

Auteur Dr. ir. R.J. (Robbert-Jan) Dikken, Peutz bv

# Lokale zonvoorspelling met hybride AI op basis van beperkte weersdata

*Het accuraat voorspellen van lokale zonstraling ter plaatse van gebouwen is met de energietransitie in toenemende mate van groot belang in het optimaliseren van het energiegebruik. Dit heeft meerdere kanten. Informatie over instantane en verwachte zonstraling kan worden gebruikt voor het optimaal regelen van klimaatinstallaties, zodat comfort voor gebruikers van een gebouw op een energie-efficiënte manier wordt gerealiseerd. Ook is informatie over verwachte zonstraling essentieel, wanneer er fotovoltaïsche (PV) panelen op het dak van het gebouw geïnstalleerd zijn, in het realiseren van optimale laadstrategieën van elektrische voertuigen (EV) op de parkeerfaciliteit van een gebouw. Met de huidige ontwikkelingen van elektrificatie in de maatschappij en de belasting die dit legt op de elektriciteitsnetten is de informatie over verwachte zonstraling, en dus opgewekt vermogen, essentieel om te optimaliseren op zelfconsumptie en reductie van piekbelastingen. Vanuit dit perspectief is het van belang om niet alleen informatie te hebben over de verwachte zonstraling, maar ook over de robuustheid van de voorspelling.*

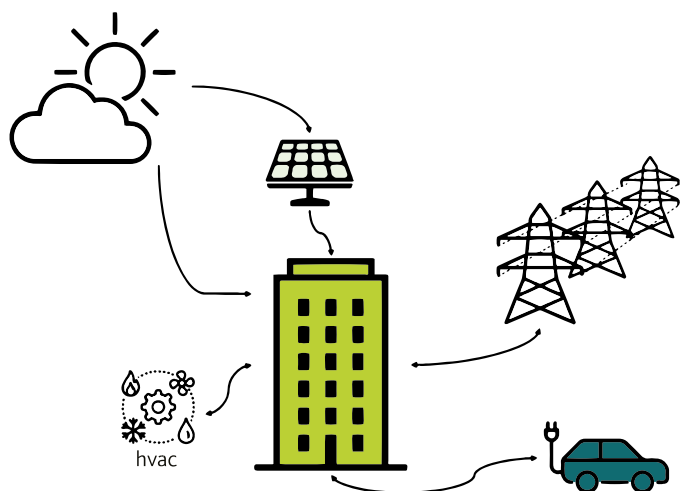
## Zon als warmtewinst en koellast

Het effect van de energie van de zon op het energiesysteem en comfortniveau van een gebouw is niet eenduidig te definiëren. In de winterperiode is zonwarmte een efficiënte manier om warmte binnen het gebouw te krijgen en kan hiermee het energiegebruik voor verwarming worden gereduceerd. Daarentegen is in de zomerperiode de zonwarmte juist een belasting, wat extra energiegebruik vraagt om zonwarmte die het gebouw in komt weg te koelen. In figuur 1 is dit schematisch weergegeven. Informatie over het verwachte lokale zonvermogen is dan ook zeer belangrijk in relatie tot het binnenklimaat van een gebouw. Deze informatie kan worden gebruikt om efficiënt het gebouw te gebruiken door bijvoorbeeld

te sturen op het gebruik van de verschillende zones van het gebouw, of optimale aansturing van klimaatinstallaties voor verwarming, koeling en ventilatie (HVAC) en zonwering. Een dergelijke optimale sturing is toestandsafhankelijk, waarbij de toestandsvariabelen bestaan uit externe factoren zoals buitenluchttemperatuur en zonvermogen en interne factoren zoals gewenst binnenklimaat en huidig binnenklimaat.[1]

## Opgewekte zonne-energie en systeemstabiliteit

Naast het effect van zon op het binnenklimaat in een gebouw is het effect van zon ook heel direct op het opgewekte vermogen van PV-panelen, die steeds vaker en in grotere omvang op de daken van gebouwen zijn geïnstalleerd. Daarnaast hebben gebouwen ook steeds vaker en in grotere omvang laadfaciliteiten voor elektrische voertuigen (EV). Het opgewekte vermogen van PV leidt bij een zuid-oriëntatie, en in beperktere mate ook bij een oost-west-oriëntatie, van de panelen tot een piek. Wanneer dit opgewekte vermogen niet direct lokaal kan worden verwerkt, leidt dit tot een belasting van het elektriciteitsnet. Omgekeerd leidt het simultaan opladen van meerdere EV ook tot een belasting van het elektriciteitsnet wanneer er geen lokaal opgewekte energie gebruikt wordt. In figuur 1 is hier



**Figuur 1:** Schematische weergave van het belang van informatie over zonstraling in verband met het energiegebruik voor HVAC en EV-laadstrategieën.

een schematische weergave van gegeven. De onbalans van deze belastingen op het elektriciteitsnet zijn op grote schaal zeer problematisch en kunnen tot instabiliteit en uitval leiden [2][3]. Informatie over het lokaal verwachte zonvermogen is dus zeer relevant voor het lokaal kunnen balanceren van aan de ene kant energieaanbod van opgewekt zonvermogen en aan de andere kant energievraag van EV en gebouwgebruik en -klimatisering.

