Vlaanderen
is sterke steden

CITY FLOWS

Rapport CityFlows

Bevindingen datafusiemodel

Yanick Van Hoeymissen en Kasper Segers

AGENTSCHAP
BINNENLANDS
BESTUUR

www.vlaanderen.be/stedenbeleid

Inhoud

Kadering 4

Executive summary 5

Inleiding 7

Input van het model 8

- Telecomdata 8
- Telraam 9
- ANPR 11
- Fietstelpalen 13
- Tijdelijke meetinstallaties 15
- Straatvinken 16
- Floating car data 17

Datafusiemodel 19

- Dataverwerking 19
  - *De input van het model* *20*
  - *De output van het model* *20*
  - *Datafusie* *21*
  - *Gewogen verdeling* *21*
  - *Het validatieproces* *22*

De resultaten van het model 23

- Uitdagingen 23
- Voorstellen voor verbetering 27

Conclusie 30

Kadering

Onderstaand document is gebaseerd op de ervaring die is opgedaan tijdens het CityFlows-project dat imec uitvoerde in opdracht van ABB. Dit document is een kritische analyse van het datafusiemodel dat voor het project gebouwd is en bevat daarnaast voorstellen ter verbetering. Het rapport is gericht op mobiliteitsonderzoekers en beleidsmedewerkers met enige kennis van data zodat ze de voorstellen voor verbetering ook kunnen toepassen.

Voor meer informatie over het aankopen van data rond stadsdrukte verwijzen we graag naar de opgestelde modelclausules die in samenwerking met de steden zijn ontwikkeld. Voor de communicatiedienst van de steden is er daarnaast een draaiboek geschreven rond het monitoren van drukte waarin de focus ligt op de gerichte communicatie naar de burger en minder op het technische aspect van databronnen. De code van de modellen en de analysetoolbox zijn openbaar gemaakt via Github.

Zoals reeds aangegeven worden er in dit rapport mogelijke tekortkomingen in ons model aangekaart en voorstellen ter verbetering daarvan aangedragen. Zoals in elk innovatief project is er namelijk altijd ruimte voor verbetering. Dit rapport moet dan ook vooral gezien worden als een kapstok om vervolgonderzoek te bevorderen en vooruit te helpen.

Disclaimer

Al de bevindingen over type sensoren van een specifiek merk of bedrijf zijn gebaseerd op bevindingen van het CityFlows team binnen de context en werking van het project. Bepaalde uitspraken/beoordelingen dienen dus ook binnen die context gezien te worden. Dit rapport werd in het voorjaar van 2023 opgesteld en werd nadien niet meer geüpdatet met nieuwe bevindingen.

Executive summary

Begin 2021 erkent het Smart Flanders-programma van het Agentschap Binnenlands Bestuur (ABB) de noodzaak om mobiliteitsstromen van alle transportmodi overzichtelijk in kaart te brengen om (de mobiliteitsexpert van) de steden te ondersteunen bij de uitwerking van hun datagedreven mobiliteitsbeleid. Het vooropgezette plan was om een totaalbeeld van de mobiliteitsstromen te ontwikkelen door de al bestaande databronnen rond mobiliteit te integreren met eenvoudige wiskundige fusies. Samen met imec werd er binnen het kader van het CityFlows-traject gezocht naar twee pilootsteden die het datafusiemodel wilden gebruiken en hieraan een use case konden koppelen.

Als pilootsteden werden Mechelen en Gent geselecteerd. In Mechelen moest de invoering van een schoolstraat getoetst worden aan de hand van het model. Er moest beoordeeld worden of en hoe het mobiliteitsgedrag was veranderd nadat de schoolstraat ingevoerd was. In Gent moest een wijkmobiliteitsplan (Zwijnaarde) geëvalueerd worden als onderdeel van het project. In beide pilootsteden moest de evaluatie uitgevoerd worden op basis van een combinatie van databronnen, zodat er niet enkel uitspraken gedaan konden worden op plaatsen waar gemeten werd, maar ook daar waar niet werd gemeten.

Het model dat hiervoor ontwikkeld werd, bouwt verder op een proof of concept (PoC) dat de universiteit van Antwerpen ontwikkelde in de Sint-Andrieswijk die de naam Antwerp Smart Zone kreeg. Hier werden verschillende technologische toepassingen getest in een echte leefomgeving. Een van die toepassingen was een datafusiemodel dat verschillende databronnen rond mobiliteit combineerde om het aantal deelnemers aan het verkeer te bepalen per modaliteit in elke straat en op elk moment van de dag. Het model dat daarvoor gebouwd was, combineerde databronnen zonder echte regels (bijvoorbeeld de regel dat mensen ’s ochtends hun woongebied verlaten en ’s avonds weer terugkomen), maar op basis van een wiskundige verdeling. Het oorspronkelijke model in de PoC had slechts een beperkt aantal modaliteiten (twee: wel en niet-gemotoriseerd) werkte slechts op een kleine geografische oppervlakte met beperkte databronnen.

De opdracht van imec was om het ontwikkelde model van de PoC op te schalen naar meer modaliteiten en meer databronnen. Er moest ook een mechanisme ontwikkeld worden dat de accuraatheid van de output van het model controleerde. imec voegde meer databronnen toe (Telraam, fietstellussen van Signco, tijdelijke tellingen en ANPR-data) en trachtte het model bruikbaar te maken op stadsniveau en voor alle modaliteiten (voetgangers, fietsers en gemotoriseerde voertuigen).

Het mechanisme dat de accuraatheid van de output van het model controleerde (validatiescore) werd ontwikkeld aan de hand van Straatvinkendata (ground truth). Bij het samenvoegen van de eerste databronnen bleek de validatiescore echter laag. Onze hypothese was dat de validatiescore zou toenemen naarmate er meer databronnen toegevoegd zouden worden. Die hypothese werd echter niet bevestigd. Om de kwaliteit van het model te optimaliseren werd ook de verdeling van de modaliteiten per straat (verder in het document aangeduid met wegingscoëfficiënten) geoptimaliseerd, maar ook hierna bleef een positief effect op de validatiescore uit.

We kunnen daaruit concluderen dat deze versie van het opgeschaalde model geen betrouwbaar middel is om de use cases in de pilootsteden op te lossen. Delen van het model zijn mogelijk wel herbruikbaar, zodat de data-uniformisatie en -ingestie. Die zullen dan ook samen met de andere componenten van het model beschikbaar worden gemaakt. Onze belangrijkste aanbeveling is dat een toekomstig model meer rekening zal moeten houden met het dynamische karakter van een stad. Dat wil zeggen: de wegingscoëfficiënten die wij hebben toegepast, zouden dynamischer moeten zijn en bedacht moeten zijn op het menselijke verplaatsingsgedrag. Artificiële intelligentie kan daarvoor mogelijk een oplossing zijn.

Nadat de lage betrouwbaarheid van het model aan het licht was gekomen, zijn de use cases met de pilootsteden herbekeken om te onderzoeken hoe die wel opgelost konden worden. Na nauw overleg met de twee centrumsteden en de stuurgroep van CityFlows heeft imec een analysetoolbox ontwikkeld die de mobiliteitsexpert in staat stelt met enkele muisklikken een bepaalde beleidsuitdaging op vlak van mobiliteit te evalueren. Op die manier creëerden we meer zekerheid dat de steden echt vooruitgeholpen waren met hun problematiek. In de aanbevelingen van dit rapport zal

daarop dieper ingegaan worden. Een korte samenvatting van het CityFlows-project wordt getoond in Figuur 1 .

Als mobiliteitsexpert wil ik in mijn stad mobiliteitsstromen in kaart brengen

- Mechelen: Schoolstraat evalueren
- Gent: evaluatie wijkmobiliteitsplan

We brengen in kaart welke mobiliteitsbronnen de steden ter beschikking hebben

Telco-data, Telraam, signco fietstellussen, tijdelijke tellingen en ANPR-data, Floating car data, …

Output datafusiemodel onbetrouwbaar:

Version	Data sources used in the model	Validation score
LC1	Cropland (telco), fietsdata	0.04435
LC2	Cropland (telco), fietsdata+ Telraam, Signco	0.04462
LC3	Cropland (telco), fietsdata, Telraam + ANPR	0.04446
Weighted	LC3 + machine learned weights	0.02749
Adapted Weights	LC3 + adapted machine learned weights	0.00971

Conclusie:

- Model is niet bruikbaar voor de steden en kan geen antwoord bieden op use case
- Werk rond data uniformisatie en ingestie is herbruikbaar voor andere modellen/partijen

Aanbevelingen:

- Antwoord op de use cases is vaak al te zien in de data (Analyse toolbox voorzien)
- Machine learning weights evalueren en verder onderzoeken met dynamiek van stad in rekening
- Nieuwe databronnen over mobiliteitsstromen onderzoeken en toevoegen
- Mogelijke nieuwe manieren van modelleren onderzoeken (agent based)

Figuur 1: Samenvatting CityFlows-traject

In dit rapport gaan we ook dieper in op de voornaamste uitdagingen van het CityFlows-project inclusief de stappen die ondernomen zijn om een oplossing voor die uitdagingen te kunnen aandragen en eventuele vervolgstappen die ondernomen kunnen worden op basis van dit project. Ten slotte zijn er - met het oog op vervolgonderzoek - ook aanbevelingen opgenomen waarmee steden aan de slag zouden kunnen gaan. Die aanbevelingen zijn voortgekomen uit interne gesprekken tijdens de evaluatie van het project en nabesprekingen met het VLOCA-team[1] en het team Stedenbeleid van ABB. Tijdens die gesprekken werd de nadruk vooral gelegd op de mogelijkheden van Telraamdata en de bijdrage daarvan aan de wegingscoëfficiënten.

[1] https://vloca.vlaanderen.be/over-vloca

Inleiding

Zoals hierboven reeds vermeld was de projectdoelstelling om twee pilootsteden te ondersteunen in hun use case. Daarvoor werd een datafusiemodel gebouwd dat de mobiliteitsstromen overzichtelijk in beeld moest brengen. De databronnen die daarbij gebruikt zijn, vertonen gaten en onnauwkeurigheden en zijn daardoor niet of slechts beperkt operationeel. De functionaliteit van elke afzonderlijke databron was voorafgaand aan het project dus eerder beperkt. Telecomdata werd bijvoorbeeld enkel gebruikt om de stadsdrukte te meten (om bijvoorbeeld de economie van de stad te monitoren).
De selectie van de pilootsteden en hun use case zit niet in de scope van dit document. Dit rapport richt zich vooral op het model en in welke hoedanigheid het wel of niet gebruikt kan worden. Dit rapport bevat een kritische analyse van de input en output en de verwerking daarvan.

Als eerste komt natuurlijk de input (de databronnen) aan bod. Elke databron wordt geanalyseerd aan de hand van een SWOT-analyse[2] waaruit een conclusie volgt met suggesties voor mogelijke verbeteringen van de desbetreffende databron. Omdat niet imec, maar de steden eigenaar zijn van de sensoren en de data, spreekt het voor zich dat de aanbevelingen hieronder slechts suggesties zijn. De aanbevelingen zijn echter wel gestoeld op ervaring in het gebruik van data in onderzoek naar oplossingen voor vraagstukken rond smart city en smart mobility.
Bij het gedeelte over het datafusiemodel wordt er dieper ingegaan op de dataverwerking. Dit deel bevat een uitleg van het datafusiemodel en aanbevelingen en mogelijke denkfouten. Ook wordt hier per pilootstad aangegeven wie de exacte eigenaar is van welke databron. De onderzoeken naar machine learning die binnen het project verricht zijn, komen ook in dit deel aan bod.

Als laatste wordt er dieper ingegaan op de output van het model en de uitdagingen die tijdens het project naar boven zijn gekomen. Voor sommige uitdagingen worden bovendien mogelijk oplossingen aangedragen. Met het oog op vervolgonderzoek worden er ook enkele aanbevelingen gedaan waar steden mee aan de slag kunnen en suggesties voor vervolgonderzoek voorgesteld. Daarbij wordt er kritisch gekeken naar de keuzes die tijdens het project gemaakt zijn.

[2] SWOT-analyse of een Sterkte-zwakteanalyse - Wikipedia

Input van het model

Zoals aangegeven in de inleiding worden de databronnen die gebruikt zijn voor het CityFlows-project met de twee pilootsteden, als eerste nader bekeken. In iedere sectie komt een andere databron aan bod. Aan de hand van een SWOT-analyse wordt elke databron afzonderlijk geanalyseerd en op basis van die analyse worden er enkele aanbevelingen voor de sensoren en databronnen gedaan.

Telecomdata

Telecomdata is data over het aantal smartphones dat aanwezig is in een specifieke zone (vierkante cellen van 500 m x 500 m, of in het geval van Mechelen één grote onregelmatige cel). De frequentie waarmee geüpdatet wordt wie aanwezig is in een cel, verschilt per data-aanbieder. Voor Mechelen was de updatefrequentie vijftien minuten, terwijl dat voor Gent een keer per uur was. Per updatefrequentie wordt bijgehouden hoe veel smartphones zich in een specifieke cel bevinden (snapshots). De locatie van een smartphone wordt, afhankelijk van wat eerst komt, elk kwartier of aan de hand van de laatste ping (de laatste connectie met het netwerk) bepaald. Het is hierbij belangrijk om te vermelden dat er in elke stad met één provider gewerkt wordt en dat die enkel data heeft over zijn eigen gebruikers. Om een telling te kunnen leveren voor alle toestellen zal de provider extrapoleren op zijn geschatte marktaandeel.

- **Sterktes**

- Deze databron geeft een beeld van het totaal aantal mensen dat aanwezig is. Dat beeld omvat dus alle modaliteiten, maar ook de achtergrond (mensen die niet deelnemen aan het verkeer, maar die hun smartphone bijvoorbeeld in huis gebruiken).
- De data is al veel gebruikt in een andere context (Toerisme).
- Goede geografische dekking: geeft niet enkel een lokaal beeld van een plek waar een specifieke sensor staat, maar geeft een beeld van een grotere zone.

- **Zwaktes**

- Er is geen opsplitsing per modaliteit mogelijk, omdat er geen beeld is van de snelheid/modaliteit per persoon.
- Ook mensen die niet deelnemen aan mobiliteit worden meegenomen in de telling.
- Door anonimisatie kan er geen link gelegd worden tussen verschillende snapshots.
- Weinig tot geen inkijk in hoe de output (snapshots) tot stand komen.
- Het is onduidelijk hoe accuraat de extrapolatie op basis van het marktaandeel is.
- Het update van de locatie van een smartphone gebeurt niet realtime (in functie van het gebruik van de smartphone). Dat zorgt voor vertraging en kan betekenen dat de persoon in kwestie inmiddels al uit de zone is, en dat er dus onjuist geteld wordt. Bovendien is er geen inschatting van hoe groot deze “fout” is.

