

Automatische generalisatie van topografische data

State-of-the-art vanuit Europees perspectief

Dr Jantien Stoter, Projectleider EuroSDR generalisatieproject tijdens aanstelling ITC, Enschede

Huidige betrekkingen:
Universitair hoofddocent, Sectie GIST, OTB, TU Delft &
Adviseur Product en Proces Innovatie, Kadaster, Apeldoorn



Andere personen die vanuit Nederland aan het project hebben bijgedragen: Nico Bakker (Kadaster), Connie Blok (ITC), Xiang Chang (ITC), Annemarie Dortland (Kadaster), Willy Kock (ITC), Peter Lentjes (Kadaster), Maarten Storm (AquaGIS, voorheen Kadaster), Harry Uitermark (ITC, voorheen Kadaster).

Sinds de jaren '80 wordt kaartproductie bij Nationale Topografische Diensten (National Mapping Agencies: NMAs) ondersteund door digitale representaties van de topografie op verschillende schalen. Deze digitale bestanden maken het in principe mogelijk om kleinere schalen automatisch af te leiden van een basis dataset. Maar in de praktijk blijkt dit nog niet zo eenvoudig. Wat zijn NMA-overkoepelende en NMA-specifieke eisen voor deze automatische generalisatie? Hoe zijn deze eisen momenteel geïmplementeerd in commerciële software? Welke onderwerpen vragen om nader onderzoek? Deze vragen stonden centraal in het EuroSDR onderzoeksproject naar automatische generalisatie dat van 2006 tot 2009 is uitgevoerd.

EuroSDR is een organisatie voor European Spatial Data Research waarbinnen onderzoeksinstituten, organisaties die voorzien in ruimtelijke data en de private sector samen onderzoek uitvoeren naar onderwerpen met een gezamenlijk belang. Het projectteam van het EuroSDR generalisatieproject bestond uit zes NMAs (Frankrijk, Engeland, Catalonië, Nederland, Spanje en Denemarken) en drie onderzoeksinstituten (Universiteiten van Zürich en van Hannover en het ITC). Vier softwareleveranciers hebben hun software laten testen: Axes Systems (software Aexpand), ESRI (software ArcGIS), Universiteit Hannover (generalisatiemodules Change, Push en Typify) en 1Spatial (software Radius Clarity). De twee hoofdvragen van het onderzoek luiden: 1. Wat zijn de capaciteiten en beperkingen van commerciële software voor automatische generalisatie met betrekking tot NMA-wensen en -eisen? 2. Welke verschillende oplossingen zijn er mogelijk voor een bepaald generalisatieprobleem en wat is de reden van deze verschillen?

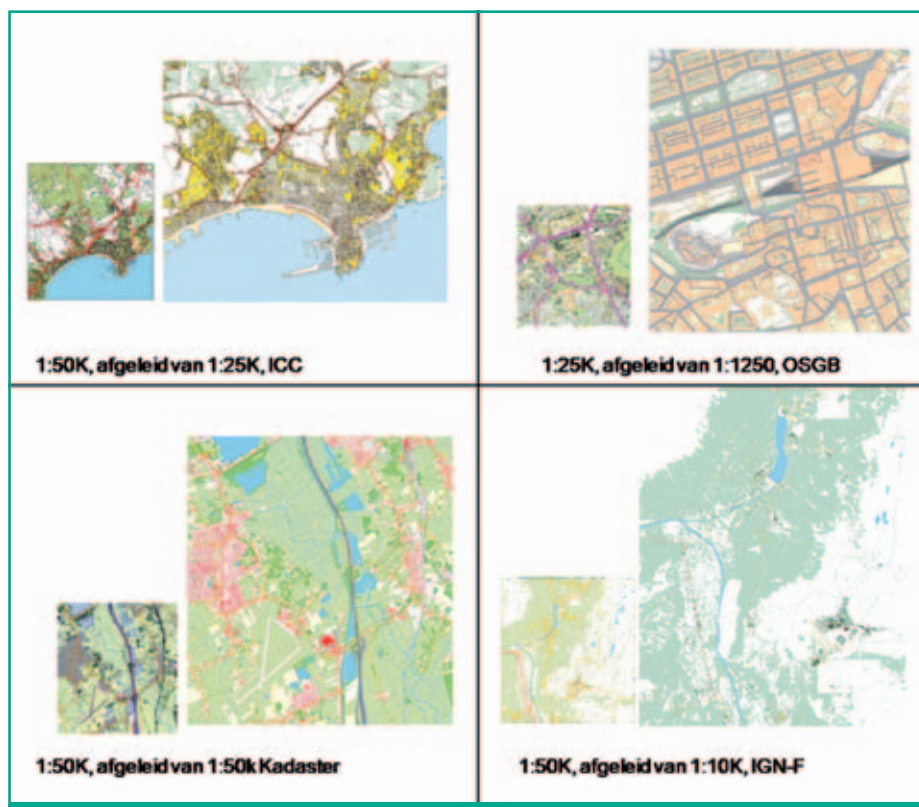


Fig. 1. Uitsnedes van de vier testcases; input en voor iedere testcase een voorbeeld van gegeneraliseerde output.

