

Het interpoleren van temperatuurgegevens

F.W.J. Salet

KNMI stageverslag

De Bilt, 2009

PO Box 201
3730 AE De Bilt
Wilhelminalaan 10
De Bilt
The Netherlands
<http://www.knmi.nl>
Telephone +31(0)30-220 69 11
Telefax +31(0)30-221 04 07

Auteur: Salet, F.W.J.



HET INTERPOLEREN VAN TEMPERATUURGEGEVENS

Verslag van een onderzoekstage bij het KNMI

F.W.J. (Frank) Salet

Student Master of Geo-Information Science, Wageningen Universiteit
Registratienummer: 85 04 01 724 120
E-mail: franksalet@gmail.com

Begeleiders:

Dr. R. (Raymond) Sluiter [KNMI]
Dr. Ir. S. (Sytze) de Bruin [WUR]

29 juli 2009
De Bilt

Centre for Geo-Information
Wageningen University and Research Centre



WAGENINGEN UNIVERSITEIT
WAGENINGEN UR

R&D Informatie en observatietechnologie
Koninklijk Nederlands Meteorologisch instituut



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Verkeer en Waterstaat

Inhoudsopgave

Introductie	3
Doelstelling	3
Outline	3
Interpolatietechnieken	4
IDW	4
Kriging met Externe Drift	4
Multiple Regressie met IDW	5
Externe variabelen	5
Data	5
Methoden	5
Resultaten	9
Discussie	11
Conclusie	11
Referenties	13
Bijlage A: Kwaliteit van het meetnetwerk	15
Spreiding van de meetstations	15
Regelmatige tussenruimte	15
De variatie in temperatuur	16
De karakteristieken van het terrein	16
Conclusie	17

Introductie

Er is zowel binnen als buiten het KNMI een grote behoefte aan datasets met ruimtelijk geïnterpoleerde klimaatgegevens. Gegridded bestanden van klimatologische grootheden worden gebruikt voor tal van doeleinden, zoals bij informatievoorziening en in impactmodellen. Het KNMI heeft als taak dit soort gegevens van het Nederlandse grondgebied in de vorm van kaarten of gegridded bestanden te produceren en aan te bieden. Om te voldoen aan de vraag naar geïnterpoleerde gegevens is het KNMI een project gestart met als doel het verbeteren van de bestaande methodiek (Sluiter, 2008). Het hier beschreven onderzoek vindt plaats binnen de context van dit project.

Het interpolatieproject richt zich op allerlei klimatologische grootheden (neerslag, temperatuur, wind, zonneschijn, enz.) en op allerlei verschillende temporele resoluties (10 minuten, dag, maand, seizoen, jaar en normalen). Het hier beschreven onderzoek behandelt enkel de temperatuur aan het oppervlak. Voor deze grootheid is onderzoek gedaan naar de optimale interpolatiemethode per tijdsresolutie, waarbij de verschillende tijdsituaties, zeven in totaal, onder te verdelen zijn in drie categorieën (zie Tabel 1). Onder de eerste categorie vallen ook de normalen, dat zijn gemiddelden over een periode van dertig jaar.

Tabel 1: Overzicht van de tijdsperiodes die onderzocht zijn onderverdeeld in drie categorieën.

Categorie	Temporele resolutie	Afnemer vergridde gegevens
1	Meerjarig jaargemiddelde Meerjarig seizoen gemiddelde Meerjarig maandgemiddelde (inclusief de normalen)	Klimaatatlas van Nederland (Heijboer en Nellestijn, 2002)
2	Jaargemiddelde Seizoen gemiddelde Maandgemiddelde	KNMI-website (KNMI, 2009a) en in diverse overzichtspublicaties
3	Daggemiddelde	KNMI-website (KNMI, 2009a; KNMI, 2009b)

Momenteel worden interpolaties van temperatuurgegevens gedaan met een Spline techniek. Deze techniek maakt gebruik van een polynoomfunctie die zo goed mogelijk een trend probeert te fitten door de samplepunten. De Spline wordt vaak een mechanische methode genoemd omdat in veel gevallen de gebruiker kiest voor de standaardparameters die door de software worden voorgesteld (Hengl, 2007). Daaropvolgend vindt in bepaalde gevallen een handmatige postprocessing plaats waardoor het proces moeilijk reproduceerbaar wordt. Voor klimaatgegevens blijkt uit onderzoek dat Spline alleen geschikt is in geval van een hoge sampledichtheid (Hengl, 2007), wat voor Nederland niet het geval is.

Doelstelling

Doel van dit onderzoek is om voor temperatuur andere interpolatietechnieken te testen om zodoende voor iedere tijds categorie een optimale interpolatiemethode te krijgen. De nieuwe techniek moet resulteren in een verbeterd ruimtelijk patroon, zodat handmatige postprocessing niet nodig is, en een verbeterde methodiek, zodat een hogere nauwkeurigheid kan worden behaald. De interpolatiemethoden die in bij dit onderzoek zijn getest zijn Inverse Distance Weighting (IDW), Multiple Regression (MRegr) en Kriging met Externe Drift (KED). In dit verslag worden alleen de resultaten van de jaarnormalen visueel gepresenteerd. De resultaten van de overige categorieën zijn digitaal beschikbaar via de stagebegeleider van het KNMI.

Outline

In dit verslag wordt eerst ingegaan op de interpolatietechnieken die getest zijn op de temperatuurgegevens. Een beschrijving van de technieken en allerlei bijkomende aspecten worden gegeven. Vervolgens wordt kort ingegaan op enkele eigenschappen van de waarneemgegevens die we als input gebruiken. Daarna komen de methoden aan bod die gebruikt zijn bij het uitvoeren van dit onderzoek. Daaropvolgend worden de

onderzoekresultaten beschreven. Het laatste onderdeel van dit verslag vormen de discussie en conclusie.

Interpolatietechnieken

Er zijn voor het maken van vergridde datasets met temperatuurgegevens veel verschillende methoden te gebruiken (Tveito *et al.*, 2007). De meest optimaal presterende interpolatiemethode varieert volgens Vicente-Serrano *et al.* (2003) als een functie van het oppervlak en de ruimtelijke schaal dat gebruikt wordt bij het vergriden. Daarnaast is ook van belang welke temporele resolutie gebruikt wordt en de karakteristieken van de klimatologische grootte (Hofstra *et al.*, 2008). Zo is temperatuur, in tegenstelling tot bijvoorbeeld neerslag, voor alle tijdsresoluties een continu element in ruimte en tijd. Andere aspecten die de keuze van interpolatiemethode kunnen beïnvloeden zijn de karakteristieken van het netwerk van waarnemstations, zoals stationsdichtheid, stationsverdeling en representativiteit van het meetnetwerk (Tveito *et al.*, 2006). Deze aspecten zijn bekeken in dit onderzoek en de resultaten daarvan worden in dit verslag gegeven. De laatste aspecten die mogelijk van invloed zijn bij het kiezen van de optimale methode zijn de belangrijkheid van geografische factoren (Perry and Hollis, 2005a), zoals landhoogte, afstand tot zee of open water, atmosferische circulatie en landgebruik. Sommige interpolatiemethoden hebben de mogelijkheid om deze geografische factoren als beschrijvende variabelen te betrekken om voorspellingen te verbeteren. Een andere mogelijkheid is om anomalieën te interpoleren (Hofstra *et al.*, 2008). Daarbij wordt het verschil met een lange termijn gemiddelde geïnterpoleerd (normaliseren) om vervolgens het interpolatieresultaat weer op te tellen bij het gemiddelde (denormaliseren). Het interpoleren van anomalieën is in de klimatologie een gebruikelijk proces om zodoende een onderliggende trend te modeleren (Tveito *et al.*, 2007).

