

Gemodelleerde energiedata

Gespreksnotitie n.a.v. overleg tussen Steven van Polen (PBL) en Jurrien Vroom (CBS) 3 december 2020

PBL dataset voor de TVW

PBL maakt nu al een dataset met gemodelleerde data die op microniveau beschikbaar is gesteld bij de publicatie van de Leidraad voor aardgasvrije buurten (september 2020). Input voor deze dataset zijn onder meer gedetailleerde kentallen van het energieverbruik van woningen die door het CBS worden gepubliceerd in combinatie met open datasets zoals de BAG. Deze dataset wordt al ruimschoots ingezet en breed gedragen en het is dan ook verstandig om deze set als uitgangspunt te nemen en van daaruit te gaan verbeteren.

Kentallen CBS

De CBS kentallensets die nu door PBL worden gebruikt zijn voor aardgaslevering aan woningen uitsplitsingen naar woningtype, bouwjaarklasse, oppervlakteklasse en labelklasse:

<http://opendata.cbs.nl/statline/#/CBS/nl/dataset/83878NED/table?dl=47282>

Voor elektriciteitslevering aan woningen betreft het uitsplitsingen naar woningtype, bouwjaarklasse en bewonersklasse:

<http://opendata.cbs.nl/statline/#/CBS/nl/dataset/83882NED/table?dl=3FD72>

De sets bevatten gemiddelden en 5 percentielklassen. Daarnaast houdt het PBL rekening met het aandeel stadsverwarming op buurtniveau (afkomstig van CBS) bij de analyse van aardgasvrije vormen van warmtevoorziening met het Vesta MAIS model.

Overige invloedsfactoren/kenmerken

De volgende zaken zijn besproken die in de huidige methodiek niet of onvoldoende aan de orde komen, en die nu (al) een significant effect op het (individueel) huishoudelijk verbruik hebben, en waarvan de invloed van de eerste drie in de toekomst sterk zal groeien. Zijn deze zaken mee te nemen bij het verkennen van het pad naar gemodelleerde microdata?

- Noodzakelijke benodigde informatie voor micro-data binnen het CBS
 - o All-electric woningen (warmtepompen). Hoe krijgen we deze in een dergelijke gemodelleerde open dataset? CBS is bezig met de opzet van een monitor voor de primaire warmtevoorziening van woningen, maar deze is met name wat betreft warmtepompen nog verre van volledig en het betreft hier op microniveau geen openbare data.
 - o Zon PV: hoe zit het met de lokale opwek? Wat is de opbrengst van een eigen ZonPV installatie, hoe groot is de levering via het openbaar net, hoe groot de teruglevering? Ook hier is sprake van een (stapsgewijs) voortschrijdend inzicht bij het CBS op microniveau als per verslagjaar 2020 het standaardjaarverbruik wordt vervangen door een waarde voor standaard afname en standaard invoeding. Maar ook deze data is/komt op microniveau niet openbaar beschikbaar.
 - o Elektrisch vervoer. Is er sprake van een aansluiting bij een woning voor EV, en zo ja, wat is het laadgedrag? De eerste is al onbekend, de tweede al helemaal. Dit pleit voor ten minste een register zoals CERES waarin een aansluiting voor het opladen van een elektrische auto op een (woon-) adres wordt geregistreerd.
- Mogelijke verbeteringen van de huidige openbare kentallen

- De huidige kentallen zijn nationale totalen die geen rekening houden met stedelijkheid. En dat terwijl de woningpopulatie (en huishoudenspopulatie) er in het landelijk gebied heel anders uitziet dan in het stedelijk gebied. Onderzocht kan worden in hoeverre kentallen als functie van stedelijkheid fluctueren.
- Daarnaast is er geen informatie beschikbaar over het isolatieniveau van individuele woningen. Er zijn voor gas weliswaar uitsplitsingen naar labelcategorie, maar deze reflecteert maar deels het totale werkelijke isolatieniveau (met name in geval van een vereenvoudigd energielabel). In elk geval lijkt er meerwaarde te zitten in het maken van een (aanvullende) uitgesplitste kentallenset naar energielabel waarbij gesegmenteerd wordt naar de origine van het energielabel (vereenvoudigd/uitgebreid).
- Openbare informatie over huishoudenssamenstelling op microniveau ontbreekt. Daarmee is de vraag of het toch zinvol zou zijn om kentallensets te maken met uitsplitsingen naar een aantal voor het energieverbruik belangrijke huishoudkenmerken. (elektriciteit wordt trouwens al wel uitgesplitst naar aantal bewoners.)
- Koudevraag, zou met name in de toekomst met meer hete zomers in het verschiet significant kunnen worden. Hier is momenteel geen zicht op, ook niet op een databron waaruit dit zou kunnen blijken.

Voor een deel van de gegevens is het eigenlijk onontkoombaar dat er op termijn (bij voorkeur open raadpleegbare) registers komen met relevante data zoals bijvoorbeeld een installatieregister van Techniek Nederland, een volledig register met warmtepompen, zonPV, etc.