| Area type | Source dataset | Target dataset | Provided by | No. of feature classes | Main feature classes |
|------------------|----------------|----------------|------------------|------------------------|---|
| Urban area | 1:1250 | 1:25k | OS Great Britain | 37 | buildings, roads, river, relief |
| Mountainous area | 1:10k | 1:50k | IGN France | 23 | village, river, land use |
| Rural area | 1:10k | 1:50k | Kadaster, NL | 29 | small town, land use, planar partition |
| Coastal area | 1:25k | 1:50k | ICC Catalonia | 74 | village, land use (not mosaic), hydrography |

Tabel 1. Testcases van het EuroSD generalisatieproject

In het project is veel aandacht besteed aan de onderzoeksmethodologie. Het was namelijk niet eenvoudig om overzichtelijke en onafhankelijke evaluaties uit te voeren van generalisatie-outputs welke gegenereerd zijn door verschillende systemen en door verschillende testers op basis van zowel generieke als specifieke specificaties van NMAs. Alvorens in te gaan op de belangrijkste conclusies van dit project, zullen eerst de twee hoofdstappen van de onderzoeksmethodologie besproken worden, namelijk a) het eenduidig beschrijven van generalisatiespecificaties voor automatische generalisatie en b) het evalueren van de output.

a. Beschrijven van specificaties voor automatische generalisatie

Het project richtte zich met name op generalisatie van grootschalige tot midschalige bestanden omdat de deelnemende NMAs dit als meest tijdrovende schaalstap ervaren in huidige productielijnen. Voor deze schaalstap zijn vier testcases geselecteerd, zodanig dat de meest voorkomende generalisatieproblemen in de testcases aanwezig zijn. De vier testcases met start- en doelschalen alsook de NMA van de testcases zijn weergegeven in tabel 1. Per testcase zijn ook aanvullende lagen aangewezen om zeker te weten dat alle interessante situaties door een van de testcases gedekt zouden zijn. Figuur 1 presenteert uitsneden van de testcases en voor iedere testcase een voorbeeld van de uiteindelijk gegeneraliseerde output. Na het selecteren van de testcases, hebben we gekeken hoe generalisatiespecificaties van NMAs op een eenduidige en software-onafhankelijke manier beschreven kunnen worden. We hebben ervoor gekozen om aan te sluiten bij een benadering die veel gebruikt wordt in generalisatiemodellen. Namelijk het

nauwkeurig omschrijven van het gewenste generalisatieresultaat door middel van condities (bijvoorbeeld 'een object moet de belangrijkste vormkenmerken behouden' of 'de minimale omvang van een gebouw moet 4mmx4mm op de kaart zijn'), zonder details te geven over hoe dit resultaat verkregen zou kunnen worden, d.w.z. het proces. In een template konden de NMAs verschillende eigenschappen van de condities aangegeven, zoals het type object waar de conditie voor geldt (bijvoorbeeld 'gebouw'), de conditie zelf, het belang van de conditie vergeleken met andere condities etc (zie tabel 2).

| Characteristic of the constraints | Example on one object | Example on two objects | Example on group of objects |
|---|--|---|---|
| Constraint ID | ICC-1-22 | ICC-2-21 | ICC-3-18 |
| Geometry type | Polygon | polygon – line | Polygons |
| Feature class 1 | Quay adjacent to sea | Building | Building |
| Condition for object being concerned with this constraint | Depth of protrusion > 1 map mm | distance between building and road < 0.5 map mm | |
| Constrained property | width of protrusion/recess | orientation | density of buildings (black/white ratio) |
| Condition depends on initial value? | No | Yes | Yes |
| Condition to be respected | target width > 0.2 map mm | building must be parallel to road | target density should be equal to initial density \pm 20 % |
| Action | collapse to a line | | |
| Importance of constraint (1 to 5, 1 is less important) | 3 | 3 | 3 |
| Exception | | | |
| Schema to illustrate if needed | | | |
| Additional for constraints on two objects: | | | |
| Feature class 2 | | Road | |
| Condition for both objects being concerned with this constraint | | objects are parallel (\pm 15°) | |
| Additional for constraints on group of objects: | | | |
| Kind of group | | | Urban block |
| Kind of objects of the initial data composing the group | | | buildings surrounded by minimal cycle of roads (in urban areas) |

Tabel 2. Voorbeeldcondities (constraints in Engels) voor ICC testcase in de project template.

Ook kon een actie worden geadviseerd om de conditie te bereiken. Door deze kennis die vaak aanwezig is bij NMAs vast te leggen (naast de conditie zelf die alleen het uiteindelijke resultaat beschrijft) kon deze gebruikt worden in de tests om het beoogde resultaat te bereiken.

