National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

Comparing two approaches for estimating nitrate levels

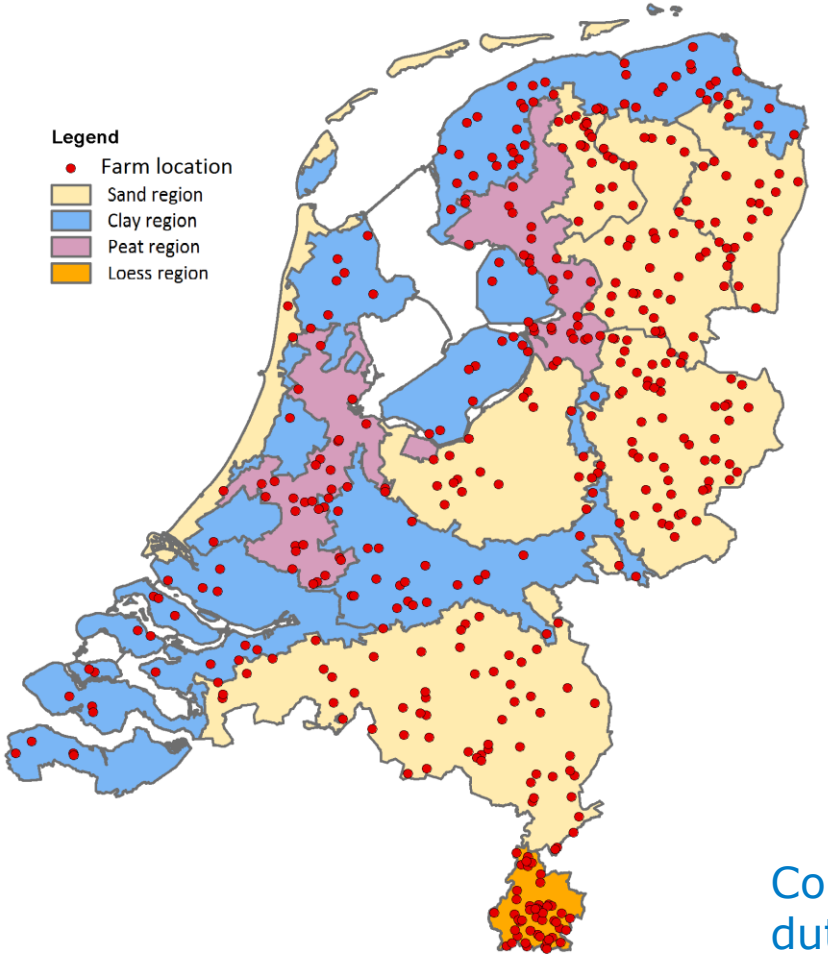
in upper groundwater in Dutch sandy soils

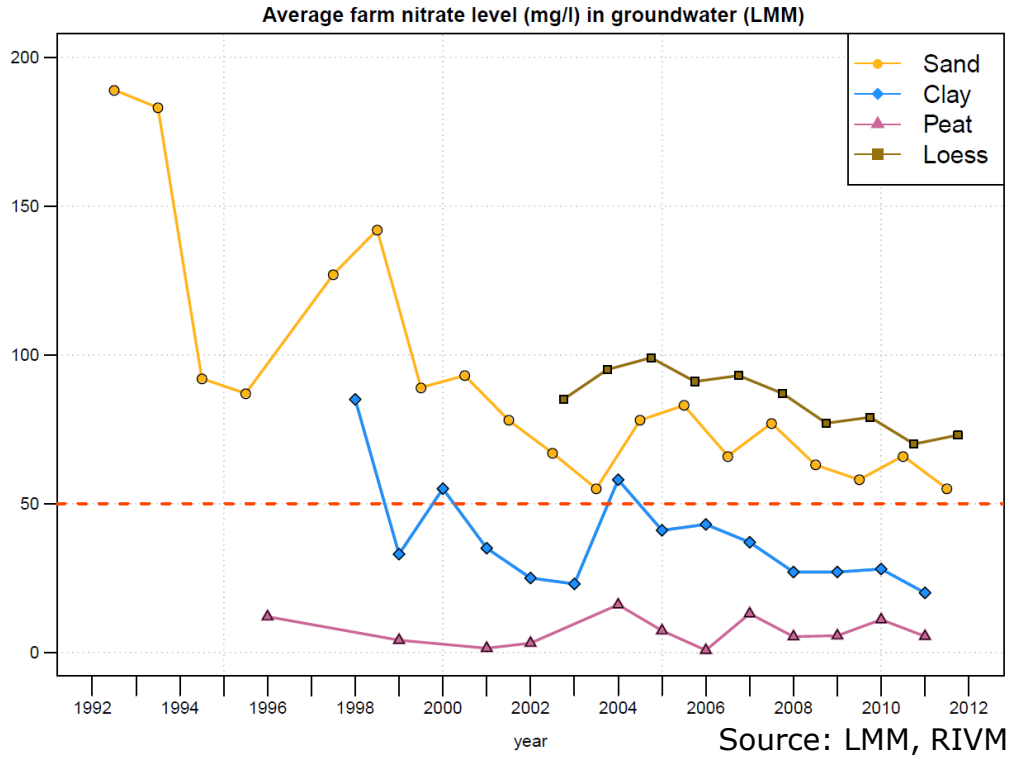
C.J. de Jong & L.J.M. Boumans

National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

Contents

Results from previous years (LMM):