- **Opportuniteiten**

- Dankzij CityFlows werden gebreken in datakwaliteit aangekaart en konden de steden die gebreken terugkoppelen aan hun leverancier. De ervaring van CityFlows geeft de steden de kans om kwaliteitsvollere data in de toekomst in een contract te laten beschrijven (zie ook modelclausules aankoop data rond drukte in de stad)
- Hogere updatefrequenties van smartphonelocaties (bij gebruik van meer apps) zorgt ervoor dat deze databron in de toekomst potentieel heeft.
- De verkoop van smartphones groeit nog steeds. Hoe meer mensen een smartphone bezitten, hoe meer mensen gedetecteerd kunnen worden.

- **Bedreigingen**

- Door de GDPR-wetgeving kan deze databron mogelijk extra beperkingen/anonimisering opgelegd krijgen.
- De verantwoordelijkheid voor het oplossen van bepaalde beperkingen het potentieel gebruik van deze data ligt bij de telecomproviders. Zij zouden de snelheid van verplaatsing of modaliteit kunnen meegeven. De vraag is of deze (vaak grote) telecomproviders het potentieel hiervan inzien en de technische uitdagingen kunnen en willen aangaan.

- **Conclusies**

Deze databron had veel potentieel, maar na het eerste gebruik gaf de databron meer problemen dan vooraf verwacht. De achtergrond (zowel van personen die niet deelnemen aan het verkeer als van smartphones die nog niet geüpdatet zijn op vlak van locatie) is groot en variabel. Bovendien is het formaat (snapshots) een beperking, want veranderingen tussen snapshots kunnen moeilijk gedetecteerd worden en de modaliteit kan daaruit ook moeilijk worden afgeleid.

Voor beide pilootsteden werd deze data aan imec overhandigd door middel van een datadump. Beide steden hadden echter wel een andere operator. Bovendien verschilde de updatefrequentie per stad (15 minuten voor Mechelen en 60 minuten voor Gent).

De interessante opportuniteiten die hieruit afgeleid kunnen worden, zijn kansen die grote telecomoperatoren moeten aannemen om hun data te verbeteren en in lijn te brengen met de behoeften van de steden. Onze aanbeveling is dan ook dat steden de kwaliteit van de data meenemen in een aanbestedingsprocedure bij een volgende aankoop van deze data. Op basis van een eerste analyse stelde Gent bijvoorbeeld al vast dat de eerste datalevering niet klopte. Na overleg met de provider werd dat echter bijgestuurd. De ervaring die is opgedaan binnen CityFlows kan dan ook helpen om het potentieel van deze databron te verbeteren voordat er wordt over gegaan tot de aankoop van dergelijke data.

Er kan bijvoorbeeld gevraagd worden om niet enkel aantallen per snapshot te leveren, maar ook om aantallen per modaliteit en per snapshot te leveren. Het is echter nog de vraag of dat technisch binnen de mogelijkheden van de telecomoperatoren valt. Er kan ook gevraagd worden naar meer inzicht in de zogenaamde “black box” die nu al achter de data zit. De steden krijgen momenteel namelijk enkel ruwe aantallen per cel en niet meer. Meer inzicht in de manier waarop de data verzameld worden en de manier waarop die geïnterpreteerd moeten worden, zal de kwaliteit en het gebruik ervan alleen maar ten goede komen.

Telraam

Telraam[3] is een Citizen Science-initiatief van het beleidsondersteunend onderzoeksbureau TML (Transport & Mobiliteit Leuven). Dat initiatief voorziet camera’s voor burgers die daarvoor betalen (127 euro op 02/03/2023) om een beeld te krijgen van de mobiliteit in hun straat. De door de burger zelf geïnstalleerde camera bevat beeldherkenningssoftware die detecteert hoeveel auto’s, fietsers en/of voetgangers er per uur passeren. TML ontwikkelde een architectuur, hardware en software voor deze sensor binnen het Europees project WeCount. Er werd een pilootstudie uitgevoerd in Kessel-Lo (Leuven) waarbij er veel Telraamsensoren gratis ter beschikking werden gesteld aan enthousiaste burgers.

- **Sterktes**

- Een lokaal beeld van verkeersvolumes (op secundaire en tertiaire wegen)
- Multimodaal

[3] https://telraam.net/nl/what-is-telraam

- Laagdrempelig en goedkoop vanuit het oogpunt van de steden, want de kost wordt gedragen door de burgers (met uitzondering van projecten waarvoor de steden de sensors kopen).
- Langdurige meting en dus een grote tijdspanne, wat deze databron geschikt maakt voor zowel use cases met kleine tijdsvensters als voor use cases met grote tijdsvensters.

➢ **Zwaktes**

- Beeldherkenningssoftware werkt niet altijd zonder problemen (dubbele telling fietsers/voetgangers).
- Bij schemering of donkere omstandigheden en 's nachts werkt de camera niet.
- Het blijft een burgerinitiatief en is dus afhankelijk van de bereidheid van burgers om de camera's te betalen, te installeren en te onderhouden (verbinding/stroomverbruik).
- De invloed die de stand van de camera (hoek tot de straat, parkeren van vrachtwagens voor de camera, invallende zon, etc.) uitoefent op de metingen is vaak onbekend.
- De camera staat op een vast punt en meet dus enkel op dat punt en niet over een grote geografische zone.

➢ **Opportuniteiten**

- Dankzij verschillende projecten kunnen deze camera's beschikbaar worden gesteld voor burgers zodat zij een constante meting kunnen opstarten en zo andere projecten kunnen ondersteunen (zoals het WeCount-project waarnaar hierboven verwezen is).
- Telraam werkt aan een verbetering van zijn sensor met een Telraam 2.0-sensor die geüpdatete beeldherkenningssoftware bevat en minder gevoelig is voor tijdelijke onderbrekingen van elektriciteit en/of internet.
- Telraam wint meer en meer aan bekendheid binnen de maatschappij, wat de bereidheid tot aankoop en installatie vergroot

➢ **Bedreigingen**

- Sterk afhankelijk van de bereidwilligheid van de bewoners om te investeren in de camera en die te blijven monitoren
- De verschillen in beschikbare Telraamsensoren per wijk/stad zijn zeer divers en bepalen in veel gevallen de functionaliteit van deze sensor voor een specifieke use case.

➢ **Conclusie**

De multimodale telling en lokale beschikbaarheid van deze sensor is een meerwaarde ten opzichte van veel andere sensoren. Het belangrijkste nadeel blijft echter de lokale context die vaak ontbreek. Dat zorgt er ook voor dat er geen algemene oplossing toegepast kan worden op alle Telraamdata. Er moet namelijk steeds lokaal nagegaan worden wat de mogelijke afwijkingen kunnen zijn, en dat zelfs voor elk tijdstip. Dat nadeel heeft een grote invloed op de nauwkeurigheid van deze sensor. Deze databron kan mogelijk verbeterd worden door bij de installatie een validatie per sensor uit te voeren. Dat kan bijvoorbeeld een kwaliteitscontrole zijn die patronen detecteren op lange en korte termijn om op die manier afwijkingen te bepalen en wiskundig te kunnen opheffen.

De Telraamdata is redelijk eenvoudig op te vragen via een publieke API die je voor elke sensor/straatsegment afzonderlijk dient op te vragen. Als je de data hebt opgevraagd, ontvang je de verwerkte data met het aantal passages per uur per modaliteit terug. Deze dataverwerking heeft reeds plaatsgevonden op de camera zelf om privacyproblemen uit de weg te gaan.

Als verbetering stellen we voor om de beeldherkenningssoftware te verbeteren en het toestel bestendiger te maken tegen tijdelijk netwerk- en stroomverlies. We zien daarvoor wel mogelijkheden in de Telraam 2.0-versie. Die moet echter wel nog uitvoerig getest worden om te kunnen bepalen of dat ook daadwerkelijk een positieve invloed heeft op de nauwkeurigheid van de resultaten. De vraag daarbij blijft ook of dat technisch gezien toch niet op beperkingen gaat botsen (slechte zichtbaarheid, ...). Ook hier zou gedacht kunnen worden aan een validatieproces per sensor. Bovendien moet er rekening gehouden worden met een overgangsperiode voordat de bestaande Telraamsensoren vervangen zijn door nieuwe exemplaren. Het gevaar dreigt daarbij dat de oude sensors evenwaardig behandeld worden als de (betere) nieuwe generatie. Daar komt bij dat het type sensor ook duidelijk weergegeven moet worden in de metadata (zie paragraaf over uitdagingen in dit rapport).

ANPR

ANPR is een afkorting die staat voor “automatic number plate recognition” (automatische nummerplaatherkenning). Sinds de terroristische aanslagen in België is deze sensor wijdverspreid in Vlaanderen. ANPR werkt via een camera die een foto maakt van de voorkant van elk voertuig en zo de passage van dat voertuig op die plaats registreert.

ANPR data werd voor de use case in Mechelen meegenomen in het CityFlows-project. Het ontsluiten van de ANPR data voor analyse doeleinden is tot op vandaag geen eenvoudige klus. Het gebruik van deze camera’s wordt geregeld door de Wet op het politieambt (WPA) en de Camerawet (21/3/2007). Het Controleorgaan op de politionele informatie (COC) kijkt toe op het naleven van deze wetgeving. Daarnaast is ook de Algemene Verordering Gegevenbescherming (AVG) van toepassing die regels op verwerking van persoonsgegevens beschrijft.

In de casus van Mechelen onderzoekt het projectteam de impact van het afsluiten van een schoolstraat. Meer specifiek wordt er onderzocht of er met de beschikbare mobiliteitsdata iets kan verteld worden over de verandering in de verkeersveiligheid, leefbaarheid en doorstroming op de straten rond de schoolstraat. Dit draagt bij tot handhaving van de openbare orde en het verzekeren van de veiligheid. Dit vormt de grond waarop de motivatie tot gebruik van deze ANPR-data stoelt.
Stad Mechelen diende een aanvraag bij het Controleorgaan op de politionele informatie COC in om de ANPR-gegevens op te vragen. De aanvraag bestond uit twee delen.

- In het eerste deel vraagt Stad Mechelen (aanvrager en primaire verwerkingsverantwoordelijke) aan de Politiezone Mechelen-Willebroek om geanonimiseerde gegevens van de ANPR-camera registraties te verkrijgen. Dit aspect valt onder de Wet op het Politieambt (WPA), meer bepaald artikel 44/11/7.
- In het tweede deel geeft Stad Mechelen de opdracht aan imec (verwerker) om de gegevensstroom van geanonimiseerde ANPR-gegevens te verwerken. Dit aspect valt onder de AVG. Daarnaast beroept Stad Mechelen zich ook op Artikel 2 uit het gemeentedecreet “De gemeenten beogen om op het lokale niveau bij te dragen tot het welzijn van de burgers en tot de duurzame ontwikkeling van het gemeentelijk gebied”. Dit is ook overeenkomstig met artikel 41 uit de Grondwet dat aangeeft dat ze bevoegd zijn voor “de aangelegenheden van gemeentelijk belang voor de verwezenlijking waarvan ze alle initiatieven kunnen nemen.”

Na goedkeuring van het COC werden de uiteindelijke gegevens uitgewisseld. In eerste instantie betreft het gegevens over plaats en tijdstip van de registratie door de camera die door de Politiezone Mechelen-Willebroek aan de stad Mechelen overgemaakt worden. Er wordt geen verrijking gemaakt met de databank van Dienst Inschrijving Voertuigen (DIV) en ook geen gepseudonimiseerde gegevens om voertuigen over meerdere camera’s te volgen. In een tweede instantie worden deze gegevens overgedragen aan imec (verwerker) die als eerste actie een aggregatie op hoger niveau doorvoert in lijn met bepalingen in het AVG (minimum 5 voertuigen per record). Door de finale aggregatie van de verschillende mobiliteitsdata is resulterende datastroom niet langer onderhevig aan de WPA en/of AVG.

In dit proces is een nauwe samenwerking tussen de verschillende actoren belangrijk. In de casus van stad Mechelen was de mobilitieitsdienst de initiatiefnemer om de ANPR data te gebruiken. Samen met de verwerker imec werd er ingeschat of deze databron een toegevoegde waarde kon brengen in de analyse. De Data Protection Officer (DPO) van de stad

evalueerde mee of de juiste evaluaties en stappen inzake gegevensverwerking gevolgd werden. Er werd ook een nauwe relatie opgezet met de Politiezone Mechelen-Willebroek omdat de verwerking van de ruwe gegevens (i.e. camerabeelden omzetten in records) in eerste instantie door hen diende opgenomen te worden.

- **Sterktes**

- Zeer nauwkeurige sensor die zo goed als al het gemotoriseerde verkeer meet.
- Link met voertuigklasse/motortype is in principe mogelijk (hoewel dat niet in de scope van CityFlows zat), dit houdt een koppeling in met DIV databank.
- Reeds op grote schaal uitgerold en beschikbaar in Vlaanderen.

- **Zwaktes**

- Wetgeving bemoeilijkt wijdverspreid gebruik van deze sensor voor analyse doeleinden. In deze casus diende een duidelijke link met de handhaving van openbare orde en verzekeren van de veiligheid aangetoond te worden.
- Niet multimodaal: enkel gemotoriseerd verkeer wordt gedetecteerd. Nieuwe generatie camera's zijn in staat zijn meerdere modaliteiten te identificeren.
- Camera staat op één vast punt en meet dus ook enkel op dat vaste punt en geeft geen data over een grote geografische zone.

- **Opportuniteiten**

- Indien het wetgevend kader zou gevormd worden (dat ANPR-data beschikbaar maakt voor analyse doeleinden) dan dienen ook de achterliggende diensten, die instaan voor verwerking van de ruwe data, ondersteund te worden. De huidige technologie staat vandaag ver genoeg om deze verwerking procedures in hoge mate te automatiseren. Dit zou een volgende stap kunnen zijn in de digitale transformatie van de publieke diensten met het oog op betere dienstverlening.
- In financieel moeilijke tijden is het voor steden en gemeenten belangrijk om bestaande infrastructuur optimaal te benutten. Die politieke wil zorgt ervoor dat er nu een uitgelezen kans is om het gebruik van deze data in mobiliteitstoepassingen in een wettelijk kader te plaatsen en al doende de beschikbaarheid van deze data te verbeteren.
- Deze sensor staat op veel plaatsen en als deze data beschikbaar wordt gesteld, is die daardoor geschikt voor verschillende use cases.

- **Bedreigingen**

- Het aanslepende debat rond het gebruik van ANPR data voor analyse doeleinden (dat aanpassingen aan huidige wetgevend kader vraagt) zorgt ervoor dat het geloof dat deze data beschikbaar gesteld kan worden, aan het afbrokkelen is.

- **Conclusie**

De sterkte van deze sensor is vooral de grote nauwkeurigheid in het meten van al het gemotoriseerde verkeer. Die sterkte is echter ook meteen het grootste nadeel: er is namelijk geen meting van niet-gemotoriseerde transportmodi. Bovendien zorgt de huidige wetgeving ervoor dat deze databron op het moment nog niet eenvoudig benut kan worden. Wij adviseren dan ook aan om die wetgeving zo spoedig mogelijk te herzien zodat er een duidelijk afgebakend toepassingsgebied kan komen voor deze sensor. Nu moet er per initiatief bekeken worden waar deze sensor wel en niet voor gebruikt mag worden. Uit de use case van CityFlows in Mechelen is gebleken dat dat veel tijd vraagt.

