

Hydroinformatik - SoSe 2024

HyBHW-S1-01-13: Maschinelles Lernen

Olaf Kolditz

¹Helmholtz Centre for Environmental Research – UFZ, Leipzig

²Technische Universität Dresden – TUD, Dresden

³Center for Advanced Water Research – CAWR

⁴TUBAF-UFZ Center for Environmental Geosciences – C-EGS, Freiberg / Leipzig

Dresden, 14.06.2024

<https://www.ufz.de/index.php?de=40416>

<https://bildungsportal.sachsen.de/opal/auth/RepositoryEntry/32518209537?10>

Übersicht

1. ML: Definition
2. ML: Anwendungsmöglichkeiten
3. ML: Methodik
4. ML: Algorithmen
5. ML in der Forschung
6. ML in der Hydrologie

Maschinelles Lernen

Maschinelles Lernen

Definition

” **Maschinelles Lernen** ist ein Oberbegriff für die „künstliche“ Generierung von Wissen aus Erfahrung: Ein künstliches System lernt aus Beispielen und kann diese nach Beendigung der Lernphase verallgemeinern. Das heißt, es werden nicht einfach die Beispiele auswendig gelernt, sondern es „erkennt“ Muster und Gesetzmäßigkeiten in den Lerndaten. So kann das System auch unbekannte Daten beurteilen (Lerntransfer) oder aber am Lernen unbekannter Daten scheitern (Überanpassung).”

Source: https://de.wikipedia.org/wiki/Maschinelles_Lernen

Maschinelles Lernen

Definition

” Maschinelles Lernen ist ein Oberbegriff für die „künstliche“ **Generierung von Wissen aus Erfahrung**: Ein künstliches System lernt aus Beispielen und kann diese nach Beendigung der Lernphase **verallgemeinern**. Das heißt, es werden nicht einfach die Beispiele auswendig gelernt, sondern es „erkennt“ **Muster und Gesetzmäßigkeiten in den Lerndaten**. So kann das System auch unbekannte Daten beurteilen (**Lerntransfer**) oder aber am Lernen unbekannter Daten scheitern (Überanpassung).”

Source: https://de.wikipedia.org/wiki/Maschinelles_Lernen

Maschinelles Lernen

Anwendungen

”Aus dem weiten Spektrum möglicher Anwendungen seien hier genannt: automatisierte Diagnoseverfahren, Erkennung von Kreditkartenbetrug, Aktienmarktanalysen, Klassifikation von Nukleotidsequenzen, Sprach- und Texterkennung sowie autonome Systeme.” Naturwissenschaften ?

Source: https://de.wikipedia.org/wiki/Maschinelles_Lernen

Maschinelles Lernen

Anwendungen

”Das Thema ist eng verwandt mit „Knowledge Discovery in Databases“ und „Data-Mining“, bei dem es jedoch vorwiegend um das Finden von neuen **Mustern** und **Gesetzmäßigkeiten** geht. Viele Algorithmen können für beide Zwecke verwendet werden. Außerdem können Methoden der „Knowledge Discovery in Databases“ genutzt werden, um Lerndaten für „maschinelles Lernen“ zu produzieren oder vorzuverarbeiten, und Algorithmen aus dem maschinellen Lernen finden beim Data-Mining Anwendung.”

Source: https://de.wikipedia.org/wiki/Maschinelles_Lernen

Maschinelles Lernen

Methoden

” Beim maschinellen Lernen spielt Art und Mächtigkeit der Wissensrepräsentation eine wichtige Rolle. Man unterscheidet zwischen **symbolischen Systemen**, in denen das Wissen – sowohl die Beispiele als auch die induzierten Regeln – explizit repräsentiert ist, und subsymbolischen Systemen wie **neuronalen Netzen**, denen zwar ein berechenbares Verhalten „antrainiert“ wird, die jedoch keinen Einblick in die erlernten **Lösungswege** erlauben; hier ist Wissen implizit repräsentiert.”

Source: https://de.wikipedia.org/wiki/Maschinelles_Lernen

Maschinelles Lernen

Algorithmen

”Die praktische Umsetzung geschieht mittels Algorithmen. Verschiedene Algorithmen aus dem Bereich des maschinellen Lernens lassen sich grob nach diesem Schema einteilen:”

- ▶ Überwachtes Lernen (engl. supervised learning)
- ▶ Unüberwachtes Lernen (engl. unsupervised learning)

Source: https://de.wikipedia.org/wiki/Maschinelles_Lernen

Maschinelles Lernen

Algorithmen - Überwachtes Lernen

”Der Algorithmus lernt eine Funktion aus gegebenen Paaren von Ein- und Ausgaben. Dabei stellt während des Lernens ein „Lehrer“ den korrekten Funktionswert zu einer Eingabe bereit. Ziel beim überwachten Lernen ist, dass dem Netz nach mehreren Rechengängen mit unterschiedlichen Ein- und Ausgaben die Fähigkeit antrainiert wird, **Assoziationen** herzustellen. Ein Teilgebiet des überwachten Lernens ist die automatische Klassifizierung. Ein Anwendungsbeispiel wäre die Handschrifterkennung.”

Source: https://de.wikipedia.org/wiki/Maschinelles_Lernen

Maschinelles Lernen

... in der Hydrologie

maschinelles lernen hydrologie

Alle News Videos Bilder Shopping Mehr Einstellungen Tools

Ungefähr 9.320 Ergebnisse (0,41 Sekunden)

FAQ Machine Learning: Was Sie über Maschinelles Lernen wissen ...
<https://www.computerwoche.de/.../was-sie-ueber-maschinelles-lernen-wissen-muessen,...>
Maschinelles Lernen wird heute mehr und mehr zum Mainstream. Waren selbstlernende Programme noch bis vor wenigen Jahren ausschließlich ein Thema für Universitäten, Forschungseinrichtungen und einige Technologieunternehmen, finden sie heute zunehmend Eingang in ganz normale Produkte und Lösungen.
Es fehlt: hydrologie

Programm — Professur für Hydrologie — TU Dresden
<https://tu-dresden.de/bu/umwelt/hydro/ihm/hydrologie/tdh2018/programm>
16.11.2017 - Messen und Modellieren. Modellbasierte Planung von Messnetzen sowie Methoden zur Regionalisierung von Messdaten. Neue Datenquellen und Methoden zur inversen Modellierung. Datenbasierte Modellierung und maschinelles Lernen ...
Du hast diese Seite 3 Mal aufgerufen. Letzter Besuch: 25.01.18

Maschinelles Lernen – Wikipedia
https://de.wikipedia.org/wiki/Maschinelles_Lernen
Maschinelles Lernen ist ein Oberbegriff für die „künstliche“ Generierung von Wissen aus Erfahrung: Ein künstliches System lernt aus Beispielen und kann diese nach Beendigung der Lernphase verallgemeinern. Das heißt, es werden nicht einfach die Beispiele auswendig gelernt, sondern es „erkennt“ Muster und ...
Es fehlt: hydrologie
Du hast diese Seite am 25.01.18 besucht.