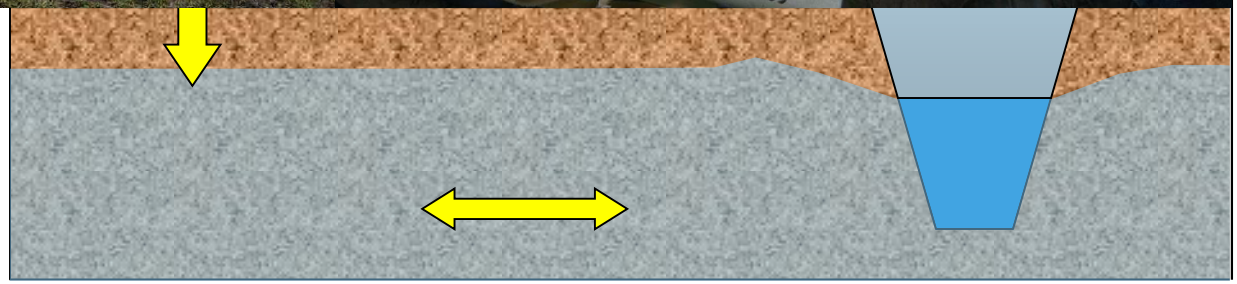




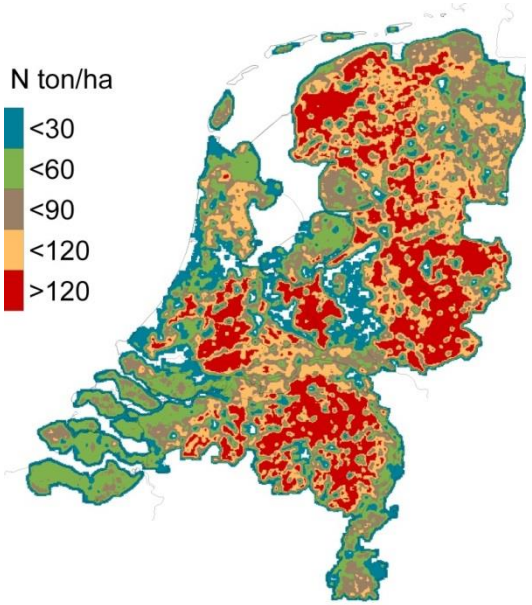
Source: LMM, RIVM

Conclusion: sandy soils form a major part of dutch soils and high nitrate levels (still) occur here

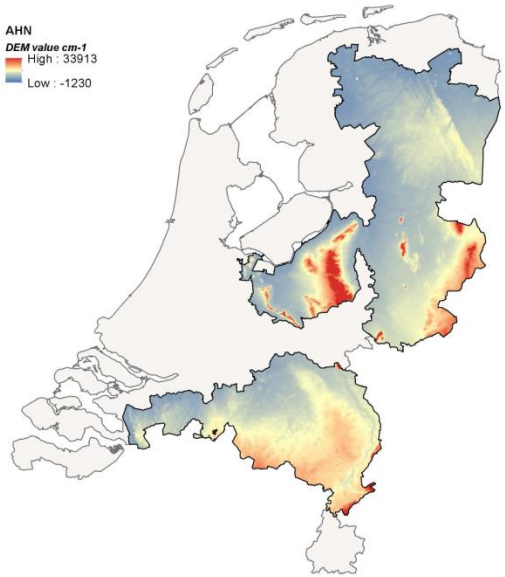
Sampling the upper groundwater



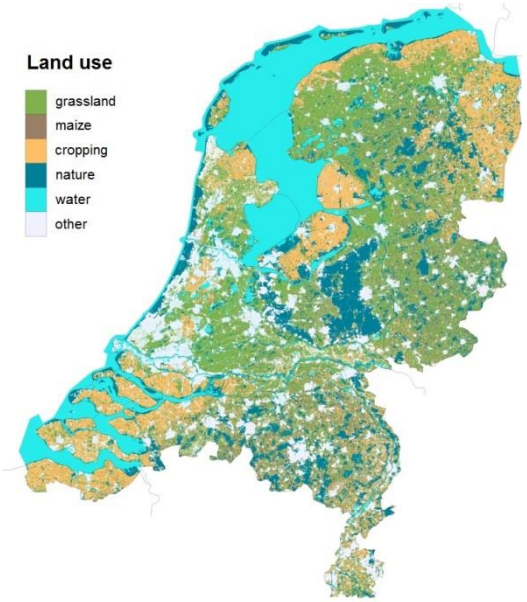
Additional data: mapped covariables



Manure/Fertilizer N



Elevation (DEM)



