsvandføringen er så høj, at én pumpe konstant er i drift. Da pumpernes vandføring i den reelle situation ikke er kendt, er det ikke muligt at bestemme tilløbsvandføring på baggrund af vandstandsvariationen i pumpeumpen. I sagens natur vil sådanne perioder være særligt interessante, da de optræder ved høje vandføringer og dermed også i perioder, hvor afløbssystemet er mest belastet.

Forudsætningen for at vandføringen kan bestemmes ud fra vandstandsvariationen i pumpeumpen er dog at vandstanden måles med en tilstrækkelig høj opløsning. Ved pumpestationen registreres pumpeumpens niveau med meget høj opløsning, da informationen anvendes til at styre udpumpningen. I projektet blev det dog erfaret, at data ikke nødvendigvis gemmes med samme høje opløsning. Ikke fordi det ikke er teknisk muligt, men fordi denne værdi af dataene ikke var taget i betragtning, da dataopsamlingen blev designet. Det typiske billede er, at data fra pumpestationer gemmes som 5 min, 10 min øjebliksværdier eller som time integrerede værdier.

For at overkomme dette, blev der i projektet af FORS leveret manuelle udtrækninger med 5 og 10 sek. øjebliksværdier for såvel niveaumålinger og pumpernes strømforbrug udenom deres normale dataopsamlingssystem, da systemet ikke på nuværende tidspunkt er i stand til at levere dette automatisk.

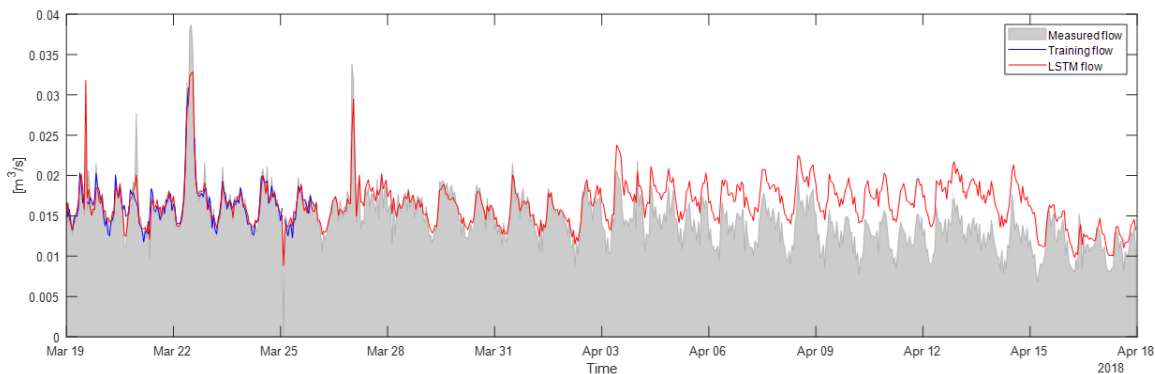
Det oprindelige koncept var at træne Machine Learnings modellen på baggrund af pumpeumpvariations vandføringen således, at den datadrevne model trænes på de tilgængelige træningsdata, og herved etablerer sammenhængen mellem pumpernes energiforbrug og vandføring. Herefter anvendes modellen til at estimere vandføringen både for perioder, hvor der ikke er tilgængelige træningsdata og nye perioder, der ikke er indeholdt i træningsdatasættet. Men i metodeudviklingen blev det konstateret, at den specifikke Machine

Learnings metode, som generelt havde det største potentiale, var et LSTM neuralt netværk (Long Short-Term Memory Neural Network) (Hochreiter og Schmidhuber, 1997). Da dette specifikke netværk er følsomt overfor 'huller' i træningsdata, blev der i metodeudvikling eksperimenteret i, hvordan perioder med manglende data kunne udfyldes med realistiske syntetiske trænings data.

Et LSTM-netværk er en specifik type af Neuralt Netværk indenfor Machine Learning, som er særligt velegnet, hvis de observationer man forsøger at estimere indeholder vedvarende- og autoregressive elementer. Dette er typisk tilfældet for en afstrømmende vandføring, hvor det er sandsynligt at hvis forrige observation er høj, så er den efterfølgende det også. LSTM-netværket har sin oprindelse inden for talegenkendelse, hvor et ords betydning i en sætning er afhængig af dets sammenhæng. LSTM-netværker indeholder et recessivt lag, som betyder, at netværkets estimat ikke kun baseres på det aktuelle input, men også estimerer fra de forrige tidsstridt. Derfor er det vigtigt, at træningsdataene indeholder realistiske tidsskalaer for datavariationen. Dette er grunden til, at det blev erfaret, at det var vigtigt, at perioder med manglende træningsvandføringer blev udfyldt med realistiske estimater for vandføringen. En række forskellige interpoleringsmetoder blev testet, men måden hvorpå det blev løst og implementeret i projektet, var ved at indbygge en foreløbig estimering baseret på et klassisk feedforward neuralt netværk. Dette netværk trænes på de tilgængelige træningsdata og anvendes til at udfylde perioder, hvor det ikke er muligt at estimere vandføringen på baggrund af niveauvariationen i pumpeumpen inden træningen af det endelige LSTM-netværk. Et feedforward neuralt netværk har mange ligheder med multi lineær regression, dog med den væsentlige forskel at sammenhængene ikke nødvendigvis er lineære. Kombinationen med et feedforward neuralt netværk til at finpudse træningsdataene inden LSTM netværkets træning, forbedrede LSTM netværkets præstations-evne betydeligt.

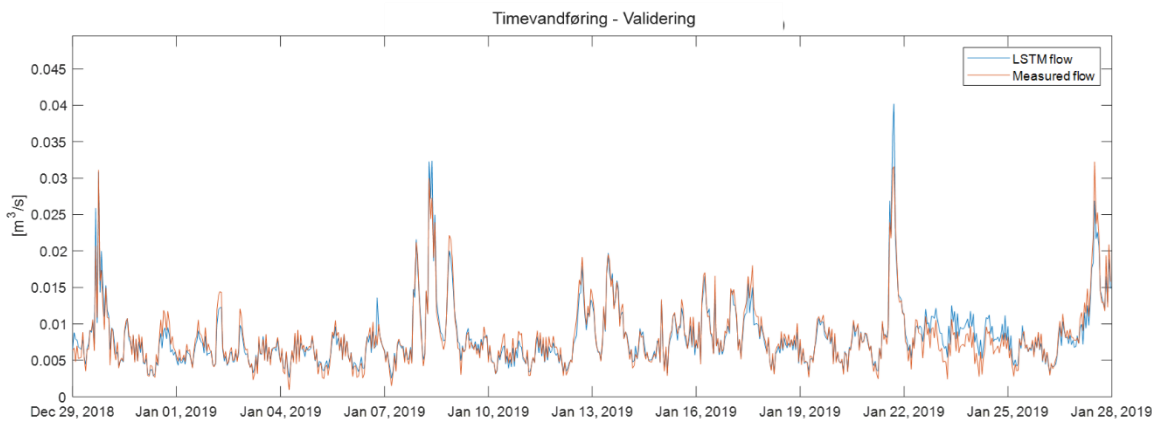
Det er kendt inden for Machine Learning, at både input og træningsdatas numeriske skalering kan have betydning for, hvor effektivt træningsalgoritmerne fungerer. Dette skyldes at træningsalgoritmerne er afhængig af, at input og træningsdataenes numeriske værdi har tilstrækkelig stor numerisk effekt på modeloutputtet førend, at træningen kan konvergere. Under metodeudviklingen blev der erfaret, at særligt vandføringstræningsdataene skalering havde betydning for LSTM netværkets træning, da vandføring med SI-enheden $m^3/sek.$ naturligt har meget lave numeriske værdier for de fleste pumpestationer. Derfor skaleres vandføringsdataene til at være imellem for 0 og 1 inden træning, da dette sikrer, at træningsalgoritmen kan konvergere effektivt. Omvendt blev det konstateret at skalering af input dataene (pumpernes strømforbrug) ikke havde nogen betydning, hvorfor der i implementeringen ikke foretages skalering af input.

