

Wijklabels

Balázs Dukai, Ravi Peters

2024-02-12

Table of contents

| | | |
|----------|---|-----------|
| 1 | Samenvatting | 2 |
| 2 | Gegevensbronnen | 2 |
| 2.1 | Spreiding van energielabels | 2 |
| 2.2 | BAG en 3DBAG | 3 |
| 2.3 | Bouwjaar | 4 |
| 2.4 | CBS buurten | 5 |
| 2.5 | EP-Online voor validatie | 5 |
| 3 | Methode | 7 |
| 3.1 | Aantal verdiepingen | 8 |
| 3.2 | Woningtypes | 8 |
| 3.2.1 | Classificatie van een Pand | 9 |
| 3.2.2 | Eengezinswoningen | 10 |
| 3.2.3 | Meergezinswoningen | 10 |
| 3.3 | Conversie van appartementtypes naar pre-NTA8800 types | 10 |
| 3.4 | Vormfactor | 12 |
| 3.5 | Het schatten van de energielabels | 12 |
| 4 | Validatie van de energielabelschatting | 13 |
| 5 | Resultaten | 13 |
| 5.1 | Woningtypes | 13 |
| 5.2 | Schatting van energielabels | 14 |
| 5.2.1 | Alleen rekening houdend met mogelijke labels | 15 |
| 6 | Conclusies | 16 |
| 7 | References | 16 |

```
import pandas as pd
import numpy as np

from wijklabels.woningtype import Bouwperiode
```

1 Samenvatting

De studie Voorbeeldwoningen 2022 beschrijft de spreiding van energielabels per vormfactorbereik voor elk woningtype uit de WoON 2018 studie. De laatste release van de 3DBAG dataset (2023.10.08) levert de oppervlakten voor het berekenen van de vormfactor voor elk Pand. Dit werk onderzoekt de mogelijkheid om de vormfactor voor elke woning binnen een Pand te berekenen en de energielabelverdelingen van de Voorbeeldwoningen 2022 studie toe te passen om de energielabelverdeling van elke buurt van Nederland te schatten.

We valideren onze geschatte energielabels ten opzichte van de labels die zijn geregistreerd in de EP-Online database. De validatie toont een variërende afwijking tussen de geschatte en EP-Online labels. De labels *A* en *B* hebben 0 mediane afwijking, de labels *A+*, *A++* zijn één label overschat, de labels *C*, *D* zijn één label onderschat. De rest van de labels heeft een afwijking van twee labels of meer van de EP-Online labels.

Er zijn twee belangrijke beperkingen die de nauwkeurigheid van de resultaten beïnvloeden. Ten eerste is de 3D locatie van appartementen niet bekend, daarom worden er sterke aannames gedaan over de indeling van een Pand, wat leidt tot een voornamelijk onnauwkeurige schatting van de appartementssubtypes. Deze onnauwkeurigheid werkt door in de schatting van de vormfactor en vervolgens in het energielabel van appartementen. Ten tweede zijn de spreidingen van de energielabels in de studie Voorbeeldwoningen 2022 onvolledig, waardoor 14% van de ingevoerde woningen geen energielabel kon krijgen en 32% van de gevalideerde woningen een energielabel heeft dat niet kan worden toegekend vanwege ontbrekende gegevens in de studie Voorbeeldwoningen 2022.

2 Gegevensbronnen

2.1 Spreiding van energielabels

De studie RVO (2022) geeft een representatieve steekproef van spreidingen van energielabels per woningtype en bouwperiode. De spreiding is tweedimensionaal, één dimensie is het energielabel, de tweede dimensie is de vormfactor van de woning. De beschreven spreidingen zijn echter niet continu. In het geval van Flatwoningen uit de periode 1965-1974 zijn er bijvoorbeeld geen gegevens in de vormfactor range 1,00-1,50 en labels A+++-D, zie Figure 1.

Dergelijke hiaten hebben een aanzienlijke invloed op de kwaliteit van onze resultaten (zie Section 5.2.1).

Als de woning een vormfactor heeft in een bereik waar geen energielabelgegevens in de spreiding zijn, krijgt de woning geen energielabel toegewezen.

De energielabelverdelingen zijn gehaald uit het Excelbestand `Illustraties spreiding Energielabel in WoON2018 per Voorbeeldwoning 2022 - 2023 01 25.xlsx`, dat we van RVO hebben ontvangen.

| Flatwoning (overig) 1965-1974 | | | | | | | 30 woningen in WoON2018 | | | | | |
|-------------------------------|-------|------|-----|----|------|------|-------------------------|-------|-------|-------|-------|--------|
| vormfactor | A++++ | A+++ | A++ | A+ | A | B | C | D | E | F | G | TOTAAL |
| 0,50 | | | | | 4,2% | | 0,5% | | 1,3% | 6,7% | 16,0% | 28,7% |
| 0,50 1,00 | | | | | 3,1% | 5,2% | 9,6% | 10,5% | | 1,8% | | 30,2% |
| 1,00 1,50 | | | | | | | | | 9,9% | 4,1% | | 14,0% |
| 1,50 2,00 | | | | | | | 3,8% | 0,8% | 3,5% | 14,1% | 2,8% | 24,9% |
| 2,00 2,50 | | | | | | | | | | | 2,2% | 2,2% |
| 2,50 3,00 | | | | | | | | | | | | |
| 3,00 3,50 | | | | | | | | | | | | |
| 3,50 | | | | | | | | | | | | |
| | | | | | 7,4% | 5,2% | 13,8% | 11,3% | 14,6% | 26,7% | 21,0% | |

Figure 1: Energielabel spreiding flatwoningen, RVO (2022)

2.2 BAG en 3DBAG

De **BAG** dataset is het officiële registratie van adressen en gebouwen in Nederland. De BAG bevat twee objecttypen die relevant zijn voor dit onderzoek, *Pand* en *Verblijfsobject (VBO)*. Elk VBO bevat informatie over het gebruik ervan in het **gebruiksdoel** attribuut. Voor deze studie worden alleen die VBO's gebruikt waarvan het gebruik **woonfunctie** bevat. De [3DBAG] (<https://3dbag.nl>) is een uitbreiding van de BAG dataset en bevat de buitenste 3D schil van de Pand objecten. De 3DBAG bevat geen informatie over de VBO's. We hebben de 3DBAG versie 2023.10.08 gebruikt. De 3DBAG 2023.10.08 is gebaseerd op het BAG 2.0 extract met een releasedatum van *08.09.2023*.

