Prognosemodel for Gudenåen

25-05-2022

Machine Learning prognose af vandstand – Proof-of-Concept

Projektet er gennemført på vegne af Silkeborg Kommune, men med deltagelse Gudenåkommunerne.

Jonas Folke Nielsen, Ørjan Heggdal, Simon Rahbek & David Getreuer Jensen

Water Resources Research

COMMENTARY
10.1029/2020WR028091

Special Section:
Big Data & Machine Learning in Water Sciences: Recent Progress and Their Use in Advancing Science

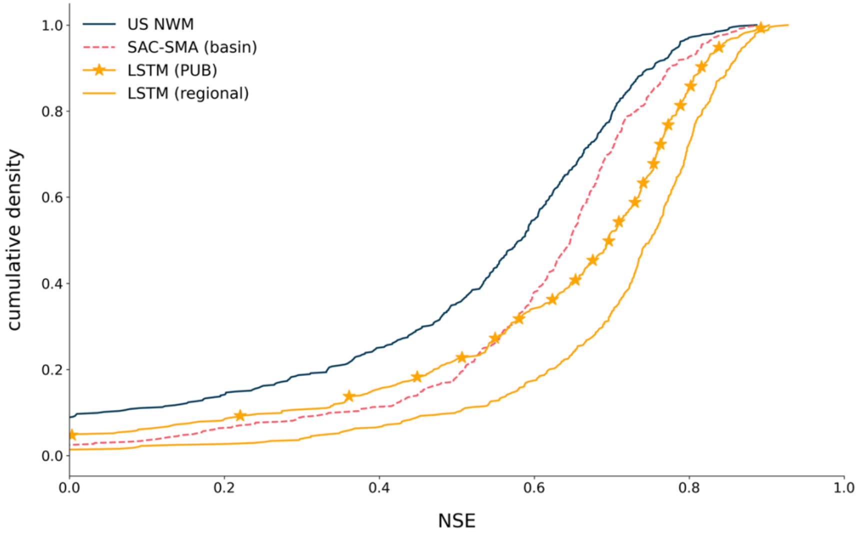
What Role Does Hydrological Science Play in the Age of Machine Learning?

Grey S. Nearing[1], Frederik Kratzert[2], Alden Keefe Sampson[3], Craig S. Pelissier[4], Daniel Klotz[2], Jonathan M. Frame[1], Cristina Prieto[5], and Hoshin V. Gupta[6]

[1]Department of Land Air & Water Resources, University of California Davis, Davis, CA, USA, [2]LIT AI Lab and Institute for Machine Learning, Johannes Kepler University, Linz, Austria, [3]Upstream Tech, Natel Energy Inc., Alameda, CA, USA, [4]NASA Center for Climate Simulation, NASA Goddard Space Flight Center, Greenbelt, MD, USA, [5]IHCantabria Instituto de Hidrulica Ambiental, Universidad de Cantabria, Santander, Spain, [6]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

Key Points:
- Hydrology lacks scale-relevant theories, but deep learning experiments suggest that these theories should exist
- The success of machine learning for hyd… pote… mod…
- It is… show… theo… fore…

Abstract This paper is derived from a keynote talk given at the Google's 2020 Flood Forecasting Meets …



… [t]he extension of laboratory scale theory to the catchment scale is unjustified and that a radical change in theoretical structure (a new paradigm) will be required before any major advance can be made in [predicting catchment-scale rainfall-runoff responses]." He proposed that two things would be necessary to push the field of surface hydrology into a new period of "normal science": (i) scale-relevant theories of watersheds ("[h]ydrology in the future will require a macroscale theory that deals explicitly with the problems posed by spatial integration of heterogeneous nonlinear interacting processes") and (ii) uncertainty quantification ("[s]uch a theory will be inherently stochastic and will deal with the value of observations and qualitative knowledge in reducing predictive uncertainty.")

Unfortunately, hydrology has not had its Einstein (with all due respect to Einstein, 1926, 1950). Nine decades from the establishment of the Hydrology section of the American Geophysical Union and after more than a half-century of computer-based hydrological modeling (Crawford & Burges, 2004), Blöschl et al. (2019) listed as one of the 23 "Unsolved Problems in Hydrology": "what are the hydrologic laws at the catchment scale and how do they change with scale?"

Nearing, Grey S., et al. "What role does hydrological science play in the age of machine learning?." *Water Resources Research* 57.3 (2021): e2020WR028091.

Machine Learning & Fysisk modellering

" Two Clouds" tale af Lord Kelvin i 1900
- Medium til at transportere lys (ether)
- Equipartition of energy

→ Einstein oversatte de "To skyer" til Relativitet & kvantemekanik (nyt paradigme)

Hydrologisk "Two Clouds" tale af Keith Beven i 1987
- Laboratorieskala
- Oplandsskala

"[t]he extension of laboratory scale theory to the catchment scale is unjustified and that a radical change in theoretical structure (a new paradigm) will be required before any major advance can be made in [predicting catchment-scale rainfall-runoff responses]."

→ Ingen Einstein og er én af de 23 "Unsolved Problems in Hydrology"

Tidligere troet årsag til at de skalaer ikke kan forenes var:
- Unikke oplande & mangel på data

→ Modbevist af Machine Learning, da man kan optræne modeller på tværs af oplande og forudsige bedre på ukendte oplande end fysisk modeller, der er tilpasset det enkelte opland. Det betyder, at der er noget teoretisk, som vi ikke forstår gående fra lille til stor skala. (Og det er ikke mangel på data, da sammenhængende kan findes af Machine Learning)

Pointe:

Indtil, at vi får en hydrologisk Einstein, så er min pointe, at fysisk baseret numerisk modellering kan noget, som Machine Learning ikke kan og omvendt.

En hydbrid tankegang er derfor at foretrække.