Machine Learning modellers præstationsevne afhænger af mængden af træningsdata, hvorfor det har været relevant at undersøge hvor meget træningsdata der skal til, førend LSMT-netværket giver et vandføringsestimat, som har værdi. Det er forventeligt, at der skal en længere periode til, inden LSTM-netværket kan betragtes som værende generel. Men som det ses af figur 5.7, kan en træningsperiode på blot en uge være nok til at give den første umiddelbare værdi. På figuren er vist en periode på en måned fra 19 marts til 19 april 2018 fra Tjebberup pumpestation. De første 7 dage er anvendt til træning af LSTM-netværket, mens de resterende dage er estimeret på baggrund heraf. Det ses at LSTM-netværket med høj præcision estimerer vandføringen den efterfølgende uge, hvorefter det dog også ses, at LSTM netværket ikke kan følge det generelle fald i vandføring som observeres. Dette demonstrerer, at opsættes applikationen for en ny pumpestation, kan LSTM netværket efter blot en uges opsamlede data begynde at estimere vandføringen for den næste umiddelbare periode. Efter yderligere en uge er træningsdatasættet fordoblet, og LSTM-netværket kan gentrænes på dette grundlag til anvendelse for den næste periode osv. Dette kan gentages indtil, at LSTM netværkets læring er generel.



Figur 5.7: LSTM netværkets præstationsevne med blot en uges træningsdata. Rød kurve er af LSTM netværkets estimerede vandføring, blå kurve er træningsvandføringen og grå kurve er den målte vandføring. Data er fra Tjebberup pumpestation.

Hvornår det kan forventes, at netværket er trænet til et generelt niveau afhænger af, hvor meget variation i vandføring træningsdataene indeholder. Men for de undersøgte pumpestationer er det erfaringen at mellem 2 til 3 måneders træningsdata er nok til, at LSTM-netværket kan estimere perioder med såvel høje som lave vandføringer. Eksempel herpå er vist på figur 5.8, hvor en periode på en måned fra den 29. december 2018 til 29 januar 2019 er estimeret på baggrund af to måneders træningsdata fra hhv. 19 marts til 19 april 2018 og 27 oktober til 27 november 2018.



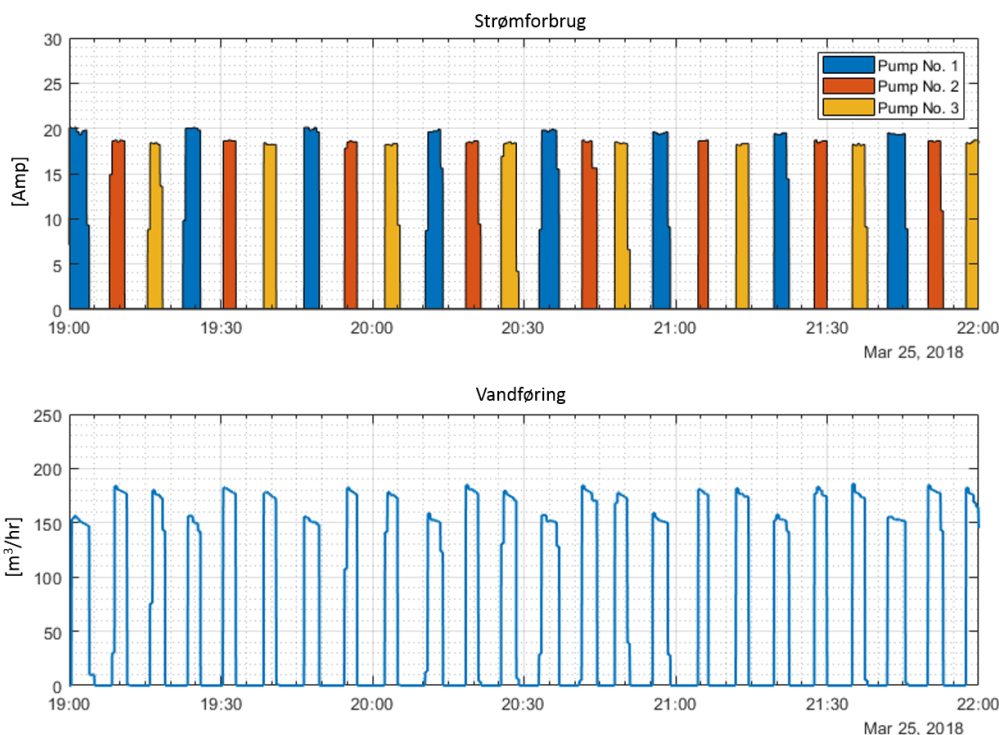
Figur 5.8: Validering af LSTM-netværk for Tjebberup pumpestation. Netværket er trænet på 2 måneders data fra 19/3 2018 – 19/4 2018 og 27/10 2018 – 27/11 2018.

Selve softwaren, som er udviklet til estimering af pumpestationers vandføring, er én samlet applikation, der indeholder samtlige elementer. Dvs. fra generering af træningsdata på baggrund af højopløselige målinger af vandstands niveau og strømforbrug til opbygning og træning af et LSTM-netværk for pumpestationen til efterfølgende estimering af timevandføring på baggrund af timeværdier for pumpernes strømforbrug. Applikationen er integreret i EnviDan's data portal, hvor den afvikles og resultater visualiseres for brugeren. Udover selve vandføringsestimatet beregner applikationen også daglige nøgle parametre KPI (Key Performance Indicators) bestående af daglig volumen og daglig strømforbrug fordelt pr pumpe.