Het BAG extract van 08.09.2023 bevat *10.877.129* Pand objecten en *10.144.833* Verblijfsobjecten. De 3DBAG versie 2023.10.08 die gebaseerd is op het genoemde BAG extract, bevat *10.360.281* Pand objecten. De scheidingsmuren dataset, die gebaseerd is op de 3DBAG, bevat *10.363.460* Pand objecten. Er zijn *5.646.848* Pand en *8.721.692* VBO met gebruiksdoel dat **woonfunctie** bevat.

In 2023 is in het project “Levering databestand 3D-BAG”, gefinancierd door RVO, de 3DBAG uitgebreid met 3D oppervlakten. De oppervlakte-informatie is opgenomen in de 3DBAG sinds versie 2023.10.08. We gebruikten echter het CSV-bestand

(3dbag_v20231008_rvo_export.csv) dat rechtstreeks aan RVO werd geleverd. De reden voor het gebruik van het CSV-bestand is puur om het voorbereiden van de gegevens te vereenvoudigen.

2.3 Bouwjaar

De RVO (2022) studie bepaalt maximaal zeven bouwjaarperioden, afhankelijk van het woningtype. We houden dezelfde perioden aan voor het categoriseren van de gebouwen op hun bouwjaar. De perioden staan in **?@tbl-telling-periode-type** samen met het aantal woningen in de resultaten.

```
df = pd.read_csv(
    '/home/balazs/Development/wijklabels/tests/data/output/labels_individual.csv',
    usecols=["bouwperiode", "woningtype", "woningtype_pre_nta8800"],
    converters={
        "bouwperiode": Bouwperiode.from_str
    }
)
total = df.count().iloc[0] # count non-NA (!) cells
pt_crosstab = pd.crosstab(
    df["bouwperiode"],
    columns=df["woningtype_pre_nta8800"],
    margins=True,
    margins_name="Totaal"
)
ct = pt_crosstab.apply(
    lambda col: list(map(lambda cnt: f"{cnt} ({round(cnt / total * 100)}%)", col))
).replace(
    "0 (0%)", ""
).reset_index(
    drop=False
)
ct.columns.name = "Woningtype"
ct["Bouwperiode"] = ct["bouwperiode"].apply(
    lambda bp: bp.format_pretty() if bp != "Totaal" else bp
)
ct.drop("bouwperiode", axis=1, inplace=True)
ct.set_index("Bouwperiode", inplace=True)
ct
```

Table 1: Aantal (en percentage) van woningen per bouwperiode en woningtype

| Woningtype Bouwperiode | 2 onder 1 kap | galerij | maisonnette | overig | portiek | rijwoning hoek | rijwoning |
|---------------------------|---------------|----------|-------------|-------------|----------|----------------|------------|
| < 1945 | | | | | 236 (1%) | 648 (3%) | 1669 (6%) |
| < 1964 | 328 (1%) | 273 (1%) | 2313 (9%) | 11007 (43%) | | | |
| 1946 - 1964 | | | | | 372 (1%) | 212 (1%) | 474 (2%) |
| 1965 - 1974 | 16 (0%) | 269 (1%) | 168 (1%) | 2394 (9%) | 50 (0%) | 193 (1%) | 629 (2%) |
| 1975 - 1991 | 10 (0%) | 70 (0%) | 285 (1%) | 1473 (6%) | 51 (0%) | 59 (0%) | 138 (1%) |
| 1992 - 2005 | 8 (0%) | | | | | 64 (0%) | 221 (1%) |
| 2006 - 2014 | 2 (0%) | | | | | 14 (0%) | 35 (0%) |
| 2015 < | 35 (0%) | | | | | 43 (0%) | 100 (0%) |
| 1992 < | | 140 (1%) | 108 (0%) | 1418 (6%) | 15 (0%) | | |
| Totaal | 399 (2%) | 752 (3%) | 2874 (11%) | 16292 (63%) | 724 (3%) | 1233 (5%) | 3266 (13%) |

2.4 CBS buurten

De buurten worden gebruikt voor het aggregeren van de energielabels van de woningen. De buurtgrenzen worden opgehaald bij het Centraal Bureau voor de Statistiek. De BAG Pand objecten worden toegewezen aan een buurt met een kruispunttest. Elk BAG Pand wordt slechts aan één buurt toegewezen.

Versie van CBS Wijken en Buurten: 2022 v1, [link naar data](#)

2.5 EP-Online voor validatie

“EP-Online is de officiële landelijke database waarin energielabels en energieprestatie-indicatoren van gebouwen zijn opgenomen.” “EP-Online” (2023) We gebruiken de energielabels van EP-Online als “ground truth” voor de validatie van onze geschatte labels. Daarom hebben we beoordeeld of we de set woningen met een geregistreerd label op EP-Online kunnen beschouwen als een betrouwbare referentie voor vergelijking.

De EP-Online database bevat energielabels die met verschillende methoden zijn bepaald. We gebruiken alleen de labels die zijn bepaald met de NTA8800 methode. Dat wil zeggen dat de `Pand_berekeningstype` NTA 8800 bevat. In totaal vormen deze records 1.284.241 VBO's.