Sluiter (2008) heeft een literatuuronderzoek uitgevoerd om kennis te vergaren over het interpoleren van klimatologische gegevens bij het KNMI. In zijn rapport worden tal van technieken en methoden geanalyseerd. Als interpolatietechniek blijken IDW, KED en MRgr in combinatie met IDW het meest veelbelovend. Voor het interpoleren van lange termijn gemiddelden (eerste categorie) wordt het interpoleren van de absolute waarden aangeraden. Voor het interpoleren van tijdsresoluties in de tweede categorie wordt het interpoleren van anomalieën geadviseerd en voor de daggemiddelden wordt de absolute waarden aangeraden. In Tveito *et al.* (2007) wordt het testen van verschillende interpolatiemethoden voor specifieke doeleinden aangeraden. Daarom worden in dit onderzoek de drie aangewezen interpolatiemethoden geëvalueerd op de temperatuurgegevens voor ieder van de zeven tijdsresoluties.

IDW

Inverse Distance Weighting modellen werken volgens de vooronderstelling dat verder weg gelegen samples een verminderde bijdrage hebben bij het voorspellen van een waarde (Isaaks and Srivastava, 1990). Deze methode is snel te implementeren en eenvoudig te begrijpen (De Smith *et al.*, 2007). IDW is niet in staat te extrapoleren waardoor een voorspelde waarde altijd tussen de hoogste samplewaarde en laagste samplewaarde ligt. De sterkte waarmee de afstand van invloed is kan worden aangepast door een kwadratenfunctie op de afstand toe te voegen. Voor de toepassing van IDW in dit onderzoek is een lineaire functie op de afstand het meest geschikt omdat we gebruik gaan maken van blokvoorspellingen. IDW wordt veelvuldig gebruikt binnen de meteorologie. Voorbeeld hiervan is het interpoleren van anomalieën binnen het REGNIE-project door de Deutscher Wetterdienst (Buishand *et al.*, 2008).

Kriging met Externe Drift

Kriging is een veel gebruikte interpolatietechniek binnen de geo-wetenschappen. Het is een stochastische methode die, net als bij regressie, een *best linear unbiased estimation (BLUE)* uitvoert. De voorspelde waarde is een lineaire combinatie van de bekende punten, op zulke wijze dat de som van de gewichten één is ('unbiased') en de standaardafwijking van de residuen van het geïnterpoleerde vlak geminimaliseerd is ('best estimation'). Het punt met de voorspelde waarde is daarom een lokale functie van de omliggende gegevens, maar bovenal is

het onderworpen aan een bepaald model dat een beschrijving geeft van de ruimtelijke variabiliteit: de semivariogram. Eén van de vele Kriging varianten is Kriging met een Externe Drift, ook wel Universal Kriging genoemd (Isaaks and Srivastava, 1990). Deze variant van Kriging is in staat één of meerdere beschrijvende variabelen bij het voorspellen te betrekken, door middel van een regressie model, om zodoende de voorspelling te verbeteren. Een voorbeeld van het gebruik van KED in de meteorologie is het interpoleren van temperatuur voor Europa in het ENSEMBLES-project (Hofstra *et al.*, 2008).

Multiple Regressie met IDW

Een lineaire regressie functie beschrijft de relatie tussen de voorspelde variabele en één of meerdere beschrijvende variabelen (De Smith *et al.*, 2007). Na het opstellen van een multiple regressiemodel (i.e. met meerdere beschrijvende variabelen) worden de residuen, dat wil zeggen het verschil tussen de samplewaarden en de voorspelde waarden, berekend. De residuen worden vervolgens geïnterpoleerd met IDW. Als laatste worden de resultaten van de interpolatie opgeteld bij de voorspelde variabelen van het regressiemodel. Deze methode wordt onder andere gebruikt door de UK Metoffice voor de productie van vergridde datasets met de gemiddelde maandtemperatuur (Perry and Hollis, 2005a).

Externe variabelen

Bij de interpolatietechnieken KED en Mregr+IDW kan gebruik worden gemaakt van beschrijvende variabelen om de voorspelling te verbeteren. Zoals in het geval van temperatuur: oost-west gradiënt, noord-oost gradiënt, terreinhoogte, afstand tot open water, afstand tot zee en gemiddelde windrichting.

Data

De temperatuurgegevens die gebruikt zijn in dit onderzoek komen uit de database van het Klimaat Informatie Systeem van het KNMI. In deze database zitten temperatuurgegevens van alle waarnemstations voor de periode dat ze in gebruik waren en deze zijn te verkrijgen in allerlei tijdsresoluties. Het netwerk van waarnemstations bevat voor de tijdsresolutie van na 1990 ongeveer 34 samples. Zie Figuur 1 voor de locaties van de Nederlandse meetstations. Voor de normaalperiode is voor een aantal meetstations de tijdreeks niet volledig. In dat geval wordt de ontbrekende waarde benaderd door lineaire vergelijking met daarin de omliggende stations op te stellen. Voor de tijdsresoluties van na 1994 zijn tevens de gegevens van ongeveer tien meetstations in Duitsland en België (grensstations) beschikbaar. Voor de eerste categorie van tijdsituaties zijn de grensstations niet te gebruiken omdat deze geen complete tijdreeks hebben.

Methoden

De uitvoering van het onderzoek vindt plaats in de programmeeromgeving van R. Dit is een opensource softwareprogramma voor een uiteenlopende reeks van statistische berekeningen en grafische technieken (R-Project, 2009). Voor de geostatistische analyses worden de volgende R-modules gebruikt: gstat, sp en maptools. Vooral gstat bevat allerlei tools voor multivariabele geostatistische modellen en voorspellers. R heeft, in tegenstelling tot bijvoorbeeld ESRI's ArcGIS, een goede experimenteeromgeving en is eenvoudig te implementeren in een automatische procesomgeving.