Overvågning af pumpestationers driftstilstand

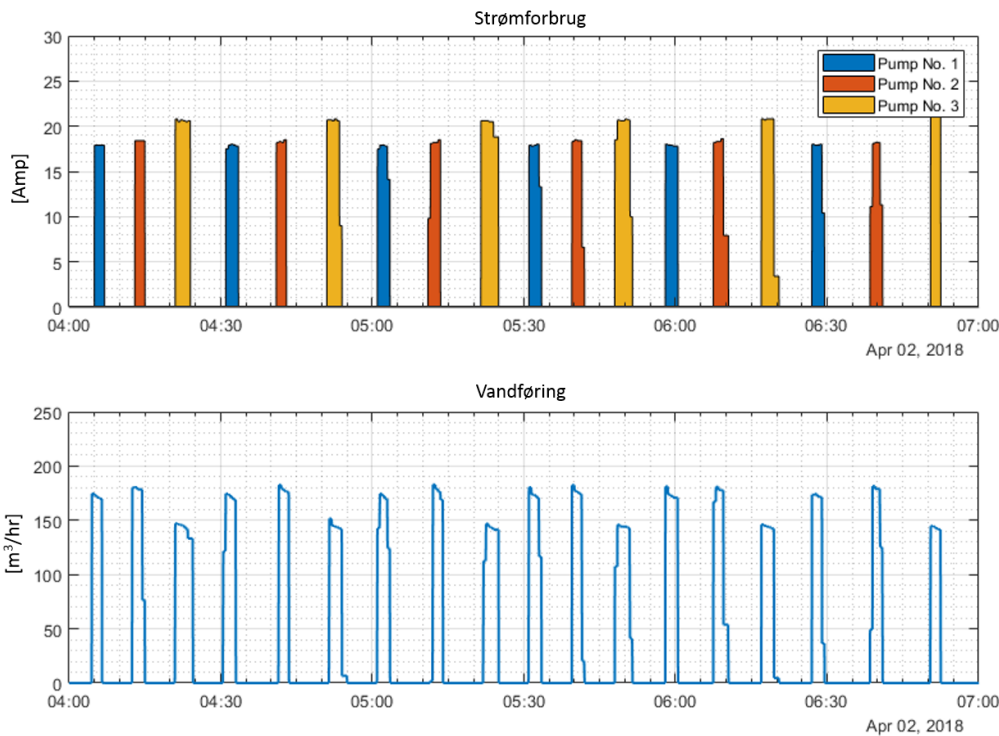
Pumperne i en pumpestation kræver løbende vedligeholdelse for at sikre, at de fungerer efter hensigten. Pumperne opererer ofte i et aggressivt miljø og pumpehjul og pumpens elektriske dele udsættes for slitage. Større afløbssystemer indeholder ofte mange pumpestationer med flere pumper i hver og driften af disse er som oftest en betydelig opgave i forsyningen.

I projektet har det været et delmål at undersøge, om det med baggrund i eksisterende data er muligt at opnå realtids viden om pumpestationernes tilstand, som kan anvendes i forsyningens daglige arbejde med drift og vedligeholdelse af pumpestationer.



Figur 5.9: Data for strømforbrug og målt pumpevandføring fra Tjebberup pumpestation den 25. marts 2018. Stationen indeholder 3 pumper som kører alternerende. Det ses at pumpe 1 bruger lidt mere strøm end de andre pumper, men yder ca. 20% mindre vandføring.

Ved at undersøge de højopløselige pumpedata og vandføringsmålinger i projektet, er det konstateret, at der er et potentiale for realtids overvågning, hvis disse data gøres tilgængelige i realtid. På figur 5.9 ses data fra den 25. marts 2018 for Tjebberup Pumpestation. Her ses det, at pumpe 1 er betydelig mindre effektiv end de øvrige. Pumpe 1 bruger lidt mere strøm end de andre, når den er aktiv, men yder ca. 20% mindre end de andre.



Figur 5.10: Data for strømforbrug og målt pumpevandføring fra Tjebberup pumpestation den 2. april 2018. Stationen indeholder 3 pumper, som kører alternerende. Det ses at det nu er pumpe 3, som bruger mest strøm og kører længere tid end de andre, samtidig med at den yder ca. 20% mindre vandføring.

Ca. en uge senere d. 2. april har situationen ændret sig. Nu er det pumpe 3, som er mindst effektiv. Af figur 5.10 ses det at pumpe 3, selvom den bruger mest strøm af de 3 pumper, leverer den mindste vandføring, og den er nødt til at køre længere end de øvrige for at tømme pumpebassinet. Hvad årsagen er til, at pumpernes ydelse varierer mellem de to perioder er ukendt, men det formodes at være mest sandsynligt, at pumpernes sugeledning har været delvist tilstoppede. Det forventes ikke, at slitage kan have så varierende effekt over denne korte periode.

Af figur 5.9 og 5.10 ses det, at der er et potentiale for overvågning af pumpernes driftstilstand i denne type af data. Det har i projektet været forsøgt, om det har været muligt at estimere vandføringen med samme høje opløsning for pumpestationer, som ikke har installeret en vandføringsmåler. Det har dog ikke været muligt med en tilstrækkelig præcision. Udfordringen ligger i at hvis pumpernes effektivitet skal evalueres, er det ændringerne i sammenhængen mellem pumpernes strømforbrug og vandføring, som undersøges. Det er således vigtigt at de to observationer forbliver uafhængige. Anvendes Machine learning til at estimere vandføringen ud fra pumpeens strømforbrug for stationer uden vandføringsmålere, forsvinder denne uafhængighed.

Et muligt alternativ til en vandføringsmåler, vil være at monitorere pumpernes individuelle driftstryk ved hjælp af tryktransducere. Pumpeens driftstryk kan da anvendes som mål for pumpeens ydelse. Dette er dog ikke undersøgt nærmere i projektet.

Selvom dataene viser, at det er muligt at identificere variationer i pumpernes effektivitet ud fra højopløselige pumpe og vandføringsdata, er det stadig meget svært på baggrund af disse data alene at vurdere, om pumpen trænger til vedligeholdelse eller bør tilses. En gren inden for Machine Learning er klassifikation, som er meget relevant til denne applikation. Dog kræver dette samholdende observationer mellem pumpeeffektivitet og fysiske tilstand (f.eks. 'ny pumpe', 'lidt slid', 'medium slid', 'meget slid', 'tilstoppet' osv.), før det er muligt at træne en klassifikationsmodel. Sådant information har ikke været tilgængelig, og det har ikke været muligt at generere et sådant dataset inden for rammerne af projektet. Det konkluderes dog med baggrund i de foretagende analyser, at der ligger et uudnyttet potentiale inden for dette område, som er relevant for fremtidige projekter.

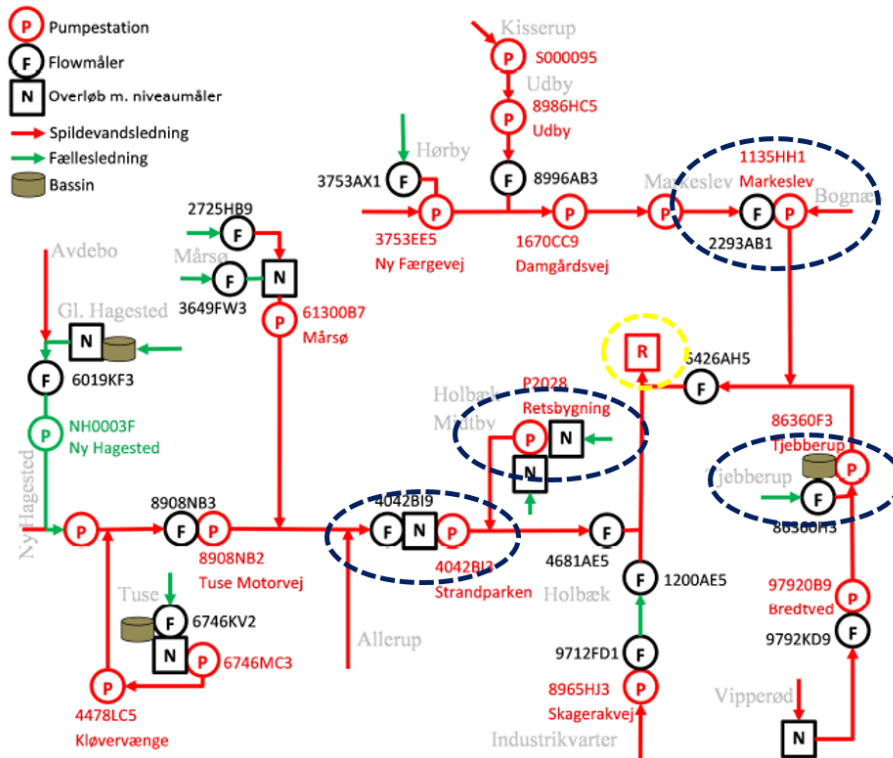
Estimering og forudsigelse af tilløbsvandføringen til Holbæk renselanlæg

Aktuel og opdateret viden om tilløbsvandføringen til renselanlæg er afgørende for renselanlæggets drift. Typisk er det den hydrauliske belastning, som er den begrænsende faktor for renselanlægget, hvorimod den stofmæssige belastning sjældent er kritisk. Derfor er der i projektet fokuseret på, om og hvordan det er muligt at anvende observationer fra pumpestationer i oplandet til Holbæk renselanlæg til både at estimere og forudsige tilløbsvandføringen til renselanlægget. Konsekvensen af hydraulisk overbelastning vil være midlertidig bypass af en del af det tilløbende spildevand til eventuelle bufferbassiner, hvor spildevandets magasiner indtil, der igen er kapacitet på anlægget, og i yderste konsekvens nødoverløb til recipienten.