In eerste instantie resulteerde het invullen van deze specificatie-templates door de vier afzonderlijk NMAs in ongeveer 250 verschillende condities. Na een eerste check bleek al gauw dat de condities van de vier NMAs vaak eenzelfde situatie betroffen, maar net iets anders beschreven werden. Daarom zijn de condities van de vier testcases geharmoniseerd, waardoor het mogelijk werd de tests eenduidiger uit te voeren en conclusies te trekken over generieke (d.w.z. testcase-overstijgende) generalisatiespecificaties. De harmonisatie leidde tot 45 generieke condities: 21 condities voor individuele objecten (bijvoorbeeld de minimale dimensies van een polygoon), 11 condities voor twee objecten (bijvoorbeeld de minimale afstand tussen twee objecten) en 13 condities voor groepen van objecten (bijvoorbeeld het behoud van de ruimtelijke verdeling van een groep van objecten). De NMAs hebben vervolgens hun eigen generalisatiespecificaties gedefinieerd als specificaties van deze generieke condities. In tabel 2 staan de

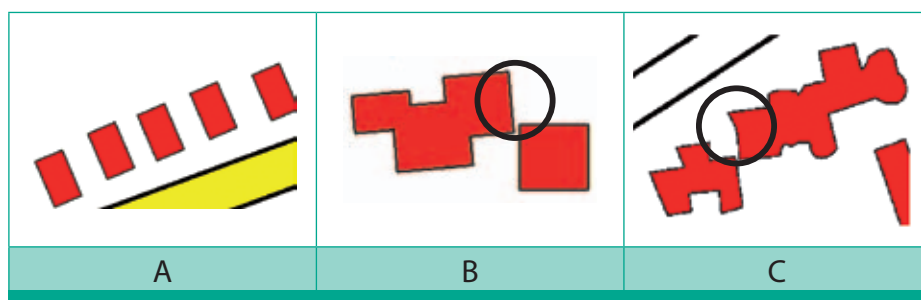


Fig. 2. Minimale afstand tussen gebouwen voorkomt kartografische conflicten in A. Maar in B en C is het maar de vraag of deze minimale afstand een kartografisch conflict voorkomt.

NMA-specifieke aspecten van de condities weergegeven in rood.

Dit harmonisatieproces resulteerde uiteindelijk in ongeveer 300 condities: 50 meer dan aanvankelijk omdat door de harmonisatie-actie NMAs ook gewezen werden op missende condities die wel door andere NMAs waren gedefinieerd.

Als laatste stap voor het eenduidig vastleggen van de generalisatiespecificaties hebben we de vier testcases met elkaar vergeleken om te bestuderen wat voor specificaties de NMAs belangrijk vinden. Hierbij moet de kanttekening geplaatst worden dat de specificaties niet volledig zijn. In de eerste plaats omdat NMAs zich beperkt hebben tot specificaties voor de testcases en tot specificaties die min of meer gemakkelijk gedefinieerd kunnen worden in een conditie. Daarnaast zijn de gedefinieerde specificaties hoofdzakelijk

gebaseerd op bestaande generalisatiespecificaties bedoeld voor kartografen, die overal in het proces zijn/haar interpretatie kan toevoegen. Er is daarbij geen daadwerkelijk automatisch generalisatieproces uitgevoerd wat zou hebben kunnen wijzen op missende of ambigue condities. Een voorbeeld van een ambigue conditie is minimale afstand tussen gebouwen afgebeeld in figuur 2. De minimale afstand tussen gebouwen zou kartografische conflicten moeten uitsluiten, maar het is de vraag of het niet voldoen aan deze conditie in figuur 2 B en figuur 2 C (in tegenstelling tot figuur 2 A) ook echt duidt op een kartografisch conflict. Bij het nader bestuderen van de gedefinieerde condities (zie tabel 3), valt in de eerste plaats op dat ICC veel meer condities gedefinieerd heeft dan de andere drie NMAs. Dit kan worden verklaard doordat

| Testcase | Total number of constraints | Number of objects | | | Constraint type | | | | | | | | Feature classes involved | | | | | | | |
|--------------|-----------------------------|-------------------|----------------|---------------------|-----------------|------------|----------------------|--------------------------------|-----------|-------------|----------|-----------|---------------------------|-----------|-----------|-----------|-----------|-----------|-----------|------------------|
| | | | | | Model | Legibility | Preservation | | | | | | | | | | | | | |
| | | on one object | on two objects | on group of objects | | | Model generalisation | Min. dimension and granularity | Position | Orientation | Shape | Topology | Distribution / Statistics | Other | Building | Land use | Road | Water | Relief | Coastal features |
| ICC | 137 | 86 | 23 | 28 | 12 | 80 | 0 | 4 | 19 | 12 | 5 | 5 | 39 | 20 | 16 | 25 | 8 | 19 | 9 | 1 |
| Kadaster | 52 | 27 | 21 | 4 | 11 | 18 | 1 | 0 | 1 | 6 | 0 | 15 | 10 | 13 | 23 | 3 | 0 | 0 | 0 | 3 |
| IGNF | 61 | 32 | 15 | 14 | 2 | 15 | 2 | 4 | 15 | 12 | 2 | 9 | 33 | 2 | 12 | 9 | 2 | 0 | 2 | 1 |
| OSGB | 49 | 24 | 13 | 12 | 2 | 16 | 1 | 0 | 0 | 8 | 0 | 22 | 24 | 1 | 8 | 1 | 8 | 0 | 2 | 5 |
| Total | 299 | 169 | 72 | 58 | 16 | 129 | 4 | 8 | 35 | 38 | 7 | 51 | 106 | 36 | 59 | 38 | 18 | 19 | 13 | 10 |

Tabel 3: Typen condities gedefinieerd in de vier testcases voor het generalisatieproject

deze dataset meer thematische lagen bevat (zie tabel 1). Er zijn bijvoorbeeld meerdere soortgelijke condities gedefinieerd voor verschillende wegtypen gerepresenteerd in meerdere lagen. In de tweede plaats valt op dat in alle testcases de meeste condities gedefinieerd zijn voor geïsoleerde objecten en de minste voor groepen van objecten. Zeer waarschijnlijk omdat het definiëren van condities voor twee en voor groepen objecten veel complexer is. Een derde opvallend aspect is dat condities voor minimale dimensies van objecten domineren in alle vier testcases, evenals condities voor gebouwen, wegen en landgebruik. De redenen hiervoor zouden kunnen zijn dat er al veel bekend is over het automatisch generaliseren van dit soort condities en dat deze objecten het meest voorkomen op kaarten en ook het meest significant zijn voor kaartgebruikers waardoor (interactieve) generalisatie het meest op dergelijke objecten plaatsvindt.