Figure 2 laat zien dat zowel de BAG als de EP-Online dataset een zeer vergelijkbare spreiding van woningen hebben over alle bouwjaarperioden. Dit geeft aan dat de geregistreerde energielabels in EP-Online goed de complete set woningen in Nederland vertegenwoordigen, als we kijken naar de bouwperiode.

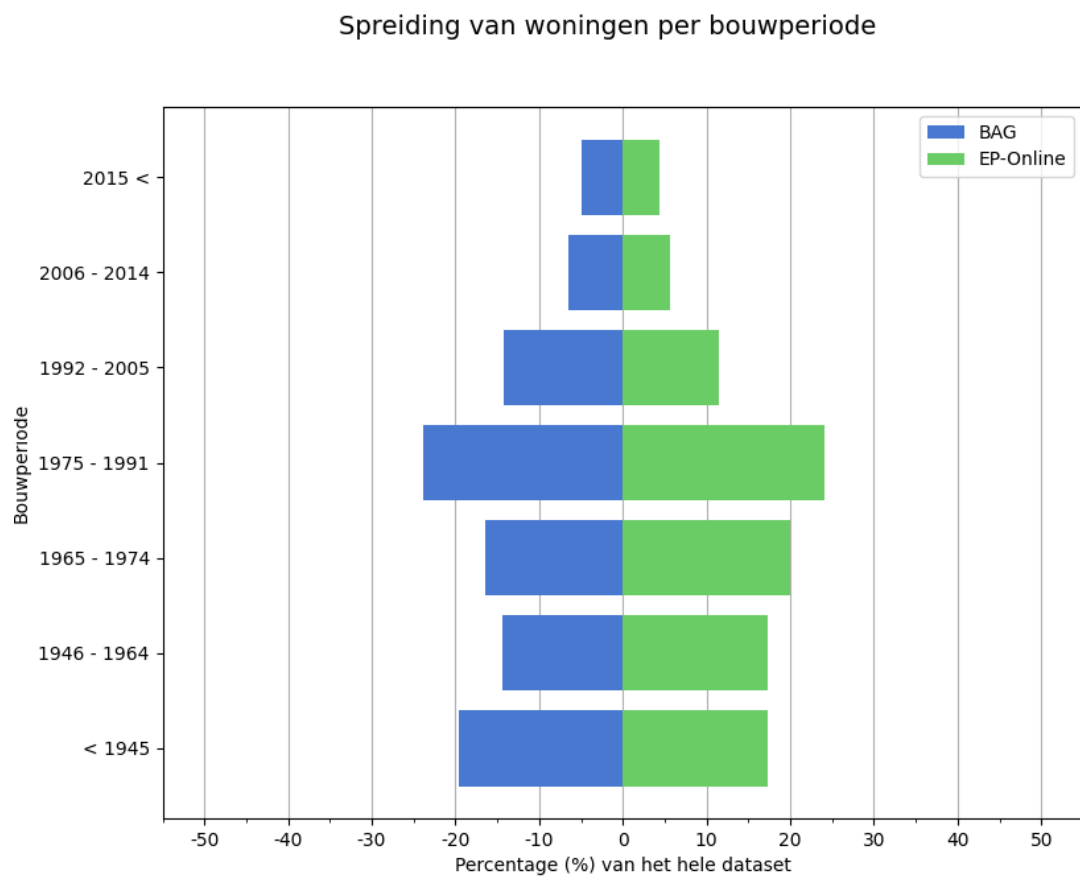


Figure 2: Spreiding van woningen per bouwperiode in de BAG en in de EP-Online gegevens

Figure 3 laat zien dat zowel de BAG als de EP-Online dataset een zeer vergelijkbare spreiding van woningen over alle woningtypen hebben. Dit geeft aan dat de geregistreerde energielabels in EP-Online de complete set woningen in Nederland goed vertegenwoordigen, als we kijken naar de woningtypen. Er zijn echter twee kanttekeningen die hierbij moeten worden geplaatst. Ten eerste omvat de vergelijking niet de subtypen appartementen, maar alleen het hoofdtype *appartement*. Dit komt doordat er geen betrouwbare methode is voor het bepalen van de appartement-subtypes voor de volledige BAG-dataset. Ten tweede zijn de woningtypen van de BAG door onszelf geschat, omdat we daar geen gezaghebbende woningtype-informatie hebben voor de hele BAG-dataset (zie Section 3.2).