Ved at estimere den aktuelle tilløbsvandføring i realtid ved hjælp af en Machine Learning baseret softwaresensor er det muligt at validere den målte tilløbsvandføring, ligesom at vandføringsestimateret også kan anvendes som den primære datakilde i tilfælde af, at vandføringsmåleren fejler eller på anden vis ikke fungerer efter hensigten.

I kortere perioder er det muligt at omstille renselanlæggets drift, sådan at anlægget kan håndtere en større hydraulisk belastning. Ved at omstille driften til 'regnstyring', som denne drift tilstand betegnes, vil anlægget i en kortere periode kunne håndtere en større spildevandmængde. Som forberedelse til regnstyring flyttes en større del af den aktive slam fra anlæggets processtanke til efterklaringstankene end normalt, hvorefter returstrømmen fra efterklaringstankene kan øges i en periode. Derved opnås en midlertidig forøget hydrauliske kapacitet i anlægget. En forudsætning for omstilling til regnstyring, er dog at den aktive slam flyttes før den øgede hydrauliske belastning optræder. Hvorfor at en effektiv regnstyring forudsætter at både den aktuelle hydrauliske belastning og belastningen i den nære fremtid kendes, således der er tid til at flytte den aktive slam. Tiden det tager at flytte den aktive slam til efterklaringstankene afhænger af anlæggets størrelse, men som oftest er den nødvendige tidshorisont 1-3 timer.

Traditionelt vil beslutning om at omstille anlægget til regnstyring bero på en radarbaseret nedbørsprognose kombineret med en hydraulisk model for oplandet til renselanlægget. I dette projekt er tilgangen en anden. Her har målet været at forudsige tilløbet udelukkende ud fra informationer fra pumpestationer i oplandet. Komplexiteten af prognosen for tilløbsvandføringen er ved denne PUFDO-tilgang væsentligt reduceret, hvilket potentielt kan give en mere robust forudsigelse. Ved at anvende de aktuelle pumpestations observationer i systemet elimineres eventuelle usikkerheder som naturligt er indeholdt i radarbaserede nedbørsforudsigelser og den hydrauliske model for oplandet, da tilløbsforudsigelsen udelukkende er baseret på systemets aktuelle tilstand og situation.



Figur 5.11: Forsimplet og konceptuel skitsering af afløbssystemet til Holbæk renselanlæg. Renselanlægget er markeret med 'R' i den gule cirkel. 'P' markerer systemets pumpestationer. Pumpestationer markeret med blå cirkler er stationerne som anvendes til at estimere og forudsige tilløbet til renselanlægget. Bemærk: 'N' og 'F' markerer midlertidige vandførings (flow) og niveau målinger, som blev fortaget i systemet af Fors i anden sammenhæng.

Figur 5.11 illustrerer afløbssystemet til Holbæk renselanlæg som både indeholder separat og fælles kloakerede deloplande. Som det ses, indeholder systemet en række pumpestationer, som tilsammen bidrager til tilløbsvandføringen til Holbæk renselanlæg, som på figuren er markeret med 'R' og en gul cirkel.

I en traditionel hydraulisk model skal tilløbet til renselanlægget beskrives ud fra samtlige delbidrag, da en sådan model er baseret på massebalance i systemet. Dette er ikke nødvendigvis et krav for datadrevne Machine Learnings modeller, de disse er baseret på datasammenhænge i systemet, hvorfor det i projektet har været muligt at basere estimering og forudsigelse af tilløbet til renselanlægget ud fra pumpestationsdata fra stationerne markeret med blå cirkler; Tjebberup, Retsbygningen, Strandparken og Markerslev pumpestation. Datagrundlaget for de datadrevne modeller er allerede eksisterende SRO-time data for de nævnte stationer kombineret med den målte time tilløbsvandføring til renselanlægget som træningsdata.

Til denne Machine Learnings applikation blev det forsøgt at anvende samme type datadrevne LSTM neuralt netværks model, som for estimering af pumpestationsvandføring. Det blev dog i projektet erfaret, at enklere datadrevne modeller var mere anvendelige til denne

applikation. Under modeludviklingen blev det identificeret at Machine Learning med regressions-træer (regression trees) (Breiman, 1996) meget effektiv er i stand til at identificere datasammenhængen i systemet til estimering og forudsigelse af tilløbsvandføringen til renseanlægget. Regressions-træer er en udbygning af de mere simple klassifikation-træer, der som udgangspunkt er udviklet til dataklassifikation. Princippet bag regression-træer er at i stedet for at klassificere data i diskrete klasser eller grupperinger, identificerer og estimerer regression-træer kontinuerte variationer ved underinddeling af såvel de drivende input (i dette tilfælde pumpestations data) og det estimerede output (i dette tilfælde tilløbsvandføring). I algoritmen foretages underinddelingen med forskellig detaljeringsgrad, således at der bag metodens estimat ligger en række regression-træer, som hver især giver et estimat for tilløbsvandføringen. Dette ensemble af estimater kombineres ved vægtning til det endelige estimat, hvorfor den anvendte Machine Learning metode også benævnes 'ensemble of regression trees'.

I applikationsudviklingen er der taget udgangspunkt i de eksisterende SRO pumpedata fra systemet, og det har været målet at arbejde med en repræsentativ og realistisk situation. Det er f.eks. ikke forventeligt for et virkeligt system, at alle pumpestationer indeholder de samme typer data. Samtidig er det også meget sandsynligt, at der for dele af det bidragende system ikke er tilgængelige observationer af den ene eller anden årsag. Som det ses af figur 5.11, løber størstedelen af spildevandet, som ender på renseanlægget gennem pumpestationerne, som er inkluderet. Men det ses også, at der er bidrag, som ikke inkluderes. Selvom metoden og applikation er udviklet på forholdene omkring Holbæk renseanlæg, har det ikke været målet at arbejde med den 'perfekte' datasituation, hvorfor det vurderes, at metoden og resultaterne heraf er generelle og kan overføres til andre systemer af tilsvarende karakter. Indenfor projektets ramme har det ikke været muligt ej heller målet at teste metoden på andre systemer men derimod at sandsynliggøre overførbare til andre systemer.

