

Analyse naar de haalbaarheid van statistische modellen die energiegebruik in woningen kunnen voorspellen op basis van bouwparameters

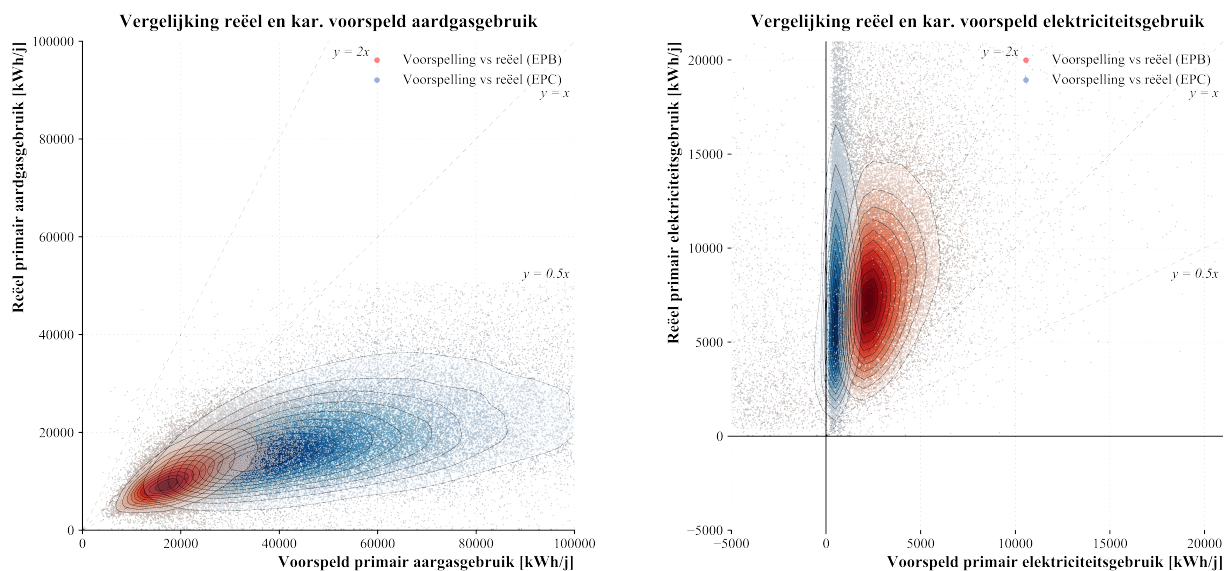
Matthias Van Hove, Marc Delghust, Arnold Janssens, Onderzoeksgroep Bouwfysica, Universiteit Gent, Juni 2021.

In deze studie rond het reëel jaarlijks energiegebruik in woningen is (i) de relatie tussen het reëel energiegebruik en het karakteristiek voorspeld energiegebruik, bouwparameters en socio-demografische parameters beschreven en (ii) onderzocht welke soort statistische voorspellingsmodellen (i.e., lineaire regressie en neurale netwerken) mogelijk en zinvol zijn en wat de te verwachten betrouwbaarheid ervan is.

De studie is gebaseerd op een zeer grote, geanonimiseerde set van Vlaamse eengezinswoningen (geen appartementen). Ze omvat 135166 EPB-wooneenheden (i.e., nieuwbouw vanaf 2006) en 122680 EPC-wooneenheden (i.e., bestaande woningbouw van voor 2006, mét EPC-opmaak tussen 2015 en 2018). De belangrijkste bouwparameters zijn gekend via hetzij de EPB-aangifte hetzij via het EPC-certificaat. Na data-cleansing, -filtering en -koppeling met de jaarlijkse gas- en elektriciteitsverbruiken van Fluvius blijven 50% van de nieuwbouwwoningen en 57% van de EPC-woningen over. Na een aanvullende koppeling met domiciliegegevens van het Rijksregister (i.e., aantal bewoners en hun geboortjaar) blijven resp. 41% en 25% van de woningen over. Ondanks de sterke datareductie blijft een betekenisvolle sample over van zeer grote omvang. De belangrijkste redenen waardoor cases wegvielen of geweerd dienden te worden uit de finale subsets zijn (i) de data-koppeling (i.e., ontbreken of niet terugvinden van wooneenheden in bepaalde datasets), (ii) tegenstrijdige data over de aanwezigheid van een PV-installatie, (iii) een gebrek aan meerdere reële jaarlijkse betrouwbare gas- en/of elektriciteitsverbruiken en (iv) het hebben van een andere energiedrager dan aardgas of elektriciteit voor ruimteverwarming en/of sanitair warm water (waarvoor geen reële energiegebruiksdata beschikbaar is voor analyses en modellering).