Uitvoeren van de tests

Nadat de specificaties voor automatische generalisatie waren vastgesteld, zijn de tests uitgevoerd. Voor elk systeem is de commerciële versie beschikbaar in juni 2007 door drie tot vier verschillende testers getest op alle vier testcases, waarbij de tester heeft geprobeerd alle gedefinieerde condities van een testcase te vertalen in de verschillende systemen. Hierbij mocht de tester alleen gebruik maken van functionaliteit die in het systeem aanwezig was. De tests zijn uitgevoerd door generalisatie-experts die zowel bekend als nieuw waren met de systemen. De gemiddelde tijdsinspanning van een test was ongeveer 1 week. Naast het testen van de juni-2007 versies, konden de softwareleveranciers parallelle tests uitvoeren tot zomer 2008 waarbij zij gebruik mochten maken van zowel zelfontwikkelde extensies als ontwikkelingen ten opzichte van juni 2007. In totaal resulteerde de tests in 35 outputs (27 gegeneraliseerd door leden van het projectteam en 8 door de softwareleveranciers) en in meer dan 700 thematische lagen.

b. Evalueren van gegeneraliseerde output

Na het testen, was de volgende uitdaging het evalueren van de gegeneraliseerde output (zie figuur 1 voor enkele voorbeelden van output). Idealiter zou de evaluatie

volledig automatisch moeten gebeuren om zoveel mogelijk subjectiviteit uit te sluiten. Er waren echter drie redenen waarom we ons in dit project niet beperkt hebben tot automatische evaluatie. Ten eerste, kartografische generalisatie bevat een bepaalde mate van subjectiviteit. Vaak zijn er meerdere oplossingen mogelijk die moeilijk kwantitatief met elkaar te vergelijken zijn. Ten tweede, niet alle condities waren voldoende formeel te beschrijven, bijvoorbeeld de conditie dat 'gebouwen hun kenmerkende vorm moeten behouden' is moeilijk automatisch te evalueren. Tenslotte is automatische evaluatie (nog) niet geschikt om condities ten opzichte van andere condities te evalueren, soms is het bijvoorbeeld gewenst om een kleine afwijking van de minimale

grootte van een gebouw toe te staan om de minimale afstand tussen twee gebouwen te kunnen behouden. In dit project is daarom gekozen om de output vanuit drie benaderingen te evalueren, waardoor het mogelijk is de resultaten van de ene benadering te valideren c.q. te ontcrachten met de resultaten van de andere benadering. De drie benaderingen zijn: 1) automatische evaluatie van een select aantal (voldoende formele) condities, 2) een evaluatie waarbij outputs van één bepaalde testcase visueel met elkaar en met de gedefinieerde condities zijn vergeleken, en 3) expertevaluatie waarbij kartografische experts de outputs in detail hebben beoordeeld.

Voor de automatische evaluatie is een prototype ontwikkeld door de Universiteit van Zürich dat berekent in hoeverre wordt voldaan aan de volgende condities: minimale omvang van gebouwen, minimale afstand tussen gebouwen en minimale afstand tussen wegen en gebouwen. Om meer inzicht te krijgen in de uitkomsten van deze evaluatie is het prototype eerst toegepast op interactief gegeneraliseerde data schaal 1:50.000 van het Kadaster. Deze test bevestigde dat veel situaties die als 'slecht' worden beoordeeld door de computer, niet altijd slechte generalisatieoplossingen hoeven zijn. Ten eerste omdat de parameters uit handleidingen komen voor kartografen.

Deze zijn bedoeld als richtlijn voor menselijke interpretatie maar worden in het automatische proces gebruikt

Kartografische generalisatie bevat een bepaalde mate van subjectiviteit

als harde waarde. Zo waren bijvoorbeeld veel gebouwen die als 'te klein' werden gedetecteerd slechts minimaal kleiner dan het gedefinieerde minimum (zie figuur 3). Blijkbaar leveren deze 'te kleine' gebouwen nog steeds een acceptabele oplossing. Een tweede reden waarom acceptabele oplossingen soms als 'slecht' worden beoordeeld door automatische evaluatie is dat de geformaliseerde condities niet voldoende rekening houden met alle situaties die zouden kunnen voorkomen, zoals aangegeven in figuur 2. Ook kan het omgekeerde gelden: goed beoordeelde situaties hoeven niet altijd goede oplossingen te zijn. Bijvoorbeeld: alle gebouwen voldoen aan de 'minimale

