



Detectie zonnepanelen regionaal

Vervolgonderzoek in het kader van VIVET

25 augustus 2022

1. Aanleiding

Het CBS en het Kadaster hebben verschillende methoden en bronnen tot hun beschikking waarmee inzicht geboden kan worden in de locatie van zonnepanelen. Dit onderzoek is een vervolg op de **Verkenning samenhang regionale zonnestroomcijfers (2020)**, waarin beide methoden en databases naast elkaar waren gelegd. De verschillen die toen waren geconstateerd hadden verschillende oorzaken. In het vervolgonderzoek is uitgezocht hoe de mismatches ontstaan, of deze weggenomen kunnen worden, en welke vervolgstappen er zijn.

Het CBS en het Kadaster hebben dit onderzoek samen uitgevoerd in het kader van [Vivet](#). De bevindingen van het onderzoek en de suggesties die hieruit zijn gerold worden door beide organisaties gedeeld middels presentaties. Deze notitie is een schriftelijke bondige vastlegging van het onderzoek ter naslag. Mocht u vragen of ideeën hebben naar aanleiding van hetgeen dat u hier heeft gelezen, dan nodigen we u graag uit om contact met ons op te nemen¹.

2. Aanpak

Het project is uitgevoerd door methodologen van het CBS en geodataspecialisten van het Kadaster. Uit de voorgaande analyse bleek dat er 2 belangrijke redenen waren voor de gevonden verschillen tussen de registerdataset van het CBS (het gaat om een combinatie van registers, hier kort het register genoemd) en Kadaster-model op basis van luchtfoto's:

- De foto's waarmee het Kadaster werkt, zijn in de winter/het voorjaar genomen, het register heeft betrekking op peildatum 31-12;
- Het duurt lang voordat de register-dataset definitief is (vanwege na-ijlers).

Daarnaast zijn er nog andere oorzaken gevonden gerelateerd aan fouten in het model en fouten in het register. In deze vervolganalyse ligt de focus op de andere oorzaken, en zijn de twee bovengenoemde oorzaken zoveel mogelijk uitgesloten. Dit is gedaan door:

¹ Krista Keller, CBS (kr.keller@cbs.nl) | Martin Tillema, Kadaster (Martin.Tillema@kadaster.nl)

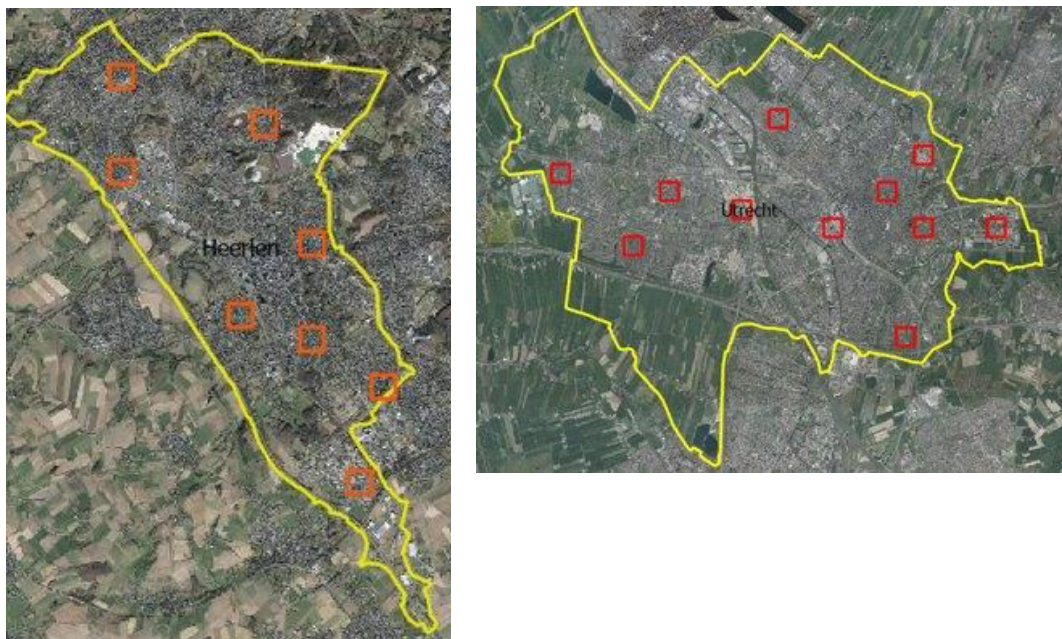
- de registersituatie af te leiden op het moment dat de foto genomen is, zodat panelen met een registratiedatum na het fotomoment in het desbetreffende verslagjaar uitgesloten worden;
- een oudere verslagperiode te kiezen (2020), omdat de kans op na-ijlers en resterende registratiefouten dan kleiner is.