Relatie reëel en karakteristiek voorspeld energiegebruik

In eerste instantie is geanalyseerd hoe het karakteristiek voorspeld energiegebruik uit de EPB- en EPC-berekeningen zich verhoudt tot het reëel energiegebruik volgens de data van Fluvius, telkens uitgedrukt in primaire energie (voor elektriciteit wordt een primaire energiefactor 2,5 gebruikt).



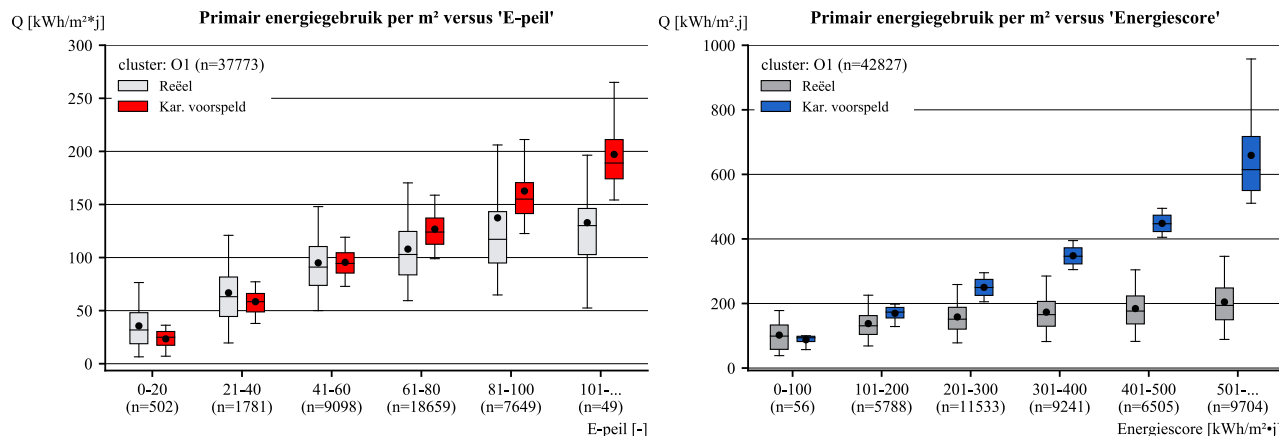
Figuur 1 - Spreidingsdiagram van het jaarlijks karakteristiek voorspeld en reëel primair aardgas- en elektriciteitsverbruik voor de EPB- en EPC-wooneenheden.

De sample voor het reëel aardgasverbruik bevat woningen waar enkel aardgas wordt gebruikt voor ruimteverwarming en sanitair warm water. De EPB- en EPC-berekeningsmethoden overschatten het reëel aardgasverbruik met gemiddeld 81% en 190% respectievelijk (Figuur 1, links). De sample voor het reëel elektriciteitsverbruik bevat woningen waar geen elektriciteit wordt gebruikt voor ruimteverwarming en sanitair warm water (alle nulverbruiken zijn geweerd uit de sample, gezien er geen informatie beschikbaar is over de hoeveelheid overproductie). Het reëel jaarlijks elektriciteitsverbruik is voor die sample véél hoger dan het karakteristiek voorspeld elektriciteitsverbruik (Figuur 1, rechts). Dit is omdat enkel het elektriciteitsverbruik voor hulpenergie (i.e., ventilatoren en pompen), ruimtekoeling en PV-opbrengst in rekening wordt gebracht in de EPB- en EPC-berekeningen van die woningen, terwijl het werkelijke

elektriciteitsverbruik ook het elektriciteitsverbruik van huishoudtoestellen en verlichting bevat. De EPB-/EPC-berekeningsmethoden overschatten het totaal reëel primair energiegebruik (*i.e.*, aardgas en elektriciteit samen) met gemiddeld 13% en 118% respectievelijk.