Land Use map (modified LGN6)

Other maps: e.g. soil maps, hydrological districts, % organic matter

From sampling to mapping: two methods

- Method A: Regression modelling (At farm level)

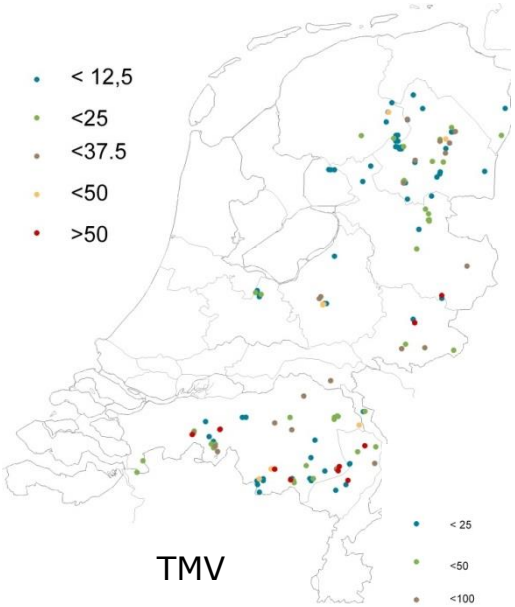
- Method B: Regression Kriging (Linear modelling + Kriging)

(at sample point level)

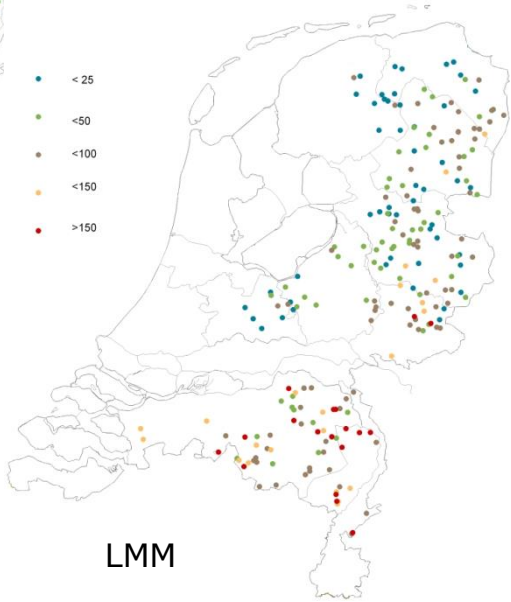
Method A

- Fortran77/GenStat
- Regression modeling
- Data used: ~200 farm location averaged data points (LMM)

  + 150 locations in nature areas (TMV)
- Time frame: 2007-2010, resulting measurements were averaged
- Land use assumed constant
- Independent variables per farm
- Prediction resolution 500 m x 500 m



