

# **Tentamen - TA2060, Data Analysis & Geostatistics**

**TN-TZ 4.25, 27 juni 2014, 14.00 - 17.00**

**Dit tentamen bestaat uit 45 vragen.**

**Je begint met 10 punten. Elk goed antwoord is 2 punten waard.**

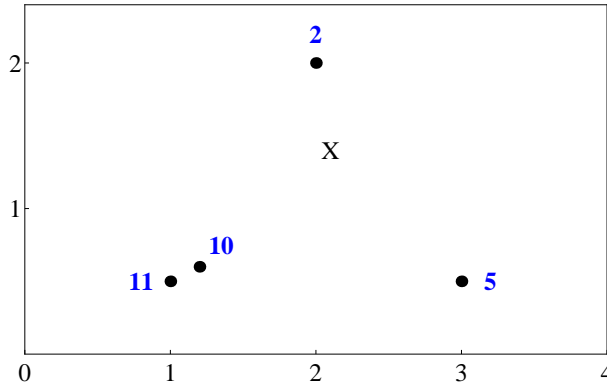
**Licht altijd je antwoord toe.**

**Technische Engelse termen staan schuingedrukt.**

**Het gebruik van een eenvoudige rekenmachine zonder geheugen is toegestaan.**

**SUCCES!**

## A. Deterministische interpolatie



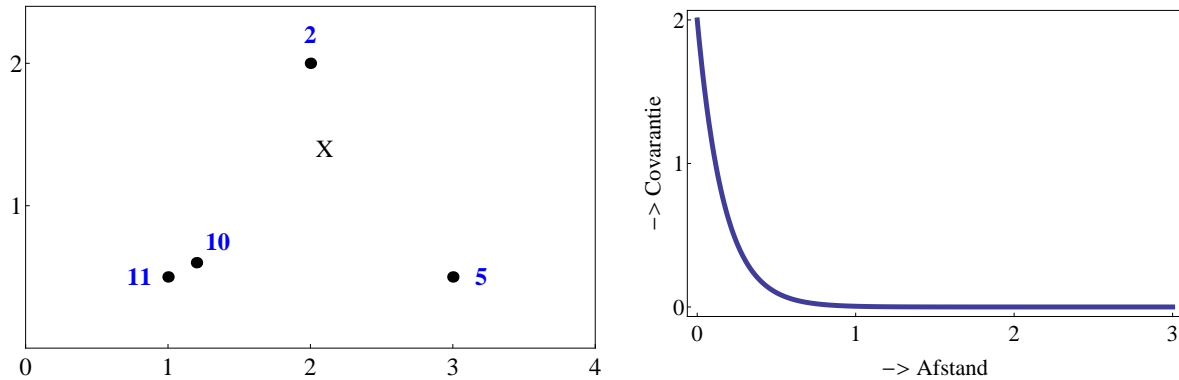
Figuur 1: Vier hoogte waarnemingen:  $h_1 = 2$ ,  $h_2 = 5$ ,  $h_3 = 11$  en  $h_4 = 10$ .

In Figuur 1 staan vier hoogte waarnemingen. Bart wil de hoogte schatten op de locatie met het kruisje.

*Opmerking: Bij het beantwoorden van Vragen 3, 5 en 9 is het niet nodig formules te gebruiken.*

1. Wat is het convex omhulsel van de vier punten in Figuur 1?
2. Hoeveel verschillende manieren zijn er om de vier punten te trianguleren?
3. Wat is ongeveer de hoogte bij het kruisje volgens *triangular interpolation*?
4. Hoe wordt het gewicht van een waarneming bepaald bij *triangular interpolation*?
5. Wat is ongeveer de hoogte bij het kruisje volgens *inverse distance interpolation* met *power 2*?
6. Krijgt waarneming  $h_3$  relatief meer of minder gewicht bij *inverse distance interpolation* met *power 3*? Waarom?
7. Met welke twee bekende methodes komt *inverse distance interpolation* met *power 0* en oneindig overeen?
8. Teken het Voronoi Diagram van de vier waarnemingslocaties.
9. Wat is ongeveer de hoogte bij het kruisje volgens *nearest neighbor interpolation*?
10. Waarschijnlijk heb je bij de vorige vragen de Euclidische afstand gebruikt tussen het kruisje en de vier punten in Figuur 1. Geef een alternatief afstandsbeleg dat een afstand tussen twee punten in Figuur 1 definieert.

## B. Stochastische interpolatie



Figuur 2: **Links.** Vier waarnemingen:  $h_1 = 2$ ,  $h_2 = 5$ ,  $h_3 = 11$  en  $h_4 = 10$ . **Rechts.** Covariantie functie  $f(x) = 2e^{-3x}$ .