#### Smart buildings en de werkelijkheid van bestaande bouw

Veel nieuwbouw wordt om bovenstaande redenen ontworpen als smart buildings. Deze gebouwen zijn volgehangen met een veelzijdigheid aan sensoren om binnenklimaat en gebruik te monitoren. De grote hoeveelheid aan sensordata geeft informatie over gebouwgebruik, energiegebruik en binnenklimaat. Deze informatie wordt veelal middels online dashboards gevisualiseerd, zodat gebouwbeheerders hierop kunnen acteren voor bijvoorbeeld het instellen van setpoints van installaties. Een stap verder is het direct kunnen aansturen van installaties via optimalisatie-algoritmes, die gevoed worden met de sensordata. Het anticiperend kunnen regelen van installaties, waarbij vraag en aanbod gebalanceerd worden, kan leiden tot een significante reductie in benodigd geïnstalleerd vermogen en een afname van het energiegebruik voor klimatisering. [1]

De werkelijkheid is echter dat nieuwbouw die als smart building ontworpen en uitgevoerd wordt, absoluut niet het merendeel van de gebouwen betreft. Het merendeel van de gebouwen is bestaande bouw, met beperkte tot geen toepassing van sensoren, en dus weinig inzicht in gebouwgebruik, energiegebruik en binnenklimaat. Sensoren die eenvoudig op een kosteneffectieve manier toe te passen zijn in bestaande bouw, zijn

bijvoorbeeld temperatuur-, CO<sub>2</sub>-, relatieve luchtvochtigheid-, en luchtdruksensoren. Wanneer deze sensoren in het gebouw worden geïnstalleerd, samen met een buitenunit, geven deze sensoren informatie over het binnenklimaat in het gebouw, en tevens een indicatief inzicht in het energiegebruik voor klimatisering. Het is dan echter nog niet evident wat lokaal het instantane zonvermogen is, en zeker niet wat het verwachte zonvermogen voor de komende uren is waarmee EV-laadfaciliteiten en HVAC optimaal aangestuurd kunnen worden.

#### Onderzoek bepaling zonvermogen met minimale data

Er zijn verschillende methodes om lokale zonstraling te bepalen. Het kan via sensoren gemeten worden, en op basis van de sensordata kan met bepaalde onzekerheid het verwachte zonvermogen voorspeld worden. Deze sensoren zijn echter niet gangbaar geïnstalleerd bij gebouwen. Op basis van een fysisch model, waar de zonhoeken in gemodelleerd zijn, kan ook een indicatie worden verkregen van het verwachte zonvermogen. Lokale en temporele invloeden, zoals bewolking, zijn hierin echter zeer slecht te incorporeren. Een methode die in deze studie onderzocht wordt, is de mogelijkheid om op basis van sensordata van eenvoudige, kosteneffectieve sensoren - zoals temperatuur-, relatieve luchtvochtigheid- en luchtdruksensoren - te bepalen wat het verwachte zonvermogen is. Data van deze sensoren verschaft in brede zin informatie over de klimatisering van een gebouw. Het kunnen extraheren van verwacht zonvermogen uit deze data is dan ook een additionele winst. Studies geven een indicatie van de mogelijkheden van het kunnen voorspellen van zonvermogen op basis van fysische modellen of meetdata. [4][5] In de meeste studies die beschikbaar zijn op dit onderwerp, wordt echter weinig aandacht besteed aan de combinatie van het effectief kunnen voorspellen van zonvermogen met beperkt aantal parameters en de betrouwbaarheid van de voorspelling. Dit laatste is zeer relevant in de bepaling van optimale regelstrategieën van klimatisering en EV-laadfaciliteiten.

Om te onderzoeken of temperatuur-, RV-, en luchtdrukdata afdoende impliciete informatie bevatten over het geassocieerde zonvermogen en betrouwbaarheid van de voorspelling wordt uurgemiddelde meetdata van het KNMI gebruikt. Deze data zijn betrouwbaar en vrij toegankelijk. [6] De meetdata van temperatuur ( $T$ ), relatieve luchtvochtigheid ( $RV$ ) en luchtdruk ( $P_{atm}$ ) wordt aangevuld met een cyclische dag van het jaar ( $N_{dag,cyclisch} = -\cos [2\pi N_{dag}/365]$ ) en een cyclisch uur van de dag ( $H_{cyclisch} = -\cos [2\pi H/24]$ ), om seizoens- en etmaaleffecten op een efficiënte manier te laten correleren met wat fysisch bekend is over de zonstraling ( $Q_{sol}$ ). In figuur 2 is een correlatiematrix gegeven van de verbanden tussen de hierboven benoemde parameters, aangevuld met de bewolgingsgraad en fysische modeldata. Etmaal- en seizoensinvloeden zijn zoals verwacht duidelijk terug te zien. Ook temperatuur, RV, luchtdruk en

bewolking zijn duidelijk gecorreleerd met de zonstraling. De fysische modeldata laat zoals verwacht ook een duidelijke correlatie zien met het uur van de dag. De correlaties tussen zonstraling en de modelparameters temperatuur, luchtvochtigheid en luchtdruk zijn negatief ten opzichte van de correlatie tussen bewolking en deze drie modelparameters, in lijn met de negatieve correlatie tussen bewolgingsgraad en zonstraling.