Maschinelles Lernen

... in der Hydrologie



M³ - Messen, Modellieren, Managen

Der 20. Tag der Hydrologie legt den Fokus auf die Schnittstellen und Wechselbeziehungen der drei Arbeitsbereiche Messen, Modellieren und Managen und richtet sich gleichermaßen an „ExperimentatorInnen“, „ModelliererInnen“ und „WasserwirtschaftlerInnen“. Präsentieren sie ihre Methoden, Ergebnisse und Erfahrungen aus Forschung und Praxis zur Integration von zwei oder drei Tätigkeitsfeldern! Besonders gefragt sind M²- und M³-Beiträge, die zeigen wie Beobachtungsmethoden, Modelle und Entscheidungstechniken voneinander abhängen und in Wissenschaft und Praxis zusammenhängend entwickelt sowie aufeinander abgestimmt angewendet werden. Gewünscht sind Beiträge, die sich mindestens einem der nachfolgenden Themenfelder zuordnen lassen:

- **Messen und Modellieren**
- **Modellieren und Managen**

• **Messen und Managen**

Maschinelles Lernen

... in der Hydrologie

TECHNISCHE UNIVERSITÄT DRESDEN Institut für Hydrologie und Meteorologie Sprache Suche

PROFESSUR FÜR HYDROLOGIE DIE PROFESSUR STUDIUM FORSCHUNG SERVICE

🏠 > TDH 2018 > **PROGRAMM**

Programm

M³ - MESSEN, MODELLIEREN, MANAGEN

Der 20. Tag der Hydrologie am 22./23. März 2018 legt den Fokus auf die Schnittstellen und Wechselbeziehungen der drei Arbeitsbereiche Messen, Modellieren und Managen und richtet sich gleichermaßen an „ExperimentatorInnen“, „ModelliererInnen“ und „WasserwirtschaftlerInnen“. Präsentieren Sie ihre Methoden, Ergebnisse und Erfahrungen aus Forschung und Praxis zur Integration von zwei oder drei

Maschinelles Lernen

... in der Hydrologie

Messen und Modellieren

- Modellbasierte Planung von Messnetzen sowie Methoden zur Regionalisierung von Messdaten
- Neue Datenquellen und Methoden zur inversen Modellierung
- Datenbasierte Modellierung und maschinelles Lernen

Modellieren und Managen

- Modellbasierte Entscheidungsunterstützung
- Entscheidungsmethoden zur Modellauswahl einschl. Kriterien zur Bewertung von Modellgüte, Unsicherheit und Generalisierungsfähigkeit

Messen und Managen

- (Neue) Mess- und Analysemethoden zur Entscheidungsunterstützung für die optimale Steuerung in der Wasserressourcenbewirtschaftung
- Umgang mit Daten- und Informationsmangel sowie Modellunsicherheit bei Entscheidungen in der Wasserressourcenbewirtschaftung

Messen, Modellieren und Managen

- Nutzung von low-cost Sensorik und Crowdsourcing zur Modellierung und Entscheidungsunterstützung
- Fusion von Beobachtungsdaten und Modellsimulation zu Umweltinformationen

Source: <https://tu-dresden.de/bu/umwelt/hydro/ihm/hydrologie/tdh2018/programm>

Maschinelles Lernen

... in der Forschung

[PDF] [Machine learning for water monitoring, hydrology and sustainability](#)
www.cs.toronto.edu/~kswersky/wp.../WorkshopPresentation.pdf ▾ Diese Seite übersetzen
Machine learning for water monitoring, hydrology and sustainability. Kevin Swersky. Joint work with Touraj Farahmand, Nando de Freitas, Michael. Osborne, Roman Garnett, Ryan Turner and others ...

[Daily streamflow forecasting by machine learning methods with ...](#)
<https://www.sciencedirect.com/science/.../S0022169411007633> - Diese Seite übersetzen
von K Rasouli - 2012 - Zitiert von: 50 - Ähnliche Artikel
11.01.2012 - Weather forecast data generated by the NOAA Global Forecasting System (GFS) model, climate indices, and local meteo-hydrologic observations were used to forecast daily streamflows for a small watershed in British Columbia, Canada, at lead times of 1–7 days. Three machine learning methods ...

[Streamflow Hydrology Estimate Using Machine Learning \(SHEM\)](#)
onlinelibrary.wiley.com/doi/10.1111/1752-1688.12555/pdf - Diese Seite übersetzen
von TR Petty - Zitiert von: 1
Streamflow Hydrology Estimate Using Machine Learning (SHEM). Journal of the American Water Resources Association (JAWRA) 1-14. <https://doi.org/10.1111/1752-1688.12555>. INTRODUCTION. Continuity and accuracy of streamflow gauge. (streamgage) data are critical for hydrological prediction systems and effective ...

[Machine learning algorithms for modeling groundwater level changes ...](#)
onlinelibrary.wiley.com/doi/10.1002/2016WR019933/full - Diese Seite übersetzen
von S Sahoo - Ähnliche Artikel

Maschinelles Lernen

... in der Hydrologie

Water Resources Research



RESEARCH ARTICLE

10.1002/2016WR019933

Key Points:

- Groundwater level change can be modeled with high accuracy using machine learning methods
- Model framework does not require subsurface parameters and simulates comparable groundwater levels to numerical models of physical flow
- Seasonal irrigation demand has the highest relevance to groundwater level change compared to climate and streamflow inputs for most wells

Supporting Information:

- Supporting Information S1

Correspondence to:

Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S.