Agenda

Formål med Proof-Of-Concept

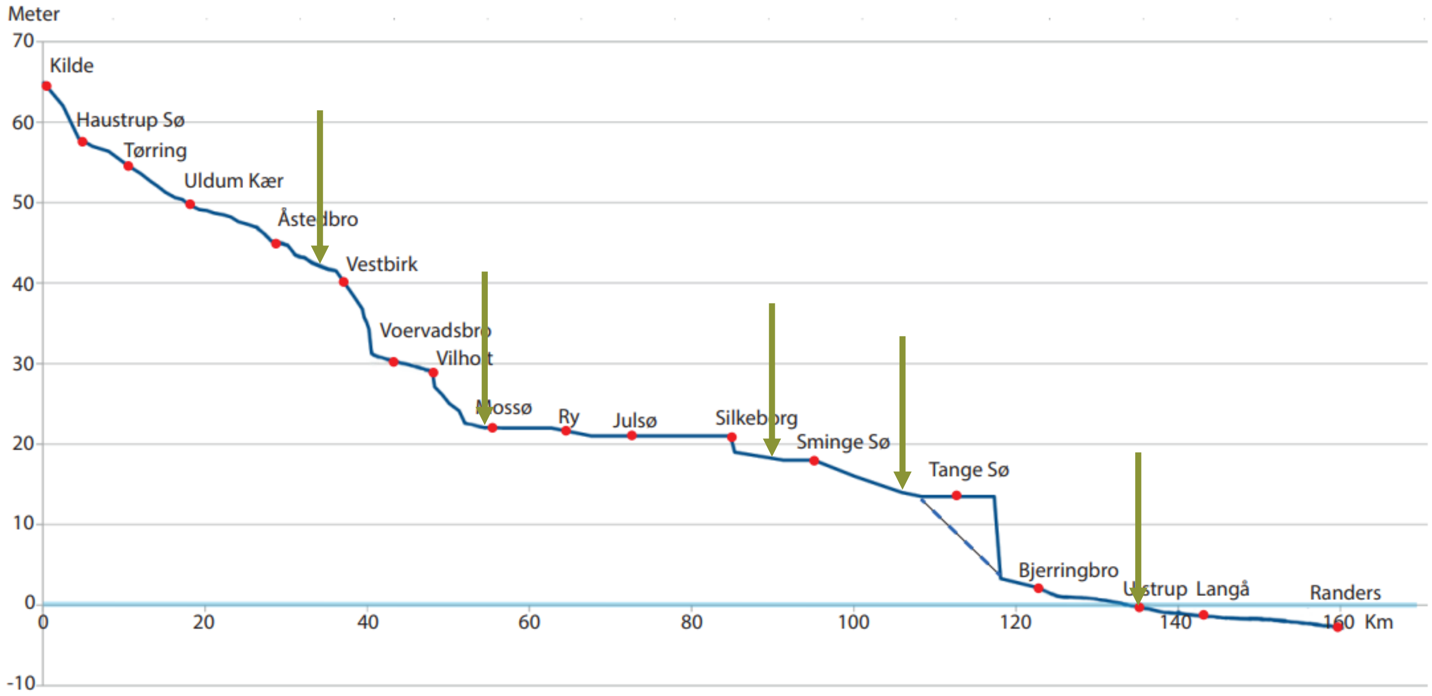
Kort omkring Supervised Learning

Gennemgang af Machine Learning model & inputdata

Metrikker & performance

Formål med Proof-of-Concept

- Afsøgning af, hvilke alternative muligheder der foreligger for forudsigelse af vandstand/flow (herunder "rene" Machine Learning modeller samt hybrid-modellering)
- En evaluering for 5 målestationer af performance for Machine Learning model til prognosticering af vandstand 72 timer ud i fremtiden
- Udpegning af målestationer, hvor stationer langt opstrøms & nedstrøms i systemet samt enkelte stationer påvirket af styring er repræsenteret