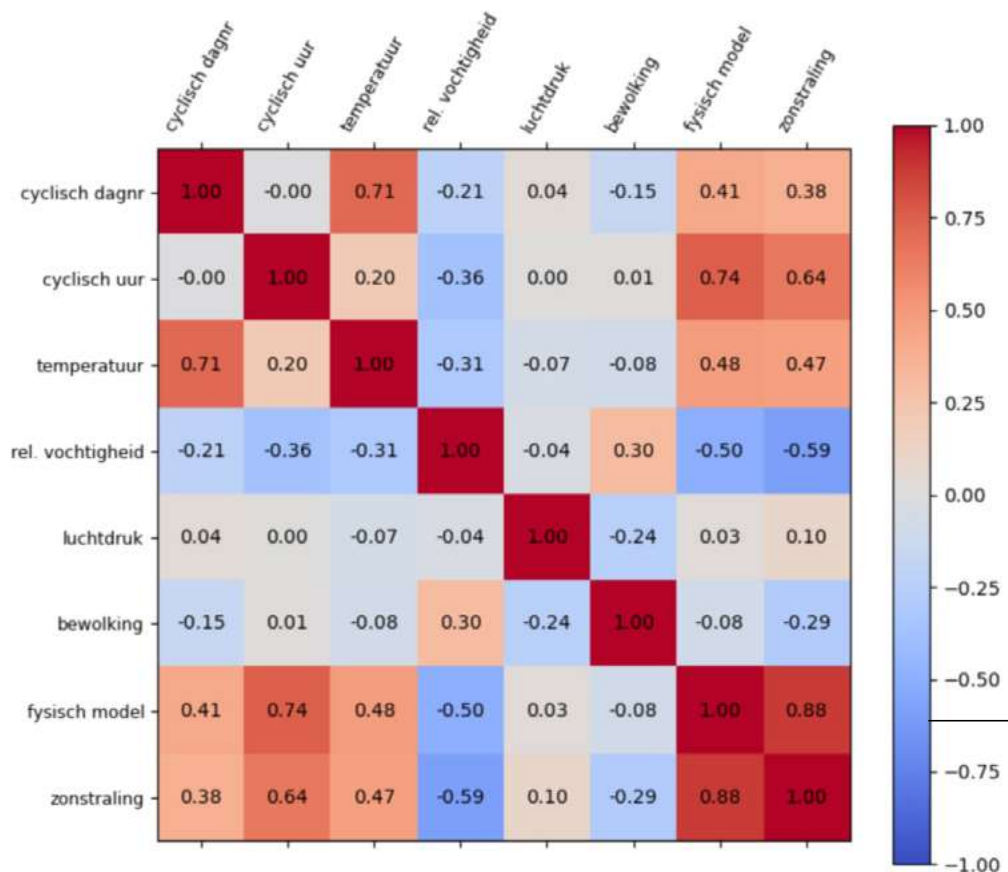
Uit dit correlatieonderzoek blijkt dat het effect van bewolking op de zonstraling op het aardoppervlak naar verwachting impliciet gerepresenteerd wordt door temperatuur, luchtvochtigheid en luchtdruk. Dit correlatieonderzoek duidt dat op basis van sensordata van temperatuur, RV en druk, waar dus kosteneffectieve sensoren op de markt van zijn die in bestaande bouw geïmplementeerd kunnen worden, indicatief inzicht kan worden verkregen in de mate van zonstraling. Hieronder zal de methodiek hiervoor beschreven worden.

Artificiële neurale netwerken en *committee machines*

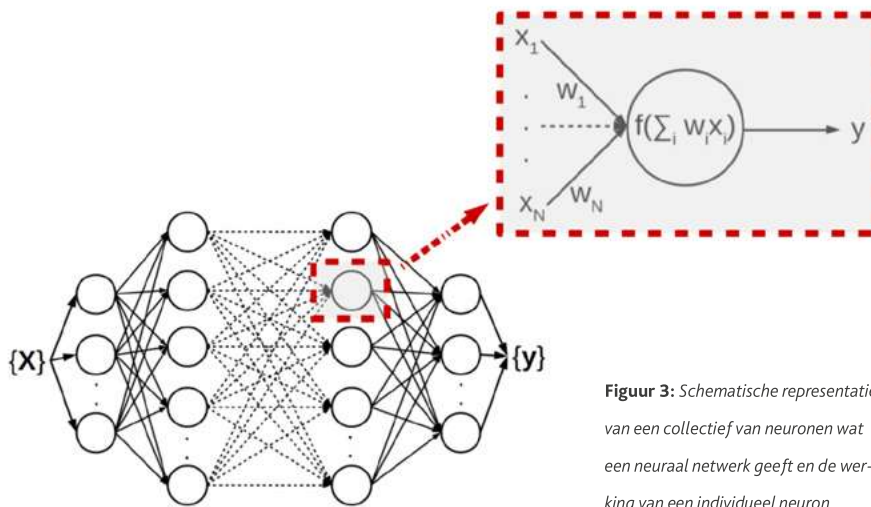
Voor datagedreven modellen zijn er verschillende machine learning-algoritmes beschikbaar. Eén van de