TMV

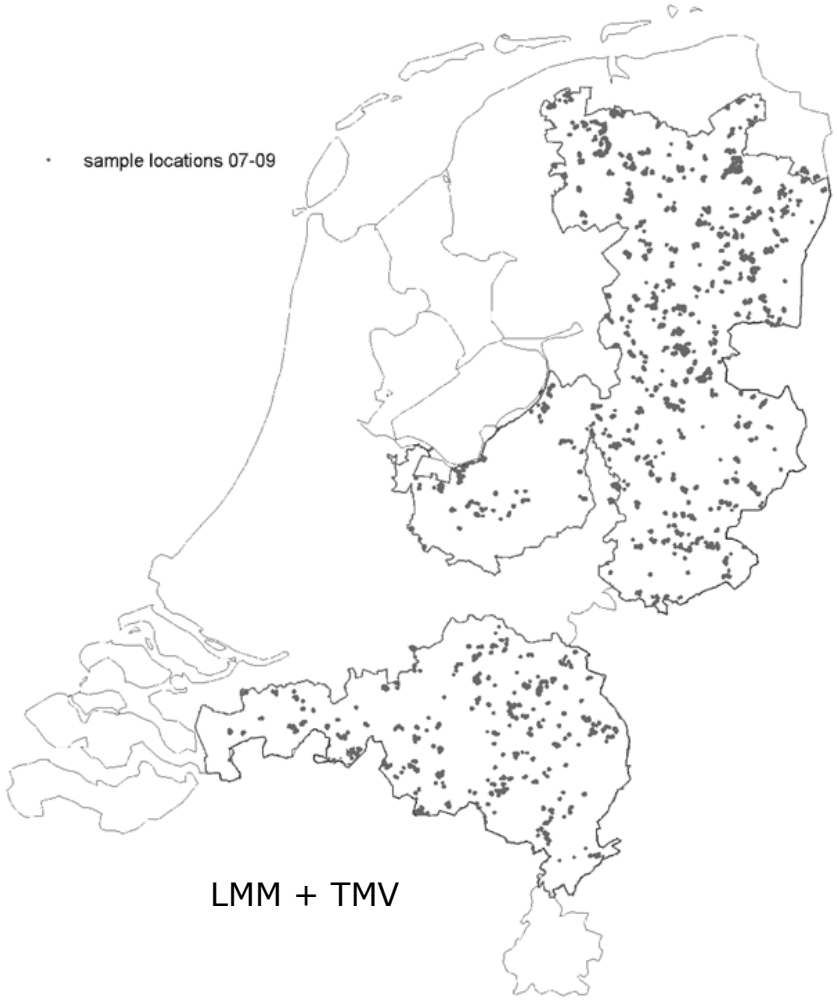


LMM

Method B

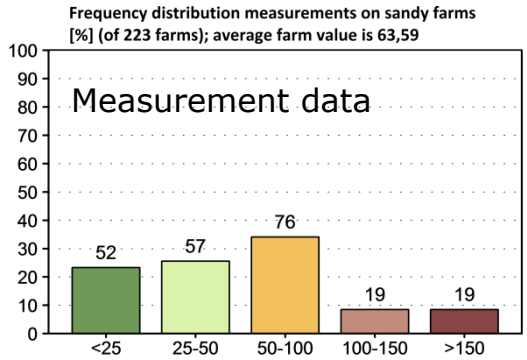
- R+Gstat/ArcGIS
- Combination of linear modelling and kriging of residual data
- Data from LMM + TMV
- 16 sample locations per farm, 10 per nature area

  (in total ~18000 sample locations)
- 2007, 2008 and 2009 are modeled separately
- Prediction resolution 25m x 25m