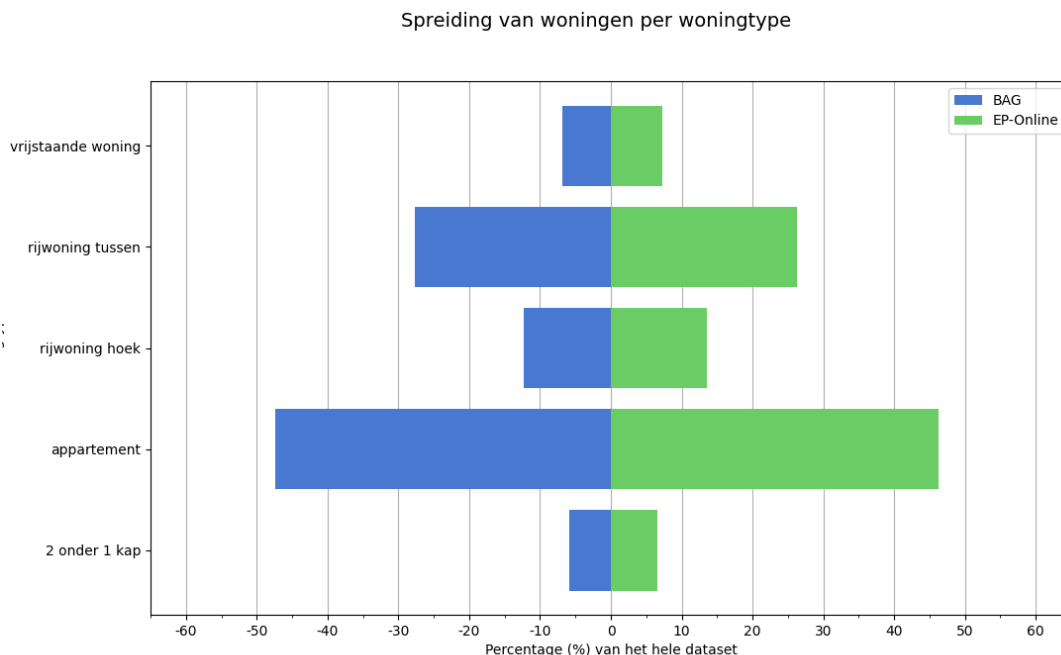


Figure 3: Spreiding van woningen per woningtype in de BAG en in de EP-Online gegevens

From the these analysis we conclude that the EP-Online database is a reliable reference for validating our estimated energy labels.

3 Methode

In dit werk wordt de vormfactor voor individuele woningen en de spreiding van energielabels in buurten in Nederland geschat. Door het gebrek aan geschikte, landelijke gegevens hebben we gewerkt met vaak strenge aannames over de parameters van een woning. De stappen voor deze aannames staan hieronder in hun volgorde in het proces van het schatten van het energielabel.

1. Schatten van het aantal verdiepingen van een Pand.
2. Het schatten van het type woning, zowel voor eengezinswoningen als voor meergezinswoningen.
3. Het omrekenen van een NTA8800 appartementtype (bijv. *appartement - hoekdak*) naar een pre-NTA8800 appartementtype (bijv. *galerij*).
4. De vormfactor van een appartement schatten. Zelfs als het exacte aantal verdiepingen en het juiste appartementtype, zoals *hoekdak*, bekend zou zijn, zou het niet mogelijk zijn om de vormfactor van appartementen exact te berekenen, vanwege de variaties in de interne indeling van gebouwen. Het zou echter wel mogelijk zijn om een veel nauwkeurigere schatting te maken.
5. Het energielabel van een individuele woning schatten op basis van de waarschijnlijkheid van energielabels die zijn gedocumenteerd in de studie Voorbeelwoningen 2022.

3.1 Aantal verdiepingen

Het aantal verdiepingen van een Pand is nodig voor de spreiding van de VBO-s over de verdiepingen in een meergezinswoning. We berekenen het aantal verdiepingen door de gebruiksoppervlakte te delen door de oppervlakte van de dakafdruk uit de 3DBAG. De gebruiksoppervlakte is de som van de **oppervlakte** van elke VBO van een Pand in de BAG. De dakprintoppervlakte is de **totale oppervlakte begane grond**, zonder ondergrondse delen. De resulterende waarde wordt naar boven afgerond op het dichtstbijzijnde hele getal, wat ons het aantal verdiepingen geeft.

3.2 Woningtypes

Woningen worden ingedeeld in verschillende typen. De huidige classificatie, zoals gebruikt door de NTA8800 methode, staat in Table 2. In het geval van appartementen verwijst hoek/tussen naar de horizontale positie van het appartement in het gebouw, terwijl vloer/midden/dak/dakvloer verwijst naar de verticale positie. Voor de eenvoud verwijzen we naar deze classificatie als **NTA8800-types**.

Table 2: Woningtypes gebruikt door de NTA8800 methode

| Eengezinswoningen | Meergezinswoningen |
|--------------------|----------------------------|
| vrijstaande woning | appartement - hoekvloer |
| 2 onder 1 kap | appartement - hoekmidden |
| rijwoning tussen | appartement - hoekdak |
| rijwoning hoek | appartement - hoekdakvloer |
| | appartement - tussenvloer |
| | appartement - tussenmidden |
| | appartement - tussendak |

| Eengezinswoningen | Meergezinswoningen |
|------------------------------|--------------------|
| appartement - tussendakvloer | |

Vóór de invoering van de NTA8800-methode werden woningen iets anders ingedeeld, zie Table 3. Dit is de classificatie die wordt gebruikt in de RVO (2022) studie. Voor het gemak verwijzen we naar deze classificatie als **pre-NTA8800-types**.

Table 3: Woningtypes gebruikt in RVO (2022)

| Eengezinswoningen | Meergezinswoningen |
|--------------------|--------------------|
| vrijstaande woning | maisonette |
| 2 onder 1 kap | galerij |
| rijwoning tussen | portiek |
| rijwoning hoek | flat (overig) |

De types van de eengezinswoningen zijn gelijkwaardig voor en na NTA8800. Aan de andere kant is er geen duidelijke relatie tussen de typen van de meergezinswoningen voor en na NTA8800.

Dit werk vereist dat we zowel het *NTA8800* als het *pre-NTA8800* type van elke woning kennen. Het *pre-NTA8800* type is nodig, want dat wordt gebruikt door RVO (2022), en het *NTA8800* type is nodig, want dat wordt gebruikt door de validatiegegevens op EP-Online.

We schatten eerst het *NTA8800* type en converteren dat vervolgens naar een *pre-NTA8800* type. De classificatie van eengezinswoningen is vrij eenvoudig en kan worden gedaan met behulp van een conventionele GIS-analyse op de BAG-polygonen. Het classificeren van de meergezinswoningen is veel moeilijker omdat dit afhangt van de binnenarchitectuur van een Pand, en daar hebben we geen kennis van uit onze invoerdatasets. Daarom moesten er een aantal sterke aannames en educated guesses worden gedaan om tot een methode te komen.

