

Gebruik van neurale netwerken voor verbeterd begrip van extreme neerslag

Joost Driebergen, Nicole Jungermann

Extreme neerslag heeft een grote impact op het watersysteem. *De afgelopen jaren is veel onderzoek verricht naar de neerslagstatistiek in Nederland (Hakvoort e.a., 2016). Het onderzoek geeft inzicht in de kans van voorkomen van extreme neerslag in Nederland. De oorzaken van regionale spreiding van de extreme neerslag zijn tot nu toe niet duidelijk.*

Conventionele methoden betrekken een selectief aantal gegevensbronnen en zijn gebaseerd op statistische modellen met beperkte geografische spreiding. In dit artikel beschrijven we de toepassing van een Self-Organizing Map (SOM) als neuraal netwerk. Deze techniek maakt patronen tussen meerdere parameters zichtbaar. Dit kan helpen om verklarende verbanden te vinden die tot lokale, extreme neerslag leiden.

We hebben dit toegepast in Nederland met behulp van open data, waarin we een conceptstudie beschrijven van parameters die gerelateerd zijn aan extreme neerslag. Op deze manier kunnen we beter zicht krijgen op de verbanden die kansrijk zijn om lokale extreme neerslaggebeurtenissen te verklaren.

We verwachten dat we patroonherkenning op basis van neurale netwerken vaker gaan toepassen om het inzicht in complexe situaties te vergroten; juist binnen ons vakgebied van hydrologie en de daarbij horende extreme neerslagstatistiek. Wel laat de toepassing van de conceptstudie zien dat de uitkomsten van neurale netwerken kritisch getoetst moeten worden aan hun inhoudelijke juistheid.

Artikel

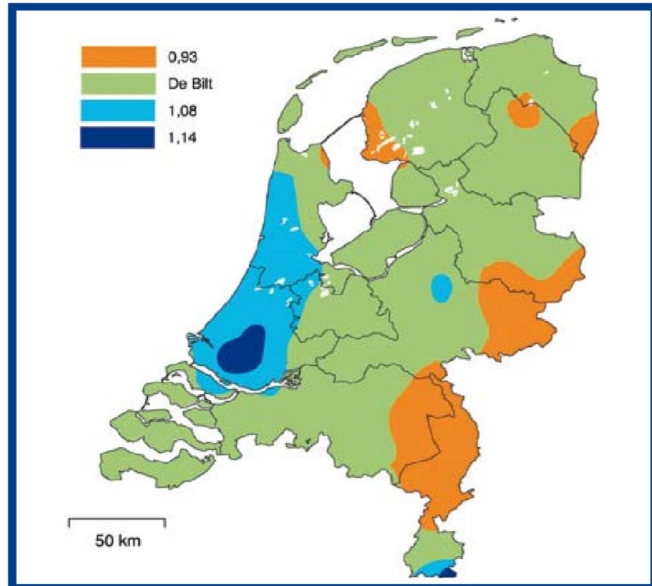
Inleiding

Bij extreme neerslag lopen in Nederland potentieel 140.000 huizen onder water, met een geschatte schade van bijna 7 miljard euro (Arcadis, RTL Nieuws). In het veranderende klimaat zullen we vaker te maken krijgen met extreme neerslag. Het aantal dagen met zware neerslag (meer dan 50 mm neerslag in 24 uur) is sinds 1951 toegenomen met 68% (CBS e.a., 2018). Hakvoort e.a. (2016) onderzochten de neerslagstatistiek op basis van tijdreeksen van meetstation de Bilt van 1906-2014. Zij concludeerden dat een gebeurtenis met een herhalingstijd van 100 jaar, oftewel 77 mm neerslag in 24 uur, twee keer zo vaak voorkomt op basis van de neerslagstatistiek uit 2014.

Verstedelijking gaat samen met een toenemende hoeveelheid verhard oppervlak. Dit zorgt ervoor dat, indien de stedelijke afwatering hier niet voor ontworpen is, steeds meer hinder ondervonden wordt van extreme neerslag.

Naast dat er tijdsafhankelijkheid is in de neerslag, zoals bijvoorbeeld verschillen in de zomer- en winterneerslag, zijn er ook ruimtelijke verschillen in de neerslag.