Som drivende input til estimering og forudsigelse af tilløbsvandføringen anvendes som beskrevet eksisterende SRO-time data fra Tjebberup, Retsbygningen, Strandparken og Markerslev pumpestation. På disse 4 stationer er der installeret vandføringsmålere. Det er valgt at anvende disse vandføringer direkte i applikationen, da de er tilgængelige, men det er vigtigt at pointere, at dette principielt ikke er en forudsætning, da den LSTM neuralt netværks baserede pumpestations vandføring lige så godt kunne have været anvendt. For Tjebberup, Retsbygningen, Strandparken anvendes også pumpernes individuelle strømforbrug, men disse data er ikke tilgængelige for Markerslev. Som et ekstra datainput anvendes også time på døgnet (angivet som heltal mellem 0 og 23), da der forventes at være systematiske variationer i tilløbsvandføringen i tørvejrperioder, som er styret af den generelle døgnvariation af spildevandsproduktionen i oplandet. Det blev erfaret, at inkludering af dette ekstra input reducerede tidlig forskydning af såvel estimat og forudsigelse af tilløbsvandføringen til renseanlægget. Tabel 5.1 opsummerer de i alt 13 forskellige anvendte datatyper.

Tjebberup	Strandparken	Retsbygningen	Markerslev	Ekstra	Træningsdata
Strøm, pumpe 1	Strøm, pumpe 1	Strøm, pumpe 1	Vandføring	Time på døgnet	Tilløbsvandføring til renseanlæg
Strøm, pumpe 2	Strøm, pumpe 2	Strøm, pumpe 2			
Strøm, pumpe 3	Vandføring	Vandføring			
Vandføring					

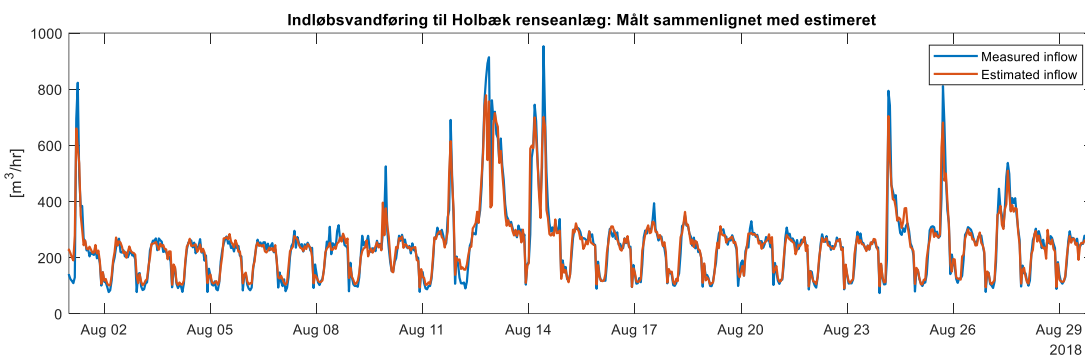
Tabel 5.1: Anvendte datatyper til estimering og forudsigelse af tilløbsvandføringen til Holbæk renseanlæg.

Til modeltræning anvendes en periode med 1 års tidedata udtrukket fra Fors SRO-system. Som sikring mod overfitting af modellen, anvendes k-fold valideringsmetoden. Dette er en anerkendt metode designet til formålet. Metoden består af at dataene som anvendes til

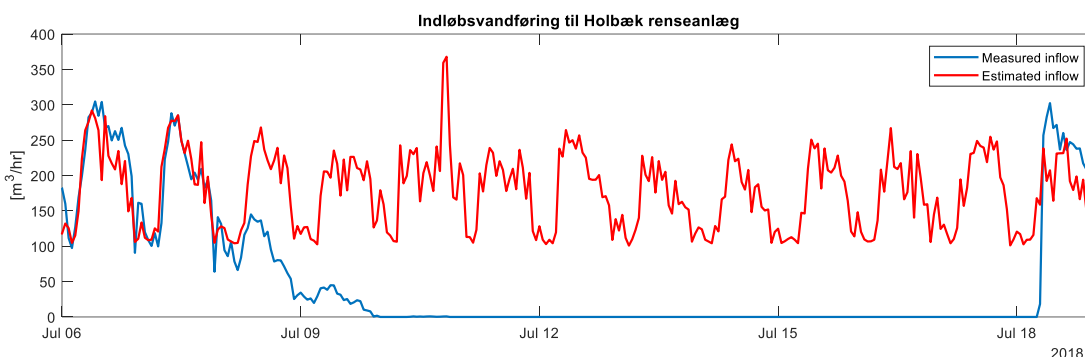
modeltræning inddeles i individuelle perioder (folds), som systematisk udvælges således at kun en del af datasættet anvendes til træning, mens en anden udvælges til validering. Udvalgelsen sker gentagende gange med forskellige kombinationer, som alle evalueres, hvilket sikrer en generel model, som har samme præcision for samtlige dataperioder.

Der opbygges i alt 6 forskellige datadrevne modeller i applikationen; én model til at estimere den aktuelle tilløbsvandføring og 5 til at forudsige vandføringen med forskellige forudsigelses horisont. Modelopbygning og træningsmetode er identisk for alle 6. Forskellen ligger i, at for forudsigelsesmodellerne er tilløbsvandføringen målingen tidsforskudt i forhold til pumpestationsdataene med hhv. 1 til 5 timer. Dette gør at i stedet for at estimere den aktuelle vandføring, estimerer forudsigelsesmodellerne tilløbsvandføringen med en forudsigelses længde på hhv. 1 til 5 timer. Når alle 6 modeller er trænet præsenteres de 6 modeller for nye pumpestationsdata, hvorved kombinationen af de 6 modeller giver et estimat af den aktuelle tilløbsvandføring og en forudsigelse 5 timer frem.

Analyser af såvel aktuelle estimater og forudsigelser viser, at der er rigtig god overensstemmelse mellem den målte vandføring og estimeret aktuelle vandføring. Figur 5.12 viser en dataperiode fra den 1. til 30. august 2018, hvor den målte indløbsvandføring er sammenlignet med aktuelle estimatet baseret på pumpestationsdata fra de førnævnte stationer. Som det ses, er applikationen på meget realistisk vis i stand til at estimere indløbsvandføringen til Holbæk renseanlæg.