3.2.1 Classificatie van een Pand

Elk BAG Pand wordt geclassificeerd door de kruisende BAG-geometrieën te clusteren. Een rij van vijf rijtjeshuizen vormt bijvoorbeeld één cluster, omdat ze een groep verbonden objecten vormen. We bepalen de typen *vrijstaand/2 onder 1 kap/rijwoning* aan de hand van het aantal gebouwen in het cluster. In het geval van een *rijwoning* bepalen we de positie *hoek/tussen* uit het aantal kruispunten met andere gebouwen in het cluster.

Misclassificatie treedt op als de snijpunten onjuist worden bepaald, meestal omdat er een kleine tussenruimte is tussen BAG polygonen die elkaar zouden moeten raken. Om de classificatie van eengezinswoningen te verbeteren, moeten de BAG-polygonen dus topologisch worden gecorrigeerd, zodat de gaten en overlappingen worden gecorrigeerd.

3.2.2 Eengezinswoningen

Als een pand slechts één VBO bevat, dan beschouwen we de VBO als een eengezinswoning en krijgt de VBO de classificatie van het Pand.

3.2.3 Meergezinswoningen

De meergezinswoningen bestaan uit één hoofdtype, *appartement*. Als een Pand meer dan één VBO bevat, dan worden al zijn VBO's geassocieerd als *appartement*.

De *appartement* subtypes worden bepaald door de verticale en horizontale positie van de VBO binnen het pand. Eerst worden de VBO's verspreid over de verdiepingen (zie ?@sec-vloeren) van het pand om hun verticale positie te bepalen. Elke verdieping krijgt hetzelfde aantal appartementen toegewezen, dat wordt berekend door het aantal VBO's te delen door het aantal verdiepingen. We noemen het *aantal appartementen per verdieping* N . Vervolgens wordt het totaal aantal appartementen in het pand zo verdeeld dat de eerste N wordt toegewezen aan de begane grond (*vloer*), de tweede N wordt toegewezen aan de bovenste verdieping (*dak*) en de rest wordt gelijkmatig verdeeld over de tussenliggende verdiepingen (*midden*). Als een N gelijk is aan het totaal aantal appartementen, dan worden de appartementen geassocieerd als *dakvloer*.

Ten tweede zijn de appartementen horizontaal verdeeld over elke verdieping. We gaan uit van twee configuraties voor de indeling van de appartementen, enkele rij of dubbele rij. De keuze tussen enkele of dubbele rij hangt af van het aantal appartementen per verdieping en een willekeurige keuze. Als het aantal appartementen per verdieping kleiner of gelijk is aan drie, dan wordt gekozen voor een indeling met enkele rij, anders is er 50% kans op een indeling met dubbele rij. Nu wordt het aantal *hoek* appartementen geschat op basis van de indeling van het pand en de eerder bepaalde indeling. Als er nog appartementen op de verdieping zijn die niet als *hoek* zijn ingedeeld, worden ze als *tussen* ingedeeld.

3.3 Conversie van appartementtypes naar pre-NTA8800 types

De RVO (2022) studie gebruikt de pre-NTA8800 woningtypes, terwijl de EP-Online database de NTA8800 woningtypes gebruikt voor de energielabels die zijn berekend met de NTA8800 methode. Dit werk vertrouwt op de resultaten van de RVO (2022) studie om het energielabel van woningen te schatten, daarom rekenen we de eerder bepaalde NTA8800 woningtypen om naar pre-NTA8800 typen. Omdat er geen directe relatie is tussen de twee classificaties, kunnen we alleen de pre-NTA8800 typen schatten.

We wijzen het pre-NTA8800 type toe aan het appartement op basis van de spreiding van pre-NTA8800 types in de EP-Online database en de bouwdatum van de woning, zie Figure 4. Als de woning bijvoorbeeld is gebouwd in de periode 1965-1974, is er een kans van 84% dat de woning het *flatwoning (overig)* pre-NTA8800 type krijgt.

Spreiding van appartementtypen in de EP-Online database

Energielabels opgenomen voor 01.01.2021.

Nr. woningen: 1541202.