Het KNMI deelt Nederland op in zogenaamde neerslagregimes, oplopend van lage (L) naar hoge intensiteit (H+), zie Afbeelding 1. In het westen van Nederland is de intensiteit van extreme neerslag het hoogst, met een piek ter hoogte van de Randstad. In het oosten van het land is de intensiteit van extreme neerslag het laagst. Deze neerslagregimes zijn afgeleid op basis van statistische relaties zoals bijvoorbeeld de jaarlijkse neerslagsom (Buishand e.a., 2010).



Afbeelding 1: Regionale verschillen in extreme neerslag (Buishand e.a., 2010).

Behalve dat de hoeveelheid neerslag verandert, zijn er ook veranderingen in landgebruik. Het oppervlak dat in beslag genomen wordt door stedelijk grondgebruik neemt toe en tegelijk gaan steeds meer mensen in de stad wonen. Volgens het CBS zal bijna driekwart van de bevolkingsgroei plaatsvinden in de grotere gemeenten. Deze veranderingen in landgebruik kunnen lokaal invloed hebben op het weer. Shepherd (2002) laat zien dat hitte-eilanden van grote steden meer neerslag te verwerken krijgen. Naast een oplopende temperatuur is het mogelijk dat extreme neerslag ook veroorzaakt wordt door andere parameters. Hierbij kunnen we denken aan landgebruik of aantal inwoners.

Naast de fysische veranderingen rondom neerslag hebben we te maken met een exponentiele groei aan hoeveelheid beschikbare data. De uitdaging van deze grote hoeveelheid aan data is om nuttige data te scheiden van onbruikbare data en vervolgens deze nuttige data te gebruiken om relevante verbanden te leggen. De mens is goed in staat om verbanden te leggen tussen een aantal (tijd)reeksen van verschillende parameters (tweedimensionale data). Bij meerdimensionale data is het menselijk brein echter niet meer in staat om verbanden te leggen tussen verschillende parameters. Daarom is in dit artikel gebruik gemaakt van Machine Learning.

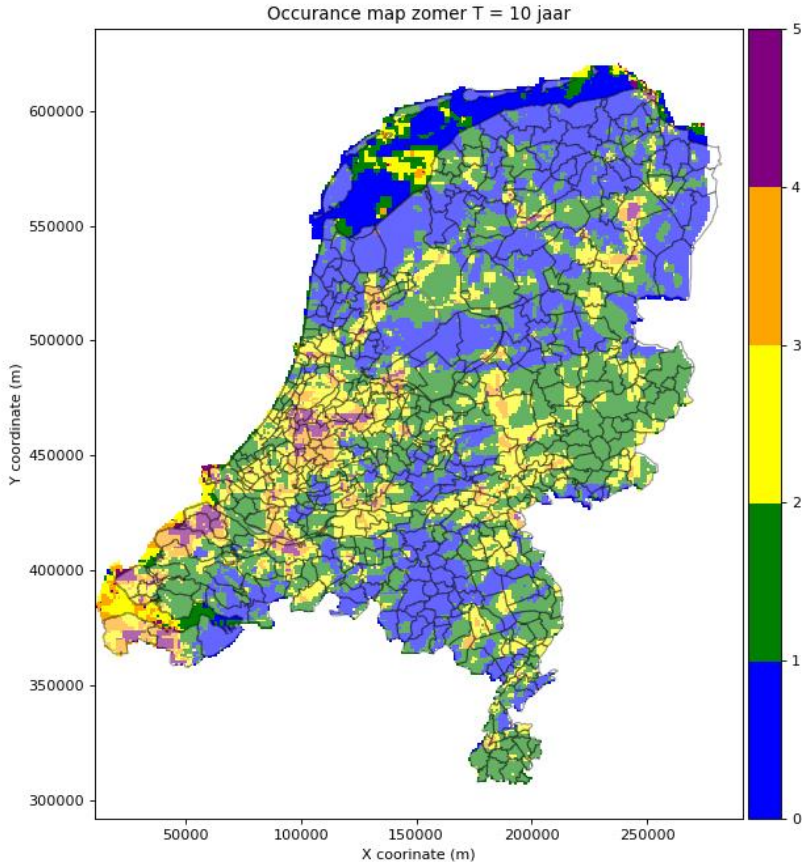
Machine Learning is een snel groeiend onderzoeksveld wat zich (onder andere) bezighoudt met classificatie, regressie en clustering; vaak op basis van grote hoeveelheden data. Een Machine Learning algoritme voeden met grote hoeveelheden data is echter niet genoeg om tot een kwalitatief goed resultaat te komen. Om een Machine Learning algoritme optimaal te laten functioneren, is het van belang dat deze gevoed wordt met de juiste data.