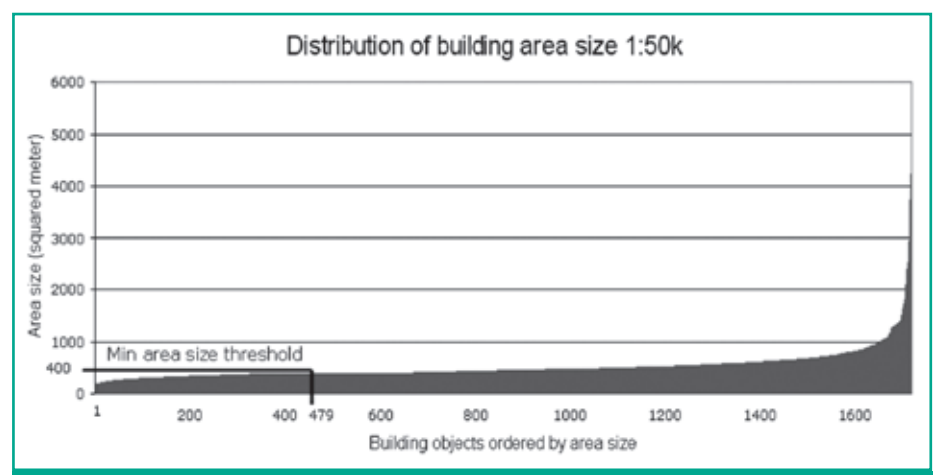


Fig. 3. Resultaat van evaluatie van de grootte van interactief gegeneraliseerde gebouwen in Kadaster 1:50.000 dataset. Een groot aantal 'te kleine gebouwen' is net iets kleiner dan de drempelwaarde.

grootte' conditie, maar een (te) groot aantal gebouwen is verwijderd (zie figuur 4). Dit laat zien dat het heel belangrijk is condities in hun context te evalueren. De meeste tijd voor de automatische evaluatie is uiteindelijk opgegaan in het ontwerp en ontwikkelen van het prototype, met name doordat het prepareren van de outputs voor deze evaluatie (wat niet te automatiseren was) veel meer tijd kostte dan verwacht. Uit de resultaten

Slechte situaties volgens de computer zijn niet altijd slechte oplossingen

konden wel enkele conclusies worden getrokken. Ten eerste, alle systemen behalen goede resultaten voor gebouwen, behalve in dichtbebouwde gebieden, ook al zaten er soms veel verschillen in de outputs van eenzelfde testcase. Ten tweede alleen CPT en Axpand scoren goed op de 'minimale afstand tussen gebouwen' conditie. Dit komt zeer waarschijnlijk doordat deze twee systemen als enige 'verplaatsing' (displacement in het Engels) ondersteunen (ook aanwezig in de nieuwe versies van Clarity en ESRI). Een andere

conclusie is dat, niet geheel onverwacht, goede resultaten voor de 'minimale afstand' conditie vaak gepaard gaan met een groot aantal verwijderde gebouwen. Ook is het zo dat de kenmerken van de originele data (bijvoorbeeld data- en informatiedichtheid) in belangrijke mate de scores van automatische evaluatie bepalen. Dit zijn belangrijke inzichten bij

het interpreteren van de (toekomstige) resultaten. Het visueel vergelijken van de

generalisatie outputs is gedaan door Cecile Duchêne, een ervaren kartograaf van IGN, Frankrijk. Zij heeft voor alle vier testcases een aantal focuszones gedefinieerd waarin ze de outputs met elkaar heeft vergeleken. Zie figuur 5 voor de focuszones die gebruikt zijn in de Kadaster-testcase. Uit deze vergelijking volgt dat geen van de vier testcases volledig is opgelost in de tests. Vooral contextuele generalisatie blijkt een heikel probleem. Daarnaast bestaan voor veel andere problemen weliswaar algoritmen, maar het blijkt zeer moeilijk de juiste parameters te

zetten die voor de gehele testcase optimaal resultaat leveren. Enerzijds komt dit omdat het effect van de parameters vooraf vaak onduidelijk is, maar anderzijds ook omdat tools ontbreken om bepaalde contexten te detecteren waarvoor context-specifieke parameters gebruikt kunnen worden, zoals 'stedelijk gebied'. In ieder geval ontbreekt vaak een match tussen de functionaliteit van de software en de door de NMAs gedefinieerde condities. Een andere conclusie van deze visuele vergelijking is dat outputs voor eenzelfde testcase heel verschillend kunnen zijn. Dit komt deels door het parameteriseringsprobleem, maar ook omdat het niet altijd mogelijk bleek om het verwachte resultaat eenduidig in condities vast te leggen. Dit leidde tot significante verschillen tussen testers die bekend waren met de specifieke testdata, en een idee hadden wat er bereikt moest worden, en testers die nieuw waren met de data. In de expertevaluatie (uitgevoerd door Connie Blok, ITC) hebben kartografische experts zowel gekeken naar globale indicatoren als de individuele condities. Ook hebben de experts de outputs gemarkeerd met voorbeelden van situaties die opvallend goed, opvallend slecht of heel verschillend waren opgelost (zie figuur 6). In tegenstelling tot de automatische generalisatie konden hierbij de individuele condities wel in hun context worden beoordeeld. Volgens de respondenten, die alleen de outputs van hun eigen NMA hebben bekeken, scoren de outputs goed op 'afwijking ten opzichte van de originele data' en 'behoud van geografische kenmerken'. Dat is niet verwonderlijk omdat ook het merendeel van de outputs als 'ondergegeneraliseerd' werd beoordeeld. Dat wil zeggen doordat er niet genoeg generalisatie was toegepast, kon de output nog sterk lijken op de originele data. De kartografische experts waren van mening dat individuele gebouwen/wegen redelijk goed tot goed waren gegeneraliseerd. De andere condities scoorden laag, vaak zelfs slecht. Een belangrijke reden hiervoor is dat de experts bij het beoordelen van de outputs de papieren kaart impliciet als referentie gebruikten. Het zou interessant zijn geweest om de experts ook te hebben geconfronteerd met een interactief gegeneraliseerde output in een blinde beoordeling. In een toekomstig project moet dit zeker worden overwogen.