Fietstelpalen

Deze databron bevat data die komt uit een tellus, een geleider die in het fietspad of wegdek geslepen is. Wanneer die geleider onder stroom gezet wordt, genereert die een magnetisch veld. Wanneer er vervolgens magnetisch materiaal (een fiets) door en over dat magnetische veld rijdt, wordt het gegenereerde magnetische veld verstoord en wordt de stroom in de geleider beïnvloed. De verwerking van deze stroomveranderingen in het aantal fietsers wordt lokaal uitgevoerd. De provincie Oost-Vlaanderen was de eigenaar van de fietstelpaal in Gent en stelde de data van deze fietstelpaal ter beschikking via een datadump. Andere instanties van fietstelpalen zoals Signco en AWV stelde hun data ter beschikking via een openbare API. In Mechelen werd er bijvoorbeeld gebruik gemaakt van een fietstelpaal van Signco.

- **Sterktes**

- De nauwkeurigheid is relatief voldoende.
- De uitrol in Vlaanderen is aan de gang.

- **Zwaktes**

- Enkel fietsers worden gemeten.
- Aangezien carbonfietsen geen magnetisch materiaal hebben, worden die niet gemeten.
- Er is nog geen eenduidige standaard voor data uit fietstelpalen.
- De paal staat op een vast punt en meet dus enkel op dat punt en geeft dus geen data voor een grote geografische zone.
- De kostprijs is nog relatief hoog, wat een verdere geografische verspreiding beperkt.

- **Opportuniteiten**

- Dankzij de ervaringen in CityFlows en andere projecten is het OSLO-traject voor verkeersmetingen gestart om datastromen uit verkeerssensoren semantisch te modelleren en de structuur van die data te standaardiseren[4]. Dat traject werd opgestart binnen het relanceproject “Data Integratiediensten voor Slimme Mobiliteit” (DIM) waarvoor imec de geleerde lessen van CityFlows ingebracht.
- De uitrol van fietssnelwegen en het groeiende belang voor de fietsers zorgen ervoor dat het aantal meetpunten toeneemt.
- Als alle organisaties die deze palen plaatsen, de data die uit die palen voortkomt, ook zouden delen (na anonimisatie), zou de geografische dekking van deze databron verbeteren.

- **Bedreigingen**

- Er is geen classificatie mogelijk tussen e-bike/speedpedelec of een gewone fiets.
- Enkel het aantal fietsers wordt gemeten.
- Bij afhankelijkheid van een API moet de voorbewerking van de data herbekeken worden wanneer er wijzigingen zijn aan de standaarden van de API.

[4] https://www.vlaanderen.be/digitaal-vlaanderen/agenda/evenementen/oslo-verkeersmetingen

Figuur 2: Fietstelpaal van AWV in Leuven

- **<u>Conclusie</u>**

De conclusies die uit de analyse van deze databron getrokken kunnen worden, zijn te vergelijken met de conclusies over de ANPR-camera. Ook deze databron meet slechts één modaliteit, niettemin met een aanvaardbare nauwkeurigheid. Het probleem is echter dat zowel AWV als de provincies voorlopig werken volgens de standaard die CROW Fietsberaad samen met UGent heeft opgebouwd. Die standaard wordt (nog) niet overal gebruikt en is ook (nog) niet vertaald in een OSLO-standaard. Op basis van de ervaring die is opgedaan in CityFlows, gaat imec samen met Digitaal Vlaanderen echter wel OSLO-verkeersmetingen opstellen waarin dat wordt meegenomen. Ons voorstel te verbetering van deze databron is dan ook het OSLO-traject waaraan imec actief meewerkt als deel van het kernteam (zie ook deel 3 van dit rapport voor meer informatie over het OSLO-traject).

Voor de use case in Gent is de data van deze soort sensoren ter beschikking gesteld via een datadump. Voor de fietspaal van Signco in Mechelen werd die als open data (via API) afgehaald. Voor telpalen van andere leveranciers (de provincies of andere spelers) moest er aparte toegang verkregen worden. Het zou een positieve verandering zijn als het verlenen van die toegang geoptimaliseerd kan worden (bijvoorbeeld met behulp van dataspaces of een andere technologie voor datadeling). Met de kennis die is opgedaan binnen het CityFlows-project is imec dan ook in het VSDS-project (Vlaamse Smart Data Space) en het CitCom.AI-project gestapt. Daarover volgt later in dit document meer.

Tijdelijke meetinstallaties

Tijdelijke meetcampagnes werden in Gent en Mechelen uitgevoerd door telslangen. In Gent werd gewerkt met een onderleverancier die de telslangen plaatste. In Mechelen werden er enkele door de politie gelegd voor een specifieke use case (politierapporten genoemd). Er zijn zowel fietstelslangen als algemene telslangen geplaatst. Telslangen werken op een gelijkaardige manier als tellussen, maar de geleider wordt bij telslangen niet in het wegdek geslepen, maar tijdelijk in een rubberen "slang" over het wegdek gelegd (zie Figuur 3).

Figuur 3: Foto van een telslang die gefocust ligt op een fietspad

De SWOT-analyse is vrij gelijkaardig voor de fietstelslangen en de algemene telslangen en daarom worden die samen geanalyseerd.

- **Sterktes**

- Relatief hoge nauwkeurigheid
- Relatief eenvoudige installatie met minimale schade aan wegdek en zonder invloed op de doorstroming

- **Zwaktes**

- Weggebruikers proberen dergelijke installaties vaak te vermijden als ze zien liggen, wat kan leiden tot onnauwkeurige tellingen.
- Carbonfietsen worden niet gedetecteerd.
- Tijdelijk van aard, en daardoor wordt het door seizoenseffecten moeilijk om trends op lange termijn te zien.
- Er bestaat geen echte standaard voor deze metingen (op basis van de geleerde lessen van CityFlows is het OSLO-traject voor verkeersmetingen opgestart).
- Bij remmend/stilstaand verkeer (file) is de werking van de lus niet optimaal.

- **Opportuniteiten**

- Op basis van de ervaringen van CityFlows en andere projecten is het OSLO-traject voor verkeersmetingen opgestart om de semantiek rond data eenduidig te krijgen.
- Deze data blijft vaak bij één lokaal bestuur en lokale bevinden worden niet gedeeld buiten de grenzen van de steden. Er is dus nog ruimte om die data- en kennisdeling te optimaliseren aan de hand van een dataspace/samenwerking.

- **Bedreigingen**

- Classificatie tussen soorten fietsers blijft moeilijk.
- Classificatie tussen wagen/bestelwagen/vrachtwagen bij traag, accelererend of remmend verkeer is niet

nauwkeurig.

- **Conclusie**

In het algemeen werd de data uit deze tijdelijke meetcampagnes binnen CityFlows als zeer waardevol gezien, omdat deze databron door de relatief hoge nauwkeurigheid als “ground truth” beschouwd kan worden en daardoor geschikt is om te vergelijken met andere databronnen. Het is aan te bevelen om deze data beter te delen buiten de grenzen van steden. Een dataspace zoals de Vlaamse Smart Data Space (VSDS) kan een mogelijke enabler zijn om dat te bereiken. Ook zouden alle vormen van tijdelijke telcampagnes een eenduidige beschrijving moeten hebben. Op basis van deze aanbevelingen heeft imec het al eerdergenoemde OSLO-traject voor verkeersmetingen opgestart.

Voor het CityFlows-project werd de data op verschillende manieren ter beschikking gesteld. Sommige leveranciers hadden daarvoor een API voorzien waarvan de data afgehaald kon worden. De politierapporten in Mechelen werden via een datadump beschikbaar gesteld.

Straatvinken

Straatvinken[5] is een burgerinitiatief dat jaarlijks burgers tellingen gedurende één uur (van 17.00 uur tot 18.00 uur) laat uitvoeren. De tellingen worden elk jaar rond hetzelfde moment uitgevoerd (in de tweede helft van mei) om ervoor te zorgen dat die statistisch gezien correct zijn. Burgers krijgen de kans om het voorbijkomende verkeer te tellen met een applicatie. Via die applicatie worden hun tellingen vervolgens verwerkt.

- **Sterktes**

- Relatief hoge nauwkeurigheid die kan dienen als ground truth
- Lage kostprijs
- Multimodaal beeld
- Kwaliteitscontrole op de tellingen door Ringland Academie[6] , Universiteit Antwerpen[7] en HIVA - KU Leuven[8]

- **Zwaktes**

- Confirmation bias speelt mogelijk bij burgers (ik zie wat ik wil zien en/of ik wil mijn gelijk bewijzen).
- Beperkt in tijd (slechts één uur op jaarbasis), dus beperkt bruikbaar voor kleinere tijdsvensters.
- De burger maakt mogelijk fouten als hij één uur lang gefocust moet blijven en daarnaast is het in een drukke straat soms moeilijk om het verkeer goed te blijven volgen en tellen. Aangezien elke telling achteraf nog geanalyseerd wordt, blijft deze foutgevoeligheid echter beperkt.

- **Opportuniteiten**

- De geografische dekking is relatief gezien voldoende (het initiatief is relatief wijdverspreid en bekend).
- Ondersteuning door universiteiten zorgt ervoor dat deze databron voldoende gecontroleerd wordt. Een transparante communicatie daarover kan het vertrouwen in de data mogelijks nog doen toenemen.

- **Bedreigingen**

[5] Wat is Straatvinken, hoe werkt het, wat is het telgebied en wie zit erachter?
[6] Ontdek de Ringland Academie | Ringland
[7] Universiteit Antwerpen | Universiteit Antwerpen (uantwerpen.be)
[8] https://hiva.kuleuven.be/en

- Deze data is mogelijk foutgevoelig, omdat de aandacht van burgers kan verslappen als ze één uur lang achter elkaar volledig geconcentreerd moeten tellen.
- Vrijwillige participatie van burgers is noodzakelijk. Het blijft daardoor onzeker of er een engagement is voor meerdere jaren zodat de lange termijn in kaart gebracht kan worden.

➢ **Conclusie**

De conclusies die we kunnen trekken uit de analyse van deze databron is beperkter in vergelijking met de andere databronnen, omdat de databron beperkter is in tijd ten opzichte van de andere databronnen. Het nadeel van deze databron is dat het gevoelig is voor menselijke fouten, ondanks controle achteraf. Daarover dient dan ook duidelijk gecommuniceerd te worden.

Het is aan te raden om deze tellingen vooral te laten uitvoeren door burgers met een Telraamsensor. De toegevoegde waarde van een dergelijke sensor is de creatie van een lokaal kalibratiepunt. De vraag is echter of burgers die een Telraamsensor in hun bezit hebben, de extra inspanning willen en kunnen leveren om een dergelijke telling binnen dit initiatief uit te voeren. Daarnaast is het ook aan te bevelen om de telling twee keer per jaar te organiseren om meer meetpunten te hebben aangezien één uur per jaar te beperkt is.

Floating car data

Floating car data is data van weggebruikers die gebruikmaken van een navigatieapp (op hun smartphone) tijdens hun verplaatsing. De data is goed te gebruiken om te weten hoeveel wagens naar een stad rijden en via welke wegen ze dat doen. Bij meer lokale use cases (bijvoorbeeld bij metingen in een schoolstraat of bij het meten van lokaal verkeer) is deze data echter vaak beperkt bruikbaar aangezien mensen voor kleine, lokale verplaatsingen vaak geen navigatie gebruiken. Hoewel het in theorie mogelijk is om op dezelfde manier data te verzamen van actieve weggebruikers (fietsers en voetgangers), is dat in de praktijk onuitvoerbaar omdat navigatieapps door deze weggebruikers beperkt of zelfs niet gebruikt worden. Als deze soort data van fietsers komt, wordt er gesproken over floating bike data. Voor het CityFlows-project werd uitsluitend floating car data (FCD) voor Mechelen ter beschikking gesteld.

➢ **Sterktes**

- Multimodaal verkeer is in theorie mogelijk.
- Het is mogelijk om een heel traject te volgen; er is dus geen beperkte puntmeting.
- Geografische dekking is direct op stadsniveau.

➢ **Zwaktes**

- Het aantal gebruikers van dergelijke navigatieapps is voor lokaal verkeer zeer laag. Voor de schoolstraat in Mechelen was er vaak geen enkel datapunt beschikbaar.
- In de praktijk worden dergelijke navigatieapps niet gebruikt door actieve weggebruikers.

➢ **Opportuniteiten**

- De uitrol van intelligente verkeerslichten (iVRI) binnen het Mobilidata-programma zal het gebruik van dergelijke navigatieapps wellicht kunnen aanmoedigen.
- Geconnecteerde wagens delen veel meer data waardoor het in de toekomst interessanter kan zijn om gebruik te maken van deze databron.

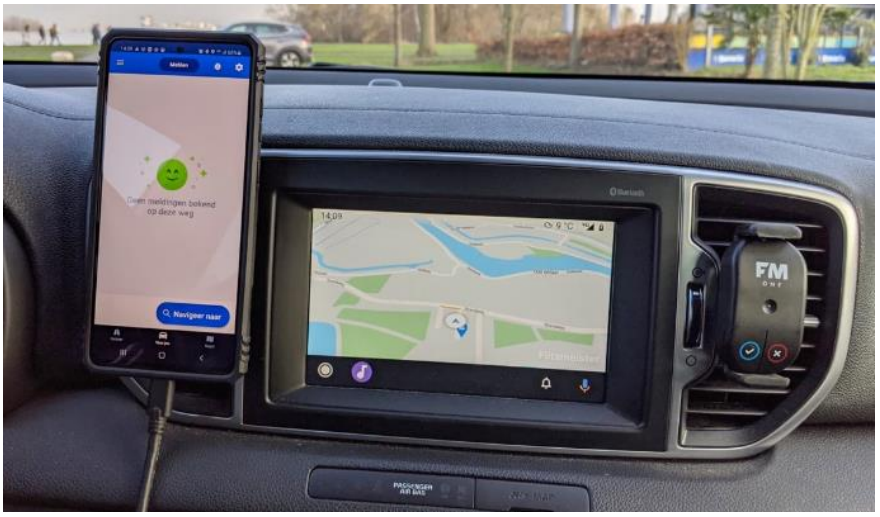
- **Bedreigingen**

- Steden nemen vaak bij één leverancier af en zijn dus ook afhankelijk van het aantal gebruikers van één app.

- **Conclusie**

De meerwaarde van deze data is de hogere geografische dekking doordat de data trajectdata bevat en niet enkel data van een lokale sensor. Het nadeel van deze data is de zeer beperkte penetratiegraad. Op primaire wegen is die penetratiegraad nog relatief hoog, maar op secundaire en tertiaire wegen is die vaak toch zeer beperkt. Voor lokale use cases zoals de evaluatie van een wijkmobiliteitsplan of het invoeren van een schoolstraat, is deze databron minder geschikt, omdat weggebruikers zelden tot nooit gebruikmaken van een navigatieapp voor wegen die ze al goed kennen. Bovendien worden dergelijke navigatieapps door actieve weggebruikers zelden gebruikt. In de praktijk is deze databron dus enkel geschikt voor gemotoriseerd verkeer.