meest flexibele methodes zijn artificiële neurale netwerken. Door de flexibiliteit van deze methode is het ook een zeer geschikte kandidaat voor modellen die zonstraling voorspellen op basis van beschikbare informatie. Deze netwerken bestaan uit neuronen die met elkaar verbonden zijn, waarbij de sterkte van de connecties (gewichten) variabel zijn en afhangen van de specifieke representatie van het netwerk [7]. Ieder neuron heeft een relatief eenvoudige wiskundige functie en geeft een uitgangswaarde op basis van de gewogen inputsignalen op het specifieke neuron. Het collectief van neuronen is uiteindelijk verantwoordelijk voor het kunnen mappen van complexe functies tussen input en output. In Figuur 3 is een schematische representatie van een neuraal netwerk en de werking van een individueel neuron gegeven.

Een neuraal netwerk wordt getraind door de individuele gewichten tussen neuronen initieel willekeurig te definiëren. Vervolgens wordt met een trainingsdataset het neuraal netwerk van input voorzien. Door de output van het netwerk te meten en te vergelijken met de bekende output in de dataset, kan de fout hierin gebruikt worden om middels een optimalisatie-algoritme supervised de weegfactoren tussen de neuronen zo aan te passen dat de fout iteratief geminimaliseerd wordt. Zo ontstaat uiteindelijk een netwerk dat de verbanden in de dataset



Figuur 2: Correlatiematrix om inzicht te geven in de verbanden tussen seizoens-, en etmaaleffecten, temperatuur, RV, luchtdruk, bewolking, fysische modeldata en zonstraling.



**Figuur 3:** Schematische representatie van een collectief van neuronen wat een neuraal netwerk geeft en de werking van een individueel neuron.

zo goed mogelijk representeert, wat op een onafhankelijke dataset getest kan worden om te verifiëren dat het netwerk kan generaliseren en niet gebonden is aan de trainingsdataset.

Het proces van trainen van een neuraal netwerk is uiteraard sterk afhankelijk van de kwaliteit en volledigheid van de dataset. Beperkingen in de dataset leiden tot beperkingen in een model. Een methode om het effect van beperkingen in kwaliteit en/of volledigheid van de dataset te reduceren is door het toepassen van een ensemble-methodiek [8], schematisch weergegeven in figuur 4. Door meerdere netwerken met verschillende random initialisatie te trainen en deze netwerken vervolgens in een comité te laten opereren, als een *committee machine*, kan in veel gevallen een betere voorspelling verkregen worden. De ensemble-middeling reduceert het effect van gebrekkige data.

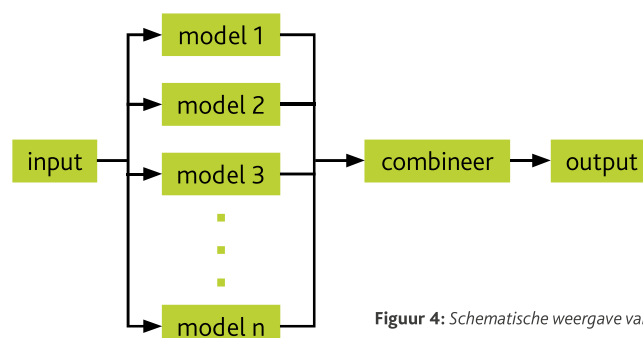
De ensemblemethodiek biedt tevens de mogelijkheid om een indicatie van de betrouwbaarheid van de voorspelling te krijgen, door naar de spreiding in de individuele voorspellingen te kijken.