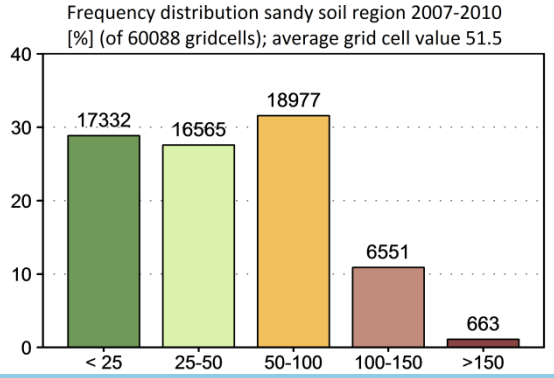


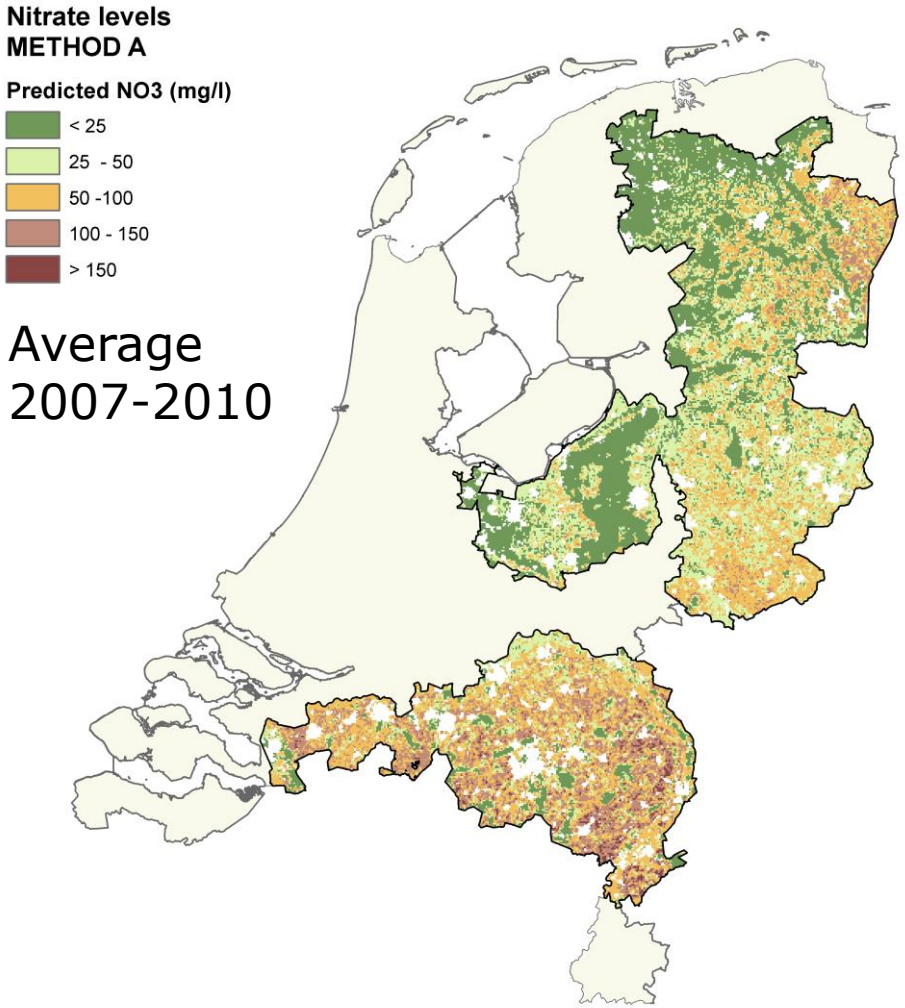
LMM + TMV

Results method A (regression only)



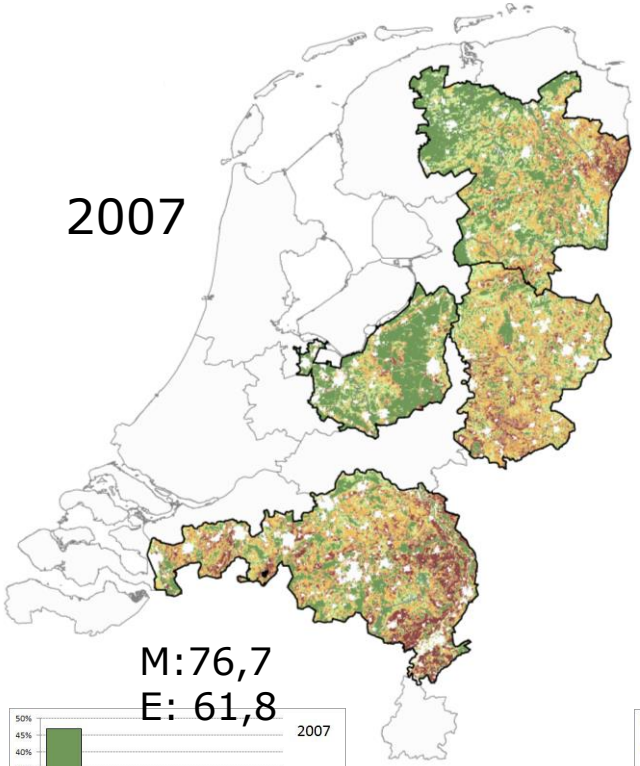
Average measured value 63,6*
Average estimated value 51,5*
*For cells with agriculture



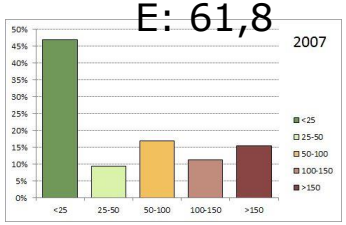


Results method B (regression kriging)

