

# Measuring the performance of geosimulation models by map comparison

## Citation for published version (APA):

Hagen-Zanker, A. (2008). *Measuring the performance of geosimulation models by map comparison*. Universiteit Maastricht.

## Document status and date:

Published: 01/01/2008

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# Samenvatting

Geosimulatie is een tak van geografie die ruimtelijke patronen en dynamiek probeert te doorgronden als het gevolg van de wisselwerking van kleine eenheden zoals huurders, landbezitters, weggebruikers, bomen, enzovoorts. Geosimulatiemodellen bepalen enkel hoe deze eenheden zich gedragen en de patronen op grote schaal zoals segregatie, verstedelijking, verkeersopstopping en bosbrand zullen spontaan ontstaan als het gevolg van de kleinschalige wisselwerkingen. Geosimulatie heeft een grote ontwikkeling doorgemaakt onder invloed van de mogelijkheden die steeds snellere computers bieden. Een eenvoudige bureaucomputer kan dienen als een virtueel laboratorium waarin onderzoekers hun eigen steden, verkeerssystemen, plattelandsgemeenschappen en natuurgebieden kunnen kweken.

Geosimulatiemodellen worden in toenemende mate gebruikt voor gebiedsspecifieke toepassingen. De onderzoekers creëren hun virtuele geografische systeem niet langer uit het niets, maar nemen een startsituatie op basis van werkelijke informatie. Van daaruit kunnen mogelijke toekomstige ontwikkelingen doorgerekend worden. Deze gebiedsspecifieke toepassingen zijn niet enkel wetenschappelijk van aard, maar ze worden ook gebruikt in beslissingsondersteunende systemen die inzicht bieden in te verwachten ontwikkelingen en de consequenties van verschillende opties.

De nieuwe generatie van geosimulatiemodellereurs ziet zich geconfronteerd met het probleem het realiteitsgehalte van de modellen te beoordelen. Richtlijnen voor 'goed modelleren' schrijven een aantal analytische stappen voor. Met name de stappen calibratie en validatie vereisen fitmaten die aangeven hoe goed de modelsituatie overeenkomt met de werkelijke situatie. Aangezien de uitvoer van de modellen overwegend kaarten zijn, ligt het voor de hand om ze te beoordelen op basis van kaartvergelijking. Het karakter van geosimulatiemodellen werpt echter een aantal obstakels op.

Eén probleem is dat de resolutie waarop het model is gedefinieerd niet overeenkomt met de schaal waarop de resultaten worden geïnterpreteerd. De interesse in de modellen zit juist in de ruimtelijke structuren die opborrelen uit de wisselwerkingen tussen de kleine eenheden. Het idee van complexiteit speelt hierbij een belangrijke rol. Wanneer het gedrag van elementen in het model wederzijds afhankelijk is ontstaan terugkoppelingsprocessen die kunnen leiden tot zelforganisatie, maar ook tot gevoeligheden voor kleine variaties en daarbij chaos en onvoorspelbaarheid. Het gevolg is dat zelfs een geosimulatiemodel dat

de werkelijke processen perfect beschrijft niet in staat zal zijn om kaarten te produceren die tot in detail met de werkelijkheid overeenstemmen. Het is daarom niet afdoende om modellen te beoordelen op plek-tot-plek overeenstemming maar ook de patronen die ontstaan, moeten beoordeeld worden. Desalniettemin, bij gebiedsspecifieke toepassingen reikt het te ver om modellen enkel op globale patronen te beoordelen, ook de plaatsgebonden verdeling is van belang.

Aldus wordt een balans gezocht tussen het vinden van realistische patronen en het vinden van deze patronen op de juiste plek. Geosimulatiemodellen worden geacht ruimtelijke structuren te creëren die realistisch zijn en ongeveer op de juiste plaatsen te vinden zijn. Bestaande vergelijkingsmethoden vinden een dergelijke balans niet. Zij zijn of lokaal, gebaseerd op plek-tot-plek overeenstemming, of het ander uiterste namelijk globaal, gebaseerd op maten die het hele landschap in een enkel getal samenvatten. Alhoewel formele methoden ontbreken, kan een expert een dergelijke vergelijking vrij eenvoudig maken door de kaarten te bekijken. Het is daarom niet vreemd dat in de dagelijkse praktijk geosimulatiemodellen vaak worden beoordeeld op basis van een dergelijke zichtvalidatie.