0-1964: 449754

1965-1974: 230556

1975-1991: 328142

1992-9999: 525479

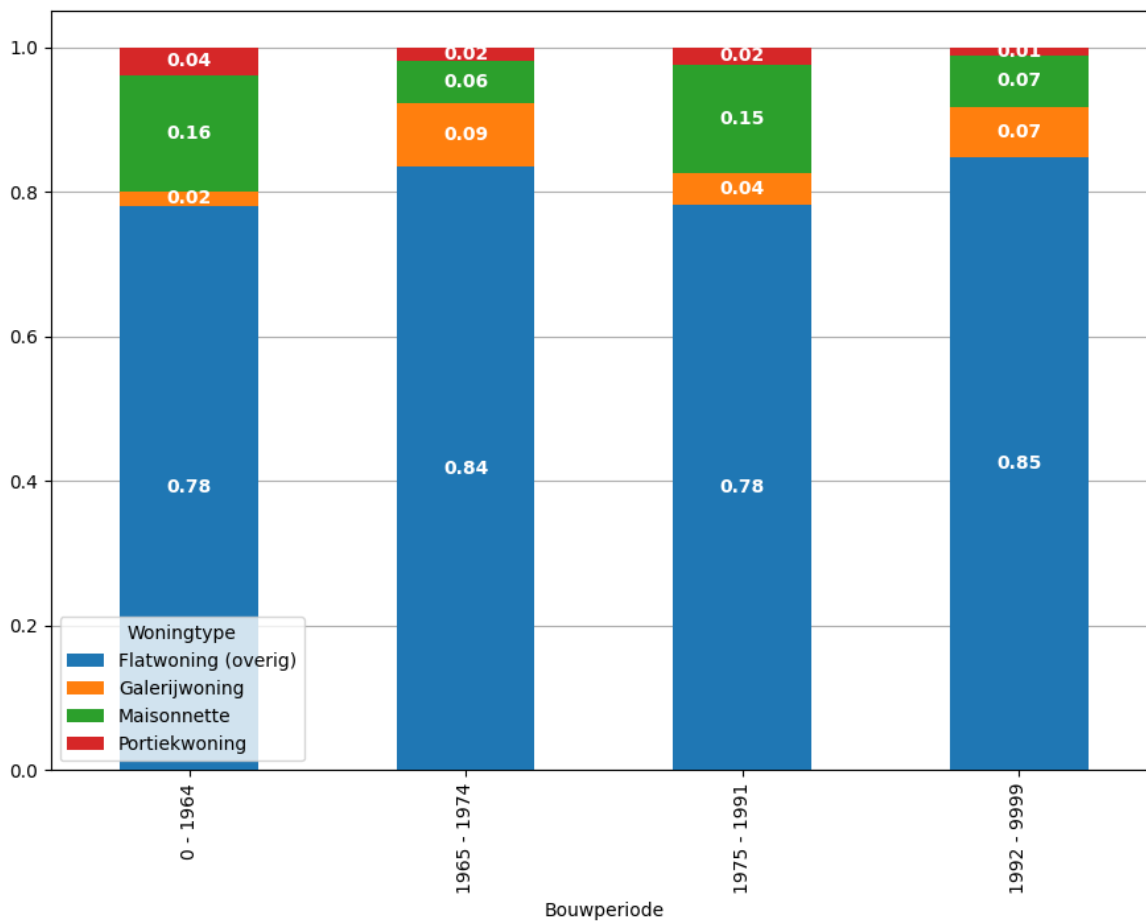


Figure 4: Spreiding van pre-NTA8800 meergezinswoningtypen in de EP-Online database, “EP-Online” (2023)

3.4 Vormfactor

De vormfactor wordt berekend als de fractie van de verliesoppervlakte en gebruiksoppervlakte. De verliesoppervlakte is de som van alle oppervlakken die de woning omhullen, behalve de oppervlakken die worden gedeeld met een andere woning. De gebruiksoppervlakte is de oppervlakte die voor de VBO is geregistreerd in de BAG (de *oppervlakte* van de VBO).

We berekenen de vormfactor voor elk Pand. De vereiste oppervlakten voor het berekenen van de verliesoppervlakte maken deel uit van de 3DBAG sinds versie 2023.10.08. De vormfactor van een eengezinswoning is gelijk aan de vormfactor van het Pand.

Voor appartementen wijzen we een deel van de oppervlakten toe aan elk appartement, afhankelijk van het type (zie *?@sec-meergezinswoningen*). Het totale dakoppervlak wordt gelijk verdeeld over de appartementen op de dakverdieping. Slechts 95% van het totale muuroppervlak wordt gebruikt om rekening te houden met muuroppervlakken die gangen en andere niet-woonruimten in het pand bedekken. Het muuroppervlak wordt vervolgens verdeeld over de appartementen, zodanig dat de appartementen op de hoek ongeveer 3x het muuroppervlak van een tussenappartement krijgen. Het totale grondoppervlak wordt gelijk verdeeld over de appartementen op de begane grond.

3.5 Het schatten van de energielabels

Om het juiste energielabel voor een woning te bepalen, moet de woning persoonlijk worden geïnventariseerd door een gekwalificeerde professional. Dit is niet haalbaar om op nationale schaal te doen. We kunnen echter wel een schatting maken van het type woning (Section 3.2) en we kennen de bouwperiode uit de BAG. Daarnaast kunnen we de vormfactor voor eengezinswoningen nauwkeurig berekenen en de vormfactor voor meergezinswoningen schatten (zie Section 3.4). Tot slot hebben we uit RVO (2022) een spreiding van energielabels voor elke combinatie van de drie parameters, woningtype, bouwperiode en vormfactor. Vervolgens wordt het energielabel van een individuele woning geselecteerd op basis van de waarschijnlijkheid van alle labels over de drie variabelen.

Bijvoorbeeld, gegeven een woning met de parameters woningtype *vlakke woning*, bouwperiode *1965-1974* en een vormfactor in het bereik van *1,50-2,00*, hangt het geschatte energielabel voor de woning af van het bereik van beschikbare labels, *C-G*, en hun waarschijnlijkheid (zie Figure 1).

Door de hiaten in de energielabelverdelingen die worden gepresenteerd in RVO (2022) (zie Section 2.1), zijn er situaties waarin het onmogelijk is om het juiste label voor een VBO te schatten. De VBO heeft bijvoorbeeld een geregistreerd label *B* in EP-Online, maar RVO (2022) heeft geen gegevens voor label *B* en de parameters van de VBO. We beschouwen dergelijke labels als *onmogelijk*, omdat het met onze methode en de beperkingen van de gegevens van RVO (2022) niet mogelijk is om het juiste label toe te wijzen. Door de *onmogelijke* labels te

meten kunnen we de invloed van de hiaten in de RVO (2022) gegevens en de gevoeligheid van onze methode evalueren.

De individuele labels worden per buurt samengevoegd om de spreiding van de energielabels voor elke buurt in Nederland te schatten.

4 Validatie van de energielabelschatting

We valideren onze energielabelschatting aan de hand van de energielabels in de “EP-Online” (2023) database. We vergelijken de geschatte labels met de EP-Online labels in termen van *deviatie*. De *afwijking* wordt berekend als de numerieke, getekende afstand van een label tot het referentielabel. Waarbij de afstand van een slechter label tot een beter label positief is. Bijvoorbeeld, de afstand van $A++++$ tot G is -10 , de afstand van G tot $A++++$ is 10 .

Het validatieproces meet de volgende aspecten.