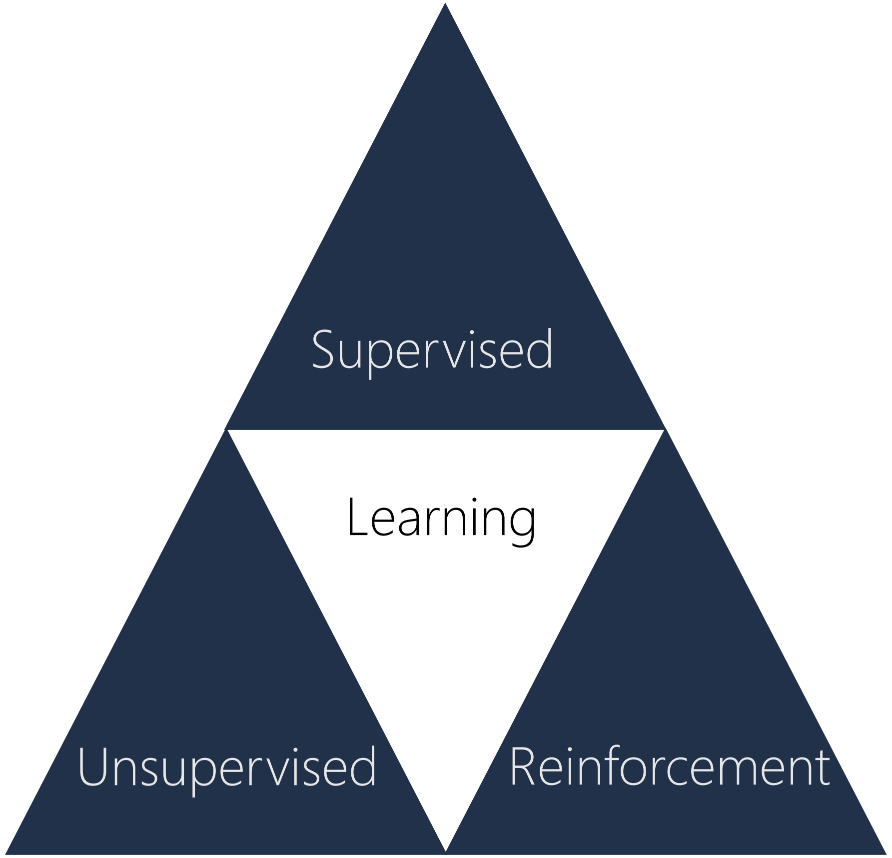


Figurer fra videnomgudenaaen.silkeborg.dk

Machine Learning

Supervised Learning

- Labeled data
- Direkte feedback
- Prædiktion af udfald/fremtid



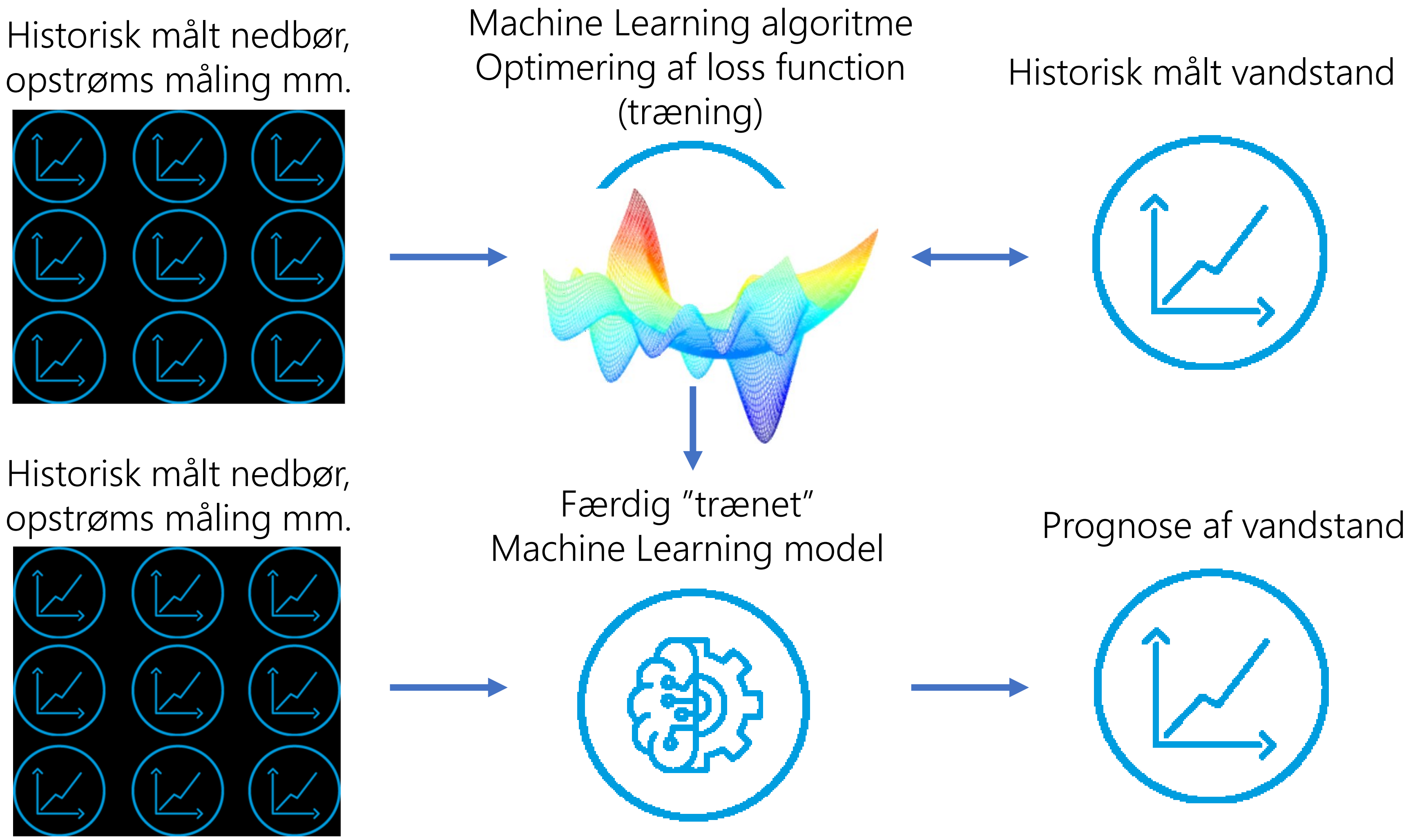
Unsupervised Learning

- Ingen labeled data
- Ingen feedback
- "Finde en gemt struktur i data"

Reinforcement Learning

- Beslutningsproces (styring)
- Kumulativ gevinst
- Læring fra en række af aktioner (optimering af kumulativ gevinst)

Supervised Learning

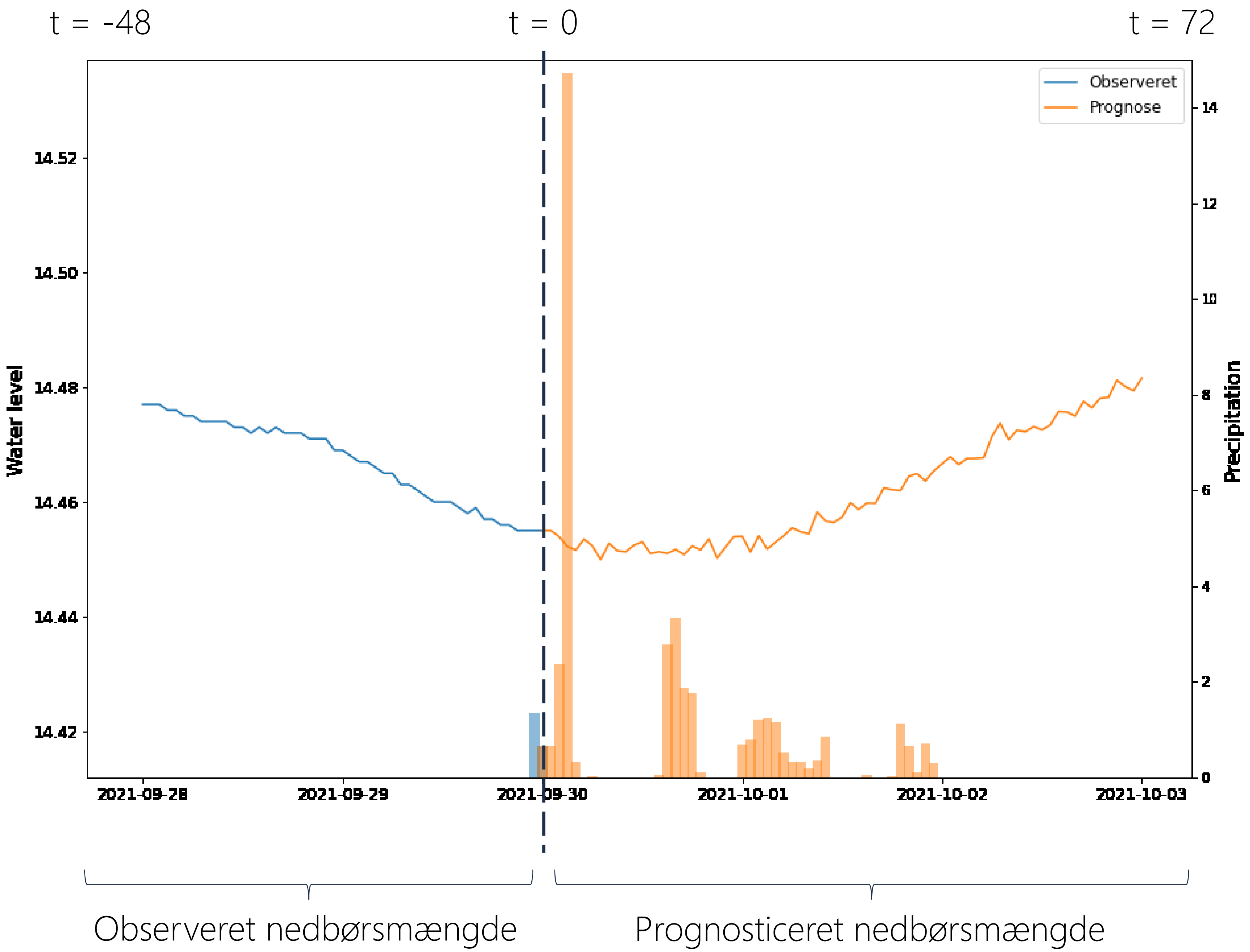


Hvad påvirker variationerne i vandstand ved en 72-timers horisont?