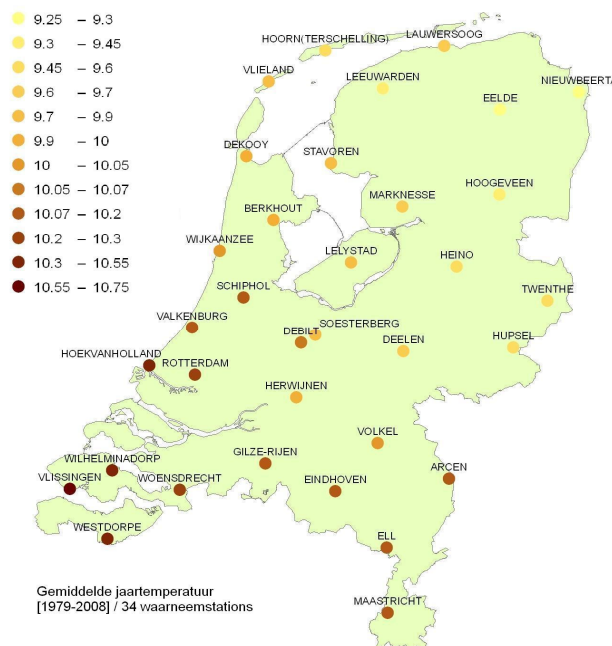
Wanneer de anomalieën in plaats van de absolute waarden gebruikt worden wordt er genormaliseerd door het verschil te berekenen tussen de absolute waarden en een tienjarig gemiddelde van de betreffende periode (i.e. jaar, seizoen, maand). Om vervolgens na het interpoleren te denormaliseren is een vergridde dataset nodig van het desbetreffende tienjarig gemiddelde. Deze dataset valt onder de categorie met de meerjarige gemiddelden en wordt daarom geïnterpoleerd volgens de methode die voor die categorie het meest optimaal is.

Iedere interpolatiemethode brengt een aantal randvoorwaarden met zich mee. Zo zijn de basismodellen van Kriging en IDW niet in staat te extrapoleren, zodat een voorspelde waarde altijd binnen het bereik van de samplewaardes ligt. Het toevoegen van een regressiemodel (zoals met KED en MRegr+IDW) daarentegen maakt dit wél mogelijk. De semivariogram die voor Kriging nodig is wordt in ons onderzoek automatisch gefit door een functie in de Automap-package dat specifiek bedoeld is voor het opzetten van samplevariogrammen. Voor het onderzoek gaan we uit van tweedeorde stationariteit zodat één samplevariogram per Krigingsysteem gebruikt wordt. Ook hebben we voor bepaalde gevallen de samplegegevens getest op de aanwezigheid van anisotropy.

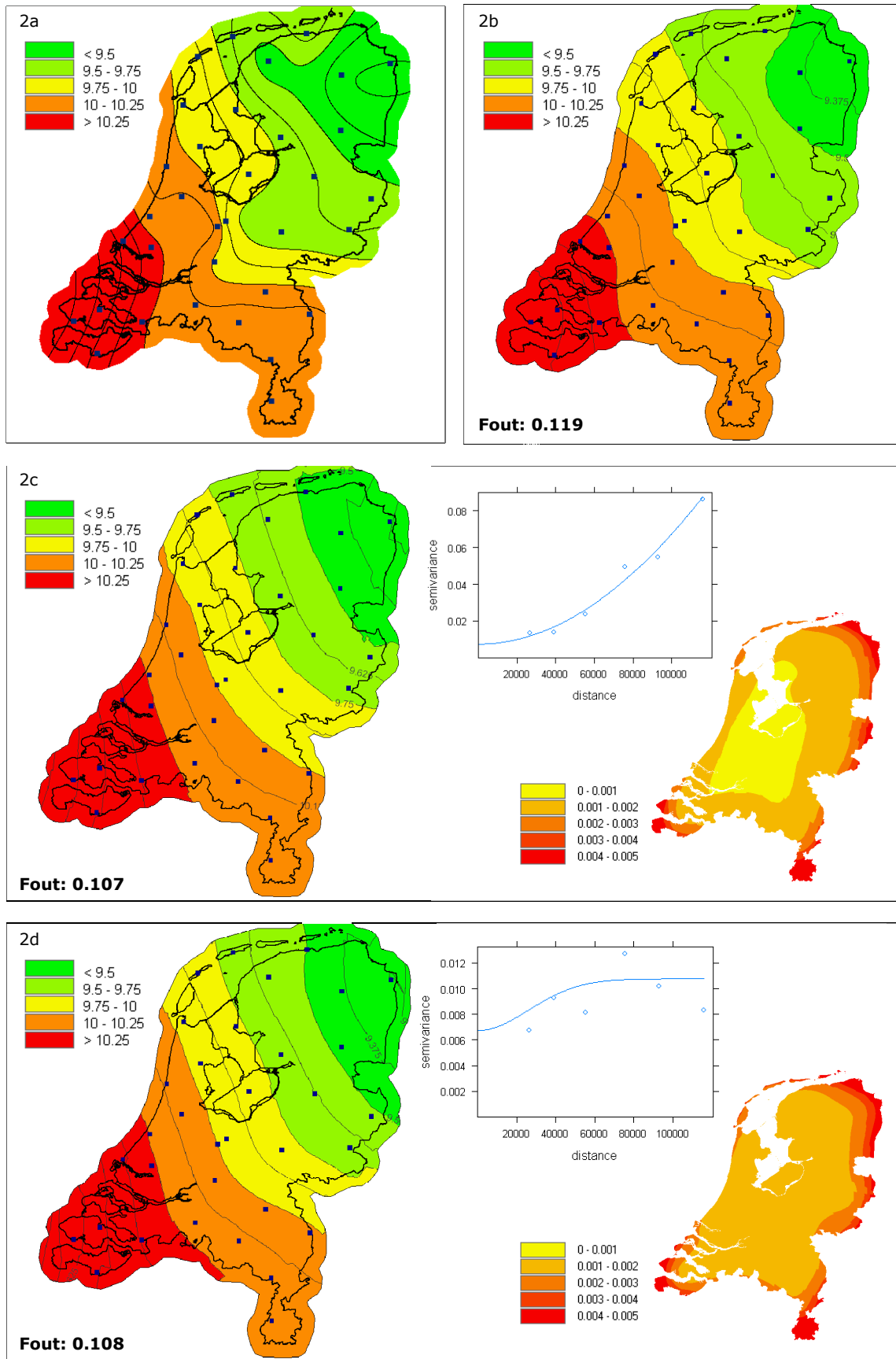
De uitkomsten van de verschillende interpolatiemethoden zijn onderworpen aan een kwantitatieve verificatie en een kwalitatieve verificatie. De kwantitatieve verificatie bestaat uit het berekenen van de foutwaarde door middel van het LOOCV-principe (Leave-one-out cross validation) (De Smith *et al.*, 2007). Dit levert voor ieder resultaat een RMSE-waarde (Root Mean Square Error) op. De kwalitatieve verificatie bestaat uit een beoordeling van het geproduceerde patroon door experts. Dit gebeurt door het laten becommentariëren van het resultaat of het visueel vergelijken van het resultaat met een getekend patroon dat door de experts verwacht wordt.