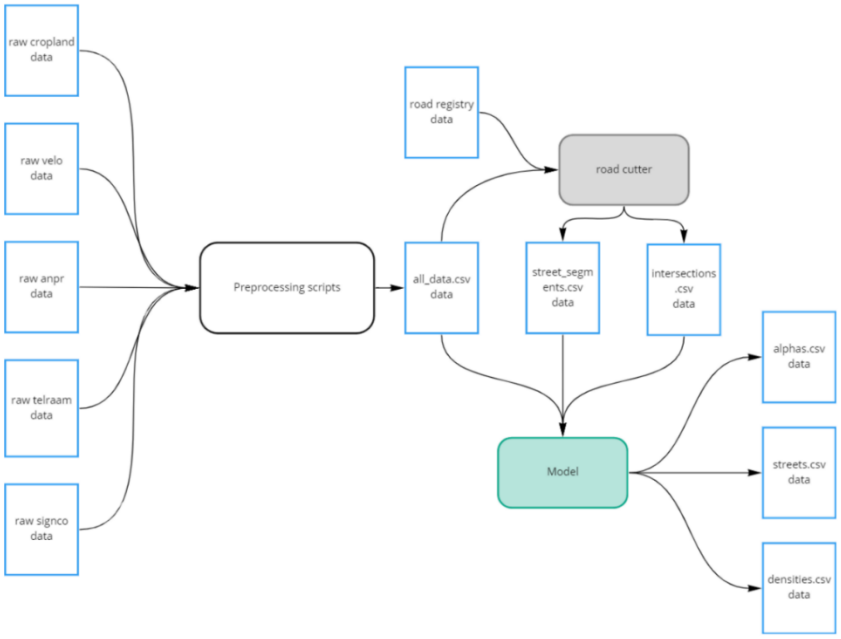
Deze databron werd enkel gebruikt in Mechelen. Er werd gewerkt met leverancier Be-Mobile. Deze databron is doorgaans relatief duur vergeleken met de mogelijke functionaliteiten die de databron oplevert (althans binnen de context van het CityFlows-project). De data werd ter beschikking gesteld via een datadump (een statisch CSV-bestand dat doorgestuurd werd).



Figuur 4: Een navigatieapp die gebruikt wordt in een wagen en die op deze manier FCD genereert

Datafusiemodel

In dit deel wordt beschreven hoe er vanuit de data, de dataverwerking en het model naar de output gegaan wordt. Aangezien dit deel technischer van aard is, wordt in dit deel soms verwezen naar een PowerPointpresentatie. Met beelden kunnen zaken soms namelijk veel beter overgebracht worden dan met een lange tekst. De algemene dataflow is te zien in Figuur 5. Vanuit die dataflow is dit deel van het rapport opgebouwd. Deze flow wordt in de volgende delen van de tekst verklaard.



Figuur 5: Dataflow van het CityFlows-project

Dataverwerking

In het eerste blok van de dataflow worden de datadumps of API-bevragingen uit het eerste deel van dit rapport gecontroleerd op hun kwaliteit. Er wordt bekeken hoe het signaal van de sensoren zich gedraagt om eventuele gaten of hiaten in de data te kunnen signaleren. De meetfrequentie en opbouw van de data wordt ook in deze fase gecontroleerd. Het formaat waarin de data binnenkomt, wordt getransformeerd naar een uniform formaat dat specifiek is voor het CityFlows-model. Als elke bron individueel bewerkt is tot het gewenste formaat, wordt alle data samengevoegd in één bestand (het “all_data.csv”-bestand in Figuur 5). Tabel 1 is toegevoegd om een overzicht te geven van de plaats van elke databron en het processing script.

Tabel 1 : Databronnen met de contactpersoon per pilootstad

Databron + verwerking	***Stad + contactpersoon***
Telecomdata (Mechelen)	Mechelen: Dienst strategie en ontwikkeling
Telecomdata (Gent)	Gent: Dienst data en informatie
Telraam	Mechelen en Gent: Dienst mobiliteit
ANPR	Mechelen: Dienst strategie en ontwikkeling
Fietstelpalen	Mechelen en Gent: Dienst mobiliteit
Tijdelijke metingen	Gent: Dienst mobiliteit
Politierapport	Mechelen: Dienst strategie en ontwikkeling
Floating car data	Mechelen: Dienst strategie en ontwikkeling

Het handling script voor een bepaalde databron begint telkens met het opkuisen van lege velden en het verwijderen van onnodige of lege kolommen. Voor een databron die voor een punt intensiteiten (aantal passages per tijdseenheid)

teruggeeft, is de volgende stap om deze intensiteiten om te zetten naar dichtheden (het aantal weggebruikers per lengte-eenheid). Hiervoor geldt de volgende formule:

dichtheid = aantal passages / (tijdsperiode * snelheid)

Voor deze formule is het nodig om het gegeven "snelheid" te kennen. Als die niet gekend is (omdat de databron dat niet meet), wordt er een bepaalde snelheid aangenomen. Voor databronnen die de data direct in een snapshotformaat aanleveren, is dat uiteraard niet nodig.

Als meettoestellen een puntmeting uitvoeren, moet het punt gelinkt worden aan het corresponderende straatsegment in het Vlaamse wegenregister. Als de meting uitgevoerd wordt voor een groter geografisch gebied, kan dat uiteraard niet gelinkt worden aan een straat, en zal het gehele gebied beschreven moeten worden. In dat geval beschrijft de kolom voor de geometrie een oppervlak en geen lijn (zie paragraaf over input van het model in dit rapport).

In de laatste stap krijgen alle kolommen een andere naam en worden ze gestructureerd tot ze overeenkomen met het formaat van het "all_data.csv"-bestand (zoals beschreven is in sectie 2.2.1).

Model

Er werd een model ontwikkeld dat data van verschillende databronnen benut om een beeld te geven van de dichtheid van verschillende modaliteiten. Het ontwikkelende model bouwt verder op het rapport "Data fusion model for CityFlows" van de Universiteit Antwerpen. Dat model werd ontwikkeld in een PoC in de smart zone van Antwerpen. In de smart zone (waar veel databronnen aanwezig waren) werkte het model voor twee modaliteiten (wel en niet-gemotoriseerd). Binnen dat traject werd onderzocht of het model ook op grotere schaal (stadsniveau) en voor meer modaliteiten kan werken.

De input van het model

Het model heeft in essentie maar drie elementen nodig om te werken:

- "street_segments.csv" – Straatinformatie
- "intersections.csv" - Kruispuntinformatie
- "all_data.csv" - Data met de tellingen van de verschillende gegevensbronnen

Het model dient die input aan te krijgen als CSV-bestanden met een specifieke structuur. Op GitHub staat ook een voorbeeld van de roadcutter procedure. Dat is een procedure om de eerste twee bestanden (met straat- en kruispuntinformatie) te creëren. Als basis zal daarvoor het Vlaamse Wegenregister gebruikt worden. De roadcutter heeft als input ook het "all_data"-bestand nodig.

Voor elke beschikbare meting van elk van de niet-weerhouden databronnen is er een lijn in het databestand. Die lijn bevat ook informatie over de regio van de meting. Dat is belangrijk voor de roadcutter procedure. Daarnaast staan er in de lijn nog het tijdstip en enkele indexen.

De gewichten van de gewogen verdeling is optioneel. Die input is ook een CSV-bestand. De structuur van dat bestand is simpel: elke straat heeft een lijn en in de kolommen staan de gewichten per modaliteit. Als deze optionele input niet ingegeven wordt in het model zullen alle gewichten gelijkgesteld worden aan 1 en zal het model in feite dus een uniforme verdeling maken.

De output van het model

De outputdata van CityFlows kan worden aangeleverd in het NGSI-LD formaat TrafficFlowObserved (zie deze link voor

meer details). Enkele attributen zijn daaraan toegevoegd om tot een werkbare standaard te komen voor CityFlows. Het bovenstaande dataformaat staat het toe om data in te geven:

- voor elk straatsegment in de afgebakende zone;
- voor elke vijf minuten - op voorwaarde dat er voldoende databronnen bestaan die deze mate van detail kunnen aanleveren;
- voor elke modaliteit - indien er voldoende sensoren zijn die de betrokken modaliteiten meten. In het kader van de voorgestelde scope zullen dat de volgende gemodelleerde modaliteiten zijn: voetgangers, fietsers, gemotoriseerd verkeer en de achtergrond (de personen die geteld zijn door telecomdata maar die niet actief deelnemen aan het verkeer).

Het model bepaalt de verkeersdichtheid (het resultaat van de modelberekening) voor elk van de bovenstaande gegevens. Daarmee kan men vervolgens de use cases per stad evalueren.

Datafusie

Het model is een datafusiemodel dat zo min mogelijk effecten of menselijk gedrag opgelegd kreeg om ervoor te zorgen dat het gedrag enkel in kaart gebracht zou worden door de data. Zonder a priori kennis van het wegennet bleek dat echter niet mogelijk. De wiskundige insteek bleek onvoldoende om het menselijke gedrag weer te geven. Het wiskundige kwadratische optimaliseringsprobleem[9] is hier beschreven. Die uitleg is gebaseerd op de eerder vermelde paper.

Stapsgewijs werd er meer data aan het model toegevoegd om de hypothese dat het model betrouwbaarder zou worden naarmate er meer databronnen aan toegevoegd zouden worden, te kunnen bewijzen. Zoals eerder aangegeven, waren de resultaten daarvan teleurstellend.

Nog een ander effect was zichtbaar. De uniforme verdeling die gebruikt werd, was nog te zien in de output. Daarom werd er gezocht naar een gewogen verdeling (zie paragraaf over gewogen verdeling in dit rapport). Die gewogen verdeling gaf visueel een beter beeld, maar zorgde uiteindelijk niet voor een verbetering van de resultaten op zich.

Gewogen verdeling

In eerste instantie probeerde het model ervoor te zorgen dat elke straat in de gegevensbroncel (gebied van deze gegevensbron) een dichtheid toegekend werd die dicht bij de gemiddelde dichtheid lag. Dat zorgde ervoor dat resultaten met gelijkmatige verdelingen op straten in dezelfde gegevensbron cel voorgetrokken werden, waardoor er zones van constante dichtheid ontstonden.
Om dat te vermijden werd de uniforme verdeling (die zo dicht mogelijk bij de gemiddelde dichtheid lag) overboord gegooid en werd er plaatsgemaakt voor een gewogen verdeling om te kunnen “fitten”. Elke straat krijgt per modaliteit een uniek gewicht toegekend dat de waarschijnlijkheid dat die modaliteit aanwezig is in het straatsegment reflecteert. Een straat met breed fietspad heeft bijvoorbeeld een zwaarder gewicht voor de modaliteit fietsers dan een straat zonder fietspad. Deze gewichten werden bepaald aan de hand van een machine learning model. De straatvinken dataset werd gebruikt als trainingsset. Door een grote hoeveelheid aan data te linken aan de straatvinken dataset kan het machine learning model een predictie doen voor straten waar geen data voor aanwezig is. Op deze manier wordt een exhaustieve set van gewichten voor het hele stratennetwerk gecreëerd voor de drie modaliteiten. Om dit te bereiken was er een grote hoeveelheid data nodig. Het exploreren en begrijpen van de gebruikte datasets werd daardoor behoorlijk arbeidsintensief. Dat proces is samengevat in een analytical base table (zie bijlage 1).

Voordat er een model getraind kon worden, moesten de data begrijpelijk worden gemaakt en feature engineering (het proces om met domeinkennis bepaalde features (karakteristieken, eigenschappen en attributen) te extraheren van ruwe data) toegepast worden. Dat werd niet gedaan voor één model; voor elke modaliteit van de Straatvinkendataset werden er vier modellen getest. De uitkomst van de oefening bleek niet zonder gebreken en werd achteraf nog lichtjes bijgeschaafd

[9] Optimaliseringsprobleem - Wikipedia

(zie bijlage 2).

Figuur 6 geeft de evolutie van de modeloutput weer voor de gemotoriseerde modaliteit. De drie varianten van het model laten (van links naar rechts) de output zien als er gebruik gemaakt werd van de uniforme verdeling, van de gewogen verdeling en van de aangepaste gewogen verdeling.

Figuur 6: Verandering in de verdeling van de modaliteiten door weights

Op de eerste afbeelding zien we de uniforme verdeling duidelijk terugkomen in de rechthoekige broncellen van de telecomgegevens. In het tweede beeld is de uniforme verdeling vervangen door de gewogen verdeling en lijkt er al een correcter beeld te ontstaan. De hoge dichtheid aan auto's bij de zoo en het stadspark kan echter niet juist zijn. Daarom is ook de gewogen verdeling uiteindelijk aangepast, wat resulteerde in de derde en laatste afbeelding. Daarop is te zien dat er vooral op drukkere straten een hoge dichtheid aan auto's is en dat er in de parken een lage dichtheid aan auto's is. Een hoge dichtheid aan auto's vooral te zien op drukkere straten.

Het validatieproces

In het validatieproces (zie bijlage 3) voor het model werden de data van Straatvinken gebruikt als ground truth.

Het model geeft zijn uitkomsten weer in dichtheden. Om de dichtheden van het model te kunnen vergelijken met die van de Straatvinkendataset moet het resultaat van die dataset eerst omgezet worden van intensiteiten (het aantal personen per tijdseenheid) naar dichtheden (het aantal personen per meter) volgens de formule in sectie 2.1. De Straatvinkentelling houdt het gegeven "snelheid" echter niet bij. Aangezien dat gegeven noodzakelijk is voor het omzetten van de data wordt er een bepaalde snelheid verondersteld (zoals uitgelegd in sectie 2.1). Om één enkele validatiescore te krijgen, wordt de gewogen som genomen met de relatieve straatlengtes als gewichten. De validatiescore die daaruit volgt, is de gewogen verhouding van het aantal straten waarvan het resultaat ligt in een geaccepteerde zone (tussen de 66% en 150% van de Straatvinkendata). Zoals hierboven reeds vermeld zijn er stapsgewijs extra databronnen aan het model toegevoegd. Een dergelijke stap wordt aangeduid met een "learning cycle". In Tabel 2 staan de validatiescores van de verschillende learning cycles en van de gewone en de aangepaste gewogen verdeling.

Tabel 2: Validatiescores van het CityFlows-model

Versie	Gebruikte gegevensbronnen in het model	Validatiescore [%]
LC 1	Cropland (telco), Fietsdata	4,435
LC 2	Cropland (telco), Fietsdata + Telraam, Signco	4,462
LC 3	Cropland (telco), Fietsdata, Telraam + ANPR	4,446
Gewogen verdeling	LC 3 + ML gewichten	2,749
Aangepaste gewogen verdeling	LC 3 + aangepaste ML gewichten	0,971

Uit de tabel blijkt dat het model een validatiescore heeft die grofweg tussen de 1% en 5% accuraatheid varieert, afhankelijk van het aantal gebruikte databronnen. Volgens de validatiehypothese levert het model dus in 1 tot 5% van de berekende dichtheden op straatniveau een betrouwbaar resultaat, een zeer laag resultaat. De hypothese dat de score zou stijgen zodra er meer bronnen toegevoegd zouden worden, is bij dezen dus verworpen. Ook de gewogen verdelingen leidden niet tot een verbetering van de validatiescore. Dit suggereert ook dat het design van het CityFlows validatieproces niet exact meet wat het zou moeten meten.