Zichtvalidatie is niet zonder problemen, het meest fundamenteel is het gebrek aan objectieve herhaalbaarheid. Een praktisch bezwaar is dat voor sommige taken, zoals calibratie, grote aantallen consistente beoordelingen nodig zijn. Een menselijke beoordelaar kost dan veel tijd en geld en zal waarschijnlijk te veel inconsistenties introduceren.

Dit proefschrift behelst een raamwerk voor het evalueren van modelprestaties. Het bestaat uit een aantal vergelijkingsmethoden die georganiseerd zijn aan de hand van twee assen. De eerste as maakt een onderscheid dat typisch wordt gemaakt in geografische informatiewetenschappen. De as is gebaseerd op het schaalniveau van de analyse-eenheden en loopt van lokaal, via focaal naar globaal. De tweede as is gebruikelijk in landschapsecologische toepassingen en onderscheidt of de vergelijking wordt gemaakt op basis van ruimtelijke structuur of de aanwezigheid van ruimtelijke elementen.

De Kappa statistiek wordt veel gebruikt om de overeenstemming in paarsgewijze categorische observaties uit te drukken. Het houdt geen rekening met ruimtelijke relaties, behalve plek-tot-plek (rastercel-tot-rastercel) overeenstemming, en is niet specifiek bedoeld voor het vergelijken van kaarten. De statistiek schaal het percentage van overlap naar de verwachte overlap, gegeven het totale aantal observaties per categorie. Hierdoor vermengt de statistiek lokale (plek-tot-plek overeenstemming) en globale (de verwachte overeenstemming aspecten van de observaties). Dit proefschrift stelt een nieuwe uitsplitsing van de Kappa statistiek voor die Kappa uitdrukt als het product van twee componenten. De eerste component, Klocation, is een maat voor

overeenstemming van de *lokale aanwezigheid*, en de tweede component, Khisto, is een maat voor de overeenstemming van de *globale aanwezigheid*.

In een verdere uitbreiding op de Kappa statistiek wordt nabijheid in rekenschap genomen. Onbestemdheid (fuzziness) van lokatie wordt in de analyse opgenomen zodat gelijke categorieën die op ongeveer dezelfde plek liggen een zekere mate van gelijkheid inhouden. Modellen die worden geëvalueerd op basis van de nieuwe Fuzzy Kappa worden aldus positief aangerekend wanneer ze categorieën op ongeveer de juiste locaties plaatsen. De Fuzzy Kappa is daarmee een maat van gelijkheid voor *focale aanwezigheid*.

Naast onbestemdheid van locatie houdt de Fuzzy Kappa statistiek ook rekening met de onbestemdheid van categorieën. Het vindt sommige combinaties van categorieën meer gelijkend dan andere. Dit aspect van de Fuzzy Kappa maakt het bijzonder geschikt voor het wegen van verschillende aspecten van ongelijkheid, bijvoorbeeld fouten door omissie en commissie kunnen afzonderlijk gemeten en gewogen worden. Een positief neveneffect is dat de Fuzzy Kappa statistiek geschikt is om kaarten met ongelijke legenda's te vergelijken.

Landschapsmaten zijn veelgebruikt om structuur, in het bijzonder configuratie en samenstelling van kaarten, te beschrijven. De structuur is echter niet homogeen over de hele kaart en om ruimtelijke variabiliteit in rekenschap te nemen worden vergelijkingsmethoden voorgesteld op basis van een verschuivend kader. Deze methoden creëren een extra kaartlaag waarin iedere rastercel op de kaart de waarde krijgt van de landschapsmaat die hoort bij het kader dat gecentreerd is op die cel. Deze landschapsmaat-kaartlagen worden vervolgens vergeleken op basis van een cel-per-cel vergelijking. Door deze aanpak wordt het geëvalueerde model beoordeeld op de mate waarin (ongeveer) de juiste structuren worden gevonden op (ongeveer) de juiste locaties. De kaartvergelijking is dus van het type *focale structuur*.

