Hallo!

Hier vind je een kort overzichtje van hoe het project in elkaar zit. Specifiek wordt ingegaan op de exacte methode. Meer uitleg over de code staat in de .rmd files zelf..

# Preprocessing.R

Hier worden enkele zaken goed gezet. De detailkarteringen worden geclipt met het extent van het hoogteraster. Verder krijgen ze extra kolommen en worden ze omgezet naar shapefiles.

# Structuurmetrieken.rmd

Hier worden onafhankelijke variabelen van het model gecreerd. Gaande van begrazingsdruk tot “edgness”. Begrazing moet nog beter geschat worden. De meeste metrieken zijn nog niet zo performant. Dit scriptje is maw. een work in progress.

# Model en modelanalyse.rmd

Disclaimer: er is nog wel wat werk aan dit deel. We moeten nog fiksen dat het model subgrafen aankan (dat zou het ook een stuk sneller kunnen maken). Verder is het wellicht interessant om het model ook eens multivariaat te proberen.

Hier wordt het model gerund. We maken gebruik van INLA om ruimtelijke structuren in rekening te kunnen brengen. Omdat de data vooral veel nulwaarnemingen bevat (elke NA waarde = soort afwezig) is de data niet makkelijk verdeeld (erg zero-inflated). We werken daarom met een logistische regressie om de aan/afwezigheid van een plant te voorspellen. Daarnaast is de data >0 niet normaal verdeeld. We werken daarom met een gamma distributie. Dat heeft als implicatie dat we niet linear maar logaritmisch werken.

# Heatmap generating.rmd

Hier wordt van punten naar heatmaps gegaan. Dit script is initieel in arcpy gemaakt en daarna omgezet naar .Rmd. Ik heb nog nagekeken of alles correct werd overgezet en hier en daar foute vertalingen omgezet. Als er Daarvoor gebeuren heel wat dingen.

1. Punten worden gefilterd op > 2015
2. Dan worden ook nog de planten die als vlak gekarteerd werden toegevoegd. Als ergens punten van een soort in een vlak liggen van dezelfde soort worden deze punten verwijderd.
3. Er worden gebieden gegenereerd door DBSCAN clustering op basis van het jaar. Deze gebieden stellen voor waar er dat jaar gekarteerd is. (afgezien van “outliers”)
4. Een buffer van 1,5m wordt getekend rond de clusters
5. Deze gebieden overlappen doorheen de jaren. Bij overlap wordt gekeken voor welk jaar de meeste punten werden gezet in het overlappende stuk. Punten die in een overlappend stuk zitten worden verwijderd tenzij het overlappend polygoon met de meeste records hetzelfde jaar heeft als het onderliggende punt.
6. Gezien de karteringen in 5mx5m blokken gebeuren kunnen zullen punten idealiter in het centrum van een patch planten liggen. Het punt is geen indicatie van de abundantie op die exacte plaats. Wel een indicatie van de abundantie in dat 5x5m stuk. Daarom maken we een kernel rond dit punt waarin we de observatie uitsmeren over een oppervlkate van 5x5m volgens een gaussiaanse verdeling. We gaan er dus wel nog vanuit dat het centrum van de “patch” de meeste observaties bevat. Uiteraard zit er ook een fout op de gps-locaties. Die negeren we gezien we er redelijkerwijs vanuit kunnen gaan dat dit noise is op de data. Er zit maw. geen bias op tenzij ergens een GPS van één van de medewerkers echt kapot is.

Doordat onze observaties in meerdere rastercellen kunnen overlopen kan het lijken alsof er het aantal observaties voor het model artificieel verhoogd wordt. Maar omdat er wellicht geobserveerd wordt in meerdere cellen van ons grid vind ik dat niet onredelijk. Verder zorgt de buurtstructuur in de analyse ervoor dat naburige cellen weinig nieuws bijdragen aan het model. Geisoleerde cellen doen dat veel meer

1. Verder zetten we de resultaten van de kernelling om in een verwacht aantal individuen per pixel. Gewoon voor latere interpretatie te vergemakkelijken.
2. Dan worden de karteringspolygonen nog gerasterized (gewoon door de abundantie als burn in value te gebruiken)