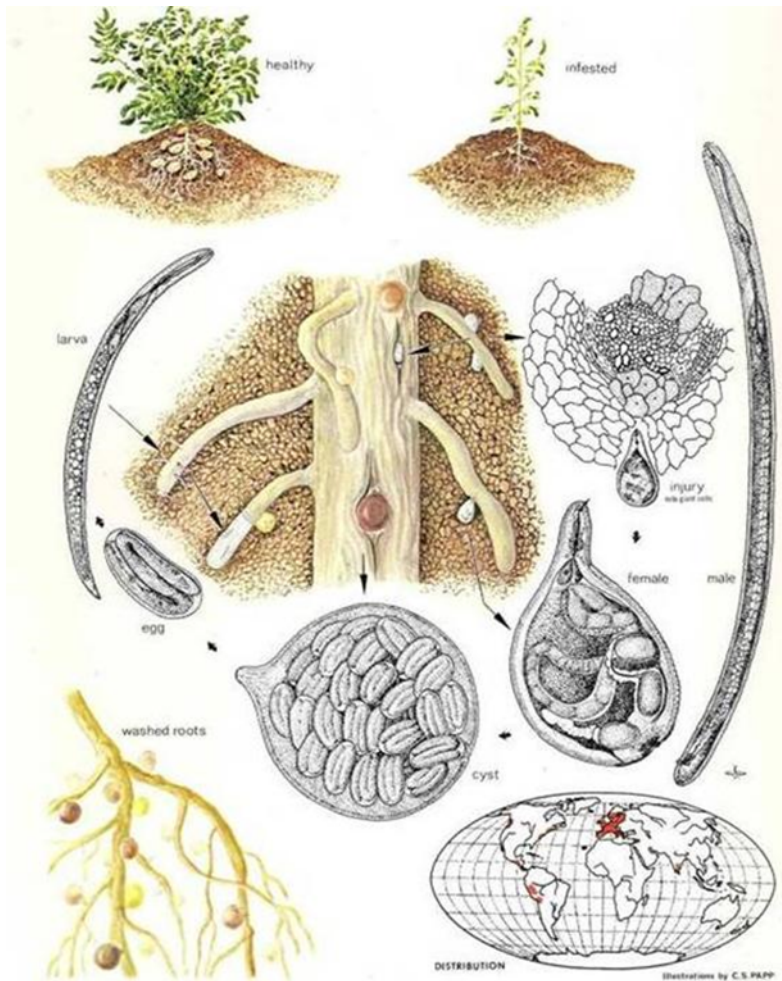




Combinatie van gdal_rasterize en numpy voor berekening aardappelmoeheid

28 juni 2017

Ronnie Lassche & Luuk Schaminée



Inhoud

- Wat is Aardappelmoehheid
- Regelgeving in een notendop
- Aanleiding
- Waarom Open Source (in een esri omgeving)
- Gebruikte componenten
- Het proces in een notedop
- Getallen
- Conclusie



Aardappelmoetheid

Aardappelmoetheid wordt veroorzaakt door 2 cystenaaltjes:

- *Globodera rostochiensis*
- *Globodera pallida*

Beide organismen zijn quarantaine organismen in de EU

Schadebeeld is identiek





EU Regelgeving (geïmplementeerd in NL wetgeving)

Richtlijn 2007/33/EG (van kracht 01-07-2010)

Belangrijkste elementen:

- Grondonderzoek voorafgaand aan teelt van bepaalde gewassen verplicht (pootaardappelen en 'licht gereguleerde gewassen') / vrijstelling van bemonstering in bepaalde situaties mogelijk
- Hoeveelheid grond op basis van perceelhistorie
 - standaard 1500 ml/ha.
 - verlaagd volume (400ml) 600 ml/ha.
- Verplichte wachtperiode na aantonen besmetting
 - 6 jaar na laatste aardappelteelt (= referentiejaar) / bij opnieuw besmetting 3 jaar.
 - bij toepassen bestrijdingsmaatregel: 3 wachtjaren / bij opnieuw besmet / direct na nemen van bestrijdingmaatregel.



Regelgeving vertaald naar 'teeltsituaties' > bereken monstervolume

Teeltsituatie is de teelthistorie in de afgelopen 12 jaar:

- 1 x pootgoed
- 2 x pootgoed
- Laatste aardappelteelt
- 6 jaar geen aardappelteelt
- 12 jaar geen aardappelteelt

Afhankelijk van de teeltsituatie geldt een ander bemonsteringsregime:
600 ml of 1500 ml grondmonster per hectare



Aanleiding

- Bestaand proces niet klaar binnen de gestelde tijdslijmiet
- Bepaalde informatie kan niet getoond worden
- Proces niet controleerbaar
- Veel kleine vlakken als gevolg van overlappende vlakken waarbij tijdens het proces de overlap een eigen vlak wordt

Voorbeeld Noordoostpolder (\pm 2000 percelen)