De eerste verkenning van een methode voor het vergelijken van veranderingspatronen in plaats van statische eindkaarten wordt gepresenteerd. Dit is een veelbelovende aanpak aangezien het geosimulatiemodellen aanspreekt op hun dynamische karakter. De methode is gebaseerd op een toestandsruimte benadering die veranderingen in verscheidene aspecten van ruimtelijke structuur registreert en vergelijkt. Het is nog niet duidelijk in hoeverre de discretisering van de toestandsruimte naar afgebakende klassen de resultaten een structurele afwijking geeft. Verder onderzoek naar dit type vergelijking is aanbevolen, niet in het minst omdat het een vergelijking is van *globale structuur*. Het globale karakter houdt in dat de dynamiek van verschillende gebieden onderling vergeleken kan worden, hetgeen de basis kan zijn voor een klassificatie van stedelijke veranderingspatronen.

Een tweede obstakel bij het beoordelen van geosimulatiemodellen op basis van kaartvergelijking komt voort uit het dynamische karakter van de modellen en de relatief geringe veranderingen die over een simulatieperiode kunnen optreden.

Hierdoor kan het voorkomen dat een hoge mate van overeenstemming tussen modelresultaat en werkelijkheid niet zozeer een gevolg is van de kwaliteit van het model, maar simpelweg een aanwijzing is dat er weinig veranderd is over de simulatieperiode. Het risico op een verkeerder interpretatie en te groot vertrouwen in geosimulatiemodellen is hierdoor aannemelijk.

Een overoptimistische interpretatie van modelresultaten kan worden voorkomen door modellen te evalueren relatief ten opzichte van referentiemodellen. Deze referentiemodellen zijn onderhevig aan dezelfde randvoorwaarden en beperkingen als het geëvalueerde model en bevatten verder geen specifieke processen behalve een weerstand tegen verandering. Daardoor zijn het neutrale modellen van landschapsverandering. Het verschil in prestatie tussen de neutrale modellen en het geëvalueerde geosimulatiemodel kan worden toegeschreven aan de processen die wel in het geosimulatiemodel aanwezig zijn, maar niet in de neutrale modellen. Een bijkomend voordeel van het gebruik van referentiemodellen is dat gelijkheidsmaten over verschillende grootheden en in verschillende eenheden onderling vergelijkbaar worden. Het is daardoor mogelijk om sterke en zwakke punten van het model te herkennen.

Alle nieuw geïntroduceerde methoden zijn toegepast op een aantal casussen, met een nadruk op de validatie van landgebruiksmodellen op basis van Cellulaire Automaten. Uit de casussen blijkt dat de beoordelingen door de nieuwe methoden goed aansluiten op zichtvalidatie, maar ook in staat zijn om patronen van verschillen te herkennen die voor de menselijke beoordelaar moeilijk te herkennen zijn. Het gebruik van referentiemodellen blijkt onmisbaar te zijn. Er is een sterke correlatie tussen de historische verandering en de modelfit en alleen met referentiemodellen is het mogelijk de kritische grenswaarden vast te stellen.

De mogelijke toepassingen van de vergelijkingsmethoden reiken verder dan de validatie van landgebruiksmodellen. Verwante problemen spelen ook in de ecologie, hydrologie en meteorologie. Behalve voor validatie kunnen de methoden ingezet worden om historische veranderingen te karakteriseren en om kaartclassificaties te beoordelen, bijvoorbeeld classificaties op basis van satelietbeelden en luchtopnamen. Uit een korte verkenning van de respons in de wetenschappelijke literatuur blijkt dat deze toepassingen inderdaad plaatsvinden.

Een tot nu toe niet verwezenlijkte mogelijkheid is de integratie van de kaartvergelijkingsmethoden in verdere ruimtelijke analyse. In principe kan iedere analyse met een vergelijking van categorische kaarten baat hebben bij de geografische nuance van de nieuwe methoden. Gezien de uitgangspunten van het onderzoek ligt integratie in calibratieprocedures het meest voor de hand. De methoden kunnen dan worden gebruikt als fitmaat maar ook als steun bij het vinden van de juiste parameteraanpassing. De verscheidene belichtingshoeken

van kaartvergelijking kunnen bijzonder geschikt zijn voor een calibratie op basis van optimaliseroutines met meervoudige doelfuncties.