Evelien gebruikt Kriging om de hoogte te schatten bij het kruisje. Daarvoor gebruikt ze een covariantie functie  $f(x)$ , zie Figuur 2.

11. Wat is de *range* en *sill* van de covariantie functie  $f(x)$ ?
12. Welke waarnemingen zijn gecorreleerd? Waarom?
13. Schat de covariantie tussen elk tweetal waarnemingen in Figuur 2, links.
14. Wat is (ongeveer) het *Kriging the mean* gemiddelde van de vier hoogte waarnemingen?

Evelien gebruikt *Ordinary Kriging* (OK) om de hoogte te schatten bij het kruisje.

15. Stel de OK *redundancy matrix* op van de vier hoogte waarnemingen.
16. Schat de covariantie tussen de locatie van het kruisje en elke waarneming in Figuur 2 en geef de *proximity vector* voor een OK schatting bij het kruisje.
17. Wat garandeert dat de *redundancy matrix* inverteerbaar is?
18. Oplossing van het OK systeem levert gewichten voor elke waarneming. Welk gewicht krijgt elk van de vier waarnemingen ongeveer?
19. Wat is de OK schatting van de hoogte op de waarnemingslokaties?
20. En wat is de OK variantie op de waarnemingslokaties?

## C. Classificatie

Band	Description	Wavelength (micrometers)	Resolution (meters)
1*	Violet-Deep Blue	0.43 - 0.45	30
2*	Blue	0.45 - 0.51	30
3*	Green	0.53 - 0.59	30
4*	Red	0.64 - 0.67	30
5	Near Infrared	0.85 - 0.88	30
6	Shortwave Infrared	1.57 - 1.65	30
7	Shortwave Infrared	2.11 - 2.29	30
8*	Panchromatic	0.50 - 0.68	15
9	Cirrus clouds	1.36 - 1.38	30
10**	Thermal infrared	10.62 - 11.19	30
11**	Thermal infrared	11.50 - 12.51	30

\* Within the visible spectrum  
 \*\* 100-meter resolution data interpolated to 30 meters



Figuur 3: **Links:** Landsat 8 banden. *Bron:* <http://www.shadedrelief.com/landsat8/introduction.html>  
**Rechts:** Luchtfoto van een deel van het interessegebied van Mien en Piet.

Mien en Piet hebben Landsat 8 data gekregen om landgebruik in kaart te brengen, zie Figuur 3. Om hun classificatie te trainen hebben ze zelf het landgebruik bepaald van 200 locaties in hun gebied.

21. Door welke Landsat banden te combineren kunnen Mien en Piet een *true color image* maken van hun gebied?
22. En met welke bandcombinatie kunnen ze een *false colour image* maken?
23. Waarom kan het nuttig zijn om twee *Shortwave Infrared* banden te gebruiken voor het in kaart brengen van landgebruik?
24. De Landsat data is gegeven in UTM coördinaten. Waarom zijn er verschillende zones voor UTM coördinaten?
25. Wat is een goed afstandsbelegrip voor de afstand tussen twee punten op aarde?
26. Welk tweetal punten is het dichtst bij elkaar:

$$A = (85^{\circ}N, 5^{\circ}W) \quad \text{en} \quad B = (85^{\circ}N, 6^{\circ}W), \quad \text{of}$$

$$A' = (5^{\circ}N, 5^{\circ}W) \quad \text{en} \quad B' = (5^{\circ}N, 6^{\circ}W),$$

waar de locaties van  $A$ ,  $B$ ,  $A'$  en  $B'$  gegeven zijn in graden noorderbreedte en westerlengte. Licht je antwoord toe.

27. Mien en Piet willen hun 200 trainingslocaties in een GIS systeem opslaan. Is het GIS raster formaat daarvoor geschikt? Waarom?