#### Instantaan zonvermogen

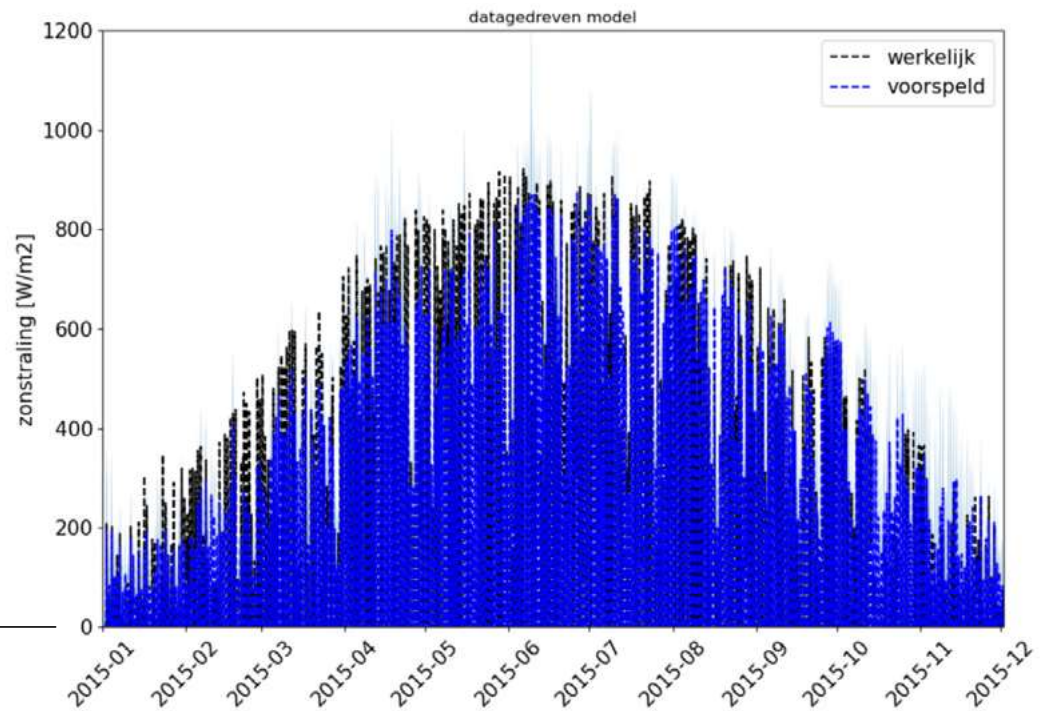
Alvorens in te gaan op de voorspelling van het verwachte zonvermogen voor de komende uren wordt allereerst gekeken naar de mate

van representatie die gehaald kan worden met artificiële neurale netwerken in een *committee machine* voor het bepalen van het instantaan zonvermogen. Hiervoor wordt een dataset van ruim 5 jaar aan observaties gebruikt en worden 50 neurale netwerken getraind op 75% van deze dataset, waarbij de initialisaties voor alle 50 netwerken willekeurig gekozen worden. De *committee machine* van de 50 neurale netwerken wordt onafhankelijk getest op de resterende 25% van de dataset. De inputparameters van de netwerken zijn de cyclische dag van het jaar en uur van de dag, de temperatuur, de relatieve luchtvochtigheid en de luchtdruk. De outputparameter van de netwerken is het geassocieerde zonvermogen in  $Wm^{-2}$ .

In figuur 5 is voor een heel jaar in de testset de gemeten zonstraling en de bepaalde instantane zonstraling weergegeven samen met de door het model bepaalde marge op de bepaling. De datagedreven voorspelling heeft een nauwkeurigheid van  $R^2=0,88$  en correleert met 0,94 met de werkelijk gemeten waarden.



**Figuur 4:** Schematische weergave van een ensemble-methodiek leidend tot een *committee machine*.



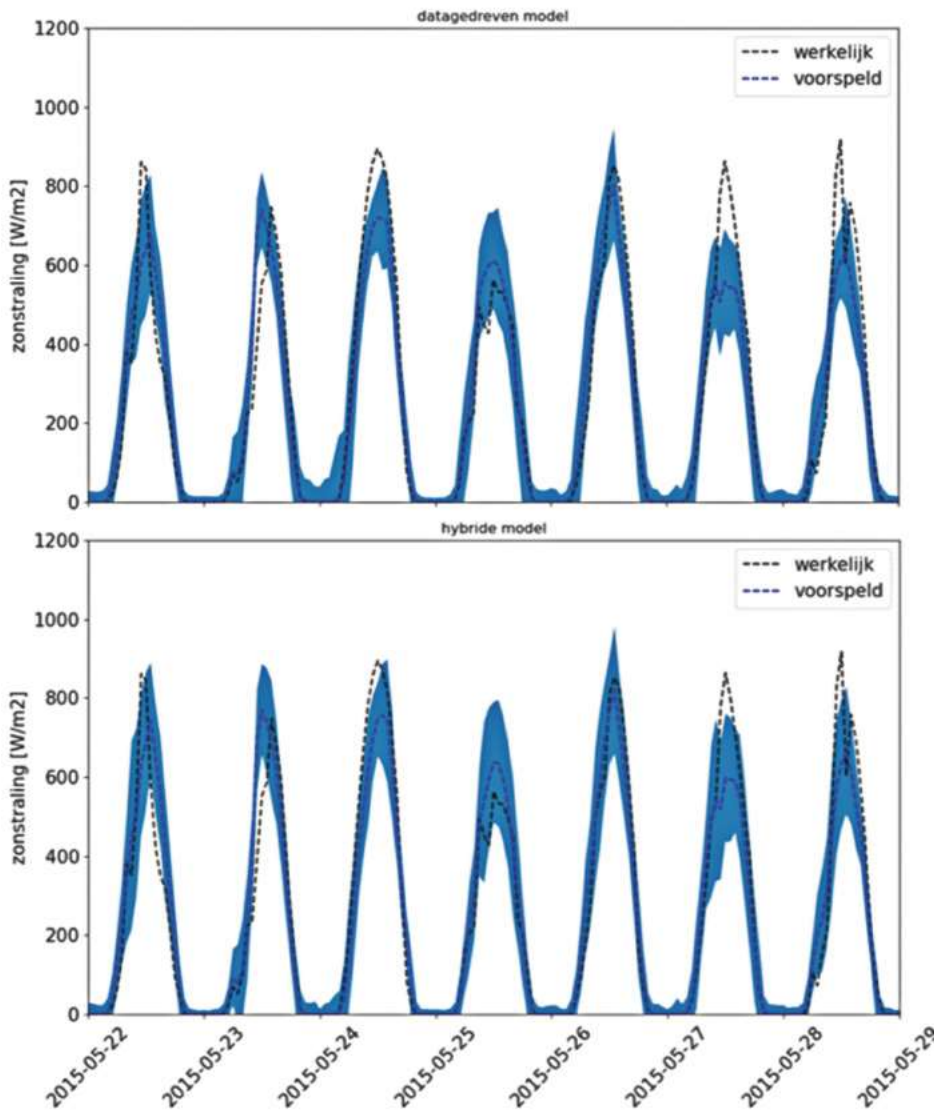
**Figuur 5:** Bepaling instantane zonstraling voor een heel jaar in de onafhankelijke testset.