Nedbørsdata

Overfladeafstrømning (nedbør)

- Afstrømning i perioden op fra t=-48 til t=0
- Afstrømning i perioden fra t=0 til t=72



Nedbørsdata

Spatielt distribueret

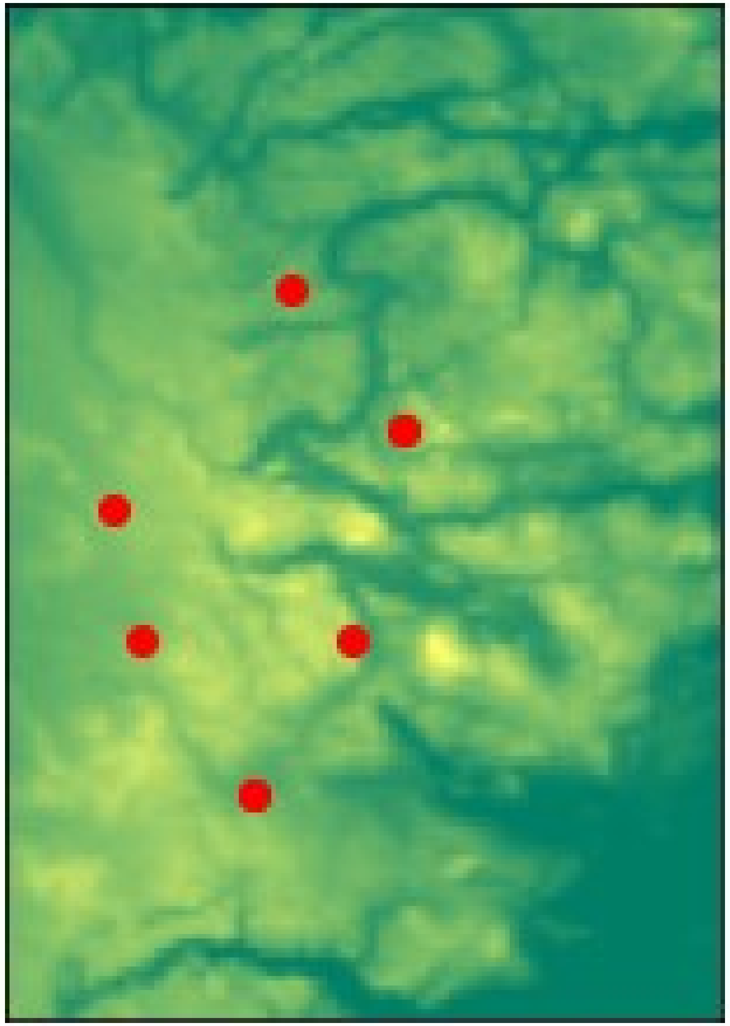
- Vejrradar
- Vejrmodeller
- Kortere tidsserier

Punktobservationer

- DMI målestationer
- Længere tidsserier

Endte med at bruge:

- Observeret nedbør fra 6 DMI vejrstationer
- Prognosticeret nedbør fra NWM (MET)



Observationer vs. prognoser

DMI vejrstationer

Tilgængelig fra ~2015

MET prognoser

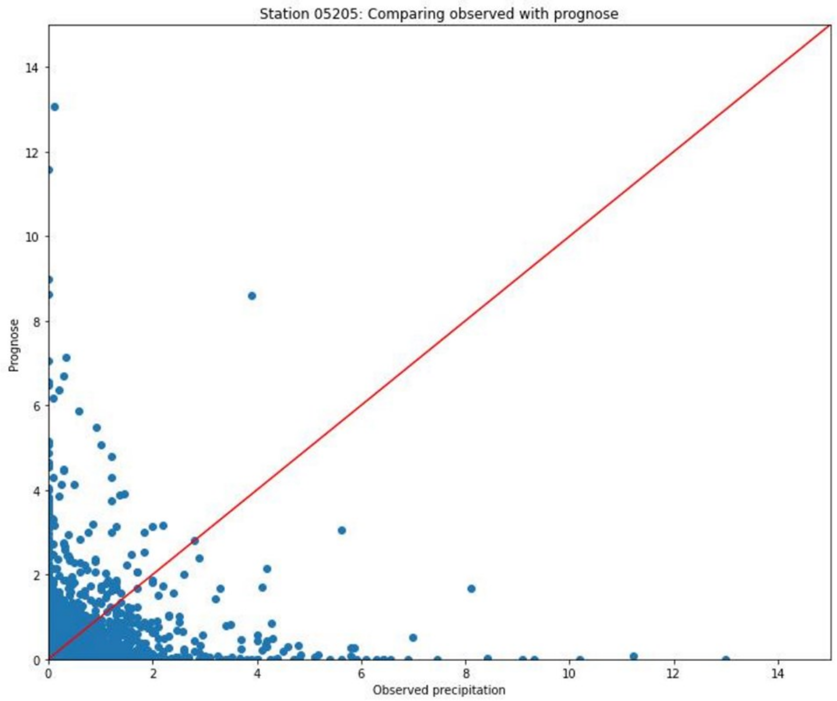
Tilgængelig fra 2019

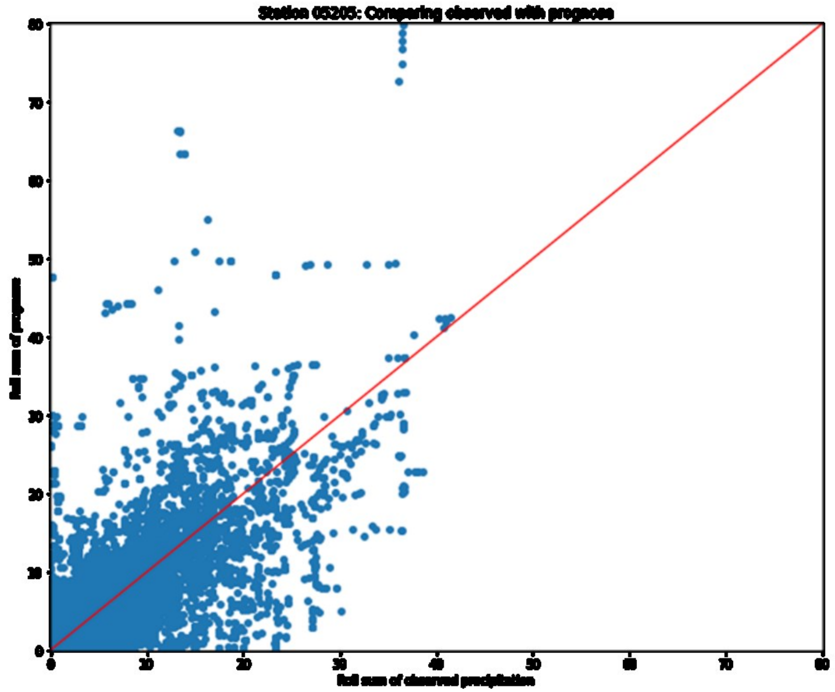
Løsning:

- Vi træner på observerede nedbørsdata som prognose
- Vi validerer på prognosticerede nedbørsdata som prognose