Het doel van deze studie is het toepassen van het Self-Organising-Map algoritme op een dataset. We hebben gekozen voor een dataset waarin extreme neerslag, landgebruik en gas- en elektriciteit zijn opgenomen om zo inzicht te geven in de relatie tussen extreme neerslag en onderliggende parameters. De keuze voor de parameters is gemaakt op basis van de beschikbaarheid. Vervolgens kunnen we aan de hand van deze relatie regionale en temporale verschillen in extreme neerslag vinden. Hierin dient rekening te worden gehouden dat ervoor gekozen is om een eenvoudige dataset te kiezen, waarvan niet alle parameters even verklarend zijn voor extreme neerslag.

De dataset

Om de werking van een SOM algoritme te beschrijven binnen het vakgebied hydrologie, is een conceptstudie gemaakt voor extreme neerslag in de periode 2000-2018, in relatie tot verschillende landgebruiksparementen. Metingen op uurbasis, met een resolutie van $1 \times 1 \text{ km}^2$, van de neerslagradar van de KNMI zijn gesommeerd tot dagelijkse neerslaghoeveelheden. Om aan te geven hoe vaak een bepaalde waarde van neerslag is overschreden, zijn overschrijdingsfrequentiekaarten gemaakt. Een overschrijdingsfrequentiekaarten voor neerslag met een herhalingstijd van 10 jaar wordt verkregen door het aantal dagen dat deze drempelwaarde wordt overschreden op te tellen (zie Afbeelding 2).

In totaal zijn er vier drempelwaardes toegepast (22, 28, 56 en 80 mm per dag). Deze waarden corresponderen met het volume bij een herhalingstijd van vijf keer per jaar, twee keer per jaar, één per 10 jaar en één per 100 jaar. Daarnaast is een onderscheid gemaakt tussen zomer- en wintergebeurtenissen. Hierbij is de tijdsperiode voor een zomergebeurtenis gedefinieerd van 21 maart tot 20 september en een wintergebeurtenis van 21 september tot 20 maart.



Afbeelding 2: Voorbeeld van een overschrijdingsfrequentiekaart voor zomerse neerslag met een herhalingsstijd van eens in de tien jaar. De kleurschaal geeft het aantal gebeurtenissen weer dat de drempelwaarde heeft overschreden.

Naast de neerslagdata zijn een aantal andere databronnen toegevoegd. Van het Centraal Bureau voor de Statistiek (CBS) zijn gegevens gebruikt over elektriciteitsverbruik, gasverbruik en het aantal omgevingsadressen per postcode. Deze gegevens zijn vervolgens aggregeert naar hetzelfde raster als de neerslagdata. Vanuit de Basisregistratie Grootchalige Topografie (BGT) is informatie over het landgebruik opgevraagd. In deze dataset wordt onderscheid gemaakt in panden, (onverhard) terrein, water en wegen. Van deze parameters is de fractie berekend dat deze parameter voorkomt per cel. Met een resolutie van 1km^2 resulteert 80.000 m^2 pand dus tot een fractie van 0.08. In dit onderzoek is de BGT van 2013 gebruikt.

Input voor de SOM is een raster gebruikt van 282 km (oost-west) bij 345 km (noord-zuid), met een resolutie van 1km^2 . In totaal zijn 22 verschillende parameterrasters gebruikt. Alleen punten op het land zijn gebruikt.

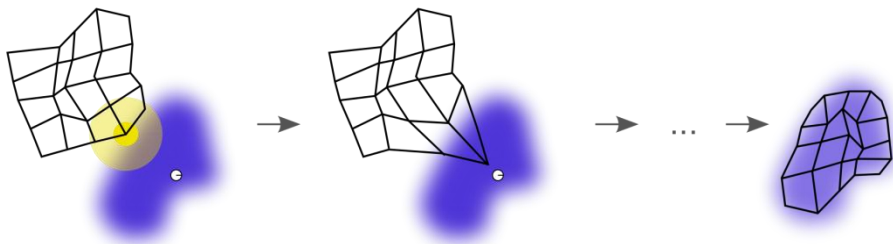
Methode

Om grote datasets te analyseren, worden vaak Machine Learning algoritmes gebruikt. Grofweg kan er onderscheid gemaakt worden tussen twee typen Machine Learning algoritmes: Supervised Learning en Unsupervised learning.

