

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

**CT 4.98 en CT 4.99, 24 juni 2013, 14.00 - 17.00**

**Dit tentamen bestaat uit 45 vragen.**

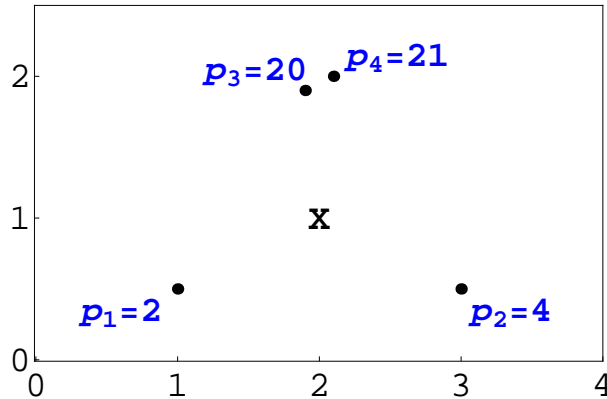
**Je begint met 10 punten. Elk goed antwoord is 2 punten waard.**

**Licht altijd je antwoord toe.**

**Technische Engelse termen staan schuingedrukt.**

**SUCCES!**

## A. Deterministische interpolatie



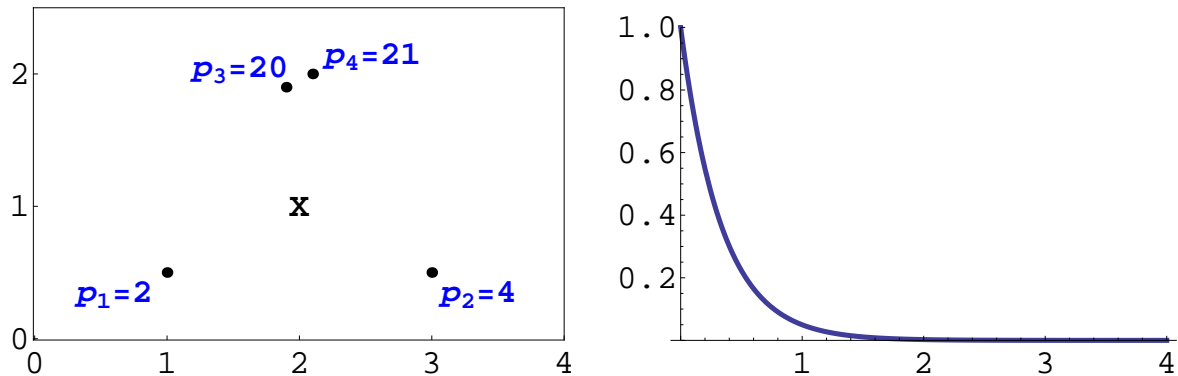
Figuur 1: Vier waarnemingen:  $p_1 = 2$ ,  $p_2 = 4$ ,  $p_3 = 20$  en  $p_4 = 21$ .

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

*Opmerking: Bij het beantwoorden van Vragen 1 tot 5 is het niet nodig formules te gebruiken.*

1. Wat is ongeveer de hoogte bij het kruisje volgens *triangular interpolation*?
2. Welk gewicht krijgt elke waarneming in dit geval ongeveer?
3. Wat is ongeveer de hoogte bij het kruisje volgens *inverse distance interpolation* met *power 2*?
4. En welk gewicht krijgt elke waarneming ongeveer in dit geval?
5. Krijgt waarneming  $p_4$  relatief meer of minder gewicht bij *inverse distance interpolation* met *power 3*? Waarom?
6. Teken het Voronoi Diagram van de vier waarnemingslocaties.
7. Wat is de hoogte op locatie (3, 2) volgens *nearest neighbor interpolation*?
8. Teken ook de Delaunay Triangulatie van de waarnemingslocaties.
9. Wat is het convex omhulsel van de waarnemingslocaties?
10. Wat is de hoogte op locatie (3, 2) volgens *triangular interpolation*?

## B. Stochastische interpolatie



Figuur 2: **Links.** Vier waarnemingen:  $p_1 = 2$ ,  $p_2 = 4$ ,  $p_3 = 20$  en  $p_4 = 21$ . **Rechts.** Covariantie functie  $f(x) = e^{-3x}$ .

Marieke gebruikt liever Kriging om de hoogte te schatten op de lokatie van het kruisje. Daarvoor gebruikt ze een covariantie functie  $f(x)$ , zie Figuur 2.

11. Wat is het gewone gemiddelde van de vier hoogte waarnemingen in Figuur 2?
12. Wat is (ongeveer) de covariantie tussen waarnemingen  $p_1$  en  $p_2$ ? En wat is (ongeveer) de covariantie tussen waarnemingen  $p_3$  en  $p_4$ ?
13. Wat is (ongeveer) het *Kriging the mean* gemiddelde van de vier hoogte waarnemingen?
14. Wat is het *Kriging the mean* gemiddelde van de vier hoogte waarnemingen als gegeven is dat de waarnemingen niet gecorreleerd zijn?

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

15. Wat is de *range* en *sill* van de covariantie functie  $f(x)$ ?
16. Stel de OK *redundancy matrix* op van de vier hoogte waarnemingen.
17. Wat is de *proximity vector* voor een OK schatting bij het kruisje?
18. Waarom is de redundancy matrix inverteerbaar?
19. Wat is de OK schatting van de hoogte op lokaties  $p_1$ ,  $p_2$ ,  $p_3$  en  $p_4$ ?
20. Schets de contourlijnen van de OK variantie als overall in Figuur 2, links, de hoogte wordt geschat.