In figuur 5 is de globale overeenkomst tussen de bepaalde zonstraling en de gemeten waarden goed te zien. Dit geeft hoofdzakelijk de seizoensgebonden effecten door het jaar heen weer. Om het effect van weersinvloeden - gerepresenteerd in het model door gemeten temperatuur, relatieve luchtvochtigheid en luchtdruk - beter inzichtelijk te maken is in figuur 6 een uitsnede voor een week gegeven, zowel puur datagedreven als hybride met additionele fysische modeldata. Door de dag van het jaar en het uur van de dag cyclisch te representeren is fysische kennis van zonstraling op het aardoppervlak deels al geïmplementeerd in de data.

Het combineren van meetdata met fysische modeldata heeft in veel gevallen toegevoegde waarde.[9] Zoals benoemd is op basis van lengte- en breedtegraad te berekenen wat de zonstraling voor een onbelemmerde doorgang door de atmosfeer zonder bewolking het jaar rond is. Dit kan als extra fysische modelparameter worden gebruikt, waardoor de fysische context wordt verbreed. Dit leidt tot een verbetering van de voorspelling. De

determinatiecoëfficiënt gaat van  $R^2=0,88$  voor de puur datagedreven voorspelling naar  $R^2=0,91$  voor de hybride methode. De voorspelling correleert met 0,95 met de werkelijk gemeten waarden. In de onderste figuur van figuur 5 is een resultaat van het hybride model gegeven, met de additionele fysische modeldata. Naast de subtiele verbetering in de voorspelling is te zien dat de spreiding in de voorspelling tevens een betere indicatie van de verwachte fout in de voorspelling geeft wanneer het hybride model met fysische modeldata wordt gebruikt. De meetwaarden vallen voor een groot deel binnen de breedte van de voorspelling.

Ter vergelijking van de methodes om zonstraling te voorspellen met een puur fysisch model, een datagedreven model en een hybride model, waarbij data en fysica worden gecombineerd, is in tabel 1 een overzicht gegeven van een aantal vergelijkende aspecten. Het is duidelijk dat over de gehele lijn het hybride model het best presteert. Wat ook opvallend is, is dat de combinatie van meetdata en fysische



**Figuur 6:** Voorbeeld van een week in de test-set waarbij de zonstraling is bepaald met een committee machine op basis van cyclisch dag van het jaar en uur van de dag, temperatuur, relatieve luchtvochtigheid en luchtdruk. De bovenste figuur representeert een resultaat van het datagedreven model met  $R^2=0,88$ . De onderste figuur representeert een resultaat van het hybride model met  $R^2=0,91$ .