Supervised learning is een methode waarbij trainingsdata wordt gebruikt om een model te trainen. Een voorbeeld van zo een methode is het eerst toepassen van classificatie om vervolgens een regressiemodel toe te passen. Unsupervised learning methoden gebruiken de ruwe inputdata om structuren te vinden. Een neurale netwerk is een algoritme wat zowel binnen Supervised learning als Unsupervised learning wordt toegepast. Deze algoritmen worden onder andere gebruikt voor spraakherkenning, het herkennen van afbeeldingen, en het diagnosticeren van ziektes in de medische zorg. Het grote voordeel van Unsupervised learning is dat het algoritme zelf structuren in de data vindt en de gebruiker geen randvoorwaarden hoeft op te geven. In dit artikel wordt gebruik gemaakt van een Self-Organizing Map (SOM) algoritme, wat gebruikt maakt van deze neurale netwerken. Het nadeel van Unsupervised learning is dat er ook verbanden gevonden kunnen worden die fysisch niet bestaan.

Een SOM wordt getraind door gebruik te maken van Unsupervised learning en produceert een tweedimensionale representatie van de multidimensionale dataset (Miljkovic, 2017). Een SOM-algoritme kent twee fasen: de trainingsfase en de mappingfase. In de trainingsfase wordt een raster van neuronen geïnitieerd. De neuronen hebben dezelfde dimensie als de inputdata. Elke neuron krijgt een willekeurige gewichtsvector, wat in feite een positie is in de input ruimte. Vervolgens wordt een deelverzameling van de input gebruikt als trainingsdata. Voor een inputvector worden alle afstanden tot de gewichtsvectoren van de neuronen berekend. **Afbeelding 4** laat dit zien.

Een neuron is het deel van het neurale netwerk dat geprogrammeerd kan worden. Het neuron met de gewichtsvector die het meest lijkt op de inputvector wordt de Best Matching Unit (BMU) genoemd. De gewichtsvector van de BMU en de dichtbijgelegen neuronen worden aangepast zodat deze beter de input vertegenwoordigen. Dit proces wordt herhaald tot de neuronen uiteindelijk een goede representatie zijn van de invoergegevens (zie Afbeelding 3). Gedurende de mappingsfase wordt een inputvector toegekend aan het neuron met de kleinste (euclideanse) afstand. Het resultaat van de mappingsfase is dat elke inputvector van de dataset wordt gekoppeld aan een neuron. Een neuron correspondeert dus met meerdere inputvectoren.



Afbeelding 4: Illustratie van het trainen van een Self-Organizing Map. De blauwe polygoon representeert de invoergegevens, het witte raster de neuronen die zich initieel willekeurig in de ruimte bevinden. Na een aantal iteraties zal het raster van neuronen de structuur van de data vertegenwoordigen.

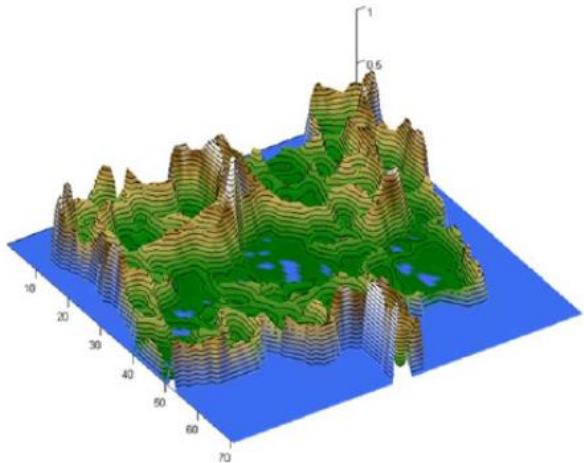
Een veel gebruikte manier om de resultaten van een SOM te visualiseren is de U-Matrix (Ultsch, 2007). De U-Matrix kan worden gezien als een representatie (samenvatting) van de multidimensionale data in lagere (en daarmee voor het menselijke oog analyseerbare) dimensies.

In een U-matrix worden de afstanden tussen de neuronen gevisualiseerd. Door deze afstanden in 3D te plotten, wordt een afbeelding verkregen zoals Afbeelding 6. Grote hoogtes in deze U-Matrix (bergen) betekenen dat er een groot verschil is in de onderliggende data voor het desbetreffende neuron. Plat gezegd hebben de onderliggende data dan niets met elkaar te maken. Lage waarden voor de U-Matrix (de dalen tussen de bergen in Afbeelding 7), houden in dat de neuronen dicht bij elkaar liggen en hetzelfde reageren op de input data. Deze

neuronen en daarmee de onderliggende data hebben dus grote overeenstemming met elkaar. Op deze manier kunnen neuronen worden onderscheiden die dicht bij elkaar liggen en daarom dezelfde onderliggende data representeren.