De grootte van de zoekstraal wordt automatisch bepaald door een radius te nemen die de kleinste crossvalidatiefout oplevert. Voor het bepalen van de 'support size' bij het interpoleren is een blokvoorspellingsmodel gebruikt. In dat model worden voorspellingen gedaan voor een aantal punten binnen het blok, de resultaten worden vervolgens gemiddeld en toegekend aan het te voorspellen punt.

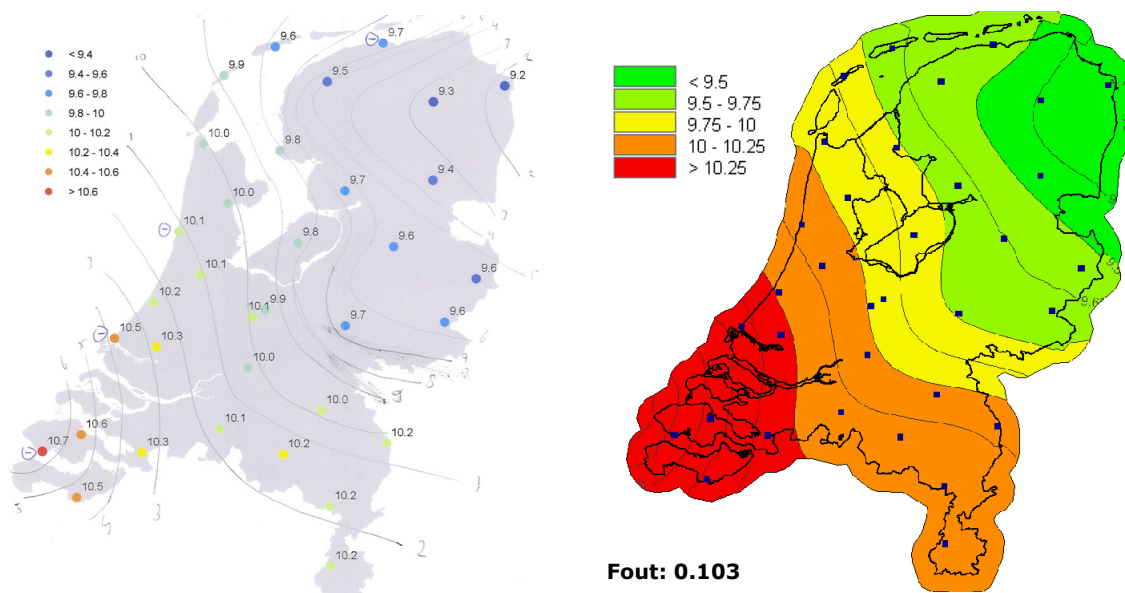
De keuze voor de meest optimale interpolatiemethode per tijdresolutie wordt gemaakt door de uitkomsten van de verificatie te beschouwen. Daarbij streven we naar een interpolatiemethode waarbij de nauwkeurigheid zo hoog mogelijk is. Echter, de beoordeling van experts wordt ook belangrijk geacht, omdat de communicatie van de temperatuurverdeling voornamelijk visueel plaats zal vinden. Het is belangrijk dat een patroon past bij de menselijke perceptie over hoe temperatuur over Nederland verdeeld is: zonder artefacten van de methode (geen onnatuurlijke vormen, haakse hoeken en eilanden). Daarnaast is het belangrijk dat de resultaten aansluiten bij de huidige vorm van communicatie zodat de consistentie van klimaatkaarten niet in het geding komen.



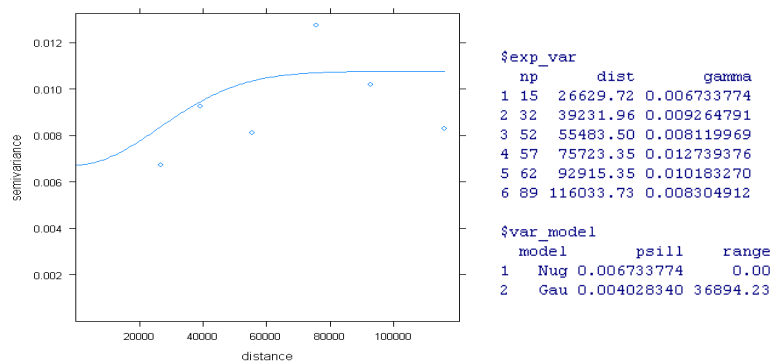
Figuur 1: Locaties van de 34 waarnemstations met per station de temperatuur normaalwaarde.



Figuur 2: Visualisaties van jaarnormalen die zijn geïnterpoleerd door middel van Spline (2a), IDW (2b), Ordinary Kriging (2c), Universal Kriging (2d). Voor beide Krigingmethoden is tevens een gefit semivariogram en de bijbehorende variantiekaart gegeven.



Figuur 3: Links het patroon voor jaarnormalen dat door de experts verwacht wordt en rechts het geproduceerde KED-resultaat.



Figuur 4: De semivariogram die automatisch gefit is en gebruikt wordt voor het interpoleren van jaarnormalen door middel van KED.



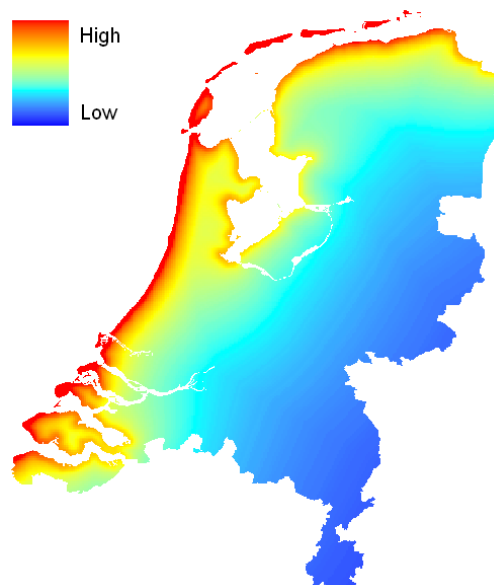
Figuur 5: De variantiekaart die hoort bij het interpoleren van jaarnormalen door middel van KED.