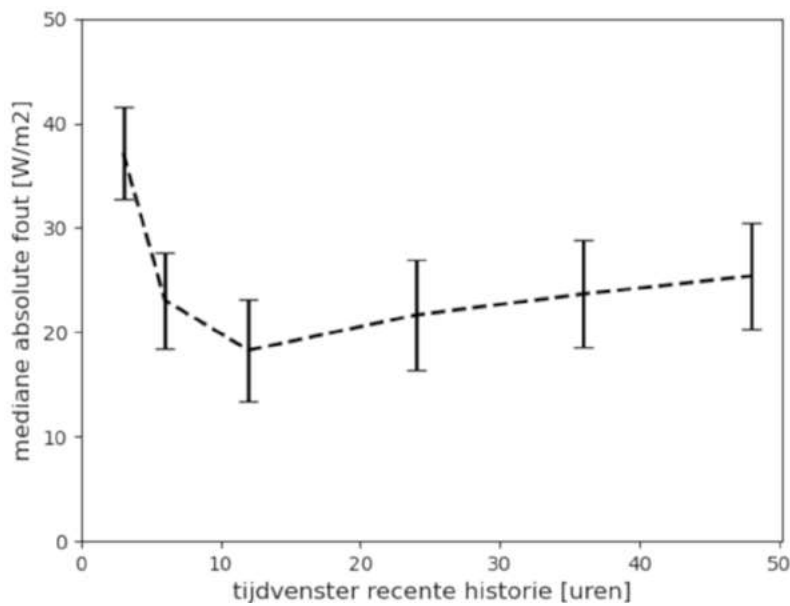
modeldata voor het hybride model resulteert in een betere correlatie tussen de spreiding in de voorspelling van de *committee machine* en de afwijking tussen de voorspelling en de meetwaarden. Dit is zeer nuttig, aangezien op basis van de spreiding in de voorspellingen dus een indicatie kan worden verkregen van de robuustheid van de voorspelling.

### Dynamische voorspellingen

In het voorgaande is beschreven hoe datagedreven en hybride modellen zeer geschikt zijn voor instantane bepalingen van zonstraling op basis van een beperkt aantal inputparameters die met kosteneffectieve sensoren verkregen kunnen worden. Voor dynamische strategieën voor EV-laadfaciliteiten en HVAC,

	Fysisch model	Datagedreven model	Hybride model
$R^2$ voorspelling	0,60	0,88	0,91
Correlatie voorspelling en meetwaarden	0,88	0,94	0,95
Mediane absolute afwijking	7,7	3,0	2,7
Correlatie spreiding en afwijking	-	0,60	0,66

**Tabel 1:** Vergelijk van verschillende modellen voor bepaling van zonstraling.



**Figuur 7:** Mediane absolute fout in de 12-uursvoorspelling als functie van verschillende tijdvensters met sequentiële meetwaarden waaruit blijkt dat een tijdvenster van 12 uur recente historie optimaal is.

bijvoorbeeld met reinforcement learning-algoritmes, is het zeer nuttig iets te kunnen zeggen over de verwachte zonstraling in de komende uren. Door een soortgelijke *committee machine* te trainen op sequenties van waarden, waarbij de neurale netwerken de associatie leggen tussen de meetwaarden van de afgelopen uren en de komende 12 uur, kan deze *committee machine* worden gebruikt om voor nieuwe situaties de verwachte zonstraling 12 uur vooruit te voorspellen. De *committee machine* wordt wederom getraind op een trainingdataset en getest op een onafhankelijke dataset. In figuur 7 is voor verschillende tijdvensters van sequentiële meetwaarden de mediane absolute fout weergegeven. Dit is meerdere keren gedaan met verschillende datasets, om een indicatie van de robuustheid te krijgen. Het blijkt dat deze methode met een meetwaarden-sequentievenster van 12 uur het beste presteert voor een voorspelling 12 uur vooruit.

In figuur 8 zijn resultaten gegeven voor de voorspelling van de zonstraling 12 uur vooruit voor zes opeenvolgende tijdvensters. De voorspelling met marge komt goed overeen met de werkelijke waarden. De nauwkeurigheid van de voorspelling is  $R^2=0.87$  en de correlatie tussen de marge die de *committee machine* geeft en de werkelijke fout is 0,82. Dit betekent dat de marge indicatief is voor de robuustheid van de voorspelling, zoals eerder ook geconcludeerd voor de instantane bepaling van zonstraling.

#### Samenvattende conclusie

In dit artikel is een onderzoek beschreven naar de mogelijkheid om met beperkte meetdata van temperatuur, luchtvochtigheid en luchtdruk een voorspelling te doen naar de instantane en verwachte zonstraling. Dit blijkt goed mogelijk te zijn door de intrinsieke verbanden tussen deze meetwaarden en bewolking. De hybride AI-benadering, waarbij gemeten waarden worden aangevuld met fysische modeldata, blijkt een goede methode om zonstraling te bepalen en voorspellen.