Het onderscheiden van deze groepen neuronen kan visueel gedaan worden, of er kan een clustering-algoritme worden toegepast om deze clusters te detecteren (Vesanto e.a., 2000). In dit artikel, waarin we de mogelijkheden van een SOM onderzoeken, is ervoor gekozen om visueel de U-Matrix te interpreteren.

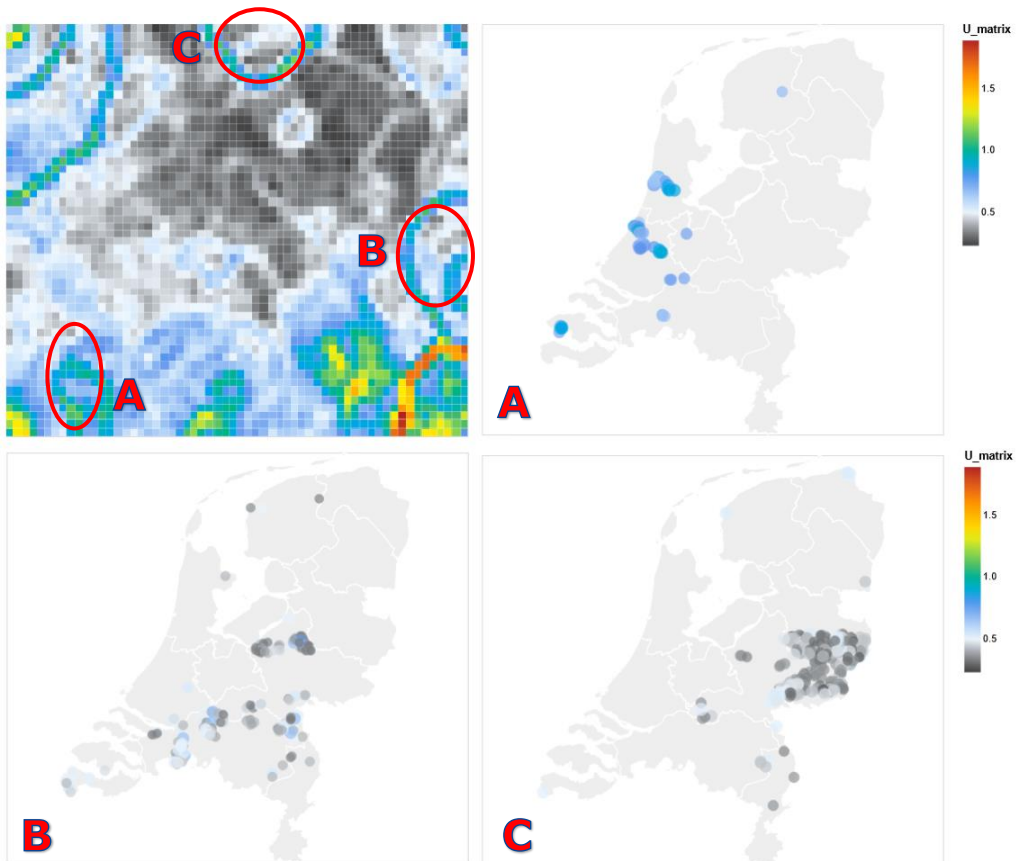
In deze studie is gebruik gemaakt van het python package "SOMPY" (Moosavi e.a., 2014). Een vierkant raster van 50 x 50 neuronen is gebruikt en de gewichten van de neuronen zijn geïnitieerd doormiddel van Principal Component Analysis. Het trainen van het netwerk duurde 31 seconden. Uniek aan de methode is dat allerlei parameters gekoppeld zijn die elk een aspect beschrijven van extreme neerslag. De resultaten van de (visueel) gedetecteerde clusters worden vervolgens gekoppeld aan de oorspronkelijke geografische locatie (de locatie is niet als input parameter gebruikt). Op deze manier ontstaat er de mogelijkheid om gedetecteerde clusters geografisch te toetsen.



Afbeelding 5: Voorbeeld van een U-Matrix waarin de afstanden tot de neuronen zijn gevisualiseerd als hoogte. Bergen representeren neuronen met een grote onderlinge afstand, dalen een kleine onderlinge afstand (Lötsch e.a., 2014).