In totaal 215.934 vlakken over 10 jaar

18.300 < 1 m²

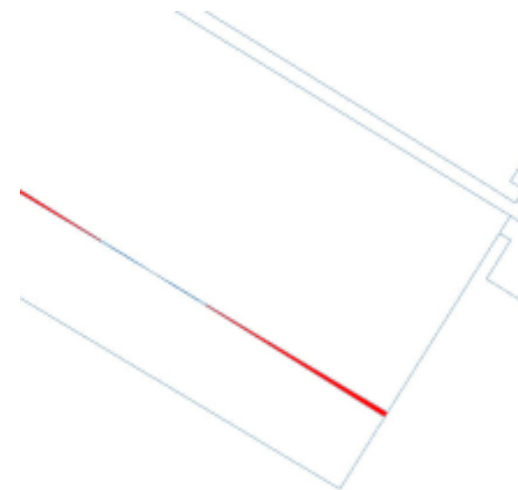
36.164 < 5 m²

46.672 < 10 m²

63.284 < 25 m²

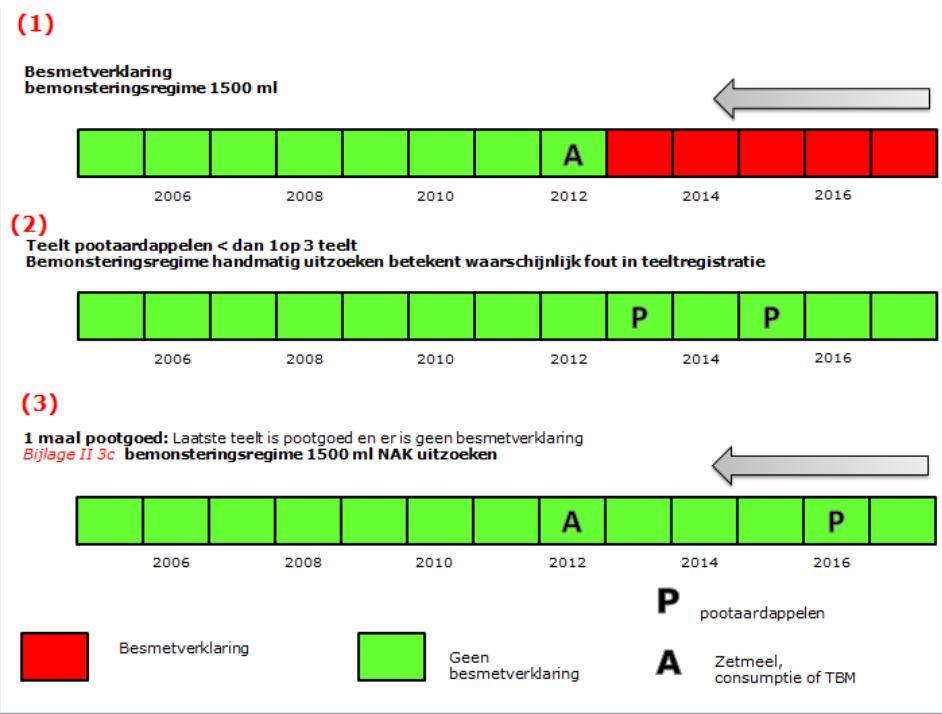
89.771 < 100 m²

46.038 > 5000 m²





Complexiteit beslisregels





Waarom Open Source

- Beschikbaarheid gdal in ArcGIS Desktop
- Beschikbaarheid Python in ArcGIS Desktop
- Mogelijkheid met raster wel tijdslimiet te halen door eenvoudig opschalen
- Mogelijkheid tot fijnmazige toegang tot raster data
- Mogelijkheid tot optimale inrichting van de code zodat het geheugen optimaal gebruikt kan worden
- Ondersteuning van multiprocessing
- Snelheid
- (Persoonlijke voorkeur)



Gebruikte componenten

- gdal_rasterize.exe (met bijbehorende dll's)
- Python:
 - arcpy
 - gdal (Python bindings toegevoegd)
 - numpy
 - multiprocessing
 - logging
 - gdal_merge.py
 - progress.py



Het proces in een notedop

- Onderverdeeld in 4 stappen
- Stap 1: Exporteren van de data (arcpy)
SDE → shapefiles
- Stap 2: Omzetten naar rasterdata en doorrekenen afgelopen 12 jaar (gdal, multiprocessing, numpy & gdal_rasterize)
Shapefiles → tiff (1 band, LZW compressie & datatype byte)
grids 25x25km met celgrootte 2m (instelbaar)
- Stap 3: Omzetten doorgerekende data naar vector
Tiff → shapefiles (gdal_merge.py & arcpy)
doorlopen gemaakte shapefile en velden invullen
- Stap 4: Importeren van de data (arpy)
shapefiles → SDE





Getallen

- Proces maakt gebruik van 2 cores
- Griddata van 12 jaar:
 - 3312 jaarbestanden
 - 92 grids met het agrarisch areaal nederland
 - 92 grids met opgetelde jaarbestanden
 - 1 samengevoegd grid voor Nederland
- 1 resulterende kaartlaag met meer dan 3 miljoen vlakken
- Initieel maken van alle data 11 uur (1 malig)
- Vernieuwen data 9 uur (wekelijks)
- '*Big data*': 546.250.000.000 cellen in 2 uur tijd → 75.868.056 cellen per seconde



Conclusie

- Robuuste software + proces robuuster te programmeren
- Vectoriseren van 1 groot tif bestand sneller dan 92 losse tif bestanden
- Zonder rechtstreeks gebruik van gdal in Python was het niet gelukt
- Mogelijkheden voor versnellen:
 - Server met meer cores
 - Exporteren data over meerdere cores verdelen
 - Doorlopen resultaten tabel over meerdere cores verdelen (pure Python code/OGR/in de database)
 - GeoPackage/Database gebruiken voor opslag van resulterende kaartlaag en hierin met SQL waarden velden vullen



Vragen