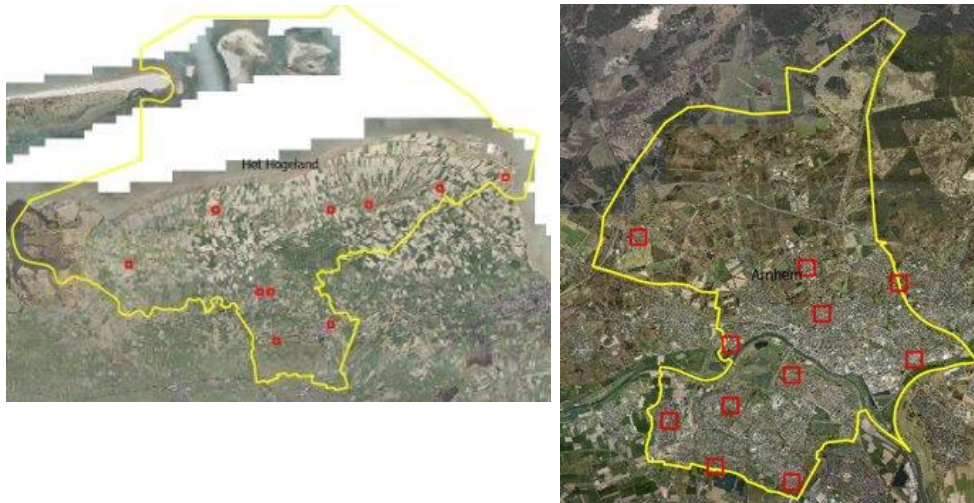
Net als in de eerste verkenning is de geografische scope van het onderzoek vier gemeenten: Utrecht, Heerlen, Het Hogeland en Arnhem. Het CBS heeft voor de deze gemeenten de aanwezigheid van zonnepanelen op x/y-coördinaten met het Kadaster gedeeld. Het Kadaster heeft deze gegevens in haar model gestopt en een verschilanalyse uitgevoerd. Om de data te kunnen en mogen delen ten behoeve van dit onderzoek, heeft het CBS haar eigen data op adresniveau omgezet naar een bestand met x/y-coördinaten en alle gevoelige informatie, zoals vermogen en bronkenmerk, eruit gehaald. Tijdens het omzetten van de data zijn er waarnemingspunten verloren gegaan, dit leidde tot een verlies van zo'n 5 procent van de records.

Het Kadaster heeft de eigen detecties en het resultaat van het matchen met de CBS-punten, incl. de fotodatum van de Kadaster detectie, BAG-id en perceel-id teruggeleverd aan het CBS. Hierop heeft het CBS statistische analyses uitgevoerd.

Het Kadaster heeft handmatige controles voor 10 procent van de zonnepanelen, verspreid over verschillende regio's binnen de vier gemeenten, uitgevoerd. Dit zijn de zogenaamde controlegebieden. Hierdoor kan er uitgerekend worden wat de kwaliteit is van de Kadaster dataset en van de CBS dataset. Daar waar er panden met panelen liggen en deze ook gevonden zijn middels luchtfoto's (Kadaster) of registers (CBS), worden deze gezien als de True Positives (TP). Panden met zonnepanelen die niet gevonden zijn, zijn False Negatives (FN). Panden met zonnepanelen die teveel gevonden zijn False Positives (FP), volgens de handmatige controle liggen daar geen panelen. De mismatches zijn gecategoriseerd naar type mismatch.

Figuur 1. Locatie van de controlegebieden (oranje/rode vierkantjes)





Voor meer uitleg over de zonnestroomdatasets van het CBS en Kadaster, zie het vorige rapport [Verkenning samenhang regionale zonnestroomcijfers \(2020\)](#).