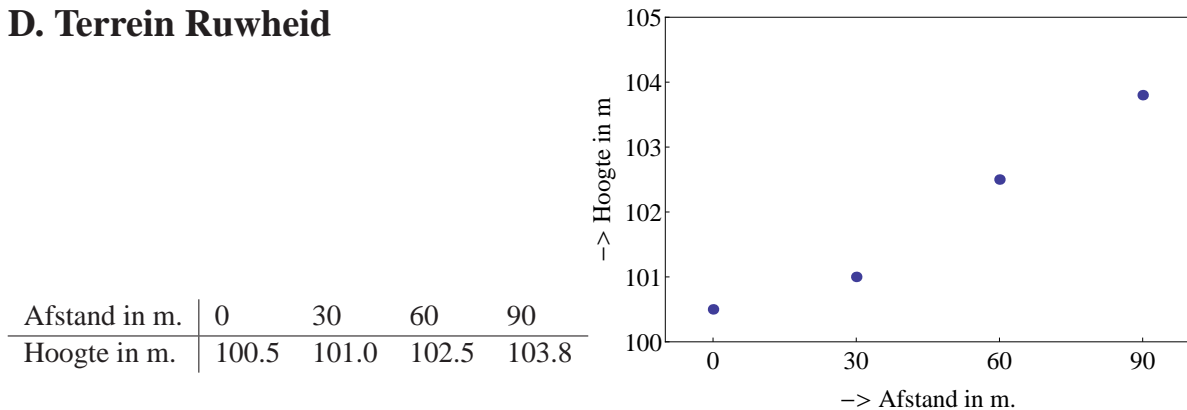
Mien en Piet besluiten alle banden in Figuur 3, links, te gebruiken voor hun classificatie. Ze hebben hun 200 trainingspunten onderverdeeld in 5 klassen: 77 punten bos, 44 punten water, 58 punten gras, 20 punten fietspad en 1 punt café.

28. Wat is de dimensie van de *feature space*?
29. Leg kort uit hoe supervised classificatie in het algemeen werkt.

Mien en Piet gebruiken *Maximum Likelihood* classificatie.

30. Hoe wordt de trainingsdata gebruikt om elke klasse te karakteriseren bij *Maximum Likelihood* classificatie?
31. Voor de klasse café is maar 1 trainingspunt aanwezig. Noem twee nadelen van een klein aantal trainingspunten.
32. Hoe wordt een Landsat pixel uiteindelijk aan een klasse toegekend bij *Maximum Likelihood* classificatie?
33. Waarom is het niet handig van Mien en Piet een klasse fietspad te kiezen als ze met Landsat data werken?
34. In het gebied van Mien en Piet komt veel stedelijke bebouwing voor. Aan welke klasse zullen de corresponderende Landsat pixels waarschijnlijk toegekend worden? Waarom?
35. Mien en Piet willen ook een kwaliteitskaart maken, waarin voor elk pixel de waarschijnlijkheid staat aangegeven, met een waarde tussen 0 en 1, dat de pixel correct geclassificeerd is door de *Maximum Likelihood* classificatie. Beschrijf een methode die deze waarden bepaalt.

## D. Terrein Ruwheid



Figuur 4: Terrein hoogtes.

Figuur 4 laat de hoogte van een terrein zien voor 4 opeenvolgende locaties langs een profiel. Het figuur suggereert dat de laatste drie punten meer op een lijn liggen dan de eerste drie punten. We gaan dit kwantificeren met de methode van de kleinste kwadraten. We beginnen met het

- Schatten van een lijn door de eerste drie punten, en het
- Schatten van een lijn door de laatste drie punten

De kleinste kwadraten oplossing van het lineaire systeem  $Ax = y$  wordt gegeven door

$$\hat{x} = (A^T A)^{-1} A^T y \quad (1)$$

36. Waarom wordt Vergelijking (1) de kleinste kwadraten oplossing genoemd?
37. Geef voor beide schattingen de vector van waarnemingen, (*observations*),  $y_1$  en  $y_2$ .
38. Welke twee parameters dienen in beide gevallen geschat te worden?
39. Geef voor beide gevallen de model matrix  $A_1$  en  $A_2$ .

De oplossingen van de twee schattingsproblemen worden gegeven door

$$\hat{x}_1 = (100.33, 0.033) \quad \text{en} \quad \hat{x}_2 = (99.63, 0.047)$$

40. Bepaal voor beide gevallen de vector van residuen  $\hat{e}_1$  en  $\hat{e}_2$ .
41. Waarom geeft de lengte van deze twee vectoren de gevraagde quantificatie?

De metingen in Figuur 4 zijn afkomstig uit het ASTER GDEM digitale terrein model.

42. Met welke techniek is ASTER GDEM gemaakt uit ASTER foto's?

43. Wat is de standaardafwijking van de ASTER GDEM data? Geef je antwoord door een geheel getal  $a$  te specificeren in "De standaardafwijking is ongeveer  $10^a$  m".
44. Bij bovenstaande schattingen is geen rekening gehouden met fouten in de waarnemingen. Schets hoe de Monte Carlo methode gebruikt kan worden om inzicht te krijgen in de onzekerheid in de bepaling van de helling van een lijn, gegeven de standaardafwijking van de ASTER GDEM data.
45. In de methode van de kleinste kwadraten is het ook mogelijk om iedere waarneming een gewicht te geven. Pas Vergelijking (1) aan door op de juiste manier een (in dit geval  $3 \times 3$ ) gewichtsmatrix  $W$  toe te voegen.