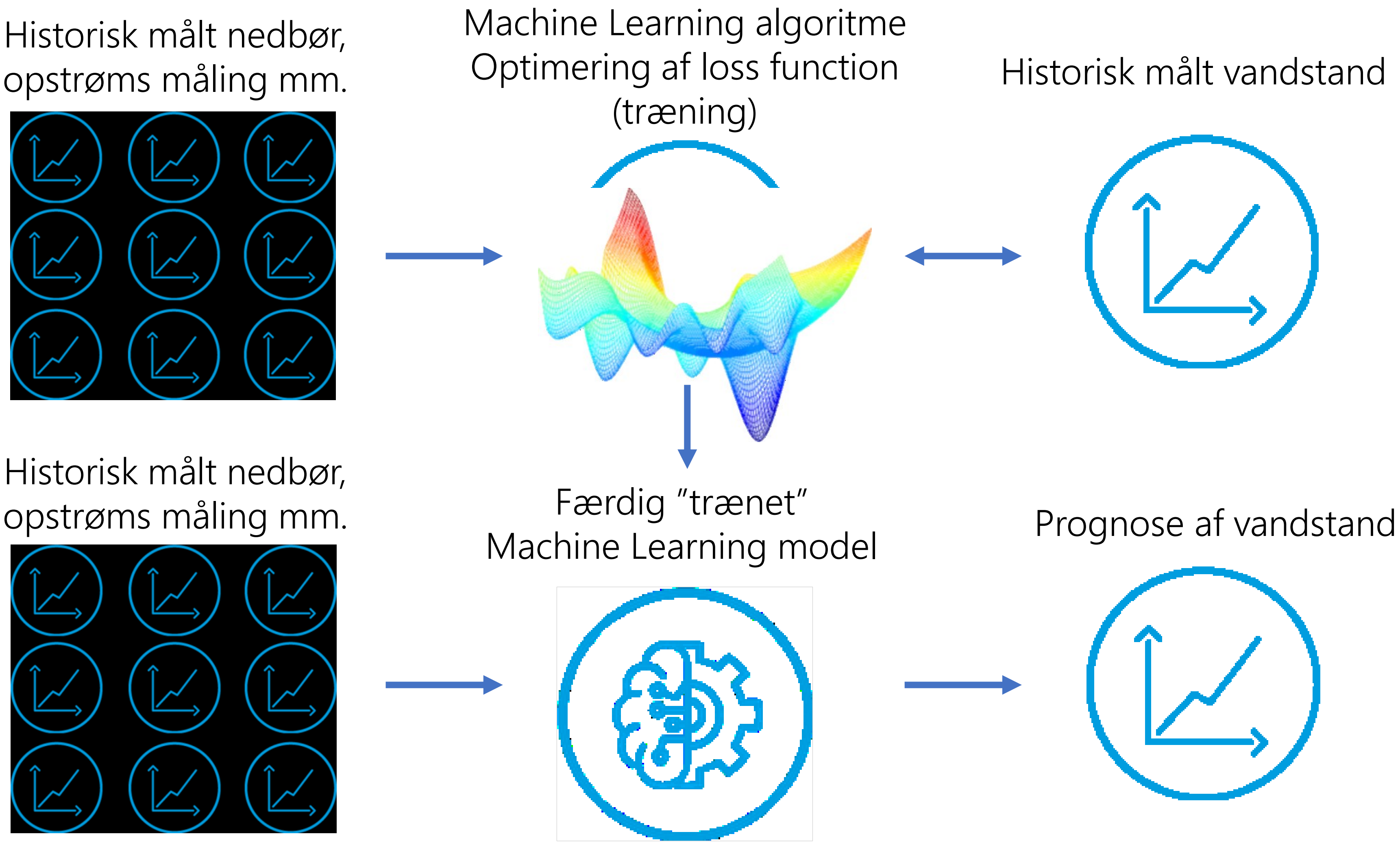
Datagrundlag

- Observeret nedbør
- Prognosticeret nedbør
- Observeret niveau
- Observeret niveau opstrøms
- Lufttemperatur
- Afledte værdier af ovenstående
    - Løbende middelværdi nedbør
    - Løbende sum nedbør
    - Deltaværdier for niveau