Self-Organizing Map (SOM)

In Afbeelding 8 (boven links) zijn de 50x50 neuronen weergegeven als de U-Matrix, die het resultaat is van de toepassing van het SOM algoritme. Kleuren zijn gebruikt om de afstanden tussen de neuronen weer te geven. Gele en rode kleuren vertegenwoordigen een grote onderlinge afstand, zwarte en lichtblauwe kleuren een kleine. Er zijn verschillende groepen neuronen te zien die een kleine onderlinge afstand tot elkaar hebben (dalen), maar omgeven zijn door neuronen met een grotere afstand (de bergen, zie rode ellipsen in Afbeelding 9). Merk op dat er meerdere datapunten behoren tot één neuron. Er zijn drie voorbeelden van clusters geselecteerd die zijn gevisualiseerd op een achtergrondkaart van Nederland (Afbeelding 11 boven rechts en afbeelding 10 onder). Hoewel deze datapunten fysiek gezien ver van elkaar afdiggen (er zijn zowel datapunten in Friesland, Noord-Holland, Zuid-Holland, Brabant en Zeeland voor het geselecteerde cluster in a), is de onderliggende data ongeveer gelijk aan elkaar. Daarom zijn deze datapunten tijdens de mappingsfase toegekend aan neuronen met vergelijkbare gewichtsvectoren.



Afbeelding 11: U-matrix als resultaat van SOM. Er zijn drie voorbeelden van clusters geselecteerd waarvan de locaties van de onderliggende data gevisualiseerd zijn in a), b) en c), respectievelijk.

Nu er een aantal clusters zijn gedefinieerd, is het interessant om te onderzoeken waarom deze datapunten een cluster vormen. Als voorbeeld wordt verder ingegaan op cluster A in Afbeelding 12. Voor cluster A blijkt dat wanneer de onderliggende data van deze neuronen wordt bekeken, deze datapunten allemaal een keer te maken hebben gehad met extreme zomerse neerslag.

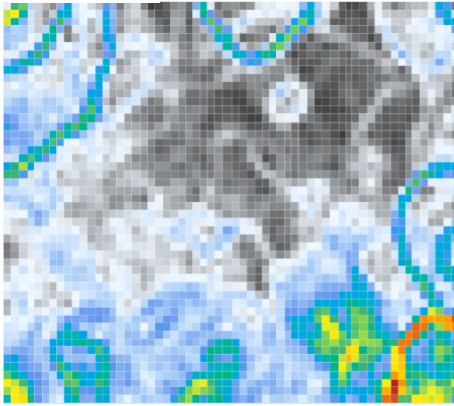
Anders gezegd: voor al deze datapunten is in de beschouwde periode (2000-2018) één of twee keer meer dan 80 mm neerslag (herhalingstijd 100 jaar) gevallen op een dag. Dit is gevisualiseerd in Afbeelding 6 (bovenste rij, rechter afbeelding).

De overschrijdingsfrequentiekaarten voor extreme zomerse neerslag is opgebouwd uit dagelijkse neerslaghoeveelheden. Wanneer naar de individuele gebeurtenissen gekeken wordt, kunnen deze worden toegekend aan een zomerse extreme neerslaggebeurtenis. De gebeurtenis in Amsterdam kan bijvoorbeeld worden herleid naar een wolkbreuk op 28 juli 2014. Delen van de A10 waren afgesloten, kelders liepen onder water, en vluchten vanaf Schiphol werden geannuleerd. De groep datapunten in Zuid-Holland kan gelinkt worden aan een extreme neerslag gebeurtenis in de provincie op 14 juli 2011. In drie dagen tijd viel in deze regio bijna 140 mm neerslag. In Friesland kreeg de stad Drachten op 26 augustus 2012 een flinke hoeveelheid neerslag te verwerken na een wolkbreuk. Straten kwamen onder water te staan omdat de riolering het overtollige water niet aankon.

In Afbeelding 6 zijn vijf parameters gevisualiseerd voor het geselecteerde cluster. Het blijkt dat de datapunten in het geselecteerde cluster allemaal a) een bovengemiddeld elektriciteitsverbruik hebben, en dat b) een relatief groot gedeelte van het oppervlak bestaat uit wegen (ongeveer 10 procent). Verder zijn er gemiddeld 10000 omgevingsadressen per vierkante kilometer (c), we classificeren dit als stedelijke gebieden. Tot slot is de fractie open terrein vrij laag voor de geselecteerde datapunten van het cluster (d).