Woningen/utiliteit

Een deel van de woningen kan niet los gezien worden van utiliteit omdat er overlap is in functie/energiegebruik. Het betreft hier bijvoorbeeld zorgwoningen (aanleunwoningen), maar ook agrarische bedrijven/boerderijen en in geval van verwevenheid van woningen en winkels of andere bedrijfsactiviteiten aan huis. Voor dergelijke mixsituaties is het al lastig om een goede uitsplitsing te maken vanuit microdata zoals op het CBS aanwezig, maar andersom is dit voor gemodelleerde data een probleempunt om vanuit kentallen naar een modelmatige inschatting te werken.

Kentallen voor utiliteit zijn overigens sowieso maar beperkt beschikbaar, voor een deel van de utiliteit in de dienstensectoren zijn kentallen gepubliceerd. De bandbreedte voor utiliteit is echter veel groter dan voor woningen, dus de vertaling naar individuele gevallen (veel) minder accuraat.

<http://opendata.cbs.nl/statline/#/CBS/nl/dataset/83376NED/table?dl=1DF04>

Gemodelleerde microdata voor de utiliteitssectoren zal daarmee naar verwachting waarschijnlijk onvoldoende betrouwbaar te construeren zijn.

Een maat voor de kwaliteit van gemodelleerde microdata is de afwijking t.o.v. de werkelijke microdata zoals op het CBS beschikbaar. In een pilot zouden deze afwijkingen op laag schaalniveau zowel voor woningen als voor utiliteit nader gekwantificeerd kunnen worden.

Een eerste stap is om de gehanteerde methoden door het CBS en PBL op elkaar af te stemmen op dit gebied.

Regionaal schalen van gemodelleerde data met werkelijke leveringscijfers aardgas/elektriciteit.

Zoals hiervoor bleek zijn er tal van invloedsfactoren die een rol spelen in het uiteindelijk verbruik en die je niet meeneemt met standaard kentallen in de huidige aanpak. Maar wellicht is het effect van

een deel van deze verschillen te minimaliseren door gebruik te maken van regionale schaling van de data. CBS beschikt namelijk over totalen levering per regio per categorie (utiliteit of woningen) en kan een regionale schalingsfactor meeleveren bij kentallensets. Een dergelijke schalingsfactor zorgt ervoor dat een lokaal totaal in de gemodelleerde data optelt tot het juiste cijfer. Het effect op de individuele afwijkingen van werkelijke microdata versus gemodelleerde zal echter in de praktijk moeten worden onderzocht om te kunnen concluderen of dit een vruchtbare aanpak zou kunnen zijn. Ook het aggregatieniveau waarop schaling plaatsvindt (gemeente/buurt/postcode6niveau?) is daarbij onderwerp van onderzoek. Hier ligt ook weer een relatie met privacy: te succesvolle schaling kan weer tot privacy issues leiden.

Eerste stappen naar verbetering

Wat zou in eerste instantie een aanpak kunnen zijn op weg naar verbetering van de huidige PBL aanpak? Het lijkt er op dat de op korte termijn meest kansrijke opties liggen op het terrein van een verdere detaillering van CBS kentallensets. De volgende opties zijn besproken, waarbij de korte termijn haalbaarheid c.q. prioritering van dit “wensenlijstje” in een vervolgoverleg kan worden bepaald.

- De inschatting van de huidige kentallen kan worden verbeterd door de lineaire regressie direct toe te passen op de CBS-data in plaats van de gemiddelde verbruiken per oppervlakteklasse.
Het PBL heeft nu een dataset waarin het verbruik van een woning afhangt van woningtype, bouwjaarklasse, energielabel en het oppervlakte van de woning. Hierbij is de relatie met het oppervlak vastgesteld op basis van een lineaire regressie tussen gemiddeld gasverbruik per oppervlakteklasse en het gemiddeld oppervlak per oppervlakteklasse. Daarbij heeft nu elke oppervlakteklasse evenveel invloed, terwijl het beter zou zijn om de lineaire regressie direct uit te laten voeren op basis van de micro-data van het CBS.
- Afstemming over de methode voor de toedeling van gebouwen aan een woon- of utiliteitsfunctie tussen CBS & PBL
- Bij de uitsplitsing van gasverbruik naar energielabel kan nog een gradatie worden aangebracht in de aard van het energielabel (vereenvoudigd of uitgebreid).
- De aanwezigheid van zonPV meenemen in de uitsplitsing van kentallen voor het gemiddeld gasverbruik van woningen. Bruikbaarheid voor het construeren van betrouwbare gemodelleerde data hangt dan nog op de beschikbaarheid van brongegevens over de aanwezigheid van zonPV.
- Kentallensets verder uitsplitsen naar huishoudkenmerken. Zoals benoemd vereist dit in de volgende stap ook weer micro inzicht in deze kenmerken. En ook hier is de vraag in hoeverre dat een verbetering van een samen te stellen gemodelleerde set zou vormen.
- Onderzoeken van de effecten van een regionale schaling van kentallen op de betrouwbaarheid van de resulterende modeldata.

Daarnaast moet toch ook een lange termijn spoor bewandeld worden die op termijn voorziet in (openbare) beschikbaarheid van microdata m.b.t. zonPV, warmtepompen, EV, etc.

Verstandig is een gefaseerde aanpak die beide routes (korte en lange termijn) in pilotvorm nader onderzoekt. Het lijkt dan ook wenselijk om daartoe vanuit Vivet naar concrete pilotprojectvoorstellen toe te werken.