De classificatie op basis satelietbeelden biedt ook veel ruimte voor integratie van vergelijkingsmethoden. De methoden kunnen bijdragen aan het onderscheiden van werkelijke veranderingen en classificatiefouten en, vooral de focale structuur vergelijkingen, kunnen nuttig zijn voor classificatiemethoden die de context van een pixel in rekenschap nemen.

# Summary

Geosimulation is a field in geography that seeks to understand geographical patterns and dynamics as the consequence of the interactions between individual entities, like tenants, land owners, car drivers, trees, etc. Geosimulation models prescribe the behaviour of these entities and as a product of the interactions between the entities large scale patterns, such as segregation, urban sprawl, road congestion, forest fire, etc. emerge. The field of geosimulation is spurred by advances in computing. As a consequence, the average desk computer can function as a virtual laboratory, where researchers can breed their own virtual cities, transport systems, rural societies, forests, etc.

Geosimulation models are increasingly finding region specific applications. Rather than using the models to let geographical structures arise from scratch, the models are fed with a real initial situation. The model is then used to elaborate possible future developments. Such region specific applications are not only used for scientific purposes. They have become part of decision support systems that help decision makers foresee future development and assess the consequences of alternative decisions.

The problem that confronts the new generation of geosimulation modellers is to assess how well their virtual worlds correspond to reality. 'Good modelling practice' prescribes different analytical steps in the modelling process. Of these, calibration and validation require the modeller to express the goodness-of-fit of the model. Since the results of the models typically are maps, it makes perfect sense to address this question by map comparison; however the nature of geosimulation models provides some particular challenges that need to be considered first.

One problem is that the resolution at which the model is defined is generally not equal to the scale at which the results are interpreted. The interest in the models lies in the geographical structures that unroll as a consequence of the interactions of the individual entities, rather than the entities themselves. The concept of complexity is relevant here. Typically the elements in the model are mutually dependent on each others behaviour. This causes feedback processes and self-organization, but also making the models sensitive to small deviations to the extent that they are chaotic or unpredictable. The consequence is that even a geosimulation model that perfectly captures the dynamics of a geographical system cannot be expected to produce maps that correspond perfectly to reality. The models should therefore not be evaluated just at the

location-to-location level, but in terms of the patterns that emerge. On the other hand, when applying a model for a particular region, one is not just interested in the global patterns, but also how the patterns are distributed in space.

A balance has to be sought between finding realistic patterns and finding them at the right location. The geosimulation model should create spatial configurations that are similar to reality and place them in approximately the right locations. Existing comparison methods do not strike such a balance. They are either local and based on cell-by-cell overlap, or global and based on metrics summarizing the whole landscape in a single value. Despite the lack of formal methods, an expert can make this kind of balanced comparison by just looking at the map. It is therefore not surprising that in practice geosimulation models are often evaluated on the basis of such subjective face validation.

There are several problems associated with face validation, most pressingly the lack of objective reproducibility. A practical concern of face validation is that for some tasks, such as calibration, large numbers of consistent assessments are required. Depending on a human judge of map similarity may be too time-consuming, costly and prone to inconsistencies.

This thesis introduces a framework to evaluate model performance. It applies a number of metrics that can be categorized according to two axes. The first axis is typical of geographical information science and is based on the spatial unit of the analysis; it ranges from local, via focal to global. The second axis is more commonly applied in landscape ecological applications and discerns whether structural aspects of landscape are considered or just the presence of landscape elements.

The Kappa statistic is a frequently used statistic expressing the agreement of pair-wise categorical observations. It does not consider spatial relations except for cell-by-cell overlap and is not specifically aimed at comparing maps; nevertheless it is often used as a map comparison method. The Kappa statistic scales the percentage of agreement between observations to the agreement that could be expected by chance given the number of observations of each class. As such, the statistic confounds local (agreement) and global (expected agreement) aspects of the observation. A new breakdown of the Kappa statistic is proposed that makes it the product of two components: The Klocation statistic that expresses the agreement in *local presence* and the Khisto statistic that expresses the agreement in *global presence* (quantity).