## C. Landgebruik

Jan en Marieke willen Landsat data en hoogtedata combineren om het landgebruik beter in kaart te brengen. Ze hebben Landsat 8 data bestaande uit 11 banden in WGS84 en een 50m raster hoogtebestand in UTM31.

21. Noem drie verschillende sensors waarmee je hoogtedata kan inwinnen.
22. Wat houdt het GIS raster formaat in?
23. Wat is het GIS vector formaat?
24. Waarom is het UTM coördinatensysteem in verschillende zones onderverdeeld?
25. Waar moet je mee rekening houden als je de afstand bepaalt tussen twee punten gegeven in WGS84?

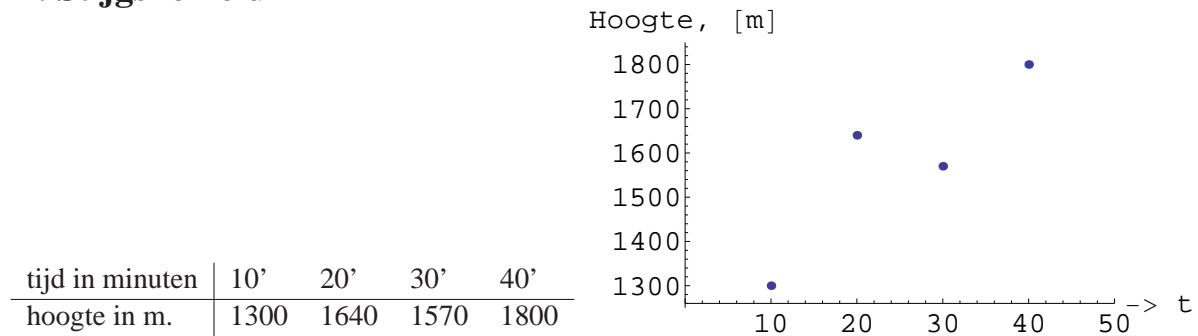
Jan en Marieke willen behalve de hoogte ook de orientatie (*aspect*) en helling (*slope*) gebruiken.

26. Geef een definitie van helling in 2D.
27. Schets een methode om voor elke rastercel de 2D helling te bepalen.
28. Schets ook een methode om de kwaliteit van de 2D hellingsbepaling vast te stellen.

Jan en Marieke gebruiken QGIS om de Landsat data naar UTM31 te converteren. Als resultaat verkrijgen ze een rasterbestand met voor elke cel 11 Landsat bandwaarden en bovendien 3 geometrische waarden, hoogte, orientatie en helling. Om hun classificatie te trainen hebben ze voor 500 locaties het landgebruik bepaald. Ze onderscheiden 5 klassen: water, bos, landbouw, stad en bergen.

29. Waarom is spectrale data relevant voor landclassificatie? En waarom is de geometrische data relevant?
30. Wat is in dit geval de dimensie van de *feature space*?
31. Gebruiken Jan en Marieke *supervised* of *unsupervised* classificatie?
32. Beschrijf hoe Jan en Marieke *Nearest Centroid* classificatie kunnen gebruiken om hun landgebruik te classificeren.
33. Jan en Marieke ontdekken dat hun trainingsdata niet altijd separabel is. Wat wordt daarmee bedoeld?
34. Geef een voorbeeld van een ommissie fout.
35. 14 waarden per pixel locatie is wat onoverzichtelijk. Hoe kan je correlatie gebruiken om uit deze 14 banden de 4 banden te bepalen die het meest relevant zijn voor de landclassificatie?

## D. Stijgsnelheid



Figuur 3: Gemeten GPS hoogtes.

Mark is een hoge berg opgefietst en heeft elke 10 minuten met GPS zijn hoogte bepaald. De resultaten staan in Figuur 3. Mark wil graag bepalen hoe snel hij steeg.

36. Mark is onderweg een skidorp met veel appartementenflats gepasseerd. Beïnvloedt dat de GPS metingen? Waarom?

Mark besluit eerst zijn stijgsnelheid te bepalen door een lijn door zijn waarnemingen te schatten met de kleinste kwadraten methode.

37. Wat is de algemene vergelijking van een lijn? Hoeveel parameters moet Mark schatten?
38. Wat is de vector van waarnemingen  $y$ ?
39. Wat is de model matrix  $A$ ?
40. Geef de formule om de parameter vector  $\hat{x}$  te schatten uit de vector van waarnemingen  $y$  en de model matrix  $A$ .

Mark rekt  $\hat{x}$  met Matlab uit en vindt  $\hat{x} = \begin{pmatrix} 16.2 \\ 1124 \end{pmatrix}$

41. Hoeveel meter steeg Mark per minuut volgens deze uitkomst?
42. Wat zijn de vereffende hoogtes (*adjusted observations*)?
43. Welke waarneming heeft het grootse residu?

Mark herinnert zich dat de weg alleen maar omhoog ging. Dit klopt niet met de tweede waarneming. Hij besluit daarom nogmaals zijn stijgsnelheid te schatten, maar nu met RANSAC.

44. Leg (zonder berekening) uit hoe Mark zijn stijgsnelheid kan bepalen met RANSAC.
45. Wat is in dit geval een goede grenswaarde (*threshold*) om te besluiten dat een waarneming niet in het model past?