De resultaten van het model

In het rapport “Data model for CityFlows” van de Universiteit Antwerpen werd verondersteld dat het model dat ontwikkeld was voor de smart zone van Antwerpen, op grotere schaal en met meer modaliteiten ontwikkeld kon worden. In het rapport werd verder aangegeven dat de validatie door imec nog uitgevoerd moest worden. Voor het controleren van de output van het model ontwikkelde imec een methode (zie paragraaf over het validatieproces in dit rapport). De validatiescore (het resultaat van het validatieproces van imec) was laag en kon niet verbeterd worden door meerdere databronnen aan het model toe te voegen. Ook het toevoegen van de gewogen verdeling leverde geen betere validatiescore op (zie paragraaf over gewogen verdeling in dit rapport).

In het rapport dat de basis vormde voor het CityFlows-project werd de hypothese naar voren gebracht dat andere databronnen toegevoegd konden worden aan het model. Die hypothese bleek te kloppen. Het is inderdaad mogelijk gebleken om databronnen te gebruiken die oorspronkelijk niet in het model zaten. Daarbij moet wel opgemerkt worden dat het belangrijk is om de context van de meettoestellen goed te kennen. De locatie van een ANPR-camera is bijvoorbeeld onvoldoende om te weten wat die camera exact meet. Het bemachtigen, uniformiseren en analyseren van de data vergde meer werk dan oorspronkelijk was voorzien. Het werk dat verzet is, is echter wel herbruikbaar voor verdere analyses van de besproken bronnen. Het kan ook gebruikt worden als een voorbeeld voor eventuele nieuwe gegevensbronnen in de toekomst.

Het kan gezegd worden dat het model vollediger is geworden. Het functioneert op stadsniveau en werkt met meerdere databronnen. Hoewel de input van de verschillende bronnen goed gefit worden, is de output van het model gelet op de lage validatiescores toch onbetrouwbaar. In het onderstaande gedeelte van dit rapport wordt er dieper ingegaan op mogelijke voorstellen voor verbetering.

Uitdagingen

- ***Uitdaging 1: Dataverzameling***

Het bemachtigen van de data in de verschillende steden duurde uiteindelijk drie keer langer dan vooraf voorzien was. Er

moest met elke stad een aparte overeenkomst opgesteld worden om op een juridisch correcte manier toegang te krijgen tot specifieke databronnen en om die vervolgens te mogen gebruiken. De stad blijft immers eigenaar van de data en de databronnen; imec verwerkt in opdracht van de steden slechts de data. Daarenboven kwam bij aanvang van het project naar boven dat de data vaak van een onvoldoende hoge kwaliteit was. Zo vertoonden de data van fietspalen fouten, waren tijdelijke tellingen soms foutief geïnstalleerd en moesten zeer vreemde telecomdata gecorrigeerd worden. Alles bijeengenomen nam het veel tijd in beslag om dat op te helderen met de leverancier.

Voor de ANPR-data te Mechelen diende ook een DPIA (Data Protection Impact Assessment) opgesteld worden, die redelijk veel tijd en inspanning vereiste. Uiteindelijk werd deze databron wel toegankelijk (zie paragraaf over ANPR in dit rapport) onder strikte anonimisatie (enkel aantallen per tijdsinterval).

- ***Oplossing***

Een datacatalogus kan de stad helpen om gerichte acties te ondernemen rond data. Bij iedere dataset zou een soort vragenlijst ingevuld kunnen worden die beschrijft wat de data inhoudt, wat er gemeten wordt, hoe frequent en kwaliteitsvol die metingen gebeuren, etc. Een dergelijke lijst kan ervoor zorgen dat er snel beoordeeld kan worden of data al dan niet ingezet kan worden in het veld. Op die manier is het niet nodig om telkens een moeizame zoektocht naar mogelijke data op poten te zetten.

Daarnaast zouden er in het kader van innovatieve projecten regulatory sandboxes[10] opgezet kunnen worden. Zulke sandboxes creëren speelruimte om het juridische kader helder te krijgen en te zoeken naar optimalisatie. Dat kan het delen van data bevorderen en ervoor zorgen dat alle betrokken partijen op de hoogte zijn van elkaars uitdagingen.

- ***Uitdaging 2: Het formaat van de databronnen***

In sectie 1 werd het al aangekaart dat elke gebruikte databron een ander formaat, tijdsinterval of semantiek had. Het lijkt vanzelfsprekend dat een fietser hetzelfde is voor een tijdelijke fietsslang, een fietstelpaal, een telraam en voor floating bike data, maar de semantiek van zulke functionele zaken ontbreekt vaak en zorgt ervoor dat een eenduidige datafusie onmogelijk is. Het is noodzakelijk om zeker te weten dat beide databronnen dezelfde “taal” spreken en zaken als “fietsers” op dezelfde manier aanduiden. Het is anders onmogelijk om alle data rond fietsers samen te brengen.

Ook het verschil in datafrequentie (per kwartier, per uur, ...) van de verschillende databronnen vormt een belemmering voor een correcte datafusie. Dat verschil zorgde er namelijk voor dat er zaken gesplitst of verondersteld moesten worden, wat resulteerde in onnauwkeurigheden. Ook dat heeft wellicht bijgedragen aan de lage betrouwbaarheid van het model.

- ***Oplossing***

De geleerde lessen uit bovengenoemde uitdagingen zijn meegenomen in de Vlaamse Smart Data Space[11] (VSDS) en het project “Data Integratiediensten voor Slimme Mobiliteit” (DIM), dat imec en Digitaal Vlaanderen samen uitvoeren. Door de VSDS zal data eenvoudig gedeeld kunnen worden op een dataspace. Bovendien is een dataspace een decentraal verzamelpunt waar toegang tot specifieke datasets makkelijk beheerd kan worden. Privacygevoelige data kunnen op die manier dus perfect beheerd worden. Het is op een dataspace ook mogelijk om voor specifieke data als floating car data een bepaald bedrag te vragen. Een dataspace kan een perfecte uitwisseling van data stimuleren tussen private partijen, steden, onderzoeksinstellingen en overheden.

In het kader van het DIM-project starten imec en Digitaal Vlaanderen samen het OSLO-traject voor verkeersmetingen om een eenduidige standaard te creëren voor de semantiek voor verkeersmetingen en -sensoren. Alle partijen die data hebben geleverd voor CityFlows worden ook uitgenodigd om mee te doen aan het OSLO-traject. Als lid van het kernteam van OSLO deelt imec de geleerde lessen uit CityFlows om daarop in het OSLO-traject verder te bouwen. De opstartmeeting

[10] Een gereguleerde testomgeving zie ook volgende link
[11] https://www.vlaanderen.be/digitaal-vlaanderen/onze-oplossingen/vlaamse-smart-data-space

vond plaats op 28 februari 2023. Het traject loopt tot juni 2023[12.]

Het belang van OSLO (Open Standaarden voor Linkende Organisaties) is voor een project als CityFlows van significant belang aangezien het bemachtigen en verwerken van de databronnen veel tijd in beslag heeft genomen. Dat kwam mede door de grote verschillen tussen databronnen. Het aanpassen of toevoegen van data leidde ook vaak tot een ander formaat waardoor de dataverwerking telkens opnieuw aangepast moest worden. Als er een eenduidige standaard in het leven wordt geroepen en bruikbaar gemaakt wordt, zou er slechts één processing script nodig zijn om databronnen in te lezen en te verwerken.

Een andere les die getrokken kan worden uit de bevindingen in dit rapport, is dat er nog heel wat juridische horden te nemen zijn om de aanwezige data van verkeersmetingen optimaal te kunnen gebruiken. Om toegang te krijgen tot bepaalde gegevens bleek bijvoorbeeld een DPIA noodzakelijk te zijn en privacy werd vaak als reden gegeven waarom data niet gedeeld konden worden. Om dat vraagstuk op te lossen heeft imec samen met CiTiP[13] (Center for IT & IP Law van KU Leuven) een project opgestart om onderzoek te doen naar het wetgevende kader voor ANPR. Het project is bedoeld om na te gaan of die ANPR wel kan functioneren binnen een regulatory sandbox en of die regulatory sandbox ook andere, mogelijk interessante databronnen rond mobiliteit beschikbaar kan stellen en bruikbaar kan maken voor onderzoeksprojecten.

- ***Uitdaging 3: Het gebrek aan data over de actieve weggebruiker***

Het valt op dat er weinig data beschikbaar is voor de actieve weggebruiker. De tijdelijke meetcampagnes, de fietstelpalen en de Telraamsensoren meten als enige fietsers. Bij een Telraamsensor is de foutgevoeligheid op actieve weggebruikers bovendien vrij groot waardoor echt betrouwbare databronnen over een groter geografisch gebied ontbreken. De uitrol van fietstelpalen van de provincies en het Agentschap Wegen en Verkeer (AWV) is dan wel aan de gang, maar is nog steeds beperkt. Fietstelpalen blijven daarnaast ook puntlocaties. Daar komt bij dat fietstelpalen vooral op fietssnelwegen staan en minder in stadskernen.

Een actieve weggebruiker is bovendien beperkt of zelfs niet-geconnecteerd. In het Mobilidataprogramma is ook vastgesteld dat actieve weggebruikers minder vaak gebruik maken van navigatieapps. Dat zorgt ervoor dat detectie via apps zeer beperkt blijft.

- ***Oplossing***

Binnen het Mobilidata-programma[14] wordt er door een van de partners binnen het consortium momenteel gewerkt aan een applicatie die de actieve weggebruiker connecteert met intelligente verkeerslichten (iVRI's). De lancering van die applicatie staat gepland binnen het Mobilidata-programma. Het doel van het Mobilidataprogramma is een grootschalige uitrol van C-ITS-toepassingen[15] in Vlaanderen. Daaronder valt ook de uitrol van iVRI's. Voor de actieve weggebruiker is er dus nog wel wat winst te behalen met het gebruik van applicaties die de actieve weggebruiker in staat stelt om deel te worden van het geconnecteerde verkeer. De actieve weggebruiker kan met dergelijke applicaties bijvoorbeeld prioriteit krijgen bij oversteekplaatsen met verkeerslichten. Floating bike data kan door zulke toepassingen beschikbaar worden gesteld. Dan rest enkel nog de voetganger die niet geconnecteerd is en dus nog vaak op een drukknop zal moeten drukken.

Die uitrol van iVRI's binnen het Mobilidata-programma zorgt er daarnaast ook voor dat er een enorme hoeveelheid data beschikbaar komt. Denk daarbij bijvoorbeeld aan data over het aantal verkeersdeelnemers (zowel gemotoriseerd als niet-gemotoriseerd) op en rond een kruispunt. Een aanzienlijke hoeveelheid kruispunten met iVRI's in Vlaanderen zou het gebrek aan data voor een groter geografisch gebied op relatief grote schaal kunnen oplossen. Met iVRI's kan er bovendien ook data worden bijgehouden voor voetgangers (zoals het indrukken van een drukknop). Om data van de iVRI's over

[12] https://www.vlaanderen.be/digitaal-vlaanderen/agenda/evenementen/oslo-verkeersmetingen
[13] https://www.law.kuleuven.be/citip/en/about
[14] https://mobilidata.be/nl/over-mobilidata
[15] Cooperative Intelligent Transport Systems

verkeersvolumes, prioriteitsaanvragen of het aantal weggebruikers te kunnen bemachtigen zal er wel telkens een DPIA uitgevoerd moeten worden. Op basis daarvan kan dan toegang verleend worden tot het archief of de interchange van Mobilidata.

- ***Uitdaging 4: Beperkte informatie over de context van sensoren***

Een probleem dat vaak vastgesteld werd is de discrepantie tussen de werkelijkheid en de beperkte beschrijving van bepaalde sensoren in de metadata. Een concreet voorbeeld daarvan is te zien in Figuur 7 en 8. Op basis van de data in Figuur 7 zou je verwachten dat de data van de ANPR-camera het gemotoriseerde verkeer in de Amsterdamstraat weergeeft. Na controle bleek dat echter niet het geval. De data gaf zeer lage meetwaarden en na een visuele check bleek dat de camera loodrecht op de straat gepositioneerd was ter controle van de voetgangerszone naast het dok (zie Figuur 8).

Figuur 7: Theoretische positionering van een ANPR-camera

Figuur 8: Effectieve positionering van een ANPR-camera

Het bovenstaande is een zeer praktisch voorbeeld dat de moeilijkheden aantoont van het ontbreken van de context van sensoren. Ook bij de Telraamsensoren werden er geregeld opvallend lage resultaten vastgesteld. Dat kwam vermoedelijk door geparkeerde vrachtwagens die tijdelijk het zicht van de sensor belemmerden. Het spreekt voor zich dat meer inzicht in de context van sensoren een belangrijke voorwaarde is voor een goed werkend model. Als de context bekend is, kan het model daarop reageren door tijdelijk databronnen uit te sluiten. Binnen het CityFlows-project was die context echter niet bekend, tenzij er op eigen initiatief een controle werd uitgevoerd.

- ***Oplossing***

Het probleem kan opgelost worden door de context voor elke bron bij de installatie te bepalen en die up-to-date te houden (bij voorkeur in de metadata). Er kan bijvoorbeeld een database bijgehouden worden van alle meettoestellen en de bijbehorende informatie (coördinatiepunten, straatnamen, de richting van de camera, etc.). Ook een afbeelding van de installatie is in veel gevallen een optie. Het OSLO traject rond verkeersmetingen zal ook de metadata meer beschrijven waardoor deze context mogelijk wel gekend is.

Voorstellen voor verbetering

In onderstaande sectie worden mogelijke voorstellen ter verbetering van het model en de beperkingen en fouten van CityFlows gepresenteerd. Het is bij een innovatief onderzoeksproject als CityFlows immers normaal dat er nieuwe inzichten ontstaan en dat er tijdens het project naar boven komt dat andere keuzes mogelijk tot een ander (beter) resultaat hadden geleid. Mede door de kritische blik van VLOCA en ABB is onderstaande sectie tot stand gekomen.

- ***Uniforme verdeling***

Zoals in eerdere secties al beschreven is, werd het model oorspronkelijk zo ontwikkeld dat de dichtheden van de modaliteiten zo uniform mogelijk verdeeld werden. Dat was ook de insteek van de PoC en het idee achter het CityFlows-model. Die insteek leverde echter niet de juiste resultaten op. Er werd geprobeerd om het resultaat te verbeteren met een gewogen verdeling, maar ook dat bleek onvoldoende.

De bedenking die tegen die aanpak ingebracht werd, is dat die aanpak plaatselijk drukke metingen van sensoren (waaraan je kan zien dat het druk is in een bepaalde regio) negeert. Het model schakelt namelijk terug naar zijn verdeling, terwijl het eigenlijk rekening zou moeten houden met de verhoogde drukte in de regio. Een Telraam of telecomdata zouden wel gebruikt kunnen worden als indicatie van de drukte. In het huidige model zit dus een denkfout die inherent is aan de oorspronkelijke insteek van het project. De conclusie die daaruit afgeleid kan worden, is dat er afgestapt moet worden van het idee van een wiskundige verdeling. In plaats daarvan zouden er wel degelijk regels ingevoerd moeten worden (zie volgende paragraaf voor meer uitleg hierover).