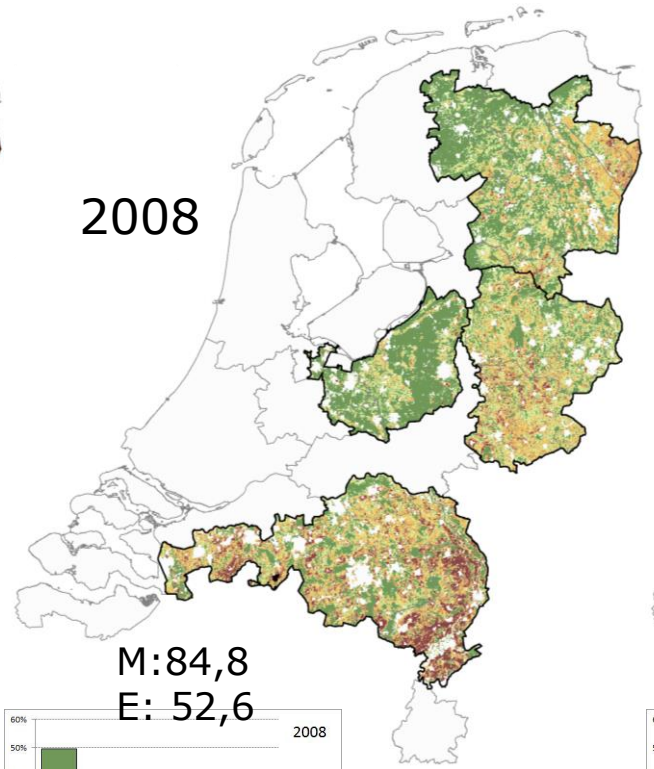
2007



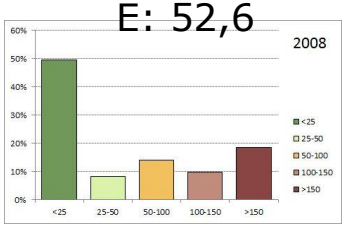
M:76,7
E: 61,8



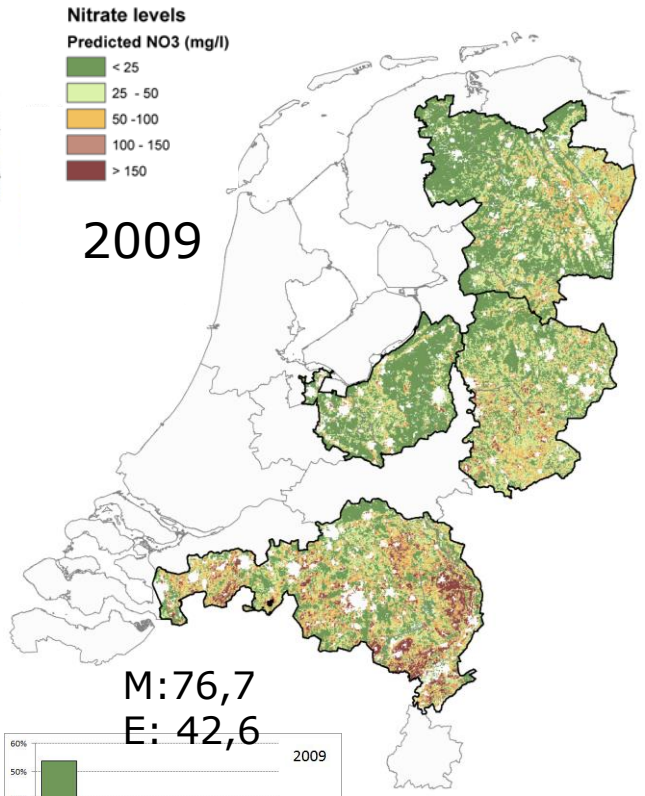
2008



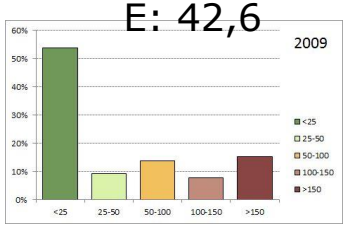
M:84,8
E: 52,6



2009