Resultaten

In een aanvullend onderzoek uitgevoerd binnen de stage hebben we met behulp van GIS-technieken aangetoond dat de kwaliteit van het meetnetwerk voor wat betreft het interpoleren van temperatuur goed is. Een aantal zwakke plekken zijn opgemerkt, zoals bevooroordeelde richting de Randstad en regelmatige tussenruimtes, maar deze vormen geen belemmering voor het interpoleren van temperatuur. Het complete onderzoek naar de kwaliteit van het meetnetwerk is terug te vinden in Bijlage A.

De meest optimale zoekstraal die met behulp van een crossvalidatie bepaald wordt ligt in alle gevallen tussen de 90 en 140 kilometer, met een gemiddelde van 120 kilometer. Voor dit onderzoek gebruiken we een onveranderlijke zoekradius van 120 kilometer. Een blokgrootte van 50 bij 50 kilometer blijkt het meest geschikt. Door gebruik te maken van deze 'support size' wordt de variatie van de samplegegevens verminderd en de distributie meer gelijk getrokken waardoor het patroon van het resultaat beter bij de visuele verwachtingen van de expert past. Daarnaast zorgt het voorspellen door middel van blokken er voor dat ook bij IDW rekening wordt gehouden met een bias in de samplewaarden.

Er is ook onderzoek gedaan naar de geschiktheid van de verschillende beschrijvende variabelen. De geringe hoogteverschillen blijken onbruikbaar wegens een slechte correlatie met temperatuur en een onrepresentatief netwerk van waarneemstations ten opzichte van de hoogtes in Nederland. Verder vertegenwoordigen de meetstations redelijk tot goed de verschillende afstanden tot zee, waardoor het gevaar tot extrapoleren minimaal is. Een logaritmisch getransformeerd grid van een gewogen combinatie van de afstanden tot de zee en de afstanden tot open water geeft de beste correlatie met temperatuur. Dit grid geeft bij het interpoleren een goede beschrijving van variaties bij de kust. In Figuur 6 is het getransformeerde grid met de afstanden tot zee en open water te zien. De implementatie van de gemiddelde wind is gedaan door de afstand te bepalen tot een subjectief punt dat de bron (het zuidwesten) van de wind weergeeft. Als vervanging voor deze beschrijvende variabele, die een kenmerkend diagonaal patroon over Nederland stimuleert, is een anisotropisch Kriging systeem een mogelijke oplossing. De laatste twee variabelen die geschikt blijken om de temperatuurverdeling globaal te beschrijven zijn de oost-west gradiënt en de noord-zuid gradiënt.



Figuur 6: De externe dataset die de sterkte van de kusteffecten representeert.

Het uitgangspunt bij ieder interpolatieproces is de set van meetwaarden die de input vormen voor de voorspelling van een onbekend punt. Per tijds categorie heeft de set van meetwaarden een andere hoedanigheid, waarbij bij de ene tijds categorie wel grensstations beschikbaar zijn en bij de andere niet. Of waarbij bij de ene tijds categorie wel genormaliseerd kan worden en bij de andere niet. Dit heeft tot gevolg dat voor iedere tijds categorie een andere interpolatiemethode als de meest optimale is gekozen (zie Tabel 2). Daarnaast is voor de categorie met de lange termijn gemiddelden de nauwkeurigheid waarmee voorspellingen worden gedaan meer significant dan voor de andere categorieën.

Tabel 2: Per tijds categorie de samplegegevens en de meest optimale interpolatiemethode.

Categorie	Temporele situatie	Data	Grensstations	Interpolatietechniek
1	Meerjarig jaargemiddelde Meerjarig seizoengemiddelde Meerjarig maandgemiddelde	Absoluut	Nee	KED
2	Jaargemiddelde Seizoengemiddelde Maandgemiddelde	Anomalieën	Ja	MRegr + IDW
3	Daggemiddelde	Absoluut	Ja/Nee	IDW

Voor de eerste categorie, de meerjarige gemiddelden, wordt geïnterpoleerd op de absolute data, omdat er geen data is om mee te normaliseren. Van de grensstations zijn geen gegevens beschikbaar voor de gehele normaalperiode dus de inputgegevens bestaan enkel uit Nederlandse meetgegevens. Naast de visualisatie van de resultaten in de Klimaatatlas en tal van andere publicaties zijn de vergridde gegevens ook gewild in tal van klimaat- en impactmodellen. Het is daarom van belang dat de interpolatie van deze categorie met een hoge nauwkeurigheid plaatsvindt. De methode die het beste voldoet aan deze voorwaarde is KED. Deze methode geeft gemiddeld genomen de kleinste fout en daarnaast, dankzij de gekozen blokvoorspelling- en zoekgebiedparameters, een patroon volgens verwachting. Voor het onderzoek naar de meest geschikte methode voor het interpoleren van jaarnormalen zijn onder andere getest: Spline, IDW, Ordinal Kriging en Universal Kriging. De resultaten van hiervan zijn te zien in Figuur 2. Figuur 4 en Figuur 5 geven de resultaten van de KED-methode voor jaarnormalen. In Figuur 3 is het resultaat van de kwalitatieve verificatie te vinden, waarbij een vergelijking is gemaakt tussen het KED-patroon en het patroon dat door experts verwacht wordt.

Voor de tweede categorie van tijdsituaties worden in plaats van de absolute gegevens de anomalieën gebruikt als input voor de interpolatie. De samplegegevens binnen deze categorie worden genormaliseerd met het meest recente tienjarig gemiddelde (i.e. het tienjarig voortschrijdend gemiddelde). Het tienjarige gemiddelde valt onder de eerste categorie en wordt daarom op zijn beurt geïnterpoleerd met de KED-methode. Voor het interpoleren van maand-, seizoen- en jaargemiddelden is de beschikking over meetgegevens van grensstations, mits de tijdsperiode waarover men de interpolatie uitvoert maximaal vijftien jaar geleden is. Uit het onderzoek blijkt dat de meest optimale interpolatiemethode voor deze techniek een multiple regressie gevolgd door IDW op de residuen is. De gemiddelde fout verschilt minimaal van de gemiddelde fout die Kriging zou voortbrengen, echter geproduceerde patronen zijn dankzij het extrapoleren van het regressiemodel veel reëler.

Voor de derde en laatste categorie met de daggemiddelden is IDW de meest geschikte interpolatiemethode. Voor deze categorie is het patroon belangrijker dan de nauwkeurigheid. Daarom is het niet nodig om daggemiddelden eerst te normaliseren. De beschikbaarheid van grenssamples is afhankelijk van tal van factoren. Het is aan te bevelen om zoveel mogelijk de grensstations bij de interpolatie te betrekken omdat deze voorspellingen in grensgebieden aanzienlijk verbeteren. IDW geeft, in combinatie met de eerder genoemde blokvoorspelling- en zoekgebiedparameters, voor gemiddelde dagtemperatuur realistische patronen, zonder eilandvorming en scherpe hoeken. Bijkomend voordeel van IDW is dat de methode eenvoudig te begrijpen is en rekenkundig snel.