Cluster A heeft daarmee extreme zomerse neerslaggebeurtenissen uit de overschrijdingsfrequentiekaarten kunnen filteren, die op een vergelijkbaar landgebruik zijn gevallen. Voor cluster B geldt dat de onderliggende datapunten allemaal een hoge fractie onverhard terrein hebben. Voor cluster C geldt, net als voor cluster B, dat de dichtheid van terrein redelijk hoog is. Echter is het elektriciteits- en gasverbruik voor de datapunten in cluster C vrijwel gelijk aan 0, terwijl voor de datapunten in cluster B het verbruik gemiddeld is.

U-Matrix



Aantal neerslag events >80mm/dag



Gem. elektriciteitsverbruik [kWh]



% wegdeel per km²



Aantal adressen



% open terrein per km²

Afbeelding 13: Ruimtelijke representatie van het geselecteerde cluster (rode ellips) voor vijf parameters. Van boven naar onder en links naar rechts: Het aantal gebeurtenissen met meer dan 80mm neerslag per dag ($T = 100$), het elektriciteitsgebruik per vierkante kilometer, fractie van oppervlakte dat uit wegen bestaat, aantal omgevingsadressen en fractie van oppervlakte dat uit terrein bestaat.

Discussie

In deze conceptstudie is een voorbeeld gegeven van hoe een neuraal netwerk in de vorm van een Self-Organizing Map (SOM) kan worden toegepast om een beter inzicht te krijgen in de (mogelijke) verbanden tussen verschillende parameters.

In deze studie zijn echter relatief eenvoudige parameters toegepast, die veelal aan elkaar gecorreleerd zijn (aantal inwoners is gecorreleerd met gas- en elektriciteitsverbruik, het oppervlak water en (onverhard) terrein zijn aan elkaar gerelateerd). In een vervolgstudie zouden andere, niet gecorreleerde, parameters kunnen worden toegevoegd (bijvoorbeeld fijnstofmetingen, dauwpunttemperatuur). Bovendien zou het interessant zijn om de tijdsafhankelijkheid van de input parameters mee te nemen.

Hierin dient opgemerkt te worden dat er niet beweerd wordt dat een bepaald aantal inwoners of een bepaalde waarde voor elektriciteitsverbruik wordt gelinkt aan extreme neerslaggebeurtenissen. Het voorbeeld dient veel meer om aan te tonen dat het toepassen van machine learning algoritmes op grote datasets snel en relatief eenvoudig mogelijk is. Het voorbeeld geeft aan dat verbanden gevonden kunnen worden, maar ook dat deze niet altijd fysisch correct zijn. Los van de beoordeling van de uitkomsten van het algoritme is het selecteren van de juiste parameter een essentiële stap.

Binnen het vakgebied hydrologie kan hier ook gebruik van worden gemaakt. In hydrologische modellen worden vaak veel parameters toegevoegd zonder altijd te weten hoe gevoelig de uitkomst van de modelberekeningen is voor deze parameters. Aan de andere kant kunnen op dezelfde manier verbanden worden gelegd tussen verschillende parameters waarvan dit niet verwacht zou worden. In dit artikel is getracht om verbanden te leggen tussen extreme neerslag en (relatief eenvoudige) parameters zoals landgebruik, aantal inwoners, gas- en elektriciteitsverbruik.

Deze methodiek biedt de mogelijkheid om onafhankelijk en onbevooroordeeld naar verbanden te zoeken. Nieuwe, onverwachte verbanden zouden daarmee gedetecteerd kunnen worden. Kortom het toepassen van unsupervised learning (in de vorm van een SOM-algoritme) is een nieuwe methodiek voor het vinden van verborgen patronen binnen het vakgebied hydrologie. Hierin dient wel opgelet te worden dat gevonden correlaties niet verwart worden met causaliteit.

Conclusie en aanbevelingen

Machine Learning is een krachtige methode om verbanden te vinden in de toenemende hoeveelheid data. In dit artikel is een conceptstudie beschreven waarin een neuraal netwerk is toegepast om patronen te vinden in gebeurtenissen met extreme neerslag. Door het toepassen van een Self-Organizing Map (SOM) algoritme konden clusters worden bepaald waarvan de datapunten een gelijksoortig karakter tonen, bijvoorbeeld een gebeurtenis met extreme neerslag in de zomer.

In deze conceptstudie zijn de tijdreeksen van de neerslag teruggebracht tot de overschrijdingsfrequentiekaarten, door het aantal dagen met een drempeloverschrijding te sommeren. Het meenemen van de complete tijdreeks zou tot een beter resultaat kunnen leiden.