3. Bevindingen en ontwikkelingen CBS en Kadaster

Deze paragraaf bevat de belangrijkste bevindingen die uit de analyses zijn gerold. Op basis van de bevindingen en de verwachte ontwikkelingen die bij het CBS en Kadaster bekend zijn, is een advies voor een vervolg opgesteld.

Bevindingen analyse

1. Waarnemingsverschil: pand (BAG) versus adres. Kadaster detecteert zonnepanelen op het niveau van BAG-panden. Het CBS houdt de locatie van zonnepanelen bij op het niveau van hoofdadres. Als er zonnepanelen op een bijgebouw liggen, bijvoorbeeld een schuur, dan worden die door het CBS niet apart waargenomen. Deze vallen dan onder het bijbehorende hoofdadres. Het Kadaster geeft deze panelen wel apart weer. Zonnepanelen die op objecten liggen die niet in de BAG staan, bijvoorbeeld woonboten, worden door het Kadaster niet meegenomen.
2. Voor- en nadelen dataset van het Kadaster:
 - Voordelen:
 - Data is locatie gebonden,
 - Geeft inzicht in waar er nog veel potentie is,
 - Het is duidelijk wat de actualiteit van de data is,
 - Mogelijkheden voor monitoring van installaties die verdwijnen.
 - Nadelen:
 - Wat niet op een luchtfoto te zien is wordt niet gedetecteerd: zonnepanelen op een gevel, geïntegreerde zonnepanelen in dakpannen etc.,
 - False positives doordat andere elementen worden aangezien voor een zonnepaneel,
 - Er is niet bekend van welke eigenaar de installatie is,
 - Het vermogen en opwek is niet bekend.

3. Voor- en nadelen dataset van het CBS:
- Voordelen:
 - Bevat meer informatie over de zonnepanelen, zoals het vermogen, gerelateerde aansluiting en subsidie, teruglevering en voor de grote installaties ook de productie,
 - Register bevat vaak ook de datum van ingebruikname.
 - Nadelen:
 - Na-ijlers in de registraties, daarom herziet het CBS jaarcijfers meerder malen (voorlopig, nader voorlopig en definitief),
 - Registerfouten,
 - Niet-geregistreerde zonnepanelen ontbreken,
 - Zonnepanelen die van het dak zijn gehaald, zullen niet ook altijd uit het register gehaald zijn.
4. Voor meer dan 70 procent komen de detecties van zonnepanelen door het CBS en Kadaster overeen. Dat is meer dan de 50 procent uit het vorige onderzoek. Dit komt vooral door de verbetering van de peildatum voor de koppeling en de verbetering van het Kadaster-model (Kausika et al., 2021). In tabel 1 is de verdeling van de detectie van zonnepanelen door beide organisaties per gemeente opgenomen. De controlebieden hebben een gelijksoortige verdeling als die van de gemeente waar ze in liggen. Daar waar de stedelijkheid hoog is lijkt er vanuit de registraties ‘te weinig’ waargenomen te worden. In minder stedelijk gebied lijkt er juist meer vanuit de registraties dan via de luchtfoto’s gedetecteerd te worden. Hier is geen eenduidige verklaring voor.

Tabel 1. Verdeling van de zonnepanelen gedetecteerd op basis van enkel het register (CBS), enkel luchtfoto’s (Kadaster) en door beiden

	CBS en Kadaster (%)	alleen CBS (%)	alleen Kadaster (%)
Arnhem	69,9	12,9	17,2
Heerlen	76,2	14,3	9,5
Het Hogeland	74,5	16,4	9,1
Utrecht	74,4	10,8	14,8

5. De volgende type mismatches komen geregeld terug in de controlegebieden:
- Geen eenduidige x/y-coördinaat koppeling met panden: meerdere x/y-coördinaten vallen binnen 1 pand-contour of liggen op de grens van twee panden, waardoor ze twee keer aan hetzelfde pand gekoppeld worden. Zie figuur 2. Deze mismatch is ontstaan doordat het CBS de gegevens op adresniveau moest omzetten naar een x/y-coördinaat. Deze omzetting gaat ten koste van de datakwaliteit.
 - Verschil in datum waarneming vanuit register en datum van de foto.
 - Pand komt niet voor in BAG.