- Aantal VBO's dat geen energielabel heeft ontvangen (door een gat in de spreiding van energielabels, zie Section 2.1).
- Percentage labels dat exact overeenkomt met de EP-Online labels. Bijvoorbeeld het geschatte label is A , dan is de EP-Online ook A voor de geselecteerde VBO.
- Percentage labels dat overeenkomt met de EP-Online labels met één labelafwijking. Bijvoorbeeld het geschatte label is A , dan kan de EP-Online een van $A+$, A , B zijn.
- De mediaan, het gemiddelde, de standaardafwijking, het minimum en maximum van de *afwijkingen* per buurt en voor de volledige gegevensverzameling.

5 Resultaten

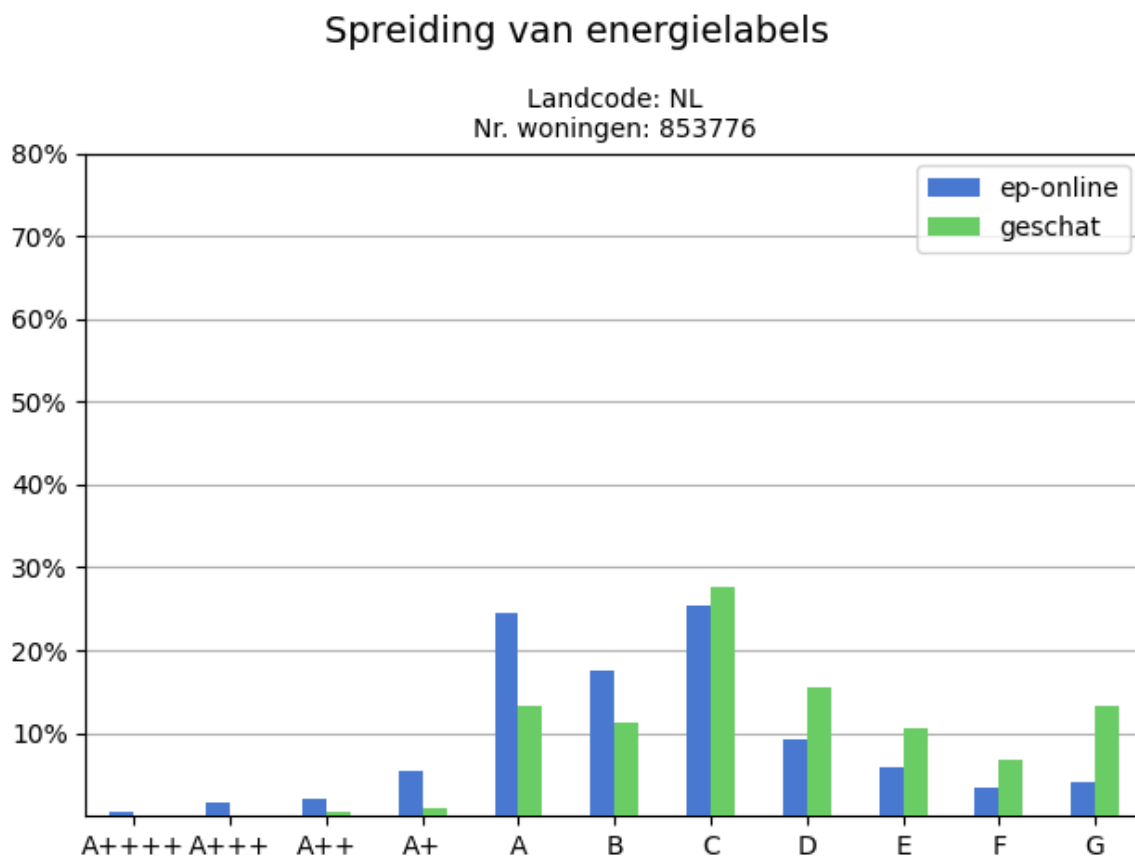
5.1 Woningtypes

Wanneer we vergelijken met de EP-Online data, laat onze woningclassificatie 87% nauwkeurigheid zien voor de eengezinswoningen, en 26% nauwkeurigheid voor de meergezinswoningen.

5.2 Schatting van energielabels

Van de beschikbare 5.385.950 Pandobjecten die zijn ingevoerd, hebben 4.847.298 een energielabel ontvangen. Dus 13% van de objecten heeft geen label ontvangen. De ontbrekende labels worden veroorzaakt door de hiaten in de energielabelverdelingen van de RVO (2022) .

We hebben onze energielabelschatting vergeleken met de labels die beschikbaar zijn in “EP-Online” (2023). **fig-nl-dist-bar** toont de spreiding geschatte (groen) en ep-online energielabels voor de woningen die geregistreerd zijn op ep-online. We kunnen zien dat het meest voorkomende geschatte energielabel *C* is, iets meer dan 25% van de labels. In de ep-online dataset vormen de labels *A* en *C* echter elk ongeveer 25% van de totale labels.



We hebben vastgesteld dat de gemiddelde afwijking van de “EP-Online” (2023) labels **-1,2** is, met een standaardafwijking van **2,1**.

Met andere woorden, gemiddeld liggen onze geschatte labels ongeveer één label onder de EP-Online labels, met een standaardafwijking van twee labels.

Als we echter kijken naar de afwijkingen per label, krijgen we een gedetailleerder beeld. Figure 5 laat zien dat onze *A++*, *A++++* labels gemiddeld 3-5 labels overschat zijn, terwijl onze

E, F, G labels gemiddeld 2-4 labels onderschat zijn ten opzichte van de EP-Online data.