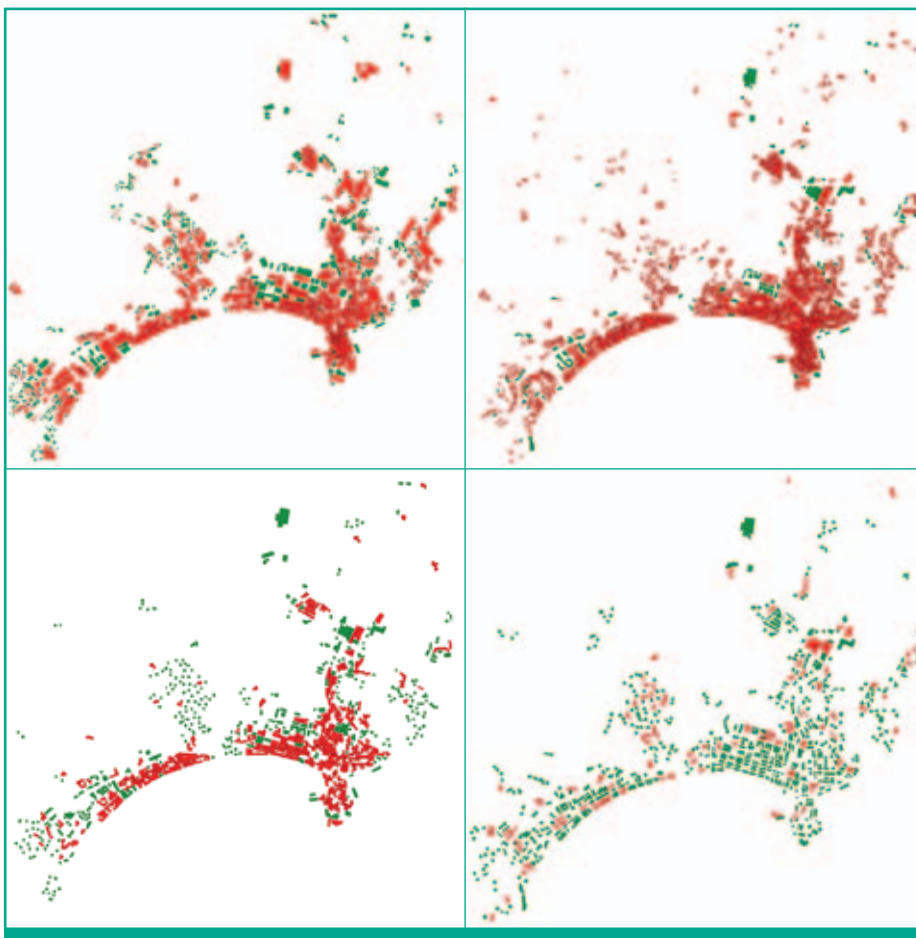


Fig. 4. Resultaat van 'te kleine gebouwen' (in rood) in vier test outputs van ICC testcase. Figuur rechtsonder laat het beste resultaat zien, maar in deze dataset zijn eveneens veel gebouwen verwijderd.