Figuur 2. Voorbeeld van een mismatch: geen eenduidige x/y-coördinaat koppeling met panden.
 De rode stippen zijn de x/y-coördinaten van de locaties van de zonnepanelen en de geel omlinjende panden zijn de panden waar zonnepanelen tijdens de controle zijn gevonden.



6. Het CBS lijkt relatief meer panelen te missen dan het Kadaster (False Negatives), maar neemt over het algemeen minder panelen ontorecht waar (False Positives). Zie tabel 2. Het kan echter mogelijk zijn dat dit grotendeels de panelen zijn die tijdens het omzetten van de data en het koppelen verloren zijn gegaan, dat het panelen zijn die op een later moment alsnog waargenomen worden (na-ijlers) en/of dat de panelen op een ander adres zijn geregistreerd. In dat laatste geval zitten ze wel in het totale cijfers voor Nederland, maar kan er ruis op lokaal niveau zijn.

Tabel 2. False positives en False negatives in controlegebied op pandniveau¹

	False positives (%)	False negatives (%)
Kadaster model	8,4	10,6
CBS dataset	3,7	14,4

¹ Bij het CBS zit hier mogelijk een overschatting, omdat het CBS op hoofdadres waarneemt en dus per definitie niet alle panden waarneemt. Hier is niet voor gecorrigeerd.

7. Zonnepanelen niet-woningen worden relatief vaker alleen door het Kadaster waargenomen, zie tabel 3. Het zou kunnen dat panelen door woningeigenaren beter geregistreerd worden in de geraadpleegde registers (voornamelijk CERES) dan door bedrijven. Of ze daadwerkelijk minder geregistreerd zijn is niet te zeggen, het kan ook zijn dat het adres waarop ze geregistreerd zijn minder goed de daadwerkelijke locatie reflecteert.

Tabel 3. Waargenomen zonnepanelen door CBS en het Kadaster, uitgesplitst naar gebruiksfunctie.

	CBS en Kadaster (%)	alleen CBS (%)	alleen Kadaster (%)
woonfunctie	76,8	12,8	10,5
andere functie	50,0	11,6	38,3
onbekend	19,1	12,4	68,4

8. Zonnepanelen op nieuwbouw na 2018 en zonnepanelen op panden met een oppervlakte groter dan 700 m² worden relatief vaker door alleen het Kadaster gezien. Het kan zijn dat

hier een ruis op de registraties zit: panelen op nieuwbouwpanden zijn mogelijk later zichtbaar in de registraties (na-ijlers) en een grotere kans dat installaties op grote panden op een ander adres worden geregistreerd. Maar het kan ook zo zijn dat hier juist veel ruis op de detectie vanuit het kadaster zit en dat hier veel False Positives voorkomen. Handmatige checks zijn vooralsnog nodig om hier meer grip op te krijgen.

Uitdagingen bij de uitgevoerde analyses

- Data en mogelijk ook matches verloren bij het omzetten van beide bestanden in x/y-coördinaten om voor deze verkenning datadelen mogelijk te maken. Op adresniveau (al dan niet versleuteld) koppelen zal een zuiverdere vergelijking opleveren, in plaats van eerst omzetten naar een puntenbestand. Nu worden er fouten geïntroduceerd door het omzetten van het bestand.
- Soms heeft het Kadaster één installatie op een gebouw gedetecteerd, terwijl het CBS meerdere losse puntjes heeft geregistreerd. Dit kunnen panelen zijn die op verschillende momenten zijn geregistreerd of op basis van de omzetting naar x/y-coördinaten net niet aan het juiste gebouw zijn gekoppeld.
- Soms ligt een installatie niet op het adres waarop deze geregistreerd is, maar op een pand ernaast of in de buurt (bijv. omdat het van dezelfde eigenaar is). Dit wordt dan als mismatch gezien. Voor statistieken op buurt of wijkniveau vallen dergelijke mismatches weg, mits de panelen in dezelfde buurt of wijk vallen.
- Ook de handmatige controles van de mismatches in de controlegebieden reflecteren niet per definitie de waarheid. Het blijft handwerk op basis van checks achter het bureau. Zo kan op basis van *deskresearch* lijken dat er ergens geen zonnepaneel aanwezig is, maar dat deze wel toch aanwezig is. Zo zijn verticale opstellingen tegen de muur niet op plaatjes van bovenaf af te leiden. Dit leidt dan in de analyses tot een onterechte False Positive in de CBS-dataset.