- ***Menselijk gedrag in rekening brengen***

Het idee om het menselijke gedrag uitsluitend te laten bepalen door de data, bleek niet werkbaar te zijn. De dekkingsgraad van de databronnen (zowel op geografisch als op multimodaal vlak) was daarvoor niet hoog genoeg. De actieve weggebruiker, en de voetganger in het bijzonder, is bijvoorbeeld ondergerepresenteerd in de beschikbare bronnen.

Om een toekomstig model werkbaar te maken zouden er aan het model regels voor menselijk gedrag opgelegd moeten worden. Dat betekent dat er gestart moet worden vanuit een volledig nieuw model. Bepaalde delen van onze methodologie kunnen daarvoor echter wel hergebruikt worden. Het menselijke gedrag kan op verschillende manieren toegevoegd worden. “Origin-destination”-gegevens, seizoensgebonden effecten, typische werkdagen (”office hours”) en “urbanism” zijn allemaal gekende fenomenen waarmee in een toekomstig model rekening gehouden kan worden. Ons voorstel is een evolutie naar een “agent-based-model” waarbij regels en interacties bepaald worden tussen onderdelen van het systeem.

Een andere weg die ook nog ingeslagen kan worden, is het dynamischer maken van de machine learning-gewichten die de basis vormden van de gewogen verdeling. Die gewichten zouden zo uitgebreid kunnen worden dat ze de verkeersdrukte zelf modeleren.

- ***Ander validatieproces***

Zoals in de voorafgaande stukken is beschreven, werd de output van het CityFlows-model hoofdzakelijk gevalideerd met behulp van de Straatvinkendata. Uit de SWOT-analyse is gebleken dat dat niet foutloos is. De vraag daarbij is welke andere databron dan wel voor de validatie gebruikt kan worden. MOW heeft onlangs een PoC laten uitvoeren naar de validatiemechanismen die noodzakelijk zijn voordat databronnen hun model kunnen binnengaan, en kwam tot de conclusie dat meer ontwikkelingen noodzakelijk zijn.

Eind 2021 heeft het Agentschap Wegen en Verkeer (AWC) een haalbaarheidsstudie uitgevoerd naar het meten van verkeersvolumes op het onderliggend wegennet. Ook uit die studie bleek dat de sensoren die momenteel geïnstalleerd zijn, tekortkomen en dat er extra magnetische sensoren gebruikt moeten worden. Uit de studie bleek bijvoorbeeld dat de

kwaliteit van de Telraam niet hoog genoeg is om het verkeersvolume te kunnen inschatten. AWV gaat die extra magnetische sensoren onderwerpen aan een test om de beste leverancier te kunnen kiezen. De mogelijkheid bestaat erin om de magnetische sensoren daarna uit te rollen op het onderliggend wegennet.

Telraam zou voor het ontwikkelde validatieproces mogelijk gebruikt kunnen worden als alternatief voor Straatvinken. Telraam heeft namelijk het voordeel dat het een veel hogere temporale dekking heeft dan Straatvinken. Die lage dekking was dan ook een van de grootste nadelen aan het gebruik van Straatvinken. De studie van AWV en onze eigen SWOT-analyse dwingen ons er toe om het gebruik van Telraam in deze context te beschouwen als een eerste aanzet voor verdere verbetering. Op termijn zijn verdere verbeteringen en andere databronnen echter noodzakelijk.

De extra sensoren die AWV mogelijk in de toekomst gaat installeren om het verkeersvolume op het secundaire wegennet in te schatten, kunnen mogelijk gebruikt worden als andere databron. Die sensoren kunnen interessant zijn voor de validatiescore. Dat is natuurlijk pas met zekerheid te zeggen als de sensoren geïnstalleerd zijn er een analyse op de data uitgevoerd kan worden. Ook de geografische dekking van die sensoren moet nog onderzocht worden. Een andere databron die eraan zit te komen en gebruikt kan worden voor de validatie, zijn de iVRI's in het kader van het Mobilidata-programma aangezien die een grote geografische dekkingsgraad hebben, multimodaal zijn en nauwkeurig zouden moeten zijn. Die databron kan echter pas gebruikt worden bij voldoende gebruikers.

- ***Model in functie van de use case***

Aan het begin van CityFlows werden de use cases gedefinieerd met het oog op een datafusiemodel. Er kan echter een vraagteken gezet worden bij de noodzakelijk van een model voor elke use case. Als bijvoorbeeld de invoering van een schoolstraat geëvalueerd wordt (de use case in Mechelen), is het nog maar de vraag of er een model nodig is die dag en nacht de dichtheid van alle modaliteiten berekent (zoals in de PoC waar volledige dagen meegenomen werden). Voor een dergelijke use case kan er beter nagegaan worden of het model niet beter enkel voor specifieke uren werkt. Dat brengt echter wel nieuwe uitdagingen met zich mee. In het huidige model worden namelijk opeenvolgende tijdstippen gekoppeld. Deze bedenking kan echter wel worden meegenomen bij andere projecten waar een model als een mogelijke oplossing gezien wordt.

Ook voor de evaluatie van een wijkmobiliteitsplan (de use case in Gent) kan de vraag gesteld worden of het model wel echt dag en nacht werkzaam moet zijn. Het antwoord op die vraag is in deze use case echter minder vanzelfsprekend, aangezien in dit geval de leefbaarheid geëvalueerd moet worden. Om de leefbaarheid te kunnen meten, moet de nacht ook geëvalueerd worden. Nachtlawaai is bijvoorbeeld niet bevorderlijk voor de leefbaarheid. Voor deze use case is het dan ook niet aan te raden om het aantal uren te beperken tot de uren waarop mensen actief zijn.

Er kan daarnaast ook bekeken worden of een datafusiemodel eigenlijk wel noodzakelijk is. Voor de schoolstraat werden er extra tijdelijke meetcampagnes uitgevoerd bovenop de bestaande Telraam-, Fietspalen- en ANPR-tellingen. Het antwoord kan vaak al gevonden worden in de opgekuiste data. Die data kan namelijk gevisualiseerd en geanalyseerd worden aan de hand van een analysetoolbox. Er moet kritisch bekeken worden welke meerwaarde een datafusiemodel heeft ten opzichte van de analysetoolbox.

De vraag blijft echter of een analysetoolbox die databronnen analyseert en valideert een oplossing is voor de huidige beleidsuitdagingen van de steden. Om die vraag te kunnen beantwoorden ontwikkelde imec een analysetoolbox (user: demo, password: demo1234; weblink slechts geldig t.e.m. januari 2024) die de mobiliteitsexpert in steden helpt om visueel en met statistische zekerheid antwoorden te geven op de huidige beleidsuitdagingen. De beschikbare databronnen kunnen toegevoegd worden en aan enkele kwaliteitscontroles onderworpen worden. Vervolgens kan er onderzocht worden of er statistisch relevante veranderingen zijn in de mobiliteit gedurende een bepaalde tijd. De toolbox kan ook voorspellingen maken op basis van gedetecteerde patronen in de databronnen.

- ***Wegingscoëfficiënten dynamisch maken***

De wegingscoëfficiënt (zie paragraaf over gewogen verdeling in dit rapport) die gebruikt werd, is gebaseerd op een

snapshot (omwille van de beperking van Straatvinken). Die coëfficiënt wordt op exact dezelfde manier toegepast in het verloop van de dag. Dat wil zeggen dat we ervan uit gaan dat de coëfficiënt niet veranderd met de tijd. Die aanname houdt geen rekening met de dynamiek van een stad. Schoolgaande kinderen zullen rond een school bijvoorbeeld meer fietsers genereren tussen 8.00 uur en 9.00 uur dan tussen 17.00 uur en 18.00 uur. Het coëfficiënt is echter niet dynamisch in functie van de tijd en locatie, wat een grote beperking is van dit onderzoek.

Het zou een verbetering zijn om die coëfficiënt wel dynamisch te maken in functie van de tijd en in functie van de dag in het jaar (in de zomer zijn er bijvoorbeeld meer fietsers dan in de winter). Binnen postdoc onderzoek van de vakgroep Geografie aan de UGent wordt daarop verder gewerkt. De kennis en bevindingen uit CityFlows-project zijn dan ook gedeeld met de UGent.

Het wegingscoëfficiënt zou ook dynamisch gemaakt kunnen worden aan de hand van artificiële intelligentie. Aan de hand van verschillende beschikbare data kan AI getraind worden om patronen te herkennen en te weten wanneer bepaalde events zullen plaatsvinden. Omdat uit het CityFlows-project is gebleken dat de wegingscoëfficiënt verbetering bracht, is imec gestart met het CityCom.AI-project. Dat project is een Europees project waar imec aan mee werkt om de uitwerking voor België te voorzien. Het plan van het project is om in Mechelen en Brussel AI Testing and Experimentation Facilities (TEF) rond het thema mobiliteit te voorzien. De bedoeling is om AI toe te gaan passen op specifieke beleidsnoden rond mobiliteit van de twee steden. De uitvoering van het project is momenteel gaande en loopt tot eind 2027. Bij de verdere uitwerking en bepaling van de use cases zijn de lessen van CityFlows meegenomen.

Als alternatief kan er ook voortgebouwd worden op de huidige machine learning-aanpak door een andere dataset te gebruiken. De huidige aanpak voor het bepalen van de wegingscoëfficiënt maakte gebruik van het wegenregister, OpenStreetMap (OSM), beelddata van straten, data van het Departement Omgeving en van Straatvinkendata. Bij deze aanpak zou het kunnen helpen om Telraamdata te gebruiken in plaats van Straatvinkendata. Telraam voorziet namelijk op langere termijn (constante) metingen, terwijl Straatvinken slechts een beperkte temporale dekking heeft. Door gebruik te maken van Telraam zou er een stabieler beeld van de modale verdeling in de straten gevormd kunnen worden. Als er voldoende data beschikbaar is (over enkele jaren), is het zelfs mogelijk om rekening te houden met seizoensgebonden effecten.

Dezelfde bedenkingen moeten echter ook geplaatst worden bij de aanbevelingen voor de validatiedataset. Telraam is namelijk een sensordataset en datakwaliteitsproblemen zijn inherent aan een dergelijke dataset. Die problemen zijn al eerder besproken, maar zullen hier kort aangehaald worden om een mogelijke oplossingen te kunnen aandragen.

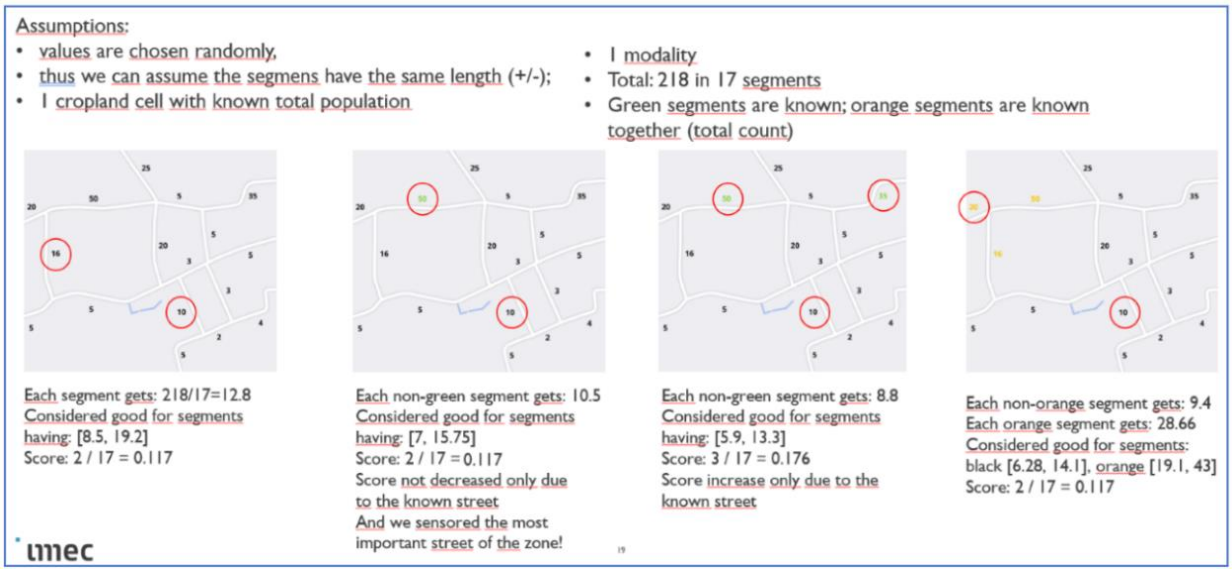
De problemen met de slechtere resultaten naarmate het daglicht afneemt, kunnen makkelijk opgelost worden door dag en nacht op te splitsen en elk als een apart geheel te zien. Daarbij moet echter wel rekening gehouden worden met de vroegere zonsondergang in de winter. In de winter kan de schemering zelfs al tijdens en voor de avondspits vallen.

Er is gebleken dat Telraam het aantal actieve weggebruikers onderschat. Die onderschatting zou wiskundig gecorrigeerd kunnen worden. Een bijkomend probleem is dat die afwijking niet uniform is voor alle sensoren. Bij de aanbevelingen van deze databron is echter al aangegeven dat dat opgelost zou kunnen worden door een eenmalige kalibratie van elke sensor. Dat is echter wel tijds- en arbeidsintensief, en zelfs in dat geval moet er verondersteld worden dat de afwijking constant blijft, er geen seizoensgebonden effecten worden gemeten en er een algemene daling in kwaliteit zal plaatsvinden door slijtage.

Het is kortom interessant om de wegingscoëfficiënten in een vervolgonderzoek te hertrainen op Telraamdata zodat er meer rekening gehouden kan worden met temporale effecten en meer dynamische gewichten bepaald kunnen worden. Daarbij moet echter wel worden opgemerkt dat deze aanbeveling slechts een eerste aanzet is. Verder onderzoek blijft noodzakelijk.

Conclusie

Het model bleek niet geschikt te zijn voor de toepassingen die de pilootsteden voor ogen hadden, omdat er te weinig vertrouwen is in de accuraatheid van het model. De oorzaak daarvan is in dit rapport aangewezen en uiteengezet. Kort samengevat komt het erop neer dat het model te statisch is en daardoor te weinig rekening houdt met het dynamische karakter van een stad. Bovendien heeft het model een hogere dekkingsgraad van databronnen nodig. Uit het gedachtenexperiment is gebleken dat bijna elke straat gemeten moet worden om met het model een voldoende accuraat beeld te krijgen (zie Figuur 9). Uiteraard moet het aantal meetpunten in verhouding staan met het nut van een model. Indien men iedere straat zou meten, is een model immers niet meer noodzakelijk, aangezien je in dat geval continu het verkeer meet.