De resultaten laten zien dat er clusters gevonden zijn die gekoppeld zijn aan verschillende gebieden in Nederland. Een voorbeeld toonde aan dat verschillende locaties in Nederland (Amsterdam, delen van Zuid-Holland, Drachten, Middelburg en Breda) een cluster vormen, gebaseerd op de onderliggende data. Deze locaties hebben een stedelijk karakter en zijn in de beschouwde periode alle getroffen door extreme zomerse neerslag. Ondanks dat deze gebeurtenissen op verschillende tijdstippen en plaatsen hebben plaatsgevonden, en het SOM-algoritme geen informatie heeft over de locatie van de punten, worden deze datapunten gegroepeerd tot één cluster.

Mogelijk interessante vervolgonderzoeken kunnen zijn:

- Het analyseren van grote datasets van meteorologische variabelen in tijd en ruimte om het optreden van extreme neerslag beter te begrijpen.
- Het analyseren van gebiedseigenschappen op basis van remote sensing-beelden om parameters van wel en niet bemeten stroomgebieden over te kunnen dragen.
- Het analyseren van tijdreeksen van verschillende waterstands- en afvoerlocaties om afwijkingen en fouten in de meetreeks op te sporen (validatie).
- Het verbeteren van tijdreeksmodellen voor bijvoorbeeld grondwaterstandsmetingen, door de beschikbare parameters met elkaar te clusteren en de bijdrage van verschillende parameters op verschillende locaties te analyseren.
- Het onderzoeken van de meerwaarde van deze techniek ten opzichte van een klassieke multivariate regressie-analyse.

Bibliography

Tijdschrift:

Buishand, T.A. en R. Jilderla en J.B Wijngaard (2010) *Regionale verschillen in extreme neerslag*, De Bilt; In KNMI wetenschappelijk rapport.

Hakvoort, H. en J. Beersma en T.Brandsma en R. Versteegh en K. Peerdeman (2016) *Nieuwe statistieken: extreme neerslag neemt toe en komt vaker voor*; in H2O.

Lötsch, J. en A. Ultsch (2014) *Exploiting the structures of the U-Matrix*; Nog te verschijnen.

Miljkovic, D. (2017) *Brief Review of Self Organizing Maps*, Zagreb; in MIPRO, pages 1252-1257.

Ultsch, A. (2007) *Emergence in Self Organizing Feature Maps*, Marburg; in Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM).

Vesanto, J. en E. Alhoniemi (2000) *Clustering of the Self-Organizing Map*; IEEE, vol 11, NO 3 pp. 586-600.

Webpagina:

Centraal Bureau voor de Statistiek (CBS) (2018) *Extreme neerslag in Nederland 1910-2017*, Den Haag; url <https://www.clo.nl/indicatoren/nl0590-neerslag-extremen>.

Shepherd, J.M. (2002) *NASA satellite confirms urban heat islands increase rainfall around cities*; in NASA/Goddard Space Flight Centre ScienceDaily, url www.sciencedaily.com/releases/2002/06/020619074019.htm.

Moosavi, V. en S. Packmann (2014) *A Python Library for Self Organizing Map (SOM)*; github, url <https://github.com/sevamoo/SOMPY>.

Summary More understanding of extreme rainfall events through the use of neural networks.

Extreme rainfall events have a huge impact on the water system. In recent years the rainfall statistics of the Netherlands were researched (Hakvoort, et al., 2016). This research provides insight in the statistics of extreme rainfall events in the Netherlands. However, the knowledge of regional occurrence of extreme rainfall events is still limited.

In conventional data science methods a limited number of data sources and parameters are analyzed with statistical models and limited geographical relations. This article presents a methodology on the use of a neural network in the form of a Self-Organizing Maps (SOM) on the analysis of extreme rainfall. Using this technique patterns between various parameters can be made visual. This can help to find patterns within the parameters which can explain the occurrence of regional, extreme rainfall.

The proof-of-concept presented in this article is based on parameters from open-datasets available for the Netherlands.

We expect to use this methodology more often to gain more insight in complex relations within hydrology and the field of extreme rainfall. However the proof-of-concept shows as well that it is important to critically analyze the results on their accuracy.

Auteurs

JOOST DRIEBERGEN
HKV lijn in water
j.driebergen@hkv.nl

NICOLE JUNGERMANN
HKV lijn in water
n.jungermann@hkv.nl