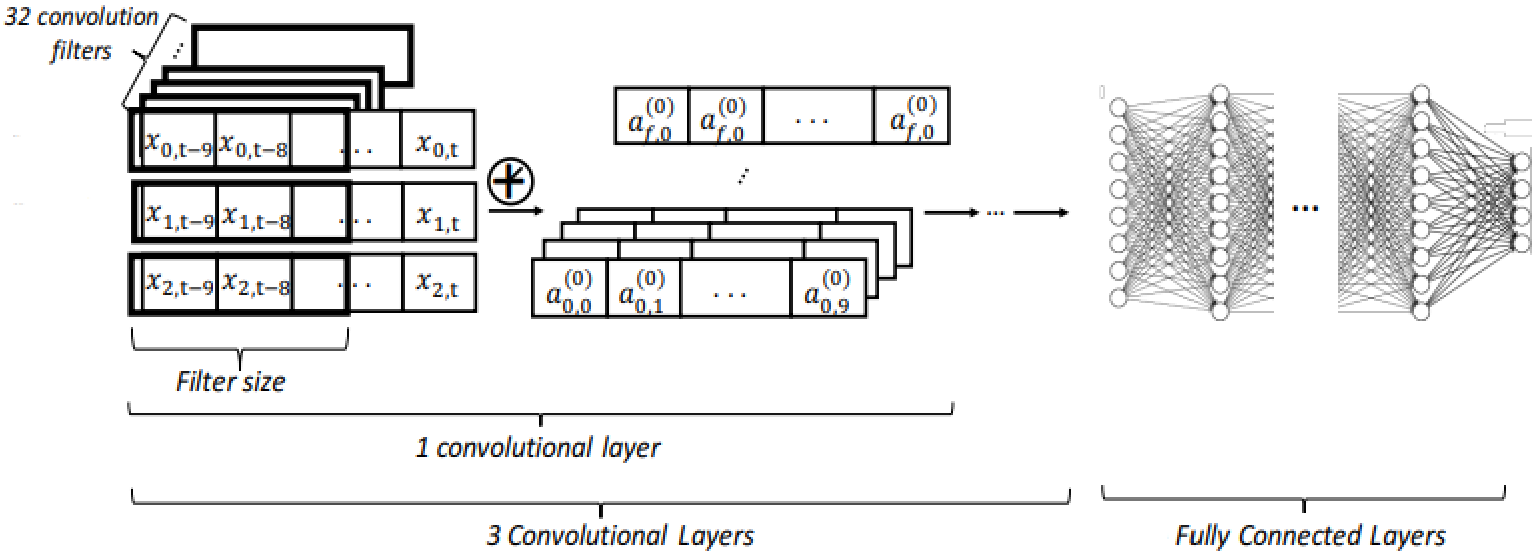


Supervised Learning



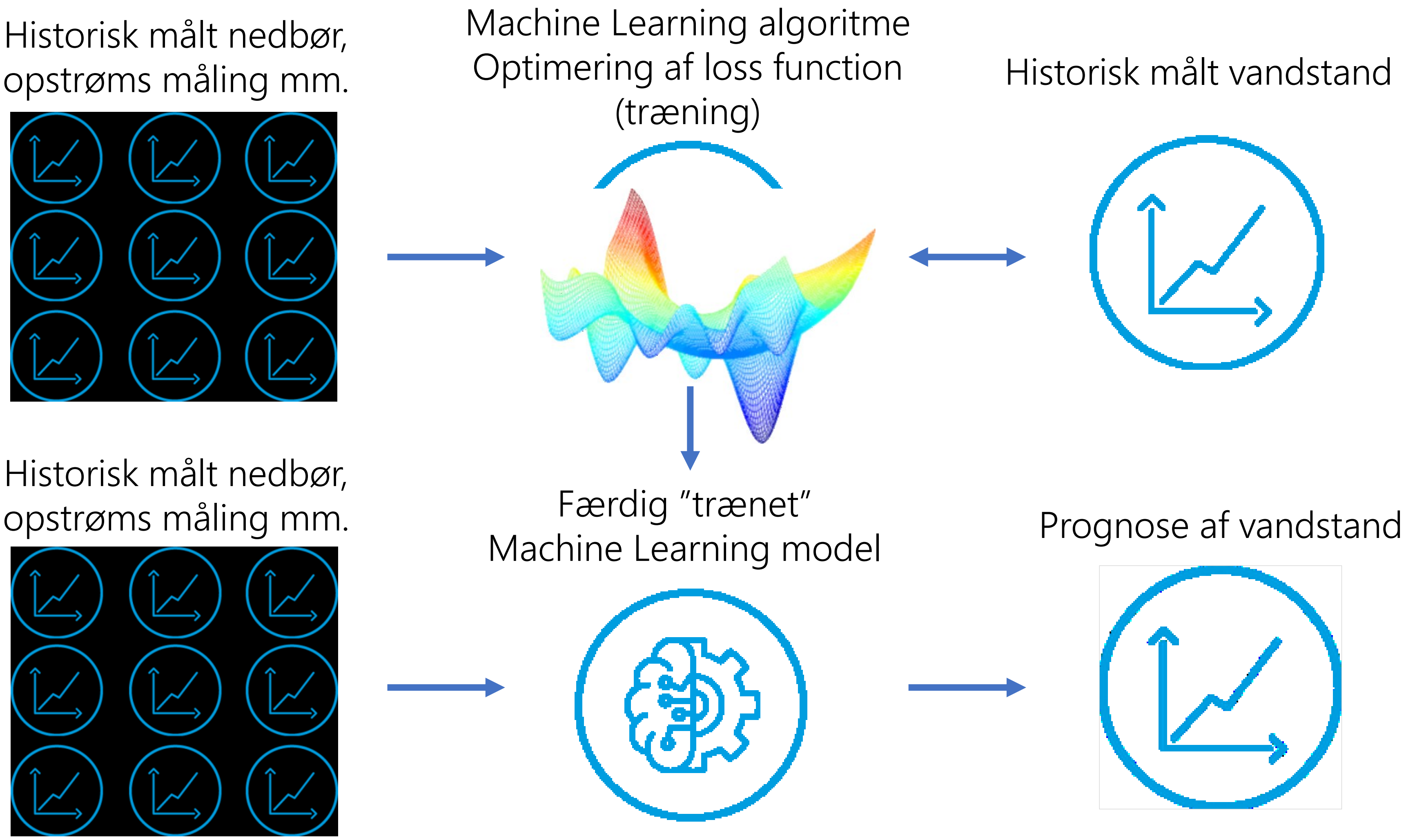
Assem, Haytham, et al. "Urban water flow and water level prediction based on deep learning." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 317-329. Springer, Cham, 2017.

1D CNN neuralt netværk



- Ny prognose for hver time med en prognose horisont på 72 timer
- Temporal opløsning er 1 time (t = 0, t = 1, t = 2 ... t= 72)

Supervised Learning



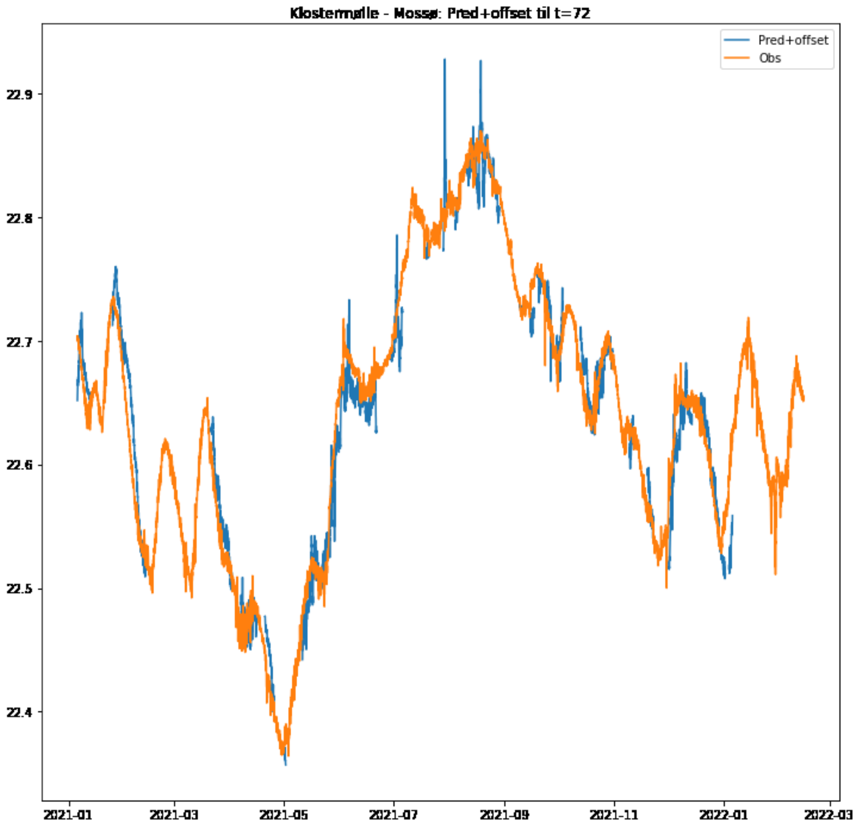
Metrikker

- R2-værdi
- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Hold-out validation (bruger seneste år til test, mens resten er henholdsvis træning & valideringsdata)