S. Sahoo¹ , T. A. Russo¹ , J. Elliott^{2,3} , and I. Foster^{2,3,4}

¹Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania, USA, ²Computation Institute, University of Chicago, Chicago, Illinois, USA, ³Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, Illinois, USA, ⁴Department of Computer Science, University of Chicago, Chicago, Illinois, USA

Abstract Climate, groundwater extraction, and surface water flows have complex nonlinear relationships with groundwater level in agricultural regions. To better understand the relative importance of each driver and predict groundwater level change, we develop a new ensemble modeling framework based on spectral analysis, machine learning, and uncertainty analysis, as an alternative to complex and computationally expensive physical models. We apply and evaluate this new approach in the context of two aquifer systems supporting agricultural production in the United States: the High Plains aquifer (HPA) and the Mississippi River Valley alluvial aquifer (MRVA). We select input data sets by using a combination of mutual information, genetic algorithms, and lag analysis, and then use the selected data sets in a Multilayer Perceptron network



Maschinelles Lernen

Grundwasserwirtschaft

Sahoo, S., T. A. Russo, J. Elliott, and I. Foster (2017), Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S., *Water Resour. Res.*, 53, 3878–3895, doi:10.1002/2016WR019933.

HPA High Plains Aquifer

MRVA Mississippi River Valley alluvial Aquifer

Source: Sahoo et al. (2017)

Maschinelles Lernen

Grundwasserwirtschaft

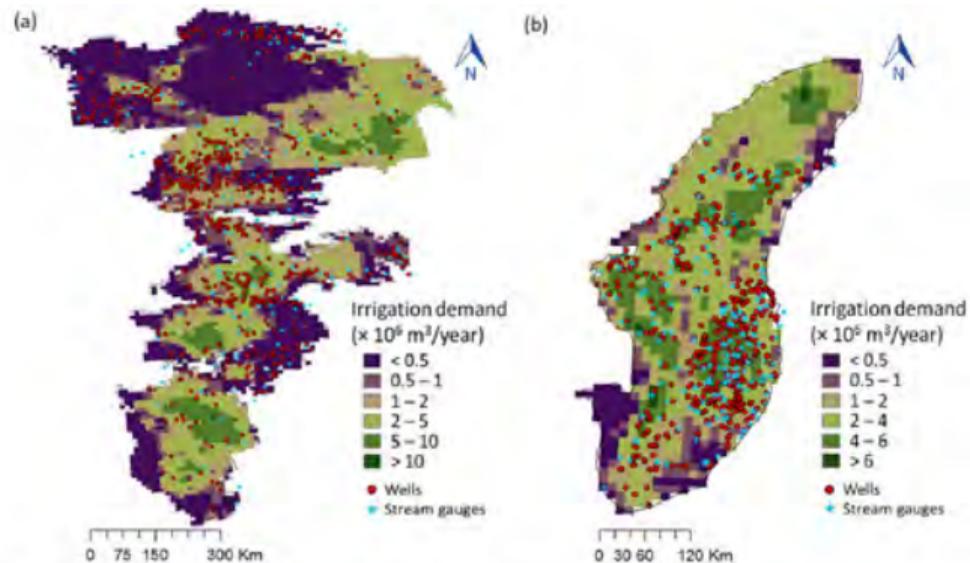
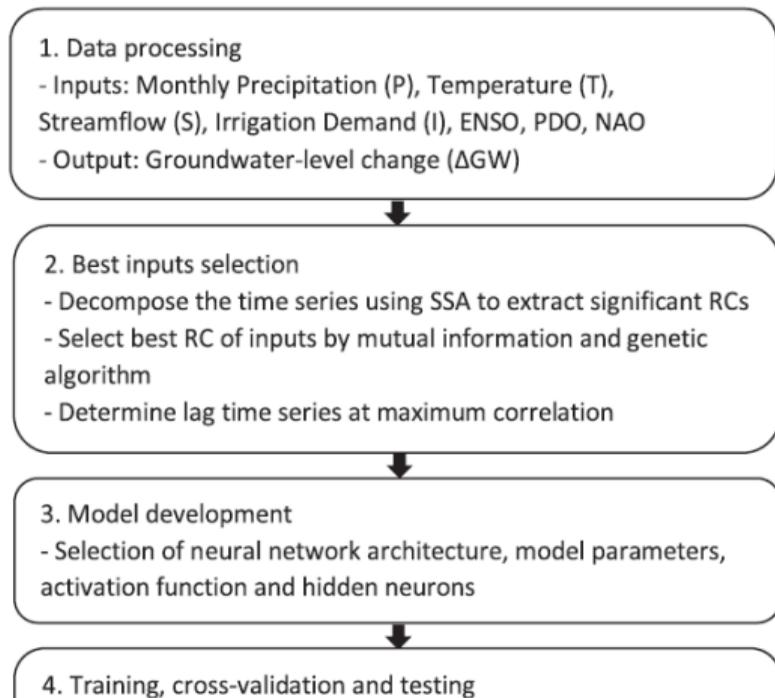


Figure 1. (a) HPA and (b) MRVA showing observation wells (red circles), stream gauges (blue stars), and DSSAT-simulated annual average irrigation demand (m^3/yr) at 5 arc min resolution. The irrigation demand in the HPA is shown for seven crops: corn, sorghum, soybean, cotton, barley, spring wheat, and winter wheat, and accounts for 83% of the total irrigated cropland. The irrigation demand in the MRVA is shown for five crops: cotton, sorghum, soybean, corn, and winter wheat, and accounts for 93% of the total irrigated cropland.