Discussie

In dit verslag hebben we beschreven wat de meest optimale interpolatiemethode is voor zeven verschillende tijdsresoluties van temperatuurgegevens voor Nederland. De zeven tijdsresoluties zijn onder te verdelen in drie categorieën: de lange termijn gemiddelden (1), de maand-, seizoen en jaargemiddelden (2) en de daggemiddelden (3). We hebben de gebruikte interpolatietechnieken geëvalueerd door te kijken naar de crossvalidatiefout en een beoordeling door klimaatexperts. Uiteindelijk is gebleken dat voor iedere categorie een andere interpolatiemethode als meest optimaal kan worden aangemerkt. Dit heeft ook te maken met de hoedanigheid van de samplegegevens, waarbij verschillen aanwezig zijn wat betreft het opnemen van grensstations en de mogelijkheid tot normaliseren.

Voor de lange termijn gemiddelden, waarbij geen grensstations beschikbaar zijn en ook niet genormaliseerd kan worden, is Kriging met Externe Drift het meest geschikt. Voor het interpoleren van maand-, seizoen en jaartemperatuurgemiddelden kan het beste IDW met multiple regressie gebruikt worden. Interpolatie binnen die categorie geschiedt op de anomalieën en de samples zijn inclusief de gegevens van België en Duitsland. Beschrijvende variabelen voor zowel KED als MRegr+IDW zijn oost-west gradiënt, noord-zuid gradiënt, en de logaritmische getransformeerde afstanden tot de zee en meren. Eventueel kan ook de gemiddelde windrichting gebruikt worden, maar in geval van Kriging heeft een anisotropisch variogrammodel de voorkeur. Daggemiddelde temperatuur kan het beste geïnterpoleerd worden door middel van IDW, waarbij normaliseren niet nodig is.

Uit het onderzoek is ook gebleken dat op alle tijdsresoluties er geen interpolatiemethode is die superieur is ten opzichte van de andere interpolatiemethoden. Zowel IDW, KED en Mregr+IDW zullen in zekere mate volstaan voor alle drie de tijds categorieën. Nadeel van het gebruik van verschillende optimale methoden voor verschillende tijdsresoluties is dat het interpoleren van temperatuur niet consistent gebeurt. Wanneer bijvoorbeeld 365 vergridde dagtemperaturen bij elkaar opgeteld worden komt er een resultaat uit dat ongelijk is aan het vergridde jaargemiddelde. In de praktijk zal dit probleem zich bijvoorbeeld voordoen bij gebruik van hydrologische modellen. Voor andere toepassingen zoals publicaties in de Klimaatatlas van Nederland is dit probleem niet van belang.

Voor het verantwoord fitten van een semivariogram is een netwerk met meer dan 50 samplepoints gewenst (Hengl, 2007). Het netwerk van waarneemstations dat temperatuur registreert bestaat slechts uit 34 stations en als we de grensstations meerekenen 44 stations. Strikt genomen voldoet het netwerk dus niet aan de vereisten die Kriging stelt. Toch is het met slechts 34 samples mogelijk gebleken om een degelijk semivariogram op te stellen. Bij het opstellen van een anisotropisch variogram levert dit wél problemen op omdat er per richting te weinig samples aanwezig zijn om een model te fitten. Een ander probleem met anisotropische variogrammen is dat het lastig is om geautomatiseerd de anisotropische parameters te bepalen.

Conclusie

- De optimale interpolatiemethode voor meerjarige temperatuurgemiddelden is Kriging met een Externe Drift. Voor maand-, seizoen- en jaargemiddelden is Inverse Distance Weighting met Multiple Regressie het meest geschikt. Interpoleren van daggemiddelde temperatuur geschiedt het beste met Inverse Distance Weighting.
- Externe gegevens kunnen gebruikt worden om de voorspelling van temperatuurwaardes te verbeteren. In dit geval zijn gebruikt de oost-west gradiënt, de noord-zuid gradiënt en de getransformeerde afstand tot zee.
- De gemiddelde windrichting is een (externe) factor van belang, die een diagonaal temperatuurpatroon over Nederland representeert maar niet objectief bepaald kan worden.

- Het toepassen van anisotropy in het Krigingsysteem is potentieel de beste manier om het diagonale patroon van de windrichting te creëren. Echter, het netwerk van temperatuursamples voldoet daarvoor niet en de implementatie in een productieomgeving is complex.

Referenties

- Buishand, T.A., J.J. Beersma, R. Sluiter & T. Kroon (2008). Definitiestudie rasterdata meteorologie. De Bilt, KNMI / RWS Waterdienst, 2008
- De Smith, M.J., M.F. Goodchild & P.A. Longley (2007). Geospatial Analysis – A Comprehensive Guide to Principles, techniques and Software Tools. Leicester: Matador.
- Eric P. J. Boer, Kirsten M de Beurs, A Dewi Hartkamp, Kriging and thin plate splines for mapping climate variables, *International Journal of Applied Earth Observation and Geoinformation*, Volume 3, Issue 2, 2001, Pages 146-154, ISSN 0303-2434, DOI: 10.1016/S0303-2434(01)85006-6.
- Ian A. Nalder, Ross W. Wein, Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest, *Agricultural and Forest Meteorology*, Volume 92, Issue 4, 4 December 1998, Pages 211-225, ISSN 0168-1923, DOI: 10.1016/S0168-1923(98)00102-6.
- Isaaks, Edward H.; Author: With Srivastava, R. Mohan, An Introduction to Applied Geostatistics, p. 580, October 1990, Oxford University Press, 0195050134.
- Heijboer, D. & J. Nellestijn (2002). Klimaatatlas van Nederland - De Normaalperiode 1971-2000. Rijswijk: Uitgeverij Elmar B.V.
- Hengl, T. (2007) A Practical Guide to Geostatistical Mapping of Environmental Variables, p. 143, http://eusoiils.jrc.ec.europa.eu/ESDB_Archive/eusoiils_docs/other/EUR22904en.pdf.
- Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei (2008), Comparison of six methods for the interpolation of daily, European climate data, *J. Geophys. Res.*, 113, D21110, doi:10.1029/2008JD010100.
- KNMI (2005). COST-719 official website. [28/07/2009] <http://www.knmi.nl/samenw/cost719/>.
- KNMI (2009a). Geografische overzichten van het weer in Nederland. [28/07/2009] http://www.knmi.nl/klimatologie/geografische_overzichten/index.cgi.
- KNMI (2009b). Weer Actueel. [28/07/2009] <http://www.knmi.nl/actueel/>.
- Perry and Hollis, 2005a M. Perry and D. Hollis, The development of a new set of long-term climate averages for the UK, *Int. J. Climatol* **25** (8) (2005), pp. 1023–1039.
- Perry and Hollis, 2005b M. Perry and D. Hollis, The generation of monthly gridded datasets for a range of climatic variables over the UK, *Int. J. Climatol* **25** (8) (2005), pp. 1041–1054.
- R-Project (2009). The R Project for Statistical Computing. [28/07/2009] <http://www.rproject.org/>.
- Tveito, O.E. (2007). The Developments in Spatialization of Meteorological and Climatological Elements In: H. Dobesch et al (Ed.) Spatial interpolation for climate data: the use of GIS in climatology and meteorology pp. 73-86. London: ISTE Ltd, intern rapport.
- Tveito, O.E., M. Wegehenkel, F. Van der Wel & H. Dobesch (2006). The Use of Geographic Information Systems in Climatology and Meteorology - Final Report COST Action 719.
- Schuermans, J.M., M.F.P. Bierkens, E.J. Pebesma, and R. Uijlenhoet, 2007: Automatic Prediction of High-Resolution Daily Rainfall Fields for Multiple Extents: The Potential of Operational Radar. *J. Hydrometeor.*, **8**, 1204–1224.