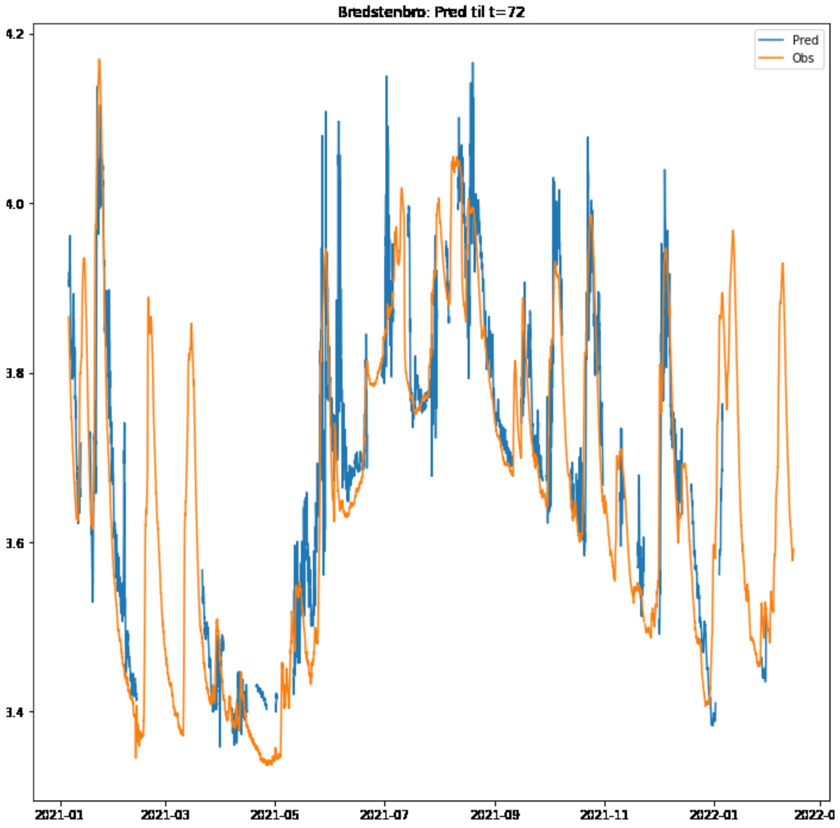
Metrikker

	Model	R^2	MAE	RMSE
Klostermølle	CNN	0.934	0.022	0.028
	CNN+offset	0.953	0.018	0.023
	Persistence	0.933	0.022	0.028
Bredstenbro	CNN	0.801	0.053	0.068
	CNN+offset	0.728	0.063	0.079
	Persistence	0.378	0.087	0.126
Kongensbro	CNN	0.733	0.049	0.056
	CNN+offset	0.838	0.037	0.043
	Persistence	0.803	0.037	0.048
Silkeborg Langsø	CNN	0.44	0.059	0.077
	CNN+offset	0.6252	0.047	0.063
	Persistence	0.842	0.034	0.041
Ulstrup	CNN	0.05	0.150	0.191
	CNN+offset	-0.151	0.165	0.21
	Persistence	-0.27	0.164	0.221

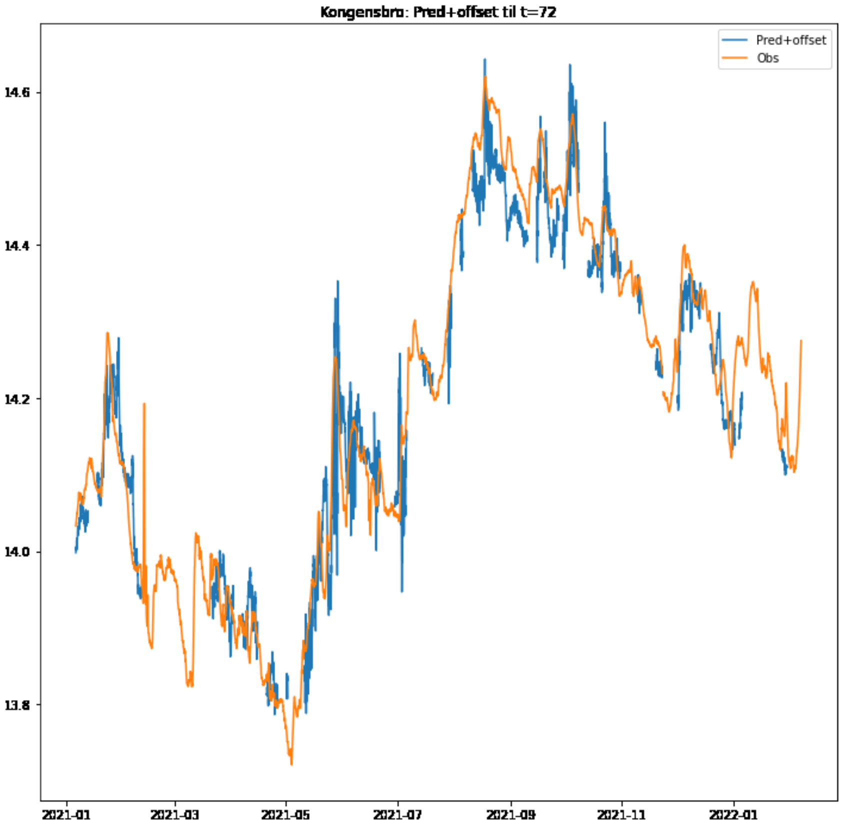
Resultater



Klostermølle - Mossø: Pred+offset til t=72
Pred+offset
Obs
22.9
22.8
22.7
22.6
22.5
22.4
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03



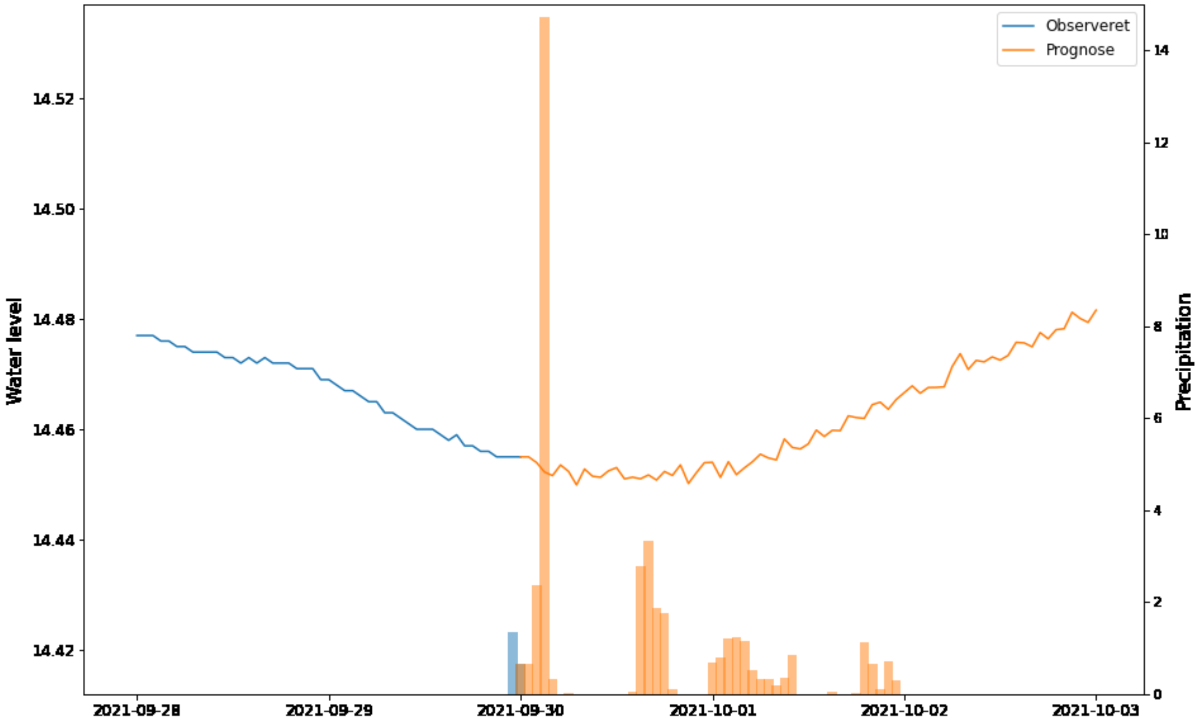
Bredstenbro: Pred til t=72
Pred
Obs
4.2
4.0
3.8
3.6
3.4
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01