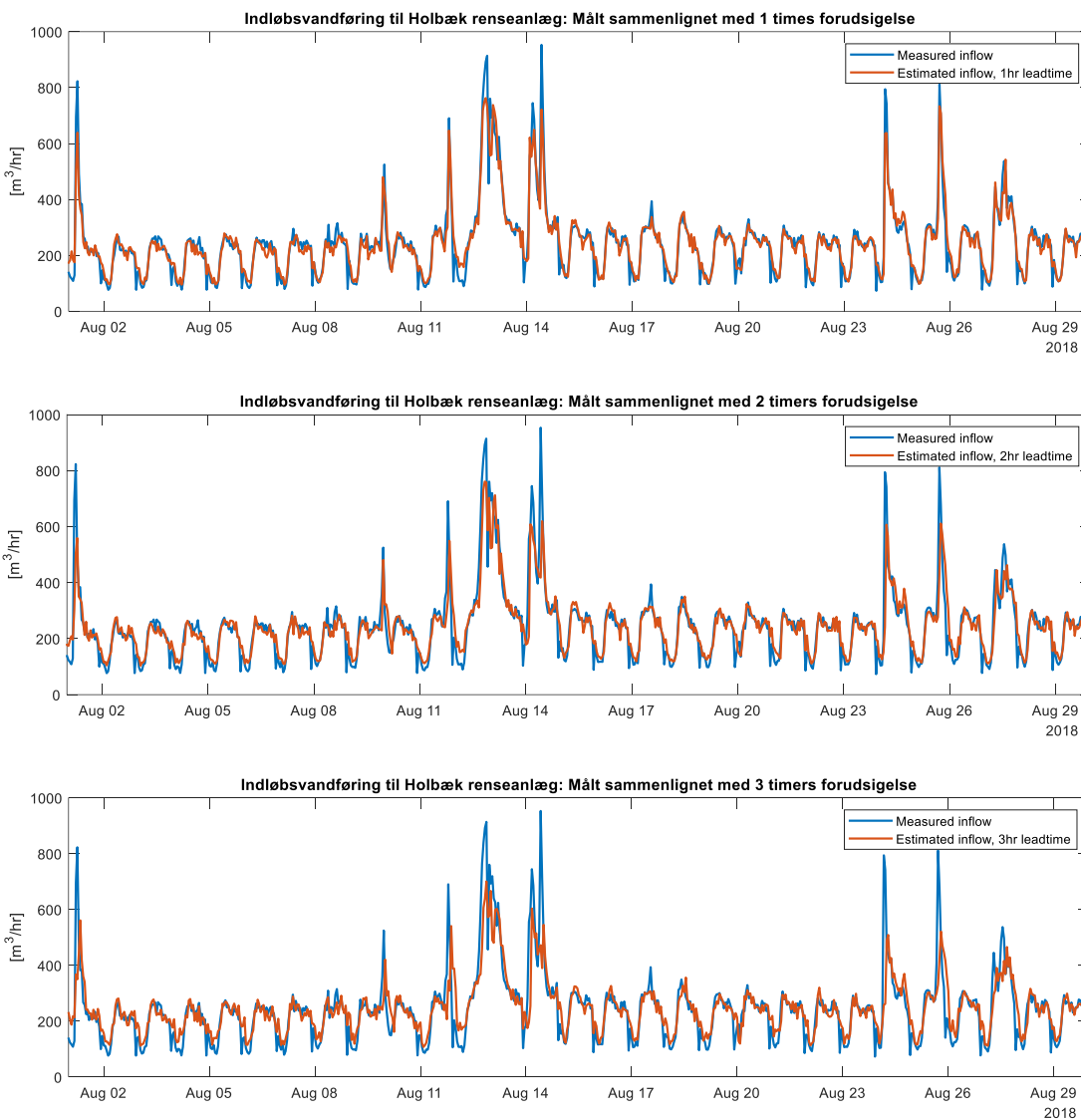


Figur 5.12: Målt og estimeret indløbsvandføring til Holbæk Renseanlæg i perioden 1. til 30. august 2018



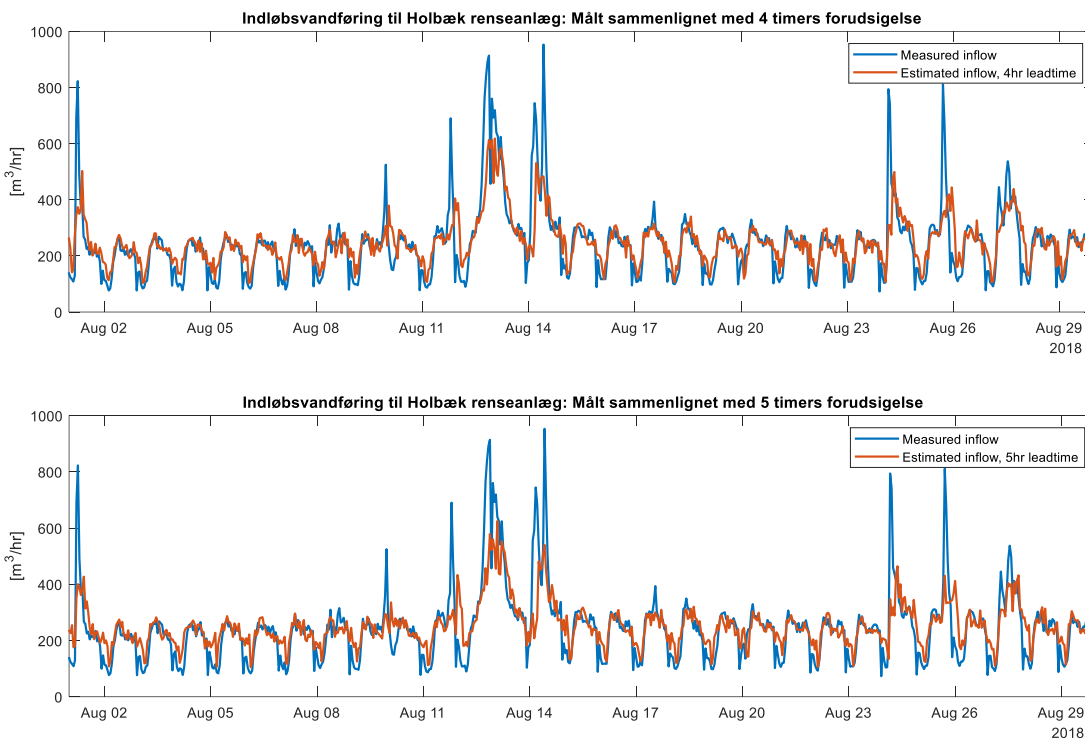
Figur 5.13: Indløbsvandføring til Holbæk renseanlæg i perioden 6. til 10. juli 2018. Perioden indeholder udfald på vandføringsmålingen, hvorunder Machine Learnings applikationen på realistisk vis estimerer tilløbsvandføringen.

Figur 5.13 illustrerer en dataperiode, hvor der er udfald på vandføringsmålingen af tilløbet til Holbæk renselanlæg. Den specifikke årsag til dataudfaldet kendes ikke, men eksemplet illustrerer tydeligt potentialet ved at have en online Machine Learning estimering af tilløbsvandføringen, der kan anvendes som backup for perioder med manglende observationsdata.



Figur 5.14: Målt indløbsvandføring og forudsigelse til Holbæk renselanlæg med hhv. 1,2 og 3 times forudsigelse. Blå kurve repræsenterer den målte indløbsvandføring, mens røde kurve repræsenterer hhv. 1,2 og 3 timers forudsigelse af indløbsvandføringen. Dvs. at enhver værdi på de røde kurve er estimeret hhv. 1,2 og 3 timer før den blå kurves observationsvandføring til samme tidspunkt.

For samme data periode som præsenteret i figur 5.12 (1. til 30. august 2018) viser figur 5.14 og figur 5.15 forudsigelsesvandføringen i samme periode. Figur 5.14 indeholder hhv. 1, 2 og 3 timers forudsigelsen, mens figur 5.15 indeholder 4 og 5 timers forudsigelsen. De blå kurver repræsenterer den målte indløbsvandføring, mens de røde kurver repræsenterer 1, 2, 3, 4 og 5 times forudsigelsen. Dvs. at enhver værdi på de røde kurver er bestemt hhv. 1 til 5 timer før den blå observationsvandføring til samme tidspunkt. Generelt ses en særlig god overensstemmelse for 1 til 3-timers-forudsigelsen. For 4 og 5-timers-forudsigelsen falder præcisionen, men det vurderes at selv 5-timers-forudsigelsen har klar værdi. Det skal i denne sammenhæng pointeres, at det er forventeligt at præcisionen vil falde med forudsigelseslængde. Men det bør også nævnes at faldet i præcision med forudsigelseslængden er væsentligt mindre end forventet.



Figur 5.15: Målt indløbsvandføring og forudsigelse til Holbæk renselanlæg med hhv. 4 og 5 times forudsigelse. Blå kurve repræsenterer den målte indløbsvandføring, mens rød kurve repræsenterer hhv. 4 og 5 timers forudsigelse af indløbsvandføringen. Dvs. at enhver værdi på de røde kurver er estimeret hhv. 4 og 5 timer før den blå kurves observationsvandføring til samme tidspunkt.