Source: Sahoo et al. (2017)

Maschinelles Lernen

Grundwasserwirtschaft



Source: Sahoo et al. (2017)

Maschinelles Lernen

Grundwasserwirtschaft

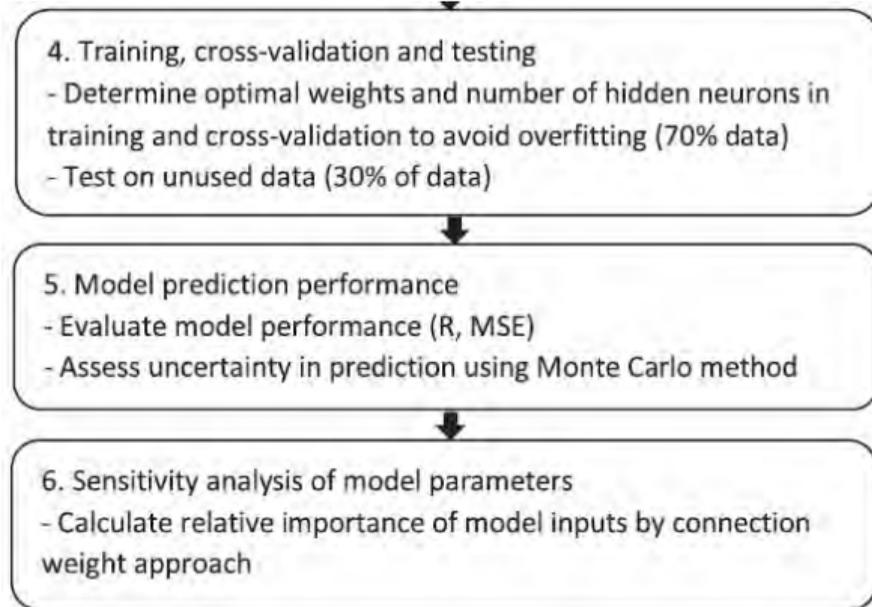


Figure 2. Methodological framework of the proposed groundwater model.

Source: Sahoo et al. (2017)

Maschinelles Lernen

Grundwasserwirtschaft

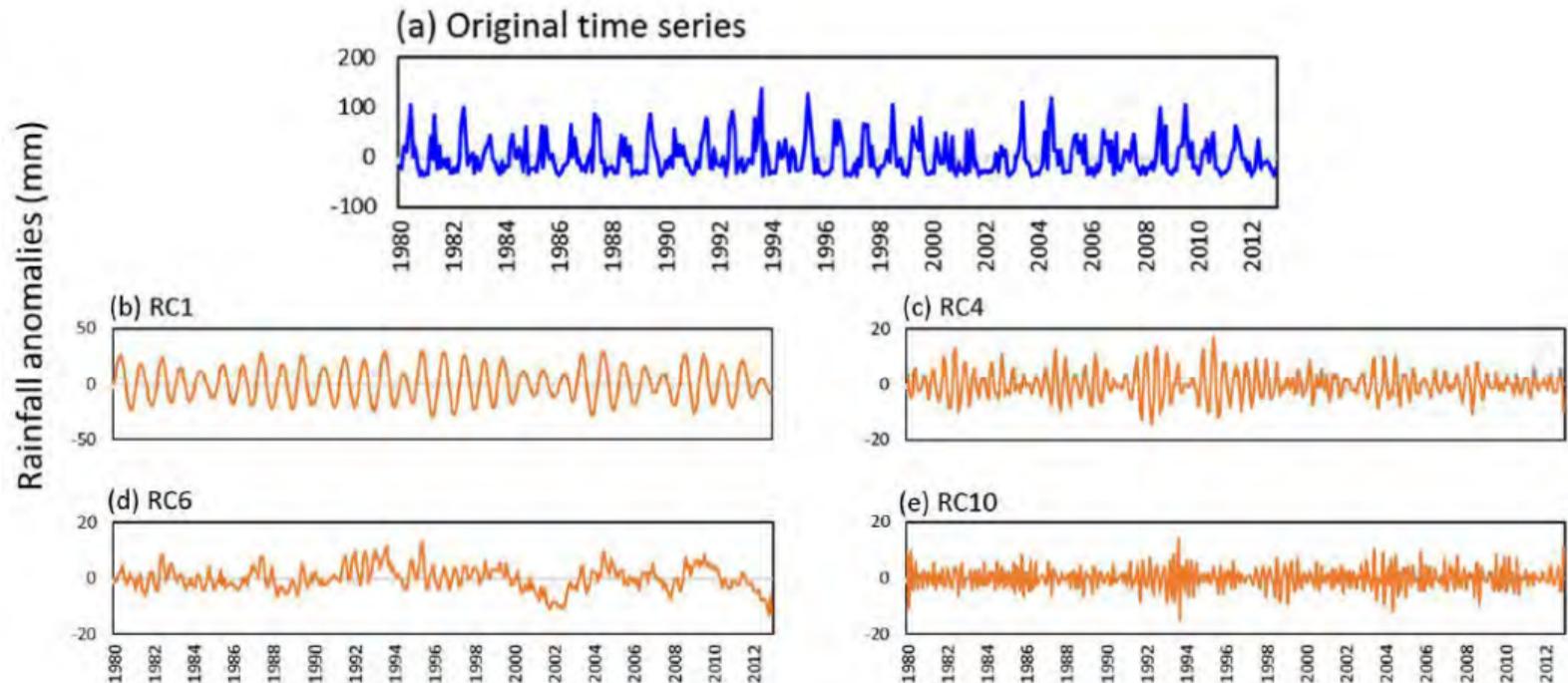


Figure 3. (a) Original time series of the rainfall anomalies. (b-e) Decomposed time series (4 RCs shown here out of 10 RCs) using SSA.