Figuur 9: Experiment over minimale datadekking

De steden verdienden echter nog steeds een betrouwbaar antwoord op hun onderzoeksvragen. Het antwoord daarop kon gevonden worden in de data zelf. Daarom heeft imec een analysetoolbox ontwikkeld die mobiliteitsambtenaren helpt om de kwaliteit van de databronnen op een automatische, data-agnostische manier te controleren. De toolbox maakt daarvoor gebruik van geavanceerde statistische methodes die op een eenvoudig manier gebruikt kunnen worden. Verder stelt de toolbox mobiliteitsambtenaren in staat om eenvoudig verschillende analyses uit te voeren. Verschillende datasets kunnen eenvoudig gecombineerd worden en kunnen tegelijkertijd geanalyseerd worden in één analyse. Met deze toolbox is het dus niet nodig om elke dataset apart te analyseren en de resultaten van die analyses vervolgens met elkaar te vergelijken om hopelijk daaruit een eenduidige conclusie te herleiden. De toolbox is bovendien niet uitsluitend bruikbaar voor een specifieke toepassing of use case, maar kan zonder aanpassingen gebruikt worden voor de analyse van eender welk project. Dat geeft de steden dus de mogelijkheid om hun volgende project zelf te analyseren.

Daarnaast zijn de geleerde lessen over de databronnen een meerwaarde voor het datagedreven mobiliteitsbeleid van de steden. Om ervoor te zorgen dat die lessen niet telkens opnieuw geleerd moeten worden, stellen we voor om niet enkel in te zeten op datadeling, maar ook op de kennisdeling die daarmee gepaard gaat. Op die manier kan vorig werk de basis vormen voor toekomstig werk.

De uitkomst is dan ook dat imec en de steden aan de slag zijn gegaan met de geleerde lessen uit het CityFlows-project. Voorbeelden daarvan zijn het pas opgestarte OSLO-traject waar veel steden aan mee werken, het VSDS-project voor de optimalisatie van datadeling en het onderzoek naar het juridische aspect daarvan dat imec samen met CiTiP uitvoert. Binnen het CitCom.AI-project wordt gekeken of de wegingscoëfficiënten en de eerste vaststellingen daarvan uit het CityFlows-project hergebruikt kunnen worden. Samen met de steden zet imec in op het delen van de geleerde lessen

zodat die voor vervolgprojecten gebruikt kunnen worden.

Bijlagen

Bijlage 1

A.1. Analytical base table

Om de gewichten van de verschillende modaliteiten te bepalen was er een nood aan een grote hoeveelheid data. Hierdoor is er veel werk gekropen in het exploreren en het begrijpen van de datasets. Dit resulteerde in de volgende ABT of analytical base table. Deze tabel omschrijft elke dataset en welke features deze levert in de 'columns' kolom.

Data provider	Dataset	Format	Granularity	Columns	Status
Ringland Academie	Straatvinken the central dataset to which all other data is coupled	CSV	coordinates (lat, long)	ID, lat, long, bike, walk, bus, car, van, truck, streetname, municipality, geometry	✓ Antwerpen (1600 samples) to be 3600+ for Flanders in 2020
Departement Omgeving	Bebouwingsdichtheid Vlaanderen	GeoTiff	ha	bedi: bebouwingsdichtheid, ss_lengte / ss_oppervl: length and area of relevant statistical sector	✓
StatBel	Population density	Excel / CSV	statistical sector	pode: population density	✓
StatBel	Cars	Excel / CSV	statistical sector	• nhh: # households • ncars: # cars • ncars_per_hh: # cars per household	✓
Agentschap Wegen & Verkeer	Wegenregister	Shapefiles & dbx files	road segment or road node (knooppunt)	• road category • road type • number of lanes • …	✓
	Street-level imagery	Panorama images (N, W, S, E)	coordinate	landscape segmentation % • sky • tree • road • building • sidewalk • (150 categories in total) feature vector: multidimensional (CNN) latent vector representation	✓
StatBel	Bevolking volgens het km²-raster (Grid) | Statbel	Raster	km2	ms_pop: population density (per km2)	✓
StatBel	Verkeersongevallen 2020			acc, acc_death, acc_death30, acc_mort, acc_ser, acc_sly	✓
StatBel	Verkeersslachtoffers 2020 | Statbel (fgov.be)			Didn't add anything to traffic accidents (above)	✗

Departement Omgeving	Ruimterapport Vlaanderen data			kc_kern_prop, kern_prop, lint_len, kc_lint_kern, bmc_prop	✓
	Onderwijsaanbod in Vlaanderen en Brussel			school_D_dist3 school_B_dist3 school_S_dist3 school_V_dist3 school_H_dist3	✓
OpenStreetMap Community	OpenStreetMap traffic elements	shapefile		trfeat_motorway_junction_dist3 trfeat_speed_camera_dist3 trfeat_crossing_dist3 trfeat_traffic_signals_dist3 trfeat_mini_roundabout_dist3 trfeat_turning_circle_dist3 trfeat_lock_gate_dist3 trfeat_fuel_dist3 trfeat_parking_underground_dist3 trfeat_stop_dist3 trfeat_parking_dist3 trfeat_slipway_dist3 trfeat_street_lamp_dist3 trfeat_marina_dist3 trfeat_weir_dist3 trfeat_parking_bicycle_dist3 trfeat_dam_dist3 trfeat_pier_dist3 trfeat_waterfall_dist3 trfeat_parking_multistorey_dist3 trfeatp_parking_dist3 trfeatp_marina_dist3 trfeatp_parking_bicycle_dist3 trfeatp_parking_multistorey_dist3 trfeatp_parking_underground_dist3 trfeatp_fuel_dist3 trfeatp_pier_dist3 trfeatp_dam_dist3 trfeatp_service_dist3 trfeatp_slipway_dist3 trfeatp_weir_dist3 trfeatp_lock_gate_dist3 trfeatp_waterfall_dist3	✓
OpenStreetMap Community	OpenStreetMap land use	shapefile		lu_park_prop lu_grass_prop lu_industrial_prop lu_residential_prop lu_meadow_prop lu_farmland_prop lu_commercial_prop lu_forest_prop	✓

A.2. Data providers en datasets

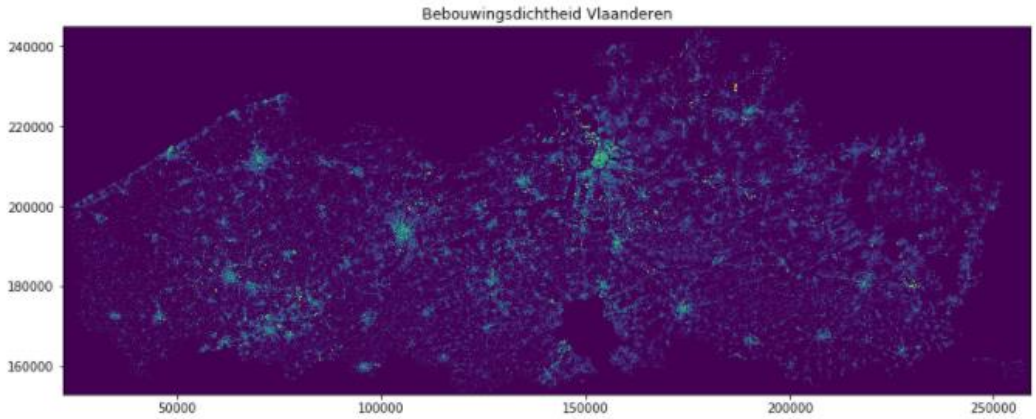
1. Ringland Academie

Ringland Academie is een koepelorganisatie over verschillende entiteiten zoals Ringland - een burgerbeweging - en onderzoekers. Het werd opgericht naar aanleiding van enkele verhitte debatten over mobiliteit in de Stad Antwerpen, er waren protesten, waarbij verschillende burgerbewegingen het heft in eigen handen namen en voorstellen deden over mobiliteit en stedelijke ontwikkeling. Ze zetten een verdrag op - het "toekomstverbond" of "toekomstverdrag" genoemd - met de stad Antwerpen en de Vlaamse regering, dat de modal shift wil bereiken, waarbij de modal split tussen personenwagens en hun alternatieven (openbaar vervoer, voetgangers en fietsen) verschuift van 70/30 naar 50/50.
Ze lanceerden het citizen science project Straatvinken.

a. **Straatvinken dataset**
De gegevens van het Straatvinken project bestaan uit de tellingen van weggebruikers per modaliteit, gekoppeld aan de breedte- en lengtegraad van de locatie waar deze tellingen zijn verkregen, en het jaar van de campagne. De regio waar de campagne plaatsvond, evolueerde in de loop van de tijd. Aanvankelijk ging het alleen om Antwerpen vervoersregio, maar tegenwoordig omvat het het hele Vlaamse gewest. De tellingen worden uitgevoerd op een gekozen donderdag in mei, tussen 17.00 uur. and 18h00. Dit is zo gekozen om de seizoens en tijdsgebonden invloeden te vermijden in de vergelijking tussen verschillende edities.

2. Departement Omgeving

a. **Bebouwingsdichtheid Vlaanderen**
De bebouwingsdichtheid Vlaanderen dataset is een open dataset die de populatie dichtheid toont per hectare in de Vlaamse regio.
Per hectare wordt berekend welke oppervlakte ‘bebouwd’ is, door de oppervlakte van alle gebouwen binnen die hectarecel op te tellen. Daarbij werden alle gebouwen uit de gebouwenlaag Gbg van GRB opgenomen (hoofdgebouwen, bijgebouwen en gebouwen afgezoomd met virtuele gevels) én alle gebouwen van CADMAP met uitzondering van de ondergrondse gebouwen (UN). Deze geodatalaag werd in 2021 vervangen door een herwerkte versie voor toestand 2013, om conform te blijven met de bijgestelde methode die gehanteerd werd bij de opmaak van toestand 2019. Voor meer details over de gehanteerde methode van toewijzing van landgebruik aan open ruimte verwijzen we naar het technisch rapport: "Poelmans Lien, Crols Tomas, Janssen Liliane, Hambsch Lorenz (2021), Indicatoren Ruimtelijk Rendement. Technische fiches", uitgevoerd in opdracht van het Vlaams Planbureau voor OmgevingDeze geodatalaag werd in 2021 vervangen door een herwerkte versie voor toestand 2013, om conform te blijven met de bijgestelde methode die gehanteerd werd bij de opmaak van toestand 2019.



b. **Ruimterapport RURA**
Het ruimterapport 2021 is de opvolger van het Ruimterapport 2018. Heel wat cijfers werden geactualiseerd, maar ook ander cijfermateriaal dat ons toelaat nieuwe tendensen en evoluties te meten, werd opgenomen.
Dit rapport geeft waar mogelijk de evolutie 2013-2019 weer en verwijst daarbij naar basisdata die

vergelijkbaar zijn. Voor een beperkte selectie van thema's en indicatoren in het Ruimterapport 2018 was het nodig de beginsituatie 2013 te herrekenen door verbeterde methodische inzichten en/of met betere data. Hierdoor kan de toestand 2013 in beperkte mate verschillen van de cijfers in het Ruimterapport 2018.

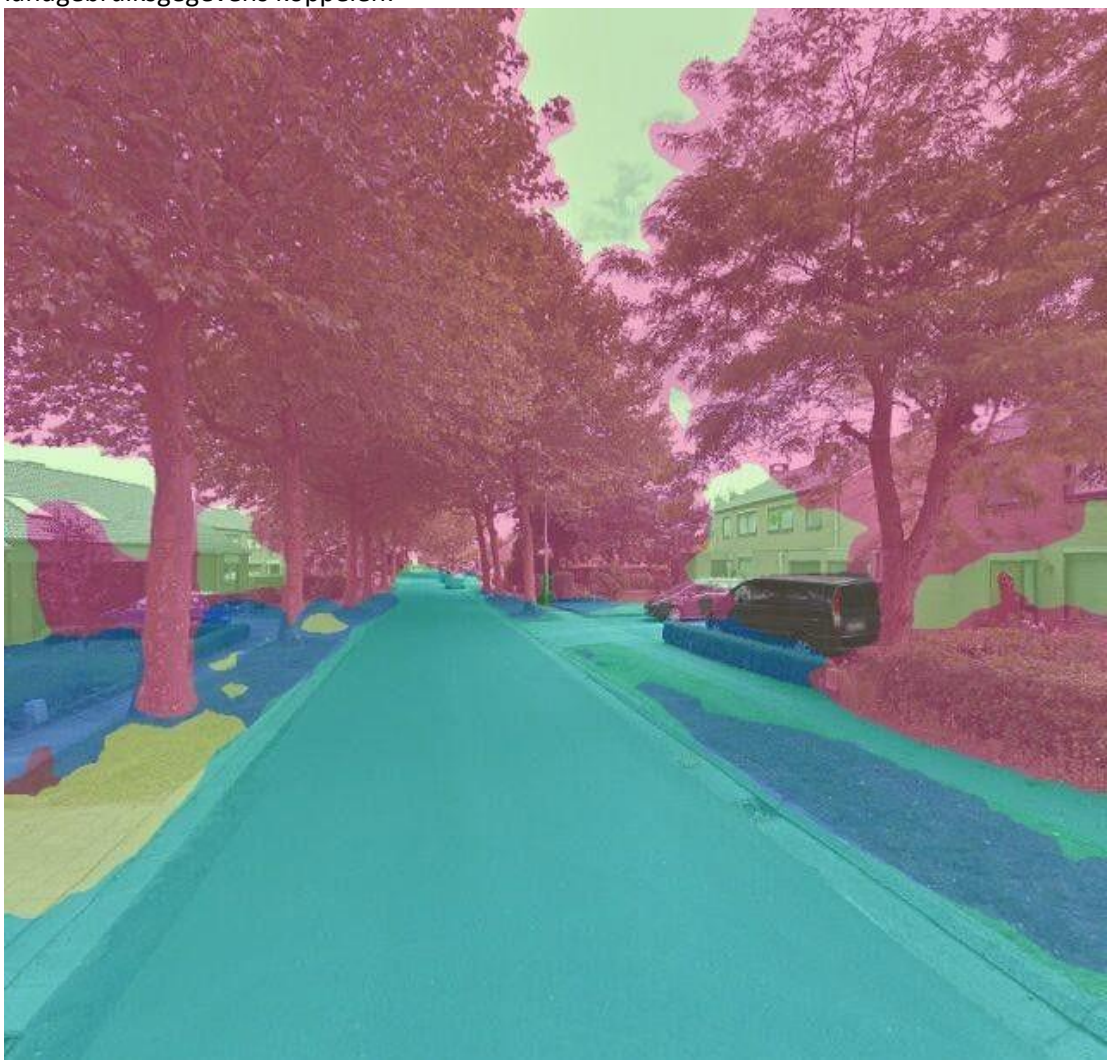
De inhoudelijke focus van dit rapport is verbreed vanuit thema's zoals ruimtebeslag en verdichting naar thema's met meer maatschappelijke opgaven zoals natuurlijk kapitaal of gezondheid. De keuzes hierin zijn gemaakt onder andere op basis van de impact van deze opgaven op onze fysieke leefomgeving en op basis van kennis en data die het departement Omgeving de afgelopen jaren door middel van onderzoek verzamelde.