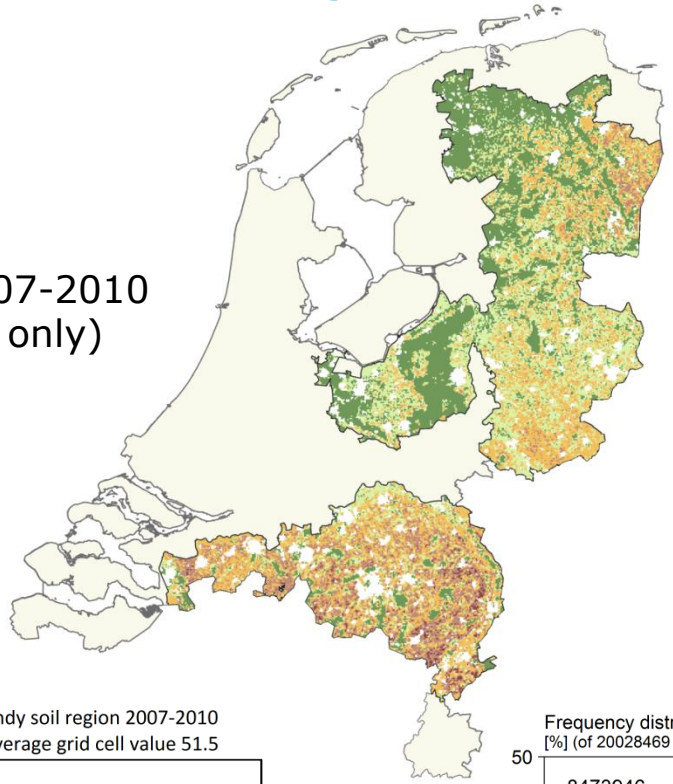


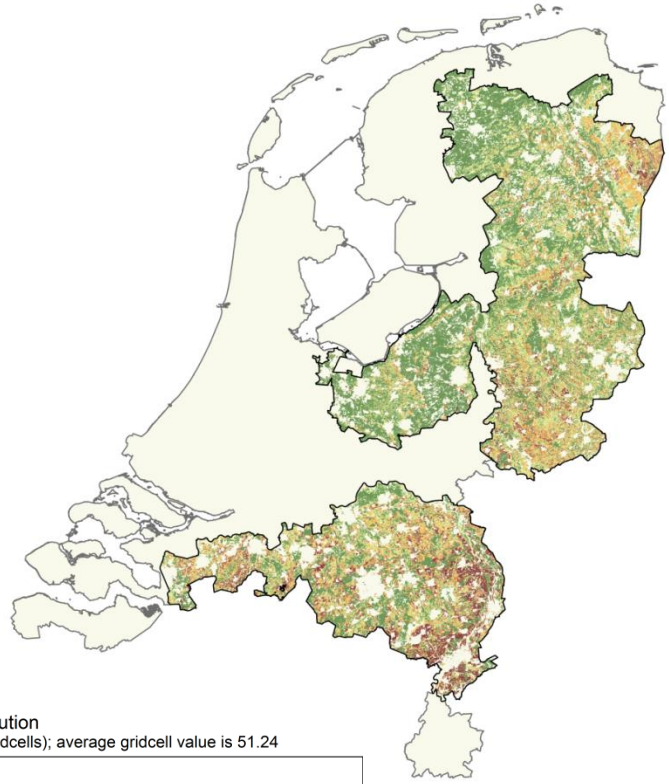
M:76,7
E: 42,6

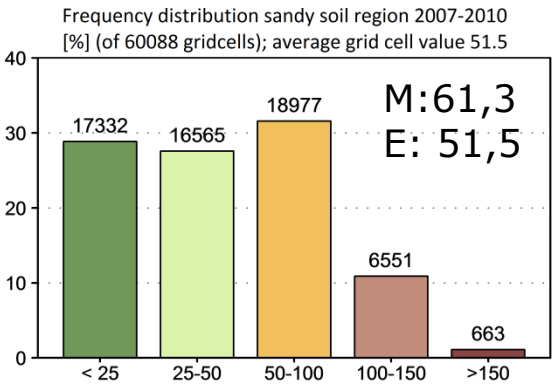


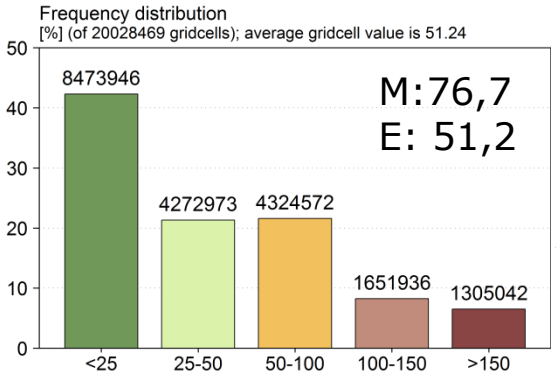
Both methods together

Method A: 2007-2010
(regression only)