Maschinelles Lernen

Grundwasserwirtschaft

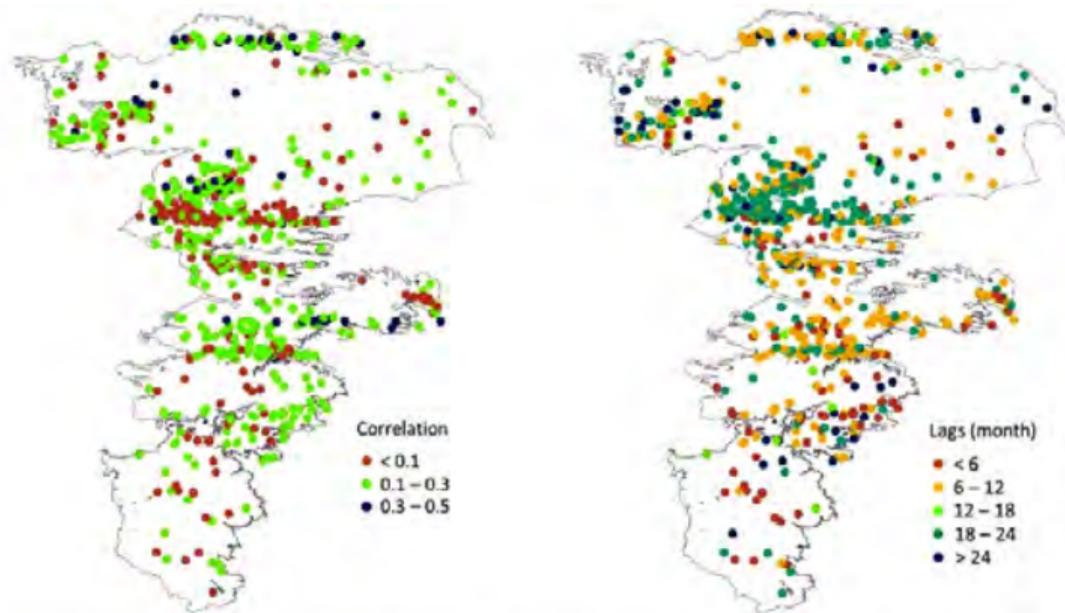


Figure 4. (a) Correlations and (b) respective time lags between monthly groundwater level and precipitation over the HPA for 687 well locations.

Source: Sahoo et al. (2017)

Maschinelles Lernen

Grundwasserwirtschaft

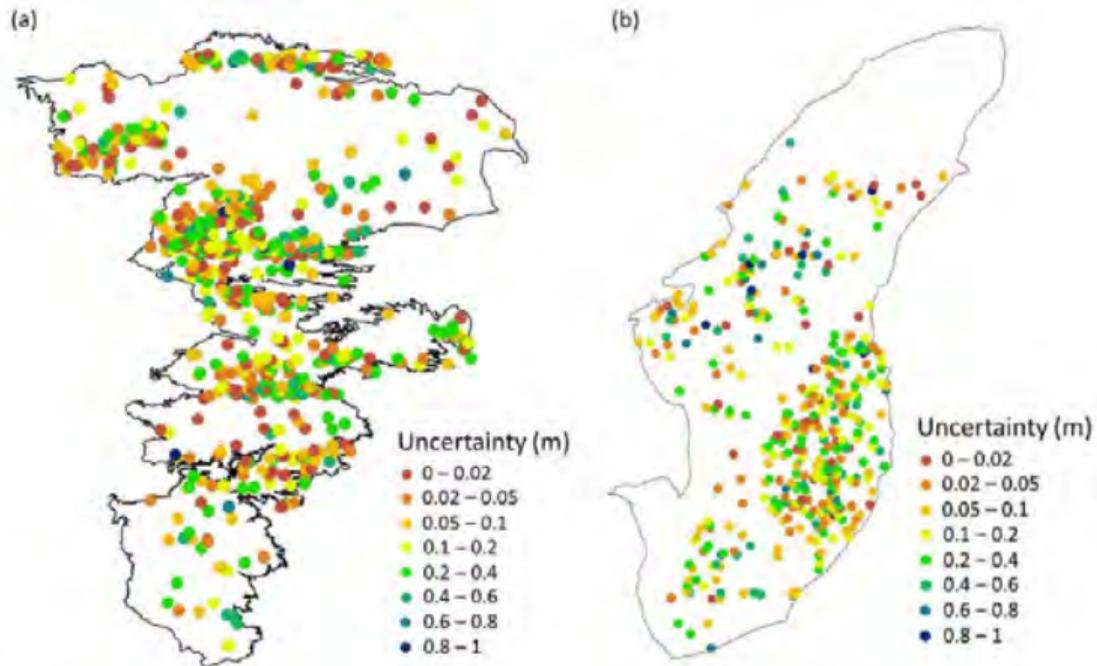


Figure 8. Average seasonal groundwater level uncertainty in model prediction for (a) 687 wells in the HPA and (b) 437 wells in MRVA calculated using HANN model and Monte Carlo method.

Source: Schaap et al. (2017)