Sluiter, R., 2008: Interpolation methods for climate data: Literature review, Version 1.0, intern rapport.

Vicente-Serrano, S.M., M.A. Saz and J.M. Cuadrat, 2003: Comparative analysis of interpolation methods in the middle Ebro valley (Spain): application to annual precipitation and temperature. *Climate Research*, **24**, 161-180.

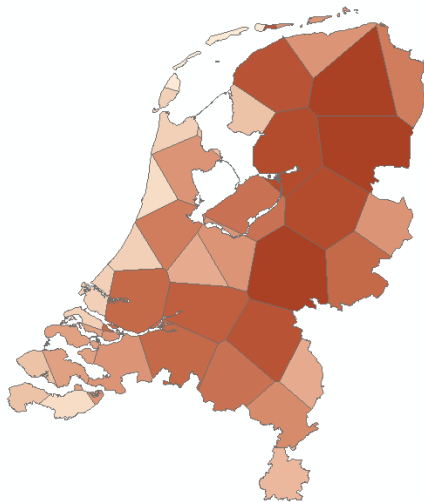
Bijlage A: Kwaliteit van het meetnetwerk

Spreiding van de meetstations

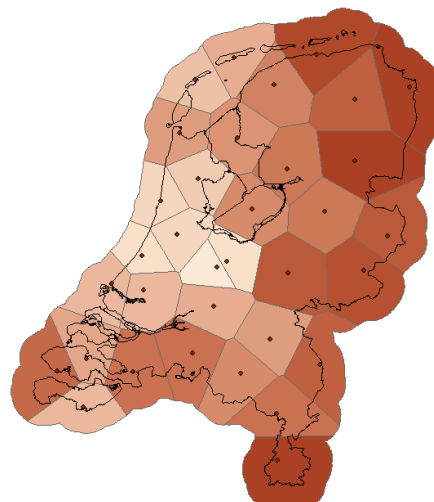
We onderzoeken de representativiteit van het meetnetwerk aan de hand van de geografische spreiding van de 34 waarnemstations. Daarvoor kijken we naar het oppervlak van het verzorgingsgebied van iedere meetstation. De gebieden zijn verdeeld door middel van de Thiessen interpolatietechniek (ook bekend als Voronoi diagrammen). Deze techniek gaat er vanuit dat een onbekende locatie de waarde krijgt van de dichtstbijzijnde bekende locatie. Door deze verdeling ontstaan 34 polygoon die de verzorgingsgebieden van de waarnemstations representeren. Vervolgens is voor iedere polygoon de oppervlakte berekend. Het resultaat daarvan is afgebeeld in Figuur 7.

De polygoon dat hoort bij het waarnemstation Hoogeveen het grootste verzorgingsgebied. Vlieland daarentegen heeft het kleinste verzorgingsgebied wat uiteraard komt doordat er weinig land aanwezig in de omgeving van dit meetstation. Want als begrenzing van het gebied is voor deze analyse het landoppervlak van Nederland gebruikt. Het is daarom ook lastig te concluderen aan de hand van bovenstaand resultaat. Daarom is dezelfde analyse nogmaals uitgevoerd met ditmaal een uitbreiding van het gebied met een buffer van vijftien kilometer om het landoppervlak van Nederland. Het resultaat hiervan is in Figuur 8 te zien.

Wederom is het verzorgingsgebied van meetstation Hoogeveen het grootst. Het kleinste verzorgingsgebied heeft station De Bilt. Door het bestuderen van beide resultaten kunnen conclusies getrokken worden over de representativiteit van de meetstations wat betreft de geografische spreiding. We zien dat over het algemeen dat de stations in het oosten van het land verder van elkaar gelegen zijn dan in het westen van het land. Ook zien we een hogere stationsdichtheid in en rond de Randstad. We kunnen daardoor concluderen dat het netwerk van waarnemstations licht bevooroordeeld is naar de westkust en naar meer bevolkte gebieden.



Figuur 7: Grootte van de verzorgingsgebieden per stations voor het Nederlands grondgebied.



Figuur 8: Grootte van de verzorgingsgebieden per stations voor het Nederlands grondgebied inclusief 15km buffer.

Regelmatige tussenruimte

Wat verder opvalt wanneer we de kwaliteit van het netwerk onderzoeken is dat de stations van zuid naar noord gezien ongeveer op één lijn liggen. De stations op deze lijnen liggen steeds met een ongeveer even grote tussenruimte van elkaar. Het gevaar hierbij is dat de grootte die we willen meten een innerlijk frequentiepatroon bevat dat precies tussen de meetstations door loopt. Mocht dit het geval zijn dan is het onmogelijk een dergelijk patroon

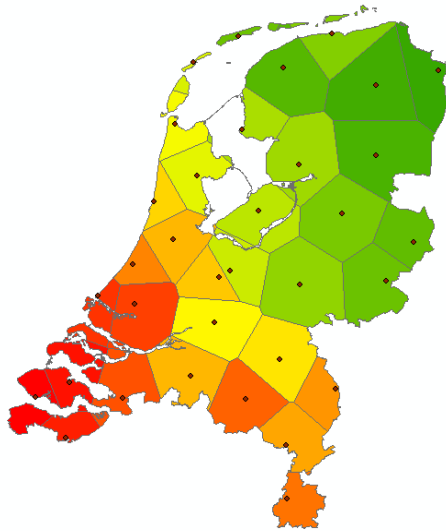
waar te nemen. Bij temperatuur zal de regelmatige opstelling van de stations geen probleem vormen omdat temperatuur een dergelijk innerlijk patroon niet heeft.