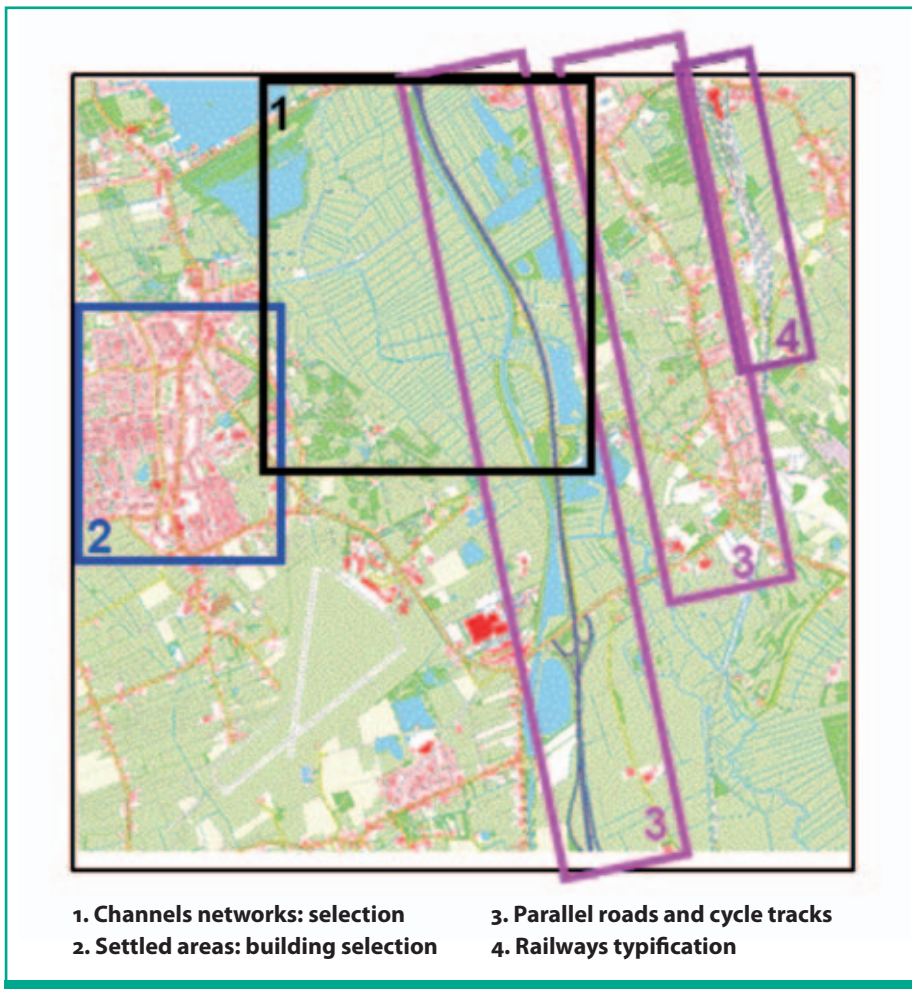


Fig. 5. Focuszones in de Kadaster-testcase die gebruikt zijn om de testoutputs kwalitatief met elkaar te vergelijken.

Conclusies: tegenvallende resultaten?

De resultaten lijken niet heel erg goed. Het is echter belangrijk om deze in het juiste perspectief te zien. In de eerste plaats waren de ambities van het project erg hoog: de generalisatiespecificaties waren geformuleerd in een groot aantal precieze condities, de testcases bevatten een selectie van bekende problemen en de focus was op de productie van papieren kaarten, vergelijkbaar met bestaande producten. Ondanks de tegenvallende resultaten, blijkt uit ons project wel dat een belangrijk deel van het productieproces kan worden geautomatiseerd, helemaal als digitale producten het eindresultaat zijn. Een andere relativering voor de tegenvallende resultaten is dat enkele gesignaleerde problemen opgelost zijn in de parallele tests van de softwareleveranciers, bijvoorbeeld het 'slim' weggooien van gebouwen (zodat het patroon behouden blijft) en verplaatsing van gebouwen en wegen in ArcGIS en Radius Clarity. Ook hebben alle deelne-

mende softwareleveranciers aangegeven blij te zijn met het project, omdat het intern tot versnelde ontwikkelingen heeft geleid. Zo zal in de volgende versie van



Fig. 6. Testoutput waarin door de kartograaf situaties zijn aangegeven welke volgens hem/haar slecht ('B'), goed ('G') of zeer verschillend ten opzichte van andere output ('D') zijn opgelost.

ArcGIS (9.4) een optimalisatietechniek ten behoeve van generalisatie beschikbaar zijn. Tenslotte is het ook niet een verrassing dat out-of-the-box software niet in staat is het automatische generalisatieprobleem van NMAs geheel op te lossen. De systemen die getest zijn, worden momenteel dan ook wel naar tevredenheid gebruikt door een aantal topografische diensten die aanzienlijk geïnvesteerd hebben in het uitbreiden van de software. Zo wordt Change, Push en Typify door verschillende NMAs in Duitsland gebruikt en door ICC, Barcelona. Radius Clarity wordt gebruikt in een aantal Duitse NMAs waarbij een basis dataset schaal 1:10.000 automatisch wordt gegeneraliseerd naar 1:50.000 database (zie <http://www.1spatial.com/pdf/adv1.pdf>). Voorbeel-

De resultaten lijken niet heel erg goed

den van gebruik Aexpand voor productie zijn te vinden op <http://www.axes-systems.com/content/gallery> en <http://www.thueringen.de/de/tlvermgeo/>. Deze laatste link verwijst naar NMA Thüringen (Duitsland) die de software gebruikt voor de productie van 1:25.000 en 1:100.000 kaarten.

Toekomstig onderzoek

Generalisatie is niet een lineair proces dat gegeven dezelfde input en specificaties leidt tot dezelfde (en geheel te verwachten) resultaten. Om er voor te zorgen dat

het resultaat zoveel mogelijk voldoet aan de verwachtingen, is het van belang gebleken dat NMAs hun specificaties formeel beschrijven, ook al blijkt uit ons project dat dit veel tijd kost. Het project heeft een eerste begin gemaakt met het formaliseren van de generalisatiespecificaties van NMAs. Op basis van de resultaten (bijvoorbeeld beschrijft de conditie de specificatie voldoende?) kunnen deze specificaties gecompliceerd, verfijnd en verbeterd worden voor een toekomstig project. In een toekomstig project is het ook belangrijk om meer inzicht te krijgen in het effect en onderlinge wisselwerking van de specificaties, zowel om het proces tussentijds bij te sturen als om het eindproduct te kunnen evalueren. Ook moet gekeken worden naar een andere manier om generalisatiespecificaties te beschrijven, omdat het gewenste resultaat niet altijd in condities gevangen kon worden. Dit leidde in ons project soms tot verschillende interpretaties en daardoor tot verschillende oplossingen. Een toekomstig project zou ook kunnen kijken naar problemen van grote datasets zoals 'computational complexity' en 'memoryoverflows' waardoor het nodig is de dataset (slim) op te delen. Een ander probleem van grote datasets is de mogelijke aanwezigheid van bijzondere gevallen waardoor sommige algoritmen vastlopen.