Selve softwaren, som er udviklet til estimering og forudsigelse af indløbsvandføringen til Holbæk renselanlæg, er én samlet applikation, der indeholder samtlige elementer. Dvs. applikationen kan på baggrund af træningsdata opbygge og træne de i alt 6 forskellige data-drevne modeller, som derefter anvendes til at estimere den aktuelle indløbsvandføring samt 1 til 5 timers forudsigelse af denne.

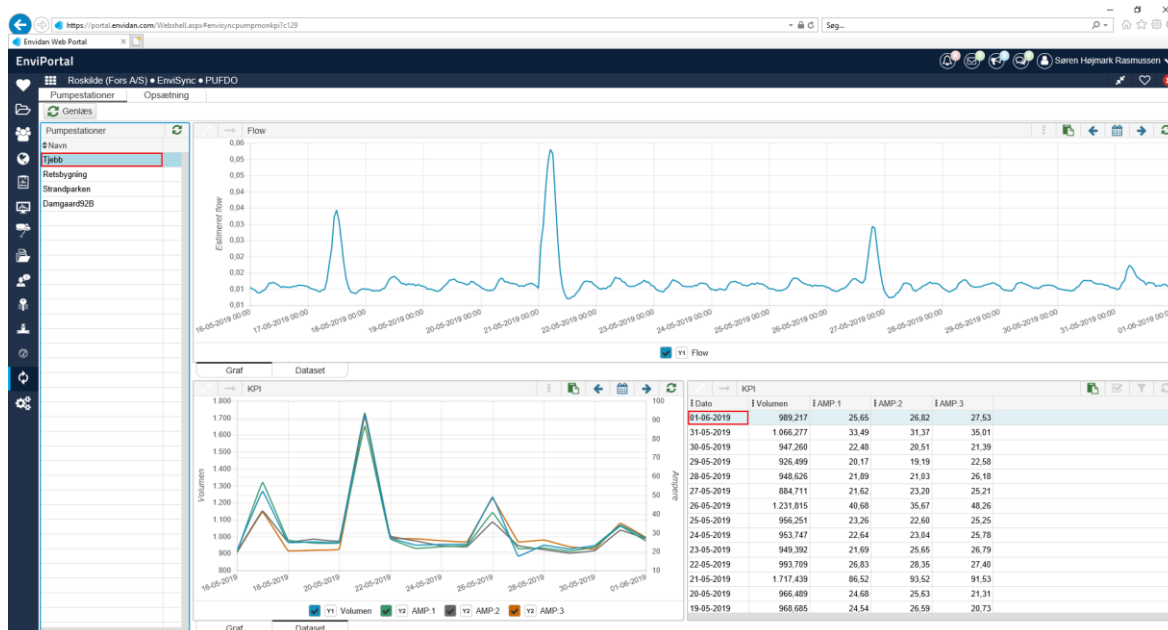
5.3 Projektresultater

Projektets slutresultat er de udviklede software applikationers implementering i EnviDans data portal, som afvikler applikationerne og visualiserer resultaterne fra de datadrevne modeller for slutbrugeren.

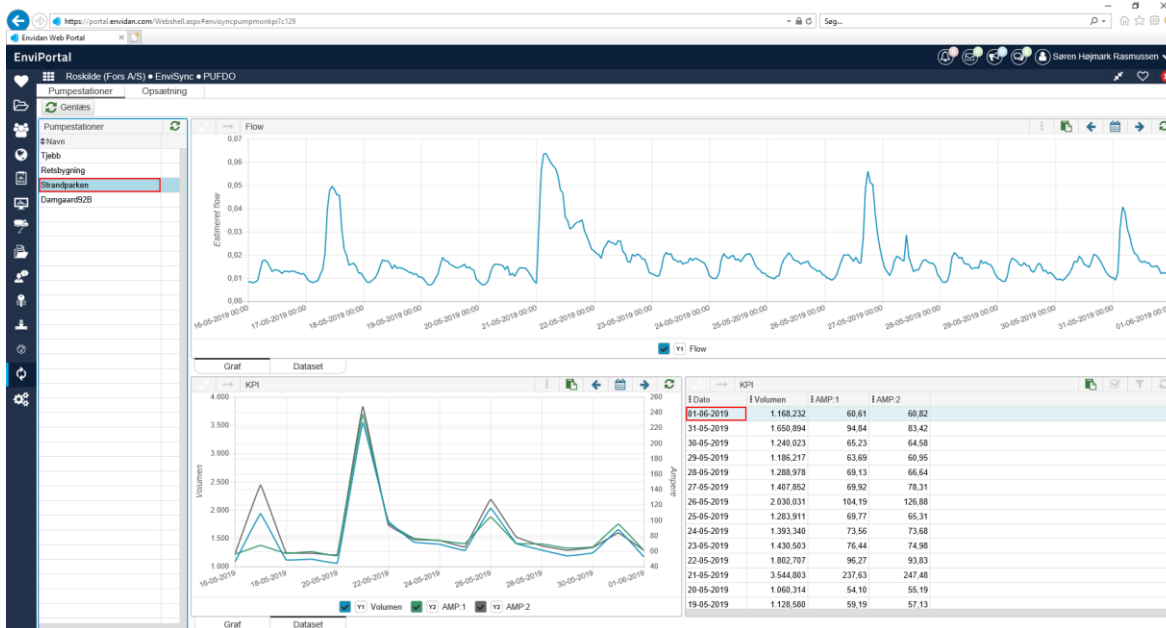
Estimering af pumpestationsvandføring er i projektet implementeret for i alt 4 pumpestationer i Holbæk renseanlægs opland: Tjebberup, Strandparken, Retsbygningen og Damgaardvej pumpestation. Tabel 5.2 opsummerer disse pumpestationer og figur 5.16 til figur 5.19 indeholder skærmpoint fra EnviDans dataportal.

Pumpestation	Antal pumper	Pumpesump areal [m ²]
Tjebberup	3	4,908
Strandparken	2	2,900
Retsbygningen	2	4,908
Damgaardvej	2	9,817

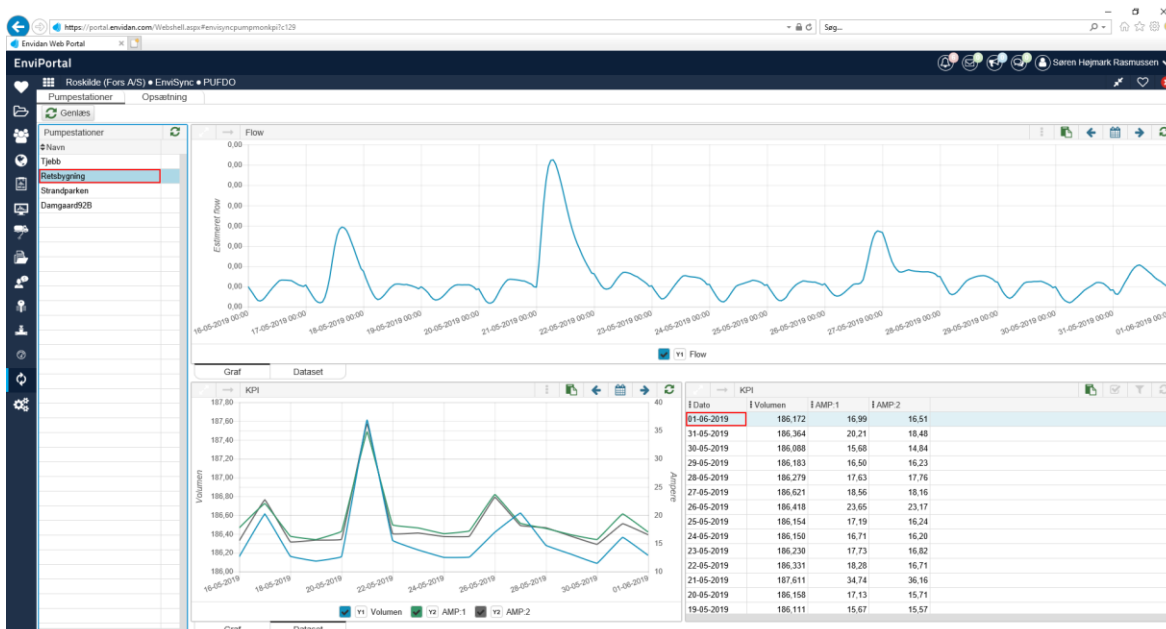
Tabel 5.2: Antal pumper og pumpesumpareal