3. **Street-level imagery**

   Dit zijn fotografische gegevens die door een auto worden vastgelegd met 360° snapshots van het uitzicht op de straat, gekoppeld aan de locatie waar deze is vastgelegd. Deze beeldgegevens kunnen worden verwerkt met behulp van machine learning-technieken om landgebruik / geospatiale functies te ontdekken, zoals:

   a. Landschapssegmentatie: % van lucht, gebouw, boom, gras, gebouw in beeld
   b. Objectdetectie: banken, auto's, parkings, bomen, fietsen, vogels, speeltuinen

Een voorbeeld usecase is dat we de streetviewpanorama's van de straatvinkengegevens vastleggen (met behulp van hun lat-lange coördinaten) en de eerder genoemde segmentatie op de afbeeldingen uitvoeren, we kunnen landgebruiksgegevens koppelen:



In het bovenstaande voorbeeld krijgen we (fictieve gegevens) dat bijna 50% bomen zijn, 30% de weg, 5% het trottoir, 3% vuil en 2% auto's, enz.

4. **StatBel**

   Statbel, het Belgische statistiekbureau, verzamelt, produceert en verspreidt betrouwbare en pertinente cijfers

over de Belgische economie, de samenleving en het territorium.
Het verzamelen gebeurt op basis van administratieve databronnen en enquêtes, het produceren gebeurt op een wetenschappelijke, kwaliteitsvolle wijze en de statistieken worden tijdig en op een klantvriendelijke wijze verspreid.
De privacy, de bescherming van vertrouwelijke gegevens en het gebruik voor louter statistische doeleinden worden hierbij gegarandeerd.

5. **OpenStreetMap Community**
   OpenStreetMap (OSM) is 's werelds grootste geografische informatieproject voor vrijwilligers. In veel opzichten demonstreert OSM de kracht van crowdsourcing van geografische gegevens aan een gebruikersgemeenschap die wereldwijd is en waar gegevens die de kaart maken die veel mensen gebruiken voor navigatie, iedereen ten goede kunnen komen.
   a. **OpenStreetMap dataset**
      OpenStreetMap is een kaart van de wereld, gemaakt door mensen zoals jij en gratis te gebruiken onder een open licentie. Mogelijk alternatief of aanvulling op wegenregistergegevens. U kunt delen van de kaart downloaden. Er moet API-tooling zijn om functies in de buurt te krijgen, enz. Het project werd opgestart met behulp van overheidsgegevens. Voor België zou dat het wegenregister (WR) en/of kadaster kunnen zijn, maar dat weten we eigenlijk niet.
      Let op: de volledigheid/nauwkeurigheid van gegevens is zeer locatieafhankelijk.

Kenmerk	OSM	WR
Verkeerslichten	✓	✗
Bushaltes/relaties	✓	✗
Max snelheid	✓	?
Wegcategorie	✓	✓
Stukken land van alle soorten landgebruik (drasland, rivier, ...)	✓	✗
Wegen van alle soorten	✓	✓

6. **Agentschap Wegen & Verkeer**
   Het Agentschap Wegen en Verkeer (AWV) is wegbeheerder van zo´n 7000 km gewest- en autosnelwegen en ruim 7700 km fietspaden. Ze werken elke dag aan een betere mobiliteit en doen dat samen met verschillende partners.
   a. Wegenregister
      Het Wegenregister is het middenschalig referentiebestand van de wegen in Vlaanderen. Het moet alle (openbaar toegankelijke) wegen van Vlaanderen bevatten, met bijbehorende attribuutgegevens. Het bestand heeft een middenschalige precisie.
      Meer info over de velden op niveau van wegsegment, wegknoop en andere attributen, kan je op hun website terugvinden. De belangrijkste voor onze toepassign zijn de volgende:

      Naam attribuut: Wegcategorie
      Code: WEGCAT
      Definitie: Wegcategorie zoals gedefinieerd in het Ruimtelijke Structuurplan Vlaanderen
      Type: Getal (5,0)
      Domein: H: hoofdweg, PI: primaire weg I, PII: primaire weg II, PII-1: primaire weg II type 1, PII-2: primaire weg II type 2, PII-3: primaire weg II type 3, PII-4: primaire weg II type 4, S: secundaire weg, S1: secundaire weg type 1, S2: secundaire weg type 2, S3: secundaire weg type 3, S4: secundaire weg type 4, L: lokale weg, L1: lokale weg type 1, L2: lokale weg type 2, L3: lokale weg type 3, -8: niet gekend, -9: niet van toepassing

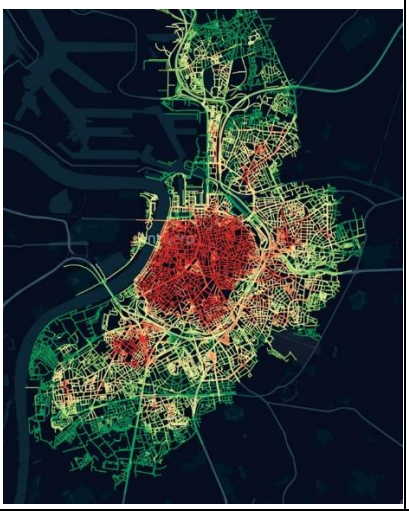
      i.

Naam attribuut: Morfologische wegklasse
Code: MORF
Definitie: Beschrijft bepaalde aspecten van de morfologische vorm die een weg kan aannemen.
Type: Getal (3,0)
Domein: 101: autosnelweg, 102: weg met gescheiden rijbanen die geen autosnelweg is, 103: weg bestaande uit één rijbaan, 104: rotonde, 105: speciale verkeerssituatie, 106: verkeersplein, 107: op- of afrit, behorende tot een nietgelijkgrondse verbinding, 108: op- of afrit, behorende tot een gelijkgrondse verbinding, 109: parallelweg, 110: ventweg, 111: in- of uitrit van een parking, 112: in- of uitrit van een

Bijlage 2

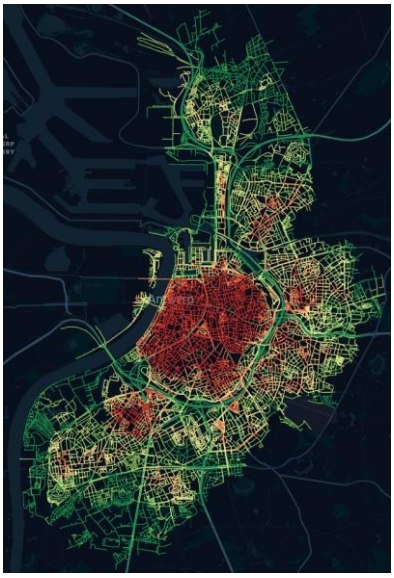
B.1. ML gewichten Voor het cityflows model

De machine learning oefening leverde volgende gewichtskaarten op:

Voetgangers	Fiets	Bus
		
Vrachtwagen	**Camionette**	**Auto**

De uitkomst van het model was uiteraard ook niet zonder fouten.
Er zijn namelijk enkele zaken die opvallen dat niet mogelijk zijn: gemotoriseerde voertuigen zoals auto's in parken en in de zoo. Dit komt omdat het machine learning model kijkt naar de omgeving en niet enkel naar de straat zelf. Om deze

problemen te verhelpen is de tabel achteraan dit document opgesteld. Deze tabel geeft weer welke modaliteiten er toegestaan zijn in een straat gebaseerd op de morfologie van de straat volgens het wegenregister van Agentschap Wegen & Verkeer.
De implementatie hield in om de modaliteiten die niet toegestaan waren op nul te zetten in de gewichten, de andere waarden bleven onveranderd. Dit leverde volgende gewichtskaarten op:

Voetgangers	Fiets	Bus
		
Vrachtwagen	Camionette	Auto

Dit levert geen gewicht voor de achtergrond, dit werd opgelost door deze gelijk aan 1 te stellen. Hierdoor bleek later in de analyse van de resultaten van 2020, dat dit niet geschikt was. Een beter alternatief was om dit ook te bepalen met machine learning technieken.

MORF		Auto	Fiets	Vracht-wagen	Bus	Voetganger
-8	niet gekend	1	1	1	1	1
101	autosnelweg	1	0	1	1	0
102	weg met gescheiden rijbanen geen autosnelweg	1	1	1	1	1
103	weg 1 rijbaan	1	1	1	1	1
104	rotonde	1	1	1	1	1
105	speciale verkeerssituatie	0	0	0	0	0
106	verkeersplein	1	1	1	1	1
107	op- en afritten van de snelweg	1	0	1	1	0
108	bypasses op het own	1	1	1	1	1
109	paralllenweg - snelweg	1	0	1	1	0
110	ventweg OWN	1	1	1	1	1
111	parking in en uitrit	1	0	0	0	0
112	in- uitrit van een dienst	0	0	0	0	0
113	voetgangerszone	0	1	0	0	1
114	wandel- of fietsweg	0	1	0	0	1
116	tramweg	0	0	0	1	0
120	dienstweg	0	0	0	0	0
125	aardeweg	0	0	0	0	1
130	veer	0	0	0	0	0
all other values		0	0	0	0	0

Bijlage 3

C.1. Validatie van het cityflows model

Straatvinken is een jaarlijkse actie waar burgers het aantal voetgangers, fietsers en auto's tellen die gedurende een uur voor hun deur (of een andere gekozen locatie) passeren. Aangezien het de vervoersmodaliteiten apart telt lijkt het of dit een goede absolute waarheid om ons model mee te staven. De telling gebeurt elk jaar in de maand mei en loopt voor 1 uur tussen 17 en 18 uur. Dit is natuurlijk maar een kleine steekproef in het tijdsdomein van onze use case (heel het jaar 2020). Er is ook geen beeld op de eventuele bias op de gemeten locaties. Het zou kunnen zijn dat omdat een bepaalde gemeenschap actiever deelneemt aan deze telling sommige wijken meer gerepresenteerd zijn dan andere. Echter is de grootste moeilijkheid om deze dataset te gebruiken als validatie het feit dat deze een intensiteit (aantal per tijdsduur) meet in plaats van een dichtheid (aantal per meter). Om deze te verenigen is er een transformatie nodig:

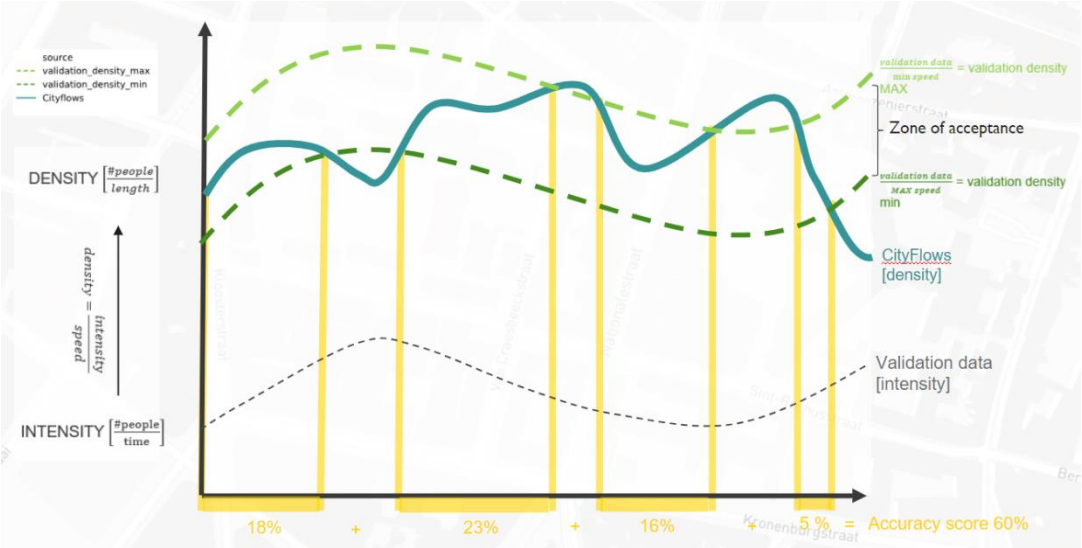
$$Intensity = Density \cdot Average\ speed$$

Hierin zien we een snelheid verschijnen, deze is echter niet gekend. Vandaar wordt er een aanname gemaakt deze is te zien in de volgende tabel:

Modaliteit	Referentie snelheid	Minimale geaccepteerde snelheid = ref snelheid * 0.66	Maximale geaccepteerde snelheid= ref snelheid* 1.5
Voetganger	5.4 km/h	3.56 km/h	8.1 km/h
Fiets	11.88 km/h	7.84 km/h	17.82 km/h
Auto	13.7 km/h	9.04 km/h	20.55 km/h
Vrachtwagen	13.7 km/h	9.04 km/h	20.55 km/h

Openbaar vervoer	13.7 km/h	9.04 km/h	20.55 km/h

Hierin zien we ook de volgende stap die gezet wordt om de straatvinken data te gebruiken als validatie dataset, namelijk een zone definiëren waarin we het resultaat accepteren. Een voorbeeld van de vergelijking hiermee ziet er dan als volgt uit, waar het geel aangeeft wat er geaccepteerd wordt:



Deze vergelijking wordt gedaan voor elke straat in de Straatvinken dataset. Om één enkele validatiescore te krijgen, wordt de gewogen som genomen met de relatieve straatlengtes als gewichten.

Deze procedure levert een validatiescore op die kan worden samengevat als een gewogen verhouding van die straten waar het resultaat tussen de 66 en 150% van het Straatvinken resultaat ligt.
De hypothese was dat door meer lokale mobiliteitsgegevensbronnen toe te voegen, we een toename van de validatiescore zouden zien en dus een betrouwbaardere output van het model zouden hebben. Als we naar de volgende tabel kijken, zien we dat de validatiescore praktisch ongewijzigd is gebleven door de verschillende learning cycles. Verder zijn de waarden ook laag.

Learning cycle	Gebruikte gegevensbronnen in het model	Validatiescore
LC1	Cropland (telco), Velo	0.0444
LC2	Cropland (telco), Velo + Telraam, Signco	0.0446
LC3	Cropland (telco), Velo, Telraam + ANPR	0.0445

Dus het verbeteren van het model door meer lokale gegevensbronnen toe te voegen, heeft het vertrouwen in de resultaten niet vergroot. Deze metriek heeft echter ook zijn eigen problemen. Als we een voorbeeld nemen en een van de geaccepteerde straten bekijken en deze handmatig vergelijken met de uitkomst van de Straatvinkentelling zien we dat deze ver van elkaar liggen.
De output van het model lag voor het gemeten uur tussen de 0.025 en 0.029 mensen per meter. Gebruik makend van de straatlengte (96m) wil dit zeggen dat er minder dan 2.8 personen/5 minuten waren. Dit komt overeen met minder dan 35 personen per uur, wat niet te vergelijken is met de Straatvinkentelling van 251.
Dus deze validatie procedure geeft mindere resultaten voor ons model maar is zelf ook niet feilloos en dat dient in het achterhoofd gehouden te worden.

Herman Teirlinckgebouw
Havenlaan 88, bus 70
1000 Brussel
stedenbeleid@vlaanderen.be