In a further elaboration of the Kappa statistic, proximity is taken into account. Fuzziness of location is incorporated in the analysis such that categories found at approximately the right location are considered similar as well. It thus evaluates agreement in *focal presence*.

Besides fuzziness of location the Fuzzy Kappa statistic also accounts for fuzziness of categories, considering some pairs of categories more similar than



others. This aspect of the comparison method is useful to weight particular aspects of similarity, for instance to separate the errors due to omission and commission. As a side-effect, the Fuzzy Kappa statistic can well be used to compare maps with different legends.

Map patterns are commonly described using landscape metrics that summarize different aspects of configuration and composition. Landscape structure is not homogenous over the map however and to acknowledge this variability, a comparison method on the basis of moving windows is introduced. In this method extra map layers are created that attribute the landscape metric pertaining to the window surrounding the cell to every cell on the map. These metric layers are then compared on a cell-by-cell basis. Effectively, the evaluated model is assessed on the degree to which that the right types of spatial structures are found at the right locations. The comparison is one of *focal structure*.

First results of a method comparing patterns of change instead of the static end-states of a simulation run are presented. This is a promising approach since it brings the analysis of model performance closer to the dynamic nature of geosimulation models which is their essential characteristic. The approach is based on a state-space representation in which the transitions of different aspects of spatial structure are registered and compared. It is not yet clear, however, how the method is affected by the discretization of state-space into classes, which possibly introduces some bias. Further research into this type of comparison is recommended, not in the least since it is a *global structure* comparison. This means that the dynamics of different regions can be compared against each other and the method can thus be the foundation of a classification of landscape change patterns.

Another problem of validating geosimulation models originates from the dynamic nature of the simulation models and the relatively small number of changes that may occur over a simulation period. A high agreement between model and reality is then not necessarily the consequence of the agreement between the processes in model and reality, but can be just an indication that not much has changed. This causes a risk of misinterpretation and false confidence in geosimulation models.

Misinterpretation of model results can be avoided if geosimulation model are evaluated relative to reference models instead. Such models should be subject to the same constraints and boundary conditions as the investigated model but represent no specific processes except for inertia, therefore they are neutral model of landscape change. The difference in performance between the neutral models and the evaluated geosimulation model can then be attributed to the model relations that are present in the geosimulation model but not in the neutral model. A further advantage of expressing model performance relative to that of a reference model is that metrics measured at various scales become

mutually comparable. Thus it becomes possible to identify strengths and weaknesses of the evaluated model.

All new methods proposed here have been tested on a number of cases, with an emphasis on the validation of Cellular Automata based land use models. In these cases it is found that the newly developed methods produce assessments that are consistent with face validation, but also have the capacity to highlight patterns of change that the human analyst may not recognize. The use of reference models proved indispensable. It appears that there is a strong correlation between the amount of historical change and geosimulation models capacity to reproduce this change. Only by using reference models it is possible to find the critical thresholds of model performance.

The potential for application is much wider than the validation of Cellular Automata land use models. There are very similar problems in the fields of ecology, meteorology and hydrology. Besides evaluating simulation models, the methods can be used to characterize historical change or to validate spatial classifications, for instance on the basis of remote sensing. A brief review of the response in the scientific literature shows that all these kinds of applications are indeed being made.

A potential that is still unfulfilled is to integrate these new map comparison method in other spatial analysis. In principle any type of analysis that includes the calculation of fit between two categorical maps has the potential to benefit from the geographically nuanced methods introduced in this thesis. Closest to the original objectives, the methods can be integrated in calibration procedures, not only by calculating a goodness-of-fit of model and reality but also by suggesting parameter values to adjust. The multi-faceted assessment of map similarity may be particularly suited for calibration procedure on the basis of multi-objective optimization.

The classification of remote sensing imagery also offers various possibilities. The methods can contribute to making the distinction between real changes in the landscape and errors in the classification. Furthermore, the focal structure comparisons can be of use in contextual classification methods that take the spatial configuration into account when classifying pixels.