Maschinelles Lernen

Grundwasserwirtschaft

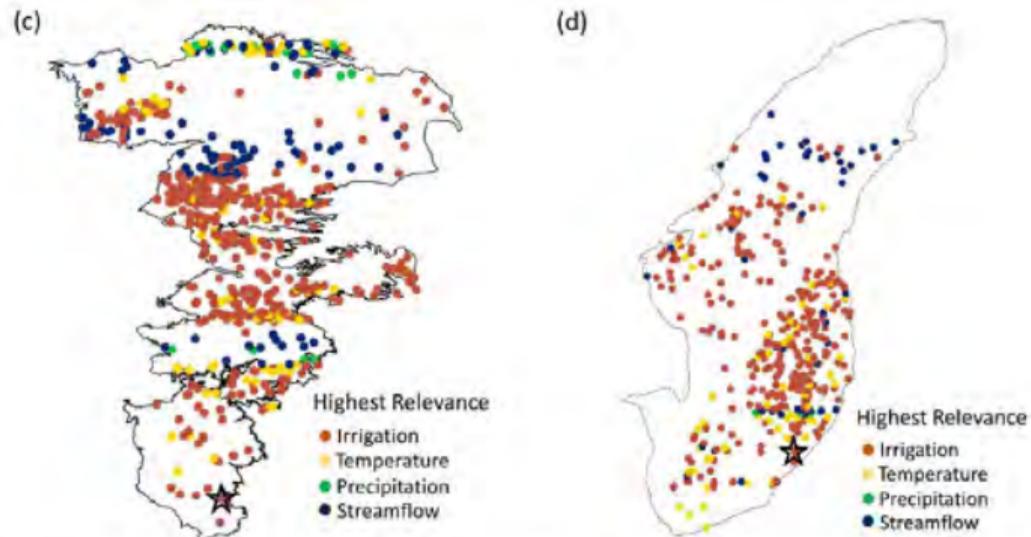
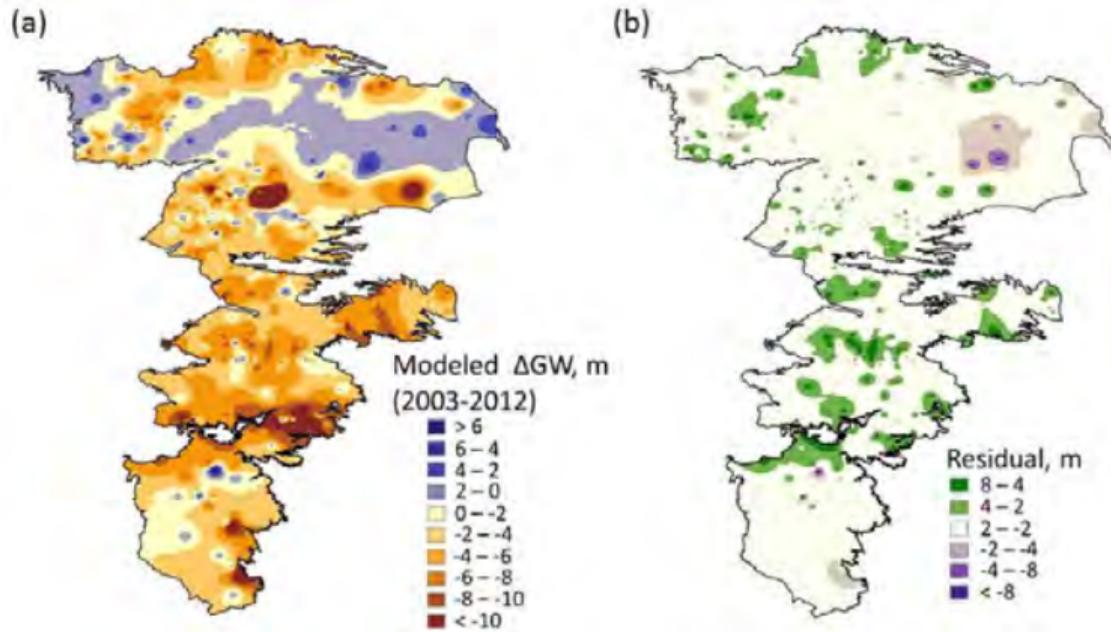


Figure 9. Relative importance of model inputs to groundwater level change prediction by connection weight approach for the example wells in the (a) HPA and (b) MRVA (I: irrigation demand, P: precipitation, T: temperature, S: stream discharge), and the input parameter with the highest relevance at each well location using the HANN model for (c) 687 wells in the HPA and (d) 437 wells in the MRVA. The black stars in the HPA and MRVA indicate the locations of two example sites. The ranking for all of the input parameters and percentage of wells that fall under each category is included in Table 2.

Source: Sahoo et al. (2017)

Maschinelles Lernen

Grundwasserwirtschaft



Source: Sahoo et al. (2017)

Maschinelles Lernen

... in den Geowissenschaften

Geophysical Research Letters

RESEARCH LETTER

10.1002/2017GL075661

Key Points:

- A new geothermal heat flux map of Greenland is obtained within ~15% accuracy using machine learning techniques
- The new map honors regional geology, tectonic settings, and ice core measurements
- Pockets of high heat flux are predicted in central-north Greenland and upstream of several fast-flowing outlet glaciers

Supporting Information:

- Supporting Information S1

Correspondence to:
S. Rezvanehbahani,
soroushr@ku.edu

Predicting the Geothermal Heat Flux in Greenland: A Machine Learning Approach

Soroush Rezvanehbahani^{1,2}, **Leigh A. Stearns^{1,2}**, **Amir Kadivar¹**, **J. Doug Walker¹**, and **C. J. van der Veen⁴**

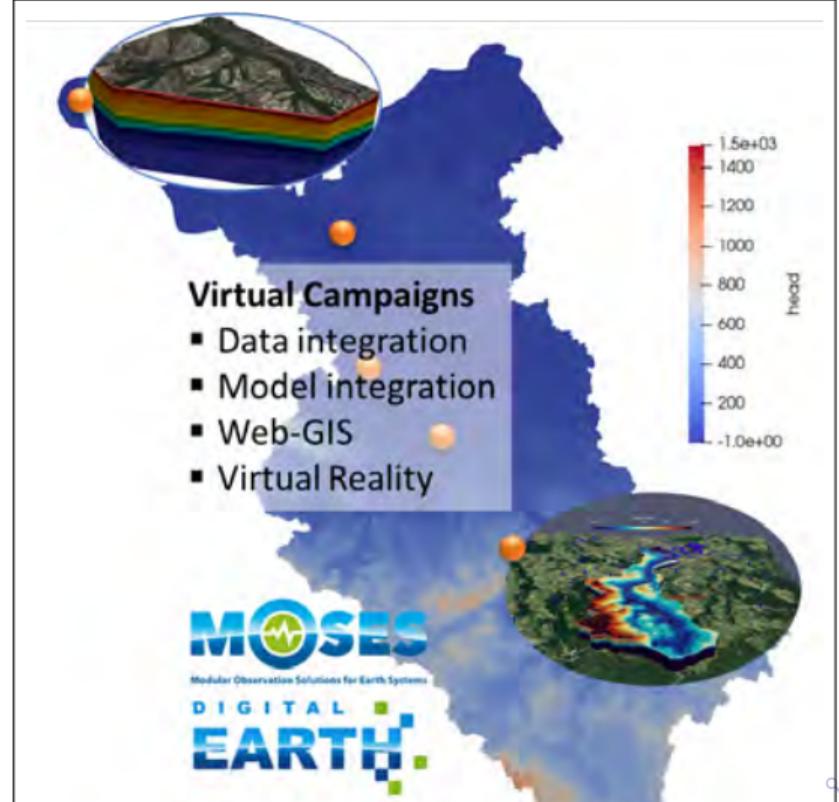
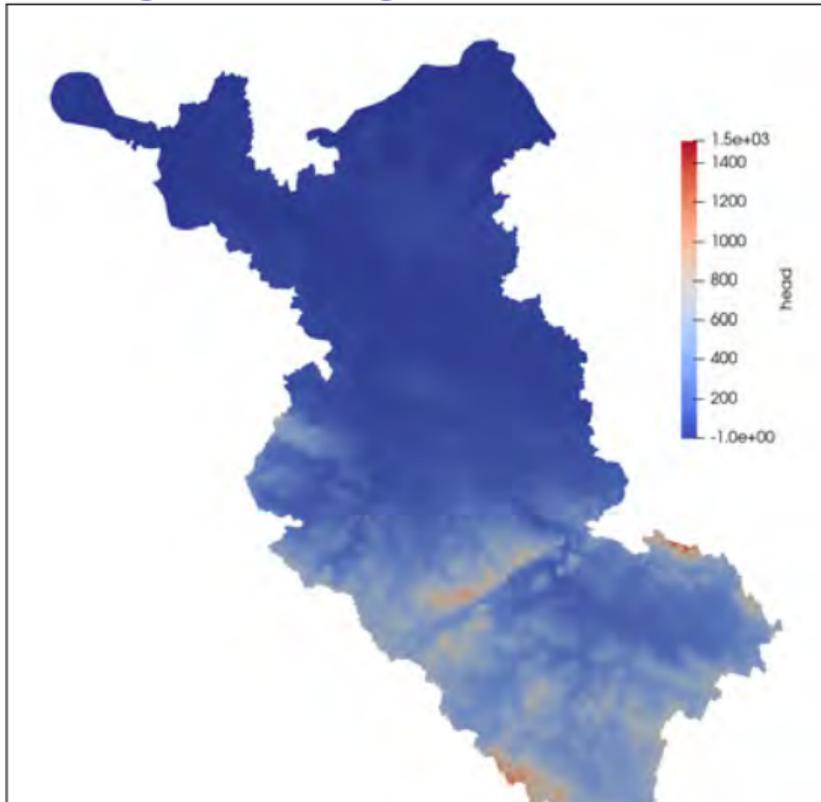
¹Department of Geology, University of Kansas, Lawrence, KS, USA, ²Center for Remote Sensing of Ice Sheets, University of Kansas, Lawrence, KS, USA, ³Department of Mathematics and Statistics, McGill University, Montréal, Québec, Canada, ⁴Department of Geography and Atmospheric Sciences, University of Kansas, Lawrence, KS, USA