Relatie tussen E-peil/EPC-kengetal en reëel en karakteristiek voorspeld energiegebruik

In het geval van de EPC-wooneenheden wordt opgemerkt dat het reëel primair energiegebruik volledig afvlakt vanaf label C (201-300 kWh/(m²•j)) naar slechtere EPC-niveaus (*Figuur 2, rechts*). Bij de EPB-wooneenheden valt op dat over de E-peilen heen het werkelijk energiegebruik veel minder toeneemt dan het karakteristiek voorspeld energiegebruik (*Figuur 2, links*). Bij de beste niveaus van de EPB en EPC (*i.e.*, E<40 en EPC < 100 kWh/(m²•j)) komt het reëel energiegebruik gemiddeld hoger te liggen dan het karakteristiek voorspeld energiegebruik. Dit zijn resultaten die overeenkomen met bevindingen uit onze buurlanden (*e.g.*, Majcen *et al.*, 2013; Sunikka-Blank & Galvin, 2012).



Figuur 2 - Boxplot van het totaal jaarlijks karakteristiek voorspeld en reëel primair energiegebruik (*i.e.*, aardgas en elektriciteit samen) in functie van het E-peil (*links*) en het EPC-kengetal (*rechts*).

Statistische voorspellingsmodellen

De predictie-resultaten uit meerdere lineaire regressiemodellen geven een veel betere overeenkomst met het reëel energiegebruik dan de theoretische EPB-/EPC-berekening. Regressiemodellen die zich enkel baseren op bouwparameters, kunnen tot ~51% en ~38% van de variatie verklaren in het totaal reëel energiegebruik van respectievelijk EPB- en EPC-woningen. Bevatten deze modellen ook socio-demografische parameters zoals het aantal bewoners en hun leeftijd, dan kan bijkomend ~6% en ~5% van de variatie extra verklaard worden. Dit is vooral omdat de socio-demografische gegevens de voorspelling van het elektriciteitsverbruik sterk verbeteren. De verklarende kracht van de regressiemodellen is relatief goed, zeker vergeleken met soortgelijke andere studies, maar de onzekerheid op een individuele voorspelling blijft groot. De onverklaarde variatie is toe te schrijven aan (*i*) parameters die niet in de aangeleverde datasets aanwezig zijn, (*ii*) parameterwaarden binnen de aangeleverde datasets die onvoldoende accuraat zijn (*i.e.*, door het gebruik van default-waarden en/of fouten in de EPB-/EPC-verslaggeving) of simpelweg (*iii*) het feit dat een lineair regressiemodel niet voldoende geschikt is om meer variatie in het reëel energiegebruik te verklaren. Het maximale voorspellingsvermogen van neurale netwerk modellen blijkt gelijkaardig aan het voorspellingsvermogen voor de beste lineaire regressiemodellen (bij overeenkomstige verklarende parameters). Bijgevolg lijken de kunstmatige neurale netwerken, gezien hun complexiteit weinig meerwaarde te bieden ten opzichte van de begrijpbare, relatief eenvoudige lineaire regressiemodellen.

Conclusie

Het verschil in energiegebruik tussen woningen met hogere versus lagere prestatieniveaus (en dus ook de potentiële energiebesparing) ligt in werkelijkheid veel lager dan de theoretische EPB- en EPC-berekeningen suggereren. Voor bestaande woningen met lagere prestatieniveaus is deze afwijking tussen theoretische en werkelijke waarden het grootst. Op individuele gebouw niveau is duidelijk dat de lineaire regressiemodellen te zwak presteren voor enige gevolgtrekking. Op gebouw patrimonium niveau daarentegen kunnen de lineaire regressiemodellen wel een bruikbare tool zijn om patrimonium eigenaars (*i.e.*, financiële instellingen zoals banken, overheden, huisvestingsmaatschappijen) te informeren en/of het opstellen van beleid te ondersteunen.