De resultaten bevestigen dat het altijd nodig zal zijn om de mogelijkheden van de systemen aan te passen aan de specifieke eisen van NMAs. Dit vraagt aan de ene kant de bereidheid van NMAs om te investeren in kennis van de software en in het formeel beschrijven van hun specificaties voor automatische generalisatie (bijvoorbeeld het definiëren van een interval rond drempelwaarden). Aan de andere kant zullen de softwareleveranciers moeten investeren in systemen die makkelijk uit te breiden zijn en die eenvoudig te tunen zijn.

Systemen zouden de gebruikers bijvoorbeeld kunnen helpen bij het kiezen van de juiste parameters door defaultwaarden aan te bieden, door gelijke contexten te selecteren die gegeneraliseerd kunnen worden met dezelfde parameters of door situatie-specifieke parameters aan te bieden. In ieder geval is er samenwerking nodig tussen onderzoekers, NMAs en softwareleveranciers, zoals gerealiseerd in dit

project, om wensen en mogelijkheden zo nauw mogelijk op elkaar af te stemmen.

Dankwoord

Graag wil ik alle personen bedanken die vanuit Nederland een belangrijke bijdrage hebben geleverd aan dit project en die bovenaan dit artikel genoemd zijn. Daarnaast wil ik alle personen uit Engeland, Denemarken, Frankrijk, Duitsland en Spanje bedanken die op enig moment tijdens dit project betrokken zijn geweest. Een aantal personen uit deze groep vormde het kernteam van het project. Hun bijdragen aan de bijeenkomsten (3 maal per jaar), de email-discussies, de tests en de afzonderlijke onderdelen van het project waren onmisbaar. Dit zijn: Blanca Baella en Maria Pla (ICC, Catalonië), Connie Blok (ITC, Enschede, NL), Dirk Burghardt (TU Dresden, voorheen Universiteit Zürich), Cécile Duchêne en Guillaume Touya (IGN, Frankrijk) en Nicolas Regnauld (Ordnance Survey, Verenigd Koninkrijk). ♡

Referenties

- Stoter, J.E., D. Burghardt, C. Duchêne, B. Baella, N. Bakker, C. Blok, M. Pla, N. Regnauld, G. Touya, S. Schmid Methodology for evaluating automated map generalization

in commercial software, Pages 311-324 In: Computers, Environment and Urban Systems Volume 33, Issue 5, September 2009.

- Stoter, J.E., State-of-the-art of automated generalisation in commercial software, Final report, in productie, zal spoedig beschikbaar komen via EuroSDR (www.EuroSDR.net).

Samenvatting

Dit artikel beschrijft het EuroSDR generalisatieproject waarin tests zijn uitgevoerd op vier testcases met representatieve generalisatieproblemen, gedefinieerd door vier Europese National Mapping Agencies (NMAs). Het doel was om te kijken in hoeverre huidige commerciële software in staat is automatische generalisatie uit te voeren. Hoewel de resultaten potenties tonen, vallen ze tegen als gekeken wordt naar de vooraf gedefinieerde generalisatiespecificaties. Dit komt vooral door de complexiteit van het vinden van de juiste parameters alsook van het voldoende formeel definiëren van generalisatiespecificaties. Systemen zouden daarom moeten verbeteren in gebruikersvriendelijkheid om functionaliteit te tunen met specifieke testcases en NMAs zouden moeten investeren om bestaande systemen uit te breiden naar specifieke NMA-behoefte.

Kamervragen over boekcensuur



PvdA-Tweede Kamerlid Lia Roefs wil opheldering van minister Camiel Eurlings (Verkeer) over een kritisch boek over Rijkswaterstaat. Onderzoeksjournalist Marcel Metze gaat in het boek *Veranderend* getij in op de reorganisatie die Rijkswaterstaat sinds 2003 heeft doorgemaakt.

Met het resultaat was de dienst niet tevreden. De organisatie wilde uiteindelijk alleen een gekuiste versie uitbrengen.

Metze ving bot bij de rechter met een verzoek om de ongekuiste versie te mogen publiceren.

Roefs noemt het vreemd dat Metze bij het begin van zijn onderzoek de vrije hand kreeg, maar dat daarna de afwijkende (negatieve) meningen uit het manuscript werden verwijderd. Het boek zou nu in aangepaste vorm zelf worden uitgegeven door Rijkswaterstaat. De PvdA-ster vraagt zich af waarom belastinggeld wordt uitgegeven aan een gecensureerd boek, terwijl het volgens haar de bedoeling was een openhartig en onafhankelijk boek te laten schrijven. Ten slotte vraagt zij de minister om het manuscript ter inzage naar de Kamer te sturen. "Ik wil graag lezen wat er fout gaat bij Rijkswaterstaat", aldus Roefs.

Bron: ANP, 10 december 2009