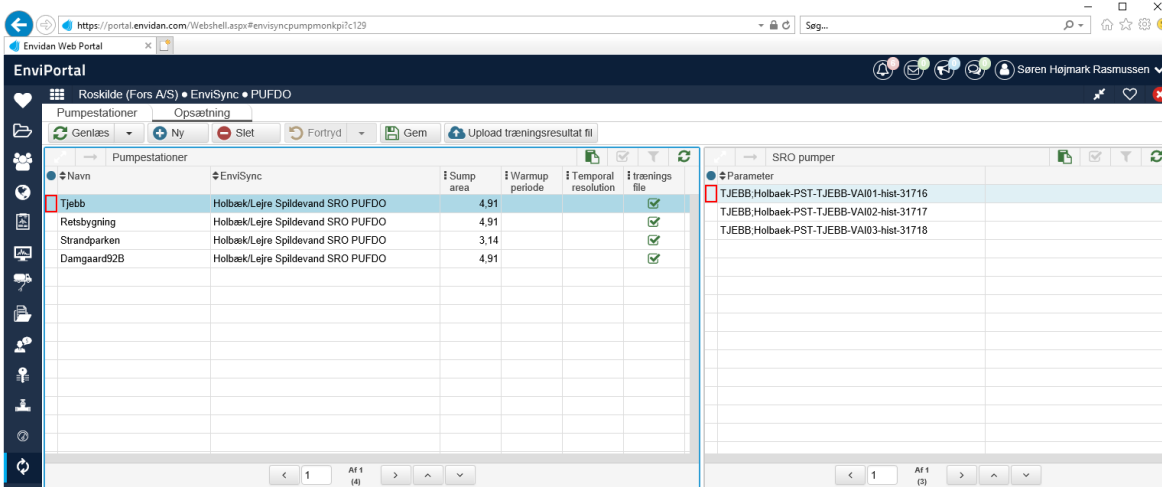
Figur 5.16: Skærmpoint fra EnviDans dataportal med visualisering af estimeret vandføring gennem Tjebberup pumpestation. Øverst tidsserie for vandføringen, nederst præsentation af KPI-parametre.



Figur 5.17: Skærmprent fra EnviDans dataportal med visualisering af estimeret vandføring gennem Strandparken pumpestation. Øverst tidsserie for vandføringen, nederst præsentation af KPI-parametre.



Figur 5.18: Skærmprent fra EnviDans dataportal med visualisering af estimeret vandføring gennem Retsbygningen pumpestation. Øverst tidsserie for vandføringen, nederst præsentation af KPI-parametre.



Figur 5.19: Skærmsprint fra EnviDans dataportal med opsætning af pumpestationer til PUFDO. Til højre angives input strømforbrug pr. pumpe på pumpestationen. Øverst i midten kan der uploades trænet LSTM-netværk.

PUFDO har vist potentiale for driftsovervågning gennem information fra højopløst data, jf. Figur 5.9 og 5.10.

PUFDO har demonstreret, at det er muligt at forudsige indløb til renseanlægget, jf. Figur 5.14 og 5.15.

5.4 Konklusion

Projektet har vist, at det er muligt at bruge primære SRO-data (primært strømforbrug) fra en pumpestation til at forudsige vandføringen ud af pumpestationen ved hjælp af Machine Learning.

Sammenlignes beregningerne med vandføringsmålinger (både direkte og indirekte vandføringsmålinger) ses en meget god overensstemmelse mellem model og målinger. Ved at bruge andre observationer, så som vandstand og start/stop tidspunkter for pumperne er det muligt at beregne estimater på vandføring, som kan bruges i Machine Learning algoritmerne, uden at man skal installere vandføringsmålere til formålet.

Projektet har også demonstreret, at man kan forudsige vandføring i centrale punkter i afløbssystemet (f.eks. indløb til renseanlægget) baseret på målinger ude i afløbssystemet. Som en interessant pointe kan det ses, at man ikke behøver at bruge data fra alle pumpestationer for at beregne tilløbsvandføringen. Dette viser, hvordan en Machine Learning tilgang kan være klassiske modeller, som Mike Urban, overlegen til denne type forudsigelse. En klassisk afløbsmodel vil altid kræve et komplet datasæt for f.eks. at estimere tilløbsvandføring.

Gennem arbejdet med de højt tidsopløst data har der vist sig potentiale til at overvåge de enkelte pumper i en pumpestation og vurdere driftstanden af hver enkelt pumpe. Det er et potentiale for at udpege at individuelle pumper fungerer mindre effektivt. Dette potentiale og datatype vil have særlig fokus til videre arbejde i kommende projekter.

Det kan konkluderes, at den anvendte LSTM-model er særdeles velegnet til at lave vandførberegninger med. Modellens evne til at medtage både foregående og nuværende observationer har vist sig meget kapabel.

Løsningen er implementeret på EnviDan's portal og leverer hvert døgn opdateret flowberegninger. Det muliggør, at man kan få et overblik over vandføringen i en række centrale punkter i afløbssystemet. Systemet kan derfor udbredes til andre forsyninger og til andre anvendelser.

Det har ligeledes vist sig, at måden hvorpå data opsamles og gemmes af forsyningerne er af meget stor vigtighed for at opnå gode resultater. Der skal gemmes data med en høj tidslig opløsning og med en så lang historik som muligt. Des flere data der er gemt, des bedre er algoritmerne til at finde sammenhængen mellem strømforbrug og vandføring. Mindst 2 måneders data har vist sig nødvendig for en god justering af modellen. Dog vil modellen løbende forbedres, når der måles – også selvom man ikke har gemt lange tidsserier.

Sidst, men ikke mindst, har metoden vist sig anvendelig til at lave forudsigelser af f.eks. tilstrømning til et renseanlæg. I det konkrete tilfælde var det muligt at levere forudsigelser på op til 3 timer før hændelsen observeres ved renseanlægget. Dette kan bruges til at forbedre driften af renseanlægget. Selv ved 5 timers forudsigelse er resultatet meget brugbart.

Litteraturliste

1. Breiman, Leo (1996) "Bagging predictors" Machine Learning 24.2, s.123-140.
2. Hochreiter, Sepp og Schmidhuber, Jürgen (1997): "Long short-term memory" Neural computation 9.8, s.1735-1780.
3. Kallesøe, Carsten Skovmose og Knudsen, Torben (2016): "Self Calibrating Flow Estimation in Waste Water Pumping Stations", 2019 European Control Conference (ECC). IEEE.