De variatie in temperatuur

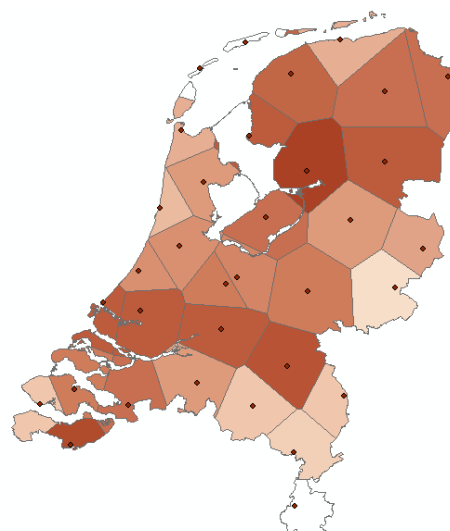
Met de Thiessen polygonen die in de vorige paragraaf aangemaakt zijn kunnen we ook een eerste resultaat geven van de interpolatie van temperatuurnormalen. Figuur 9 geeft dit resultaat.

Vervolgens kunnen we onderzoeken of het meetnetwerk representatief geacht kan worden voor het detecteren van de variatie in temperatuurnormalen. Dit doen we door voor iedere polygoon de standaardafwijking te berekenen van de waarden van alle aangrenzende polygoenen. Vervolgens is in Figuur 10 het resultaat gevisualiseerd waarbij iedere polygoon een kleurgradiënt heeft gekregen dat gelijk staat aan de berekende standaardafwijking.

Drie polygoenen (wit gekleurd) zijn in deze analyse buiten beschouwing gelaten omdat deze te weinig aangrenzende polygoenen hebben. De hoogste en laagste standaardafwijkingen zijn gevonden voor de polygoenen van respectievelijk Marknesse en Hupsel. Wat de meetstations voor temperatuurnormalen betreft is het waardevol om een hogere stationsdichtheid te hebben op plaatsen waar de variabiliteit van temperatuur hoog is. Aan de hand van bovenstaand resultaat kan geconcludeerd worden dat zowel rond het stroomgebied van Maas en Waal alsmede het grensgebied tussen Friesland en Drenthe de stationsdichtheid wat betreft het meten van temperatuurnormalen het laagst is. Een hogere stationsdichtheid op deze locaties is gewenst wanneer men met hogere nauwkeurigheid de patronen van de betreffende grootte wil detecteren.



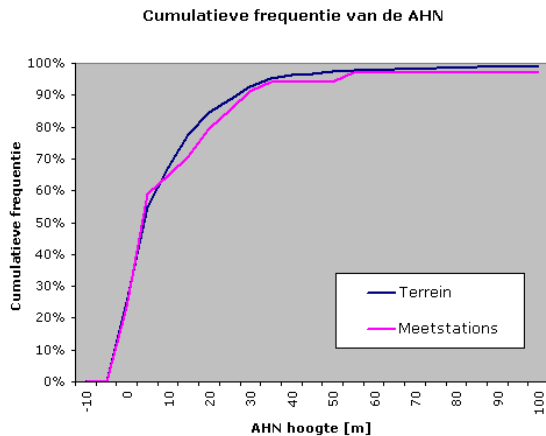
Figuur 9: Temperatuurnormalen per verzorgingsgebied van een station, met rood voor hoge temperaturen en groen voor lage temperaturen.



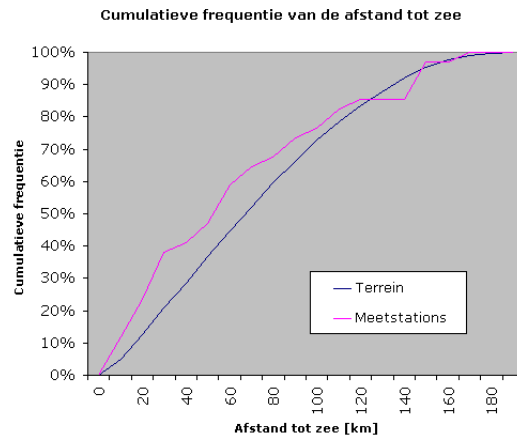
Figuur 10: Standaardafwijkingen per verzorgingsgebied van een station berekend voor de temperatuurnormalen van de aangrenzende verzorgingsgebieden.

De karakteristieken van het terrein

In deze paragraaf wordt de representativiteit van de meetstations onderzocht aan de hand van de terreinkarakteristieken. De gebruikte terreinkarakteristieken zijn in dit geval hoogte t.o.v. zee niveau en Euclidean afstand tot de zee. Aan de ene kant is het van belang dat het netwerk van meetstations dezelfde ruimte representeert als de ruimte die men daarmee wil bepalen. De analyse die we gaan uitvoeren is een vergelijking van de frequentiedistributie van een variabele zoals die in het studiegebied aanwezig is en zoals die op de locaties van de meetstations aanwezig is. In geval van een ideale verdeling van de meetstations over het studiegebied vallen de frequentiedistributie van stations en gebied samen.



Figuur 11: Cumulatieve distributiefuncties van de hoogte voor geheel Nederland en voor de meetstations



Figuur 12: Cumulatieve distributiefuncties van de afstand tot zee voor geheel Nederland en voor de meetstations

De methode die voor deze analyse gebruikt wordt is het maken van distributiefuncties (Tveito *et al.*, 2007). Voor beide terreinvariabelen is de cumulatieve distributie berekend voor zowel de 34 meetstations als het gehele studiegebied. De resultaten zijn afgebeeld in Figuur 11 en Figuur 12 in twee histogrammen met iedere histogram de resultaten van één terreinkarakteristiek. Iedere histogram bevat dus twee functies die een beschrijving geven van de verdeling van de terreinkarakteristiek ten opzichte van de meetstations (roze functielijn) en het studiegebied (blauwe functielijn).

De mate waarop de roze en blauwe lijn samenvallen geeft aan in hoeverre de indeling van het meetnetwerk overeenkomt met de verdeling van de terreinkarakteristiek in het studiegebied. Wat betreft de twee onderzochte variabelen is te zien dat met name de terrein hoogte (AHN) door de meetstations goed vertegenwoordigd wordt. De afstand tot zee presteert beduidend minder. Dicht bij de kust zijn verhoudingsgewijs teveel meetstations. Desalniettemin is het verschil tussen de frequentiefuncties gering. Bovenstaande resultaten geven aan dat de verdeling van de meetstations wat betreft de twee terreinkarakteristieken te kwalificeren is als redelijk tot goed. Uiteraard is dit geen verrassing aangezien de stationsdichtheid gemiddeld is en de variabiliteit van de terreinvariabelen zijn laag (met slechts zeer kleine hoogteverschillen en het studiegebied deelt voor meer dan de helft haar grens met de zee).

Conclusie

Uit het onderzoek naar de kwaliteit en de representativiteit van het netwerk van waarnemstations blijkt dat het netwerk voldoet voor wat betreft de interpolatie van temperatuurgegevens.