Abstract Geothermal heat flux (GHF) is a crucial boundary condition for making accurate predictions of ice sheet mass loss, yet it is poorly known in Greenland due to inaccessibility of the bedrock. Here we use a machine learning algorithm on a large collection of relevant geologic features and global GHF measurements and produce a GHF map of Greenland that we argue is within ~15% accuracy. The main features of our predicted GHF map include a large region with high GHF in central-north Greenland surrounding the NorthGRIP ice core site, and hot spots in the Jakobshavn Isbræ catchment, upstream of Petermann Gletscher, and near the terminus of Nioghalvfjærdsfjorden glacier. Our model also captures the trajectory of Greenland movement over the Icelandic plume by predicting a stripe of elevated GHF in central-east Greenland. Finally, we show that our model can produce substantially more accurate predictions if additional measurements of GHF in Greenland are provided.

Source: AGU

Maschinelles Lernen

... in der eigenen Forschung (Source: Thomas Kalbacher, Thomas Fischer, Johannes Boog, Erik Nixdorf (Umweltinformatik))



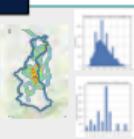
Motivation

- Soil moisture is an essential critical state variable in land surface hydrology and a key component of microclimate
- Measuring principles exist on different scales (TDR, CRNS and satellite observation)
- We integrate CRNS-derived soil moisture estimates with machine learning algorithms towards an improvement of soil moisture predictions on the catchment scale



Study Site

- Mueglitz River Basin south of Dresden (~209 km²)
- The basin is heavily impacted by heavy rain events and flash floods
- Ongoing event-driven intense watershed-scale research and monitoring efforts (bit.ly/22h9Kzi)



Methodology

Preprocessing workflow



Learner Application

- 3 learners selected for regression problem
- Trained with subset from entire CRNS dataset and tested with Mueglitz CRNS data
- 3 target variables, 2 different weather datasets, multiple hyperparameters → >62000 realizations

Learner	Parameterization
Linear Regression (LR)	WR (2x) / (2x) / (2x)
Random Forest (RF)	5-10 Features (2 steps) / 100Trees and Standardizer*
Artificial Neural Network (ANN)	3-4 hidden layers (2 steps) / 3-10 hidden units (2 steps)

Hydrological Modelling

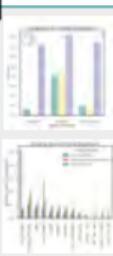
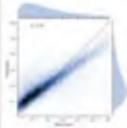
- Hydrological model of Mueglitz Basin using mHM 5.10 (Kumar et al., 2013)
- Model calibration with gauge data
- Provision of daily soil moisture estimates



ML Model Performance

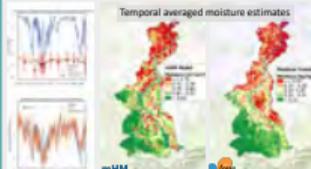
- RF outperforms LR and ANN for each target variable
- Large moisture estimates are underestimated and vice versa
- Important features are altitude, bulk density and aspect, but also neutron input

Prediction Performance of Random Forest Regression



Basin-scale Prediction

- Regionalized data-driven soil moisture estimates are compared with daily output of mHM hydrological model on a 200m grid
- Although a systematic bias between the results of both methods exists, both temporal trends and spatial patterns resemble each other



Conclusions

- Random Forest outperforms other approaches to predict CRNS products using exclusively features from open data sources
- Most relevant features for soil moisture prediction are neutron input monitored at Jungfraujoch Neutron Monitor, bulk density, altitude, slope, topographic aspect and meteorological features. This reveals that existing correction functions on CRNS data require adaptations
- Applying the trained random forest on the entire Mueglitz catchment, obtained daily moisture estimates are in the range of the CRNS rover records
- Comparing with the soil moisture product of mHM, soil moisture predicted by random forest is underestimated. Though, temporal trends as well as spatial distribution patterns of both approaches resemble each other