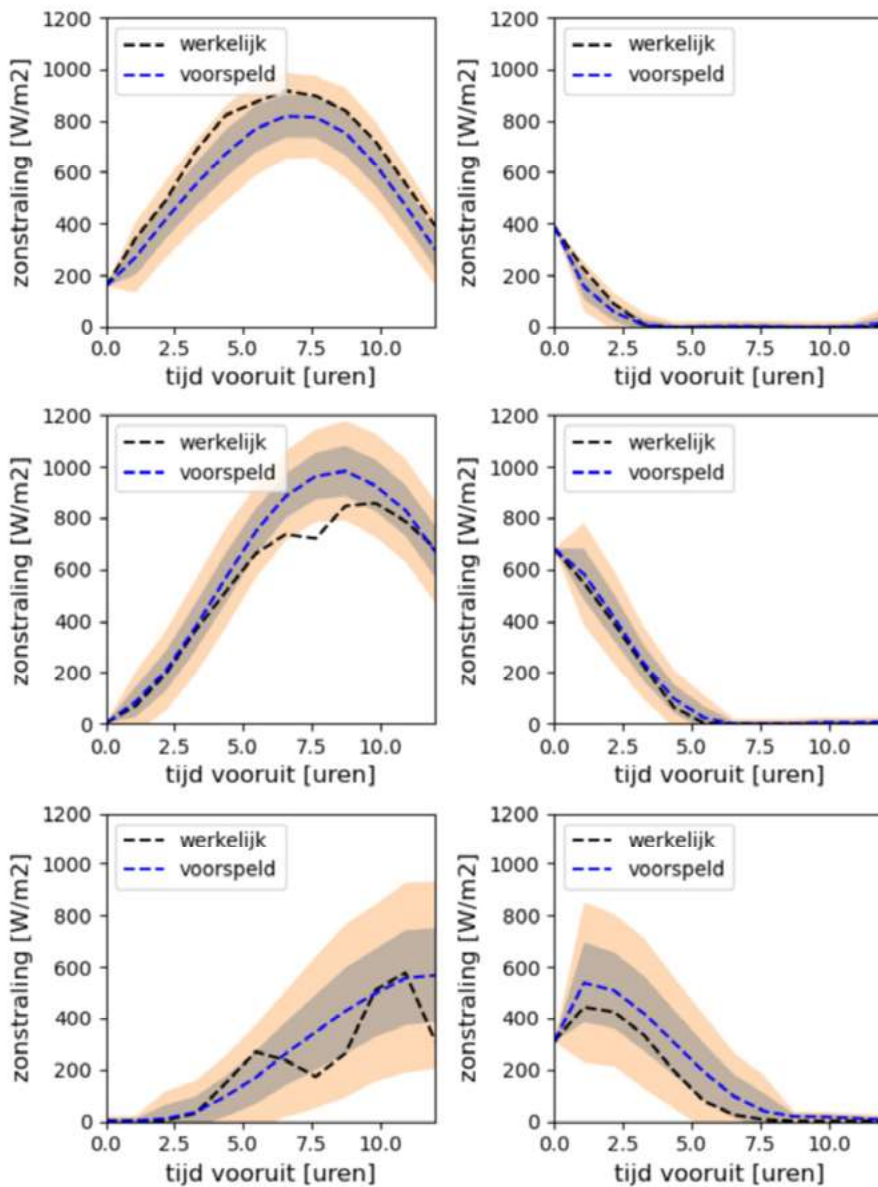
Door de toepassing van meerdere modellen in deze benadering is het mogelijk om tevens een indicatie te verkrijgen van de robuustheid van de voorspelling. Deze methodiek kan gebruikt worden om inzicht te geven in de instantane en verwachte zonstraling ten behoeve van optimalisatie van energiesystemen.

#### Vermelding

Dit onderzoek is deels uitgevoerd binnen het Brains4Buildings project met financiering vanuit de regeling Missiegedreven Onderzoek, Ontwikkeling en Innovatie (MOOI). [10]

Dr. ir. R.J. (Robbert-Jan) Dikken





Figuur 8: Zes opeenvolgende 12-uur voorspelingstijdvensters van de zonstraling, onafhankelijk getest van de trainingsdata.

Referenties

1. R.J. Dikken, Machine learning en data science voor klimaatinstallaties, TVVL Magazine 7, november 2019.
2. A. Faber, J.P.M. Ros, Decentrale elektriciteitsvoorziening in de gebouwde omgeving; Evaluatie van transities op basis van systeemopties, Planbureau voor de Leefomgeving, PBL-publicatienummer 500083011, april 2009.
3. E.J. Wiggelinkhuizen, B.H. Bulder, A.B. Schwedersky, M.P.W. van Berlo, Verkenning van toekomstige risico's voor het elektriciteitsnet, TNO 2020 R12069, 26 mei 2021.
4. D.K. Chaturvedi, Solar Power Forecasting: A Review, International Journal of Computer Applications (0975-8887), Volume 145, No.6, July 2016.
5. X.Yang et al., Simplified Method for Predicting Hourly Global Solar Radiation Using Extraterrestrial Radiation and Limited Weather Forecast Parameters, Energies, 16, 3215, April 2023.
6. <https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens>.
7. T.M. Mitchell, Machine Learning, The McGraw-Hill Companies, Inc. (1997).
8. S. Fort, H. Hu, B. Lakshminarayanan, Deep Ensembles: A Loss Landscape Perspective, arXiv:1912.02757 (2020).
9. R.J. Dikken, Voorspellen van luchtstrominggeluid met CFD en machine learning, TVVL Magazine 6, december 2020.
10. <https://brains4buildings.org/>.