Kongensbro: Pred+offset til t=72
Pred+offset
Obs
14.6
14.4
14.2
14.0
13.8
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01

Afrunding

- Generelt en god performance, når målestationerne ikke er styringspåvirket
- God til at prædiktere dynamik – mindre god ved uændrede tilstande over tid
- Kræver gerne +3 års træningsdata (Silkeborg Langsø kun ét år)
- Ekstremerne, hvor modellen skal bruges til varsling har generelt lavere performance end gennemsnitlig – fanger dog tendenser godt
- Performance "straffes", når der er perioder, hvor den ikke kan prædiktere, da de opstår mange gange (hver time over en periode)



AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

COMMENTARY
10.1029/2020WR028091

Special Section:
Big Data & Machine Learning in Water Sciences: Recent Progress and Their Use in Advancing Science

Key Points:
- Hydrology lacks scale-relevant theories, but deep learning experiments suggest that these theories should exist
- The success of machine learning for hydrological forecasting has potential to decouple science from modeling
- It is up to hydrologists to clearly show where and when hydrological theory adds value to simulation and forecasting

Correspondence to:
G. Nearing,
gsnearing@ua.edu

Citation:
Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., et al. (2021). What role does hydrological science play in the age of machine learning?. *Water Resources Research*, 57, e2020WR028091. https://doi.org/10.1029/2020WR028091

Received 5 JUN 2020
Accepted 25 OCT 2020
Accepted article online 13 NOV 2020

What Role Does Hydrological Science Play in the Age of Machine Learning?

Grey S. Nearing[1], Frederik Kratzert[2], Alden Keefe Sampson[3], Craig S. Pelissier[4], Daniel Klotz[2], Jonathan M. Frame[1], Cristina Prieto[5], and Hoshin V. Gupta[6]

[1]Department of Land Air & Water Resources, University of California Davis, Davis, CA, USA, [2]LIT AI Lab and Institute for Machine Learning, Johannes Kepler University, Linz, Austria, [3]Upstream Tech, Natel Energy Inc., Alameda, CA, USA, [4]NASA Center for Climate Simulation, NASA Goddard Space Flight Center, Greenbelt, MD, USA, [5]IHCantabria Instituto de Hidrulica Ambiental, Universidad de Cantabria, Santander, Spain, [6]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

Abstract This paper is derived from a keynote talk given at the Google's 2020 Flood Forecasting Meets Machine Learning Workshop. Recent experiments applying deep learning to rainfall-runoff simulation indicate that there is significantly more information in large-scale hydrological data sets than hydrologists have been able to translate into theory or models. While there is a growing interest in machine learning in the hydrological sciences community, in many ways, our community still holds deeply subjective and nonevidence-based preferences for models based on a certain type of "process understanding" that has historically not translated into accurate theory, models, or predictions. This commentary is a call to action for the hydrology community to focus on developing a quantitative understanding of where and when hydrological process understanding is valuable in a modeling discipline increasingly dominated by machine learning. We offer some potential perspectives and preliminary examples about how this might be accomplished.

1. Beven's Clouds

On April 27, 1900 William Thomson (Lord Kelvin) gave his "Two Clouds" speech ("Nineteenth-Century Clouds over the Dynamical Theory of Heat and Light") at the Royal Institution, in which he argued that "The beauty and clearness of the dynamical theory, which asserts heat and light to be modes of motion, is at present obscured by two clouds." The two open problems in physics that Kelvin referred to were the failure of the Michelson-Morley experiment to detect the luminous ether ("how could the earth move through an elastic solid, such as essentially is the luminiferous ether?"), and the ultraviolet paradox ("the Maxwell-Boltzmann doctrine regarding the partition of energy"). Within a decade, Einstein had proposed fundamentally novel insights that led to two paradigm shifts that define modern physics to this day—the transformation of these two "clouds" into relativity and quantum mechanics.

In 1987, Keith Beven gave what might be considered hydrology's version of the Two Clouds speech at a symposium of the International Association of Hydrological Sciences (IAHS) (Beven, 1987). He took a perspective inspired by Thomas Kuhn's theory of scientific revolutions (Kuhn, 1962) to argue that "[t]he extension of laboratory scale theory to the catchment scale is unjustified and that a radical change in theoretical structure (a new paradigm) will be required before any major advance can be made in [predicting catchment-scale rainfall-runoff responses]." He proposed that two things would be necessary to push the field of surface hydrology into a new period of "normal science": (i) scale-relevant theories of watersheds ("[h]ydrology in the future will require a macroscale theory that deals explicitly with the problems posed by spatial integration of heterogeneous nonlinear interacting processes") and (ii) uncertainty quantification ("[s]uch a theory will be inherently stochastic and will deal with the value of observations and qualitative knowledge in reducing predictive uncertainty.")

Unfortunately, hydrology has not had its Einstein (with all due respect to Einstein, 1926, 1950). Nine decades from the establishment of the Hydrology section of the American Geophysical Union and after more than a half-century of computer-based hydrological modeling (Crawford & Burges, 2004), Blöschl et al. (2019) listed as one of the 23 "Unsolved Problems in Hydrology": "what are the hydrologic laws at the catchment scale and how do they change with scale?"

Nearing, Grey S., et al. "What role does hydrological science play in the age of machine learning?." *Water Resources Research* 57.3 (2021): e2020WR028091.

Machine Learning & Fysisk modellering

" Two Clouds" tale af Lord Kelvin i 1900
- Medium til at transportere lys (ether)
- Equipartition of energy

→ Einstein oversatte de "To skyer" til Relativitet & kvantemekanik (nyt paradigme)

Hydrologisk "Two Clouds" tale af Keith Beven i 1987
- Laboratorieskala
- Oplandsskala

"[t]he extension of laboratory scale theory to the catchment scale is unjustified and that a radical change in theoretical structure (a new paradigm) will be required before any major advance can be made in [predicting catchment-scale rainfall-runoff responses]."

→ Ingen Einstein og er én af de 23 "Unsolved Problems in Hydrology"

Tidligere troet årsag til at de skalaer ikke kan forenes var:
- Unikke oplande & mangel på data

→ Modbevist af Machine Learning, da man kan optræne modeller på tværs af oplande og forudsige bedre på ukendte oplande end fysisk modeller, der er tilpasset det enkelte opland. Det betyder, at der er noget teoretisk, som vi ikke forstår gående fra lille til stor skala. (Og det er ikke mangel på data, da sammenhængende kan findes af Machine Learning)

Pointe:

Indtil, at vi får en hydrologisk Einstein, så er min pointe, at fysisk baseret numerisk modellering kan noget, som Machine Learning ikke kan og omvendt.

En hydbrid tankegang er derfor at foretrække.