Geplande ontwikkelingen zonnepanelendatasets en overige mogelijke verbeteringen

Beide organisaties hebben reeds ontwikkelingen gepland staan met betrekking tot de zonnepanelendataset. Deze zijn hieronder opgenomen, samen met de voorgestelde ontwikkelingen naar aanleiding van dit onderzoek.

Kadaster, gepland:

- Inzetten van true ortho's vanaf 2021 – test gepland voor oktober 2022;
- Twee huidige deep learning modellen onderzoeken;
- Detectie van het Kadaster verbeteren en stabiliseren is wenselijk om deze in te kunnen zetten voor de statistieken en aanvullende monitoring.

Kadaster, overige mogelijke verbeteringen:

- Ground Truth uitbreiden voor een betere training;
- Time lapsing als postprocessing inzetten;
- Het kadastermodel verder ontwikkelen om oriëntatie en hellingshoek van de installaties te schatten. Deze informatie zou een mooie aanvulling zijn op de statistiek;
- Op het moment dat het Kadaster toegang krijgt tot de volledige dataset van het CBS, heeft deze dataset veel toegevoegde waarde om de resultaten van het Kadaster te verbeteren. Doormiddel van de volledige dataset van het CBS en de toestemming om deze ook buiten het zonnestroom project te gebruiken kan het Kadaster toetsen waar de zwakte van het model zit, zijn dat bepaalde gebieden of bepaalde type woningen? In hoeverre klopt de

oppervlakte die het Kadaster detecteert, deze oppervlakten worden ook omgerekend naar mogelijke opwekcapaciteit, in hoeverre komen deze cijfers overeen? De CBS-registratie zou het mogelijk maken om het Kadaster model sneller te stabiliseren, doordat kwaliteitscontroles vele malen sneller gedaan kunnen worden.

De dataset van het CBS in combinatie met informatie uit het detectiemodel geeft waardevolle en sturende informatie voor de regionale aanpak van de energietransitie.

CBS, gepland:

- CBS heeft recent betere gegevens over terugleveringen op aansluitingsniveau ontvangen. Deze data kan goed in combinatie met de andere registerbronnen gebruikt worden om de aanwezigheid van zonnepanelen af te leiden;
- Volledige herziening van de productiestraat voor het microbestand voor zonnepaneleninstallaties, inclusief het opnieuw bezien van alle regels voor ontdebelen etc. Het kan zijn dat er onterecht panelen uit het analysebestand worden weggehaald, dus dat het ontdebelenproces te streng is.

CBS, overige mogelijke verbeteringen:

- Grote installaties die afgeleid zijn uit de btw-opgave strenger nalopen of wellicht helemaal niet meenemen, omdat deze tijdens de controle relatief vaak een False Positive gaven;
- Het oppervlak dat het Kadaster schat, kan gebruikt worden voor een plausibiliteitscontrole van het vermogen dat CBS op adresniveau uit de registers haalt;
- De verdeling zon op veld en dak is nu gebaseerd op informatie uit SDE en enkele grootvermogens worden handmatig nagelopen. De dataset van het Kadaster, met aanvullend onderscheid veld/dak, kan ter aanvulling op de SDE-informatie gebruikt worden om de verdeling completer te maken;
- De Kadasterdataset kan als aanvullende bron gebruikt worden in het productieproces om mogelijk enkele ontbrekende panelen uit de registers aan te vullen. Hier ligt wel een risico dat er ten onrechte panelen worden toegevoegd. Hier kan een nader uit te werken beslisboom bij helpen;
- De Kadasterdataset kan op termijn interessant zijn om ook weggehaalde / uit gebruik genomen panelen beter te kunnen monitoren. Dit is op basis van registers niet altijd te achterhalen.