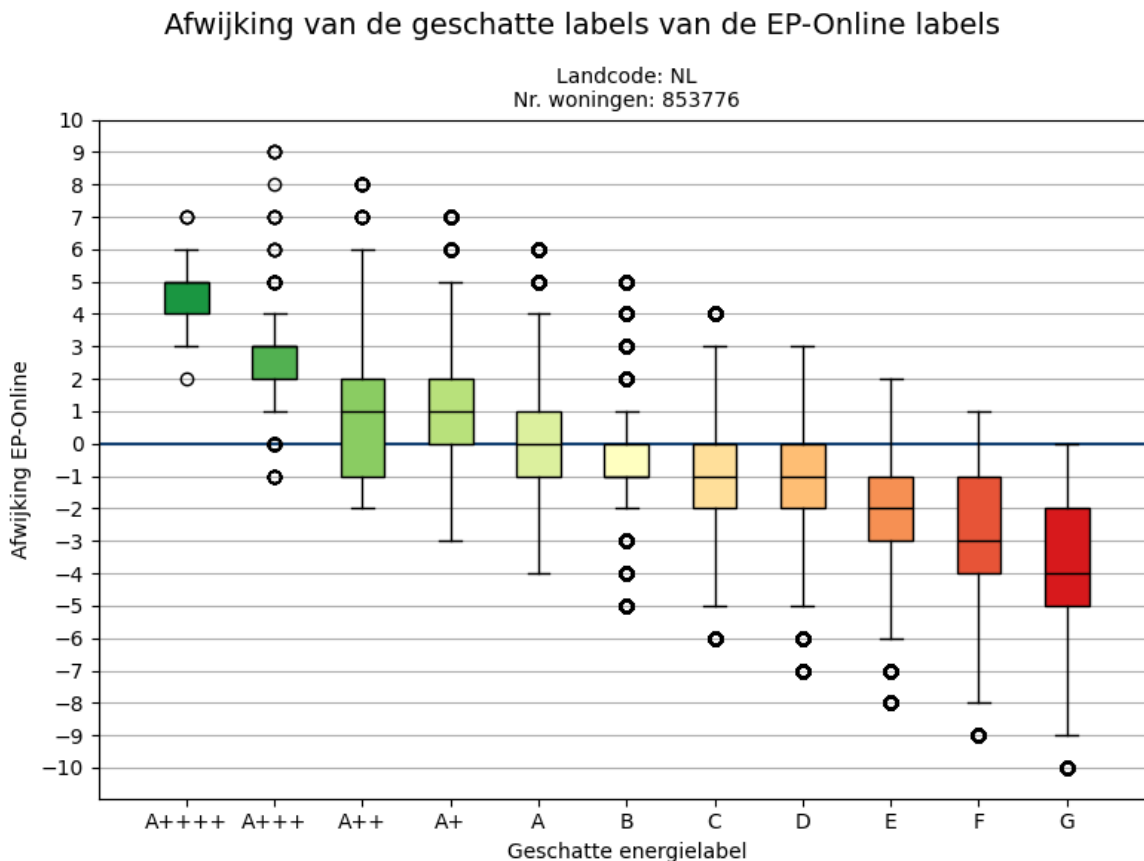


Figure 5: Afwijking van de geschatte labels van de EP-Online labels

5.2.1 Alleen rekening houdend met mogelijke labels

Door de hiaten in de energielabelverdelingen van de RVO (2022) studie heeft 32% van de woningen een label in EP-Online dat ontbreekt in de Voorbeeldwoningen 2022 gegevens (zie Section 2.1). In die gevallen is het niet mogelijk om het juiste energielabel aan de woning toe te kennen, omdat er geen waarschijnlijkheid is voor het label. Als we onze validatie beperken tot die woningen waar het juiste label wel een kans heeft in de energielabelverdelingen, krijgen we inzicht in de invloed van de hiaten in de energielabelverdelingen.

Als we alleen de mogelijke labels vergelijken, is de gemiddelde afwijking van de “EP-Online” (2023) labels **-0,6**, met een standaardafwijking van **1,6**.

6 Conclusies

De resultaten laten zien dat het meest voorkomende geschatte energielabel C is, met ongeveer $X\%$ van de woningen met dit label. Gemiddeld liggen de geschatte labels één label onder het echte label, met een standaardafwijking van twee labels. Het meest nauwkeurige label is B , met de laagste mediane afwijking van de ep-online labels, gevolgd door het label A .

Onze bevindingen geven aan dat de NTA8800 energielabels in de EP-Online database voldoende dekking hebben om in de toekomst te worden gebruikt als validatieset voor de ontwikkeling van landelijke methoden voor grote gebieden. De energielabels die zijn bepaald met de NTA8800 methode gebruiken echter een andere appartementindeling dan wat wordt gebruikt in de WoON 2018 en Voorbeeldwoningen 2022 studies. In onze methode volgen we de appartementindeling van de NTA8800 labels, daarom is een conversie tussen de twee indelingen nodig. Voor de conversie volgen we de statistische spreiding van appartementtypen die we hebben gevonden in de EP-Online dataset, omdat er voor zover wij weten geen deterministische methode is om deze conversie te doen. Daarom voegt de conversie van het appartementtype zelf meer onzekerheid toe aan de schatting van het energielabel van appartementen.

Door het gebrek aan nationale gegevens over de 3D-locatie van appartementen is onze methode voor het schatten van de horizontale en verticale positie van een appartement giswerk en dus onnauwkeurig. Daarom raden we in de huidige staat niet aan om de resultaten te gebruiken voor meergezinswoningen. In onze resultaten is 44% van de woningen een appartement.

Onze resultaten laten zien dat de hiaten in de distributies van de Voorbeeldwoningen 2022 energielabels een significante invloed hebben op de nauwkeurigheid van de geschatte labels (zie Section 5.2.1), namelijk 32% van de geschatte labels. Daarom raden we aan om de studie van Voorbeeldwoningen aan te passen zodat de distributies van energielabels zo volledig mogelijk zijn.

7 References

“EP-Online.” 2023. <https://www.ep-online.nl/>.

RVO. 2022. “Voorbeeldwoningen 2022 - Bestaande Bouw.” Technical report RVO-231-2022/BR-DUZA. Rijksdienst voor Ondernemend Nederland.