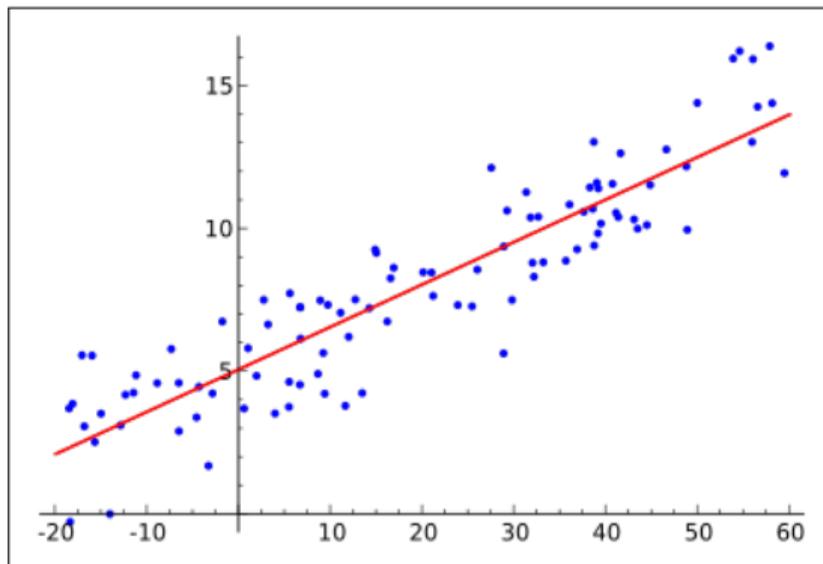
References

- A. Pedregosa et al., "Scikit-learn: Machine Learning in Python", J. Mach. Learn. Res., vol. 12, no. 81, pp. 2825–2830, 2012.
- M. Schröder et al., "Cross-River Neutron Fluxes: A New Source of Field Soil Moisture and the Influence of Roads", Water Resources Research, 2020, pp. 6022–6039, 2020.
- R. Kumar, J. Sankaranarayanan, and S. Haktanir, "Registration of distributed hydrological model parameterization on water flows at multiple scales and locations", Water Resour. Res., vol. 45, no. 1, pp. 900–919, 2009.

Maschinelles Lernen

Lineare Regression

”Die lineare Regression ist ein Spezialfall der Regressionsanalyse, also ein statistisches Verfahren, mit dem versucht wird, eine beobachtete abhängige Variable durch eine oder mehrere unabhängige Variablen zu erklären. Bei der linearen Regression wird dabei ein lineares Modell angenommen. Es werden also nur solche Zusammenhänge herangezogen, bei denen die abhängige Variable eine Linearkombination der Regressionskoeffizienten (aber nicht notwendigerweise der unabhängigen Variablen) ist.”

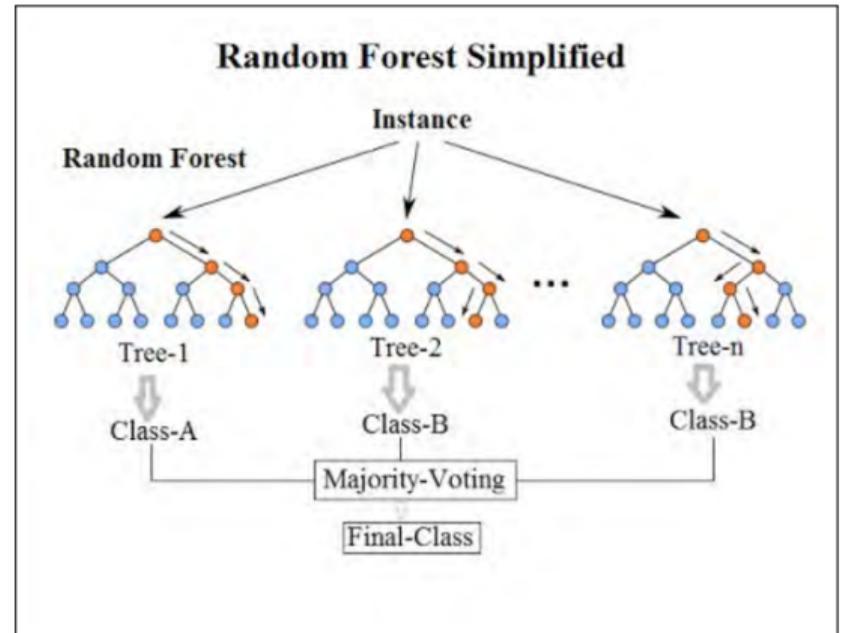


Quelle: https://de.wikipedia.org/wiki/Lineare_Regression

Maschinelles Lernen

Random Forest

„Ein Random Forest ist ein Klassifikations- und Regressionsverfahren, das aus mehreren unkorrelierten Entscheidungsbäumen besteht. Alle Entscheidungsbäume sind unter einer bestimmten Art von Randomisierung während des Lernprozesses gewachsen. ... Random Forests können auch zur Regression eingesetzt werden.“



Quelle: https://de.wikipedia.org/wiki/Random_Forest

Maschinelles Lernen

Quellen

1. Tobias Reitmaier: Aktives Lernen für Klassifikationsprobleme unter der Nutzung von Strukturinformationen. kassel university press, Kassel 2015, ISBN 978-3-86219-999-0, S. 1 (Google books).
2. Lillian Pierson: Data Science für Dummies. 1. Auflage. Wiley-VCH Verlag, Weinheim 2016, ISBN 978-3-527-80675-1, S. 105 f. (Google books).
3. Ralf Mikut: Data Mining in der Medizin und Medizintechnik. KIT Scientific Publishing, 2008, ISBN 978-3-86644-253-5, S. 34 (Google books).
4. Paul Fischer: Algorithmisches Lernen. Springer-Verlag, 2013, ISBN 978-3-663-11956-2, S. 6–7 (Google books).
5. Andreas C. Müller / Sarah Guido: Einführung in Machine Learning mit Python. Hrsg.: O'Reilly-Verlag. ISBN 978-3-96009-049-6 (oreilly.de [abgerufen am 28. November 2017]).
6. Sebastian Raschka: Machine Learning mit Python — ISBN 978-3-95845-422-4. mitp-Verlag, abgerufen am 28. November 2017

Fragen ?

Maschinelles Lernen

Fragen

1. Beschreiben sie die Methode "Maschinelles Lernen" mit eigenen Worten (3 Sätze).
2. ...