4. Conclusie en hoe nu verder?

Op basis van de bevindingen kan het volgende geconcludeerd en geadviseerd worden:

- De Kadasterdataset kan een mooie aanvullende databron zijn in het productieproces voor het samenstellen van de regionale zonnestroomstatistiek.
De CBS-dataset vormt een betrouwbare basis voor de zonnestroomstatistiek. Er kan niet eenduidig en met zekerheid aangegeven worden of zonnepanelen die het Kadaster wel ziet en CBS niet terecht toegevoegd kunnen worden aan de dataset van het CBS. Het onderzoek maakte duidelijk dat het onmogelijk is deze puzzel (geautomatiseerd) sluitend te krijgen. Waar liggen dan wel de kansen?
 - De dataset kan om te beginnen als ‘trigger’ ingezet worden voor de extra handmatige controles in het statistiekproces, onder andere m.b.t. de locatie van grootschalige zonnepanelen. Tevens biedt de Kadasterdataset de potentie om

extra informatie te geven over veld- en dakopstellingen en de uit gebruik genomen installaties.

- Op termijn kan er aan een goede beslisboom gewerkt worden om zonnepanelen (geautomatiseerd) toe te voegen of te verwijderen. De eerste stap is om te achterhalen waar de *False Negatives* bij het CBS door veroorzaakt worden. Komt dit deels door de omzetting naar x/y-coördinaten, de waarneming op hoofdadres i.p.v. pandniveau, te streng ontdebellen, registratiefouten met betrekking tot het geregistreerde adres, etc.? Als hier meer duidelijkheid over is, kan er aan een beslisboom gewerkt worden om naar een meer geautomatiseerd proces toe te kunnen werken. De kadasterdataset geeft geen informatie over het vermogen, productie of de ingebruikname datum. Hoe dergelijke informatie het best afgeleid kan worden dient dan ook nader uitgewerkt worden.
- De detectie van zonnepanelen op pandniveau door het Kadaster kan in combinatie met een zonnepotentie analyse goed gebruikt worden om (laag regionale) overheden inzicht te geven in hoeveel zonnepotentie er nog aanwezig is op daken². Deze informatie helpt de overheden om gerichte maatregelen te kunnen nemen met als doel de energietransitie te versnellen. Dit is een belangrijke maatschappelijke opgave, waarin de samenwerking tussen de dataleverende en –verwerkende partijen noodzakelijk is om de nodige betrouwbare informatie efficiënt en doelgericht op tafel te krijgen. Het model van het Kadaster kan nog wel verder getraind en verbeterd worden. Vooral de CBS-dataset is hier een belangrijke databron voor. Het delen van microdata kent al snel juridische grenzen, maar er zijn ook andere moderne data-uitwisselingsvormen denkbaar die hiervoor verkend kunnen worden. Bijvoorbeeld via *Privacy Preserving Techniques (PPT)*, hierbij blijft de data bij de bronhouder en kunnen analyses op afstand plaatsvinden.

Het CBS en Kadaster gaan in het derde kwartaal van 2022 met elkaar in gesprek over hoe een voortzetting in de samenwerking op dit thema van voldoende toegevoegde waarde is en hoe deze gerealiseerd kan worden. Denk hierbij aan waar de dataset van het Kadaster aan moet voldoen om het adequaat in te kunnen zetten voor het statistiekproces en als dat realistisch is, wat ervoor nodig is om een structurele levering mogelijk te maken. Tevens blijven beide partijen met elkaar in gesprek over de inzet van nieuwe vormen van data-uitwisseling, waarbij de privacy gewaarborgd blijft en met het gezamenlijk doel de informatievoorziening blijven verbeteren. In dit geval de aanwezigheid van zonnepanelen op pandniveau.

² Zodra deze informatie op pandniveau meegenomen wordt in landelijk monitoring van de zonnepotentie analyses, dan is het in het kader van het streven naar een eenduidige informatievoorziening voor de energietransitie van belang dat er een koppeling gemaakt wordt met het gerealiseerd vermogen en productie van zonnestroom volgens de statistiek.

Afkortingen

BAG	Basisregistratie Adressen en Gebouwen
BTW	belasting over de toegevoegde waarde (Omzetbelasting)
CBS	Centraal Bureau voor de Statistiek
FN	False Negatives
FP	False Positives
PPT	Privacy Preserving Techniques
RVO	Rijksdienst voor Ondernemend Nederland
SDE	Stimuleringsregeling Duurzame Energieproductie
TP	True Positives
VIVET	Verbetering van de Informatievoorziening voor de Energietransitie