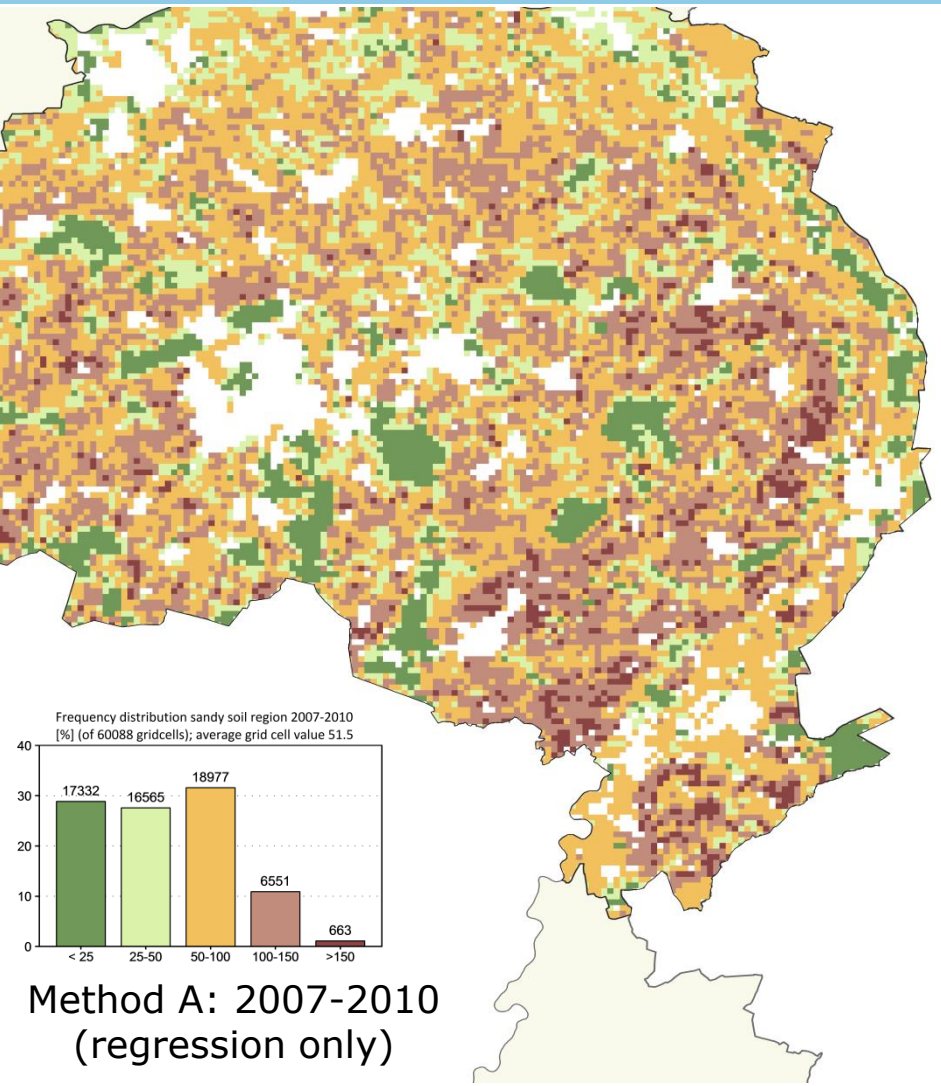




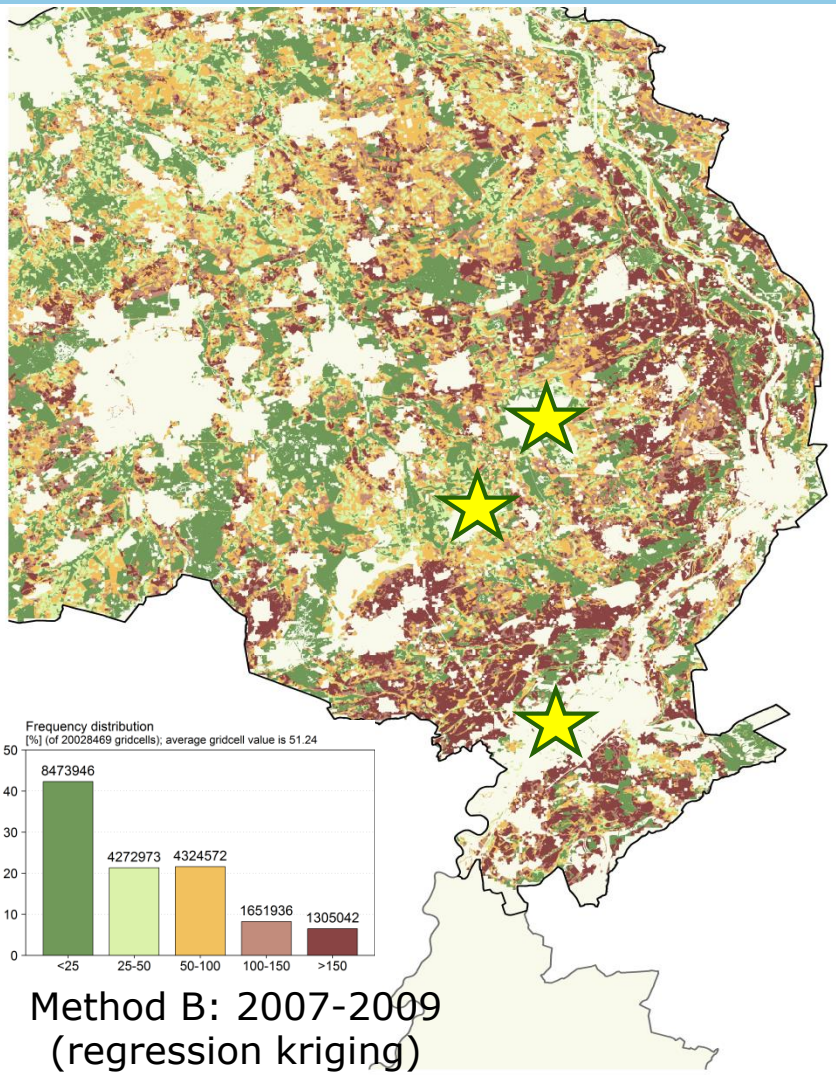


Method B: 2007-2009
(regression kriging)
Averaged from three years

5. Comparison



Method A: 2007-2010
(regression only)



Method B: 2007-2009
(regression kriging)

Differences between methods

	Method A (regression only)	Method B (regression kriging)
INPUT	Certified laboratory analyses of mixed samples per farm or nature area, over 4 years (2007-2010)	16 field sample measurements per farm, 10 per nature area, for three years (2007, 2008, 2009)
INPUT	Covariable maps: 4 categorial + 2 continuous value	15 continuous maps, 13 categorial maps
METHOD	Some gaps in input map were filled	No map modifications
METHOD	Separate models for measurements in nature and agriculture	One model for all data, per year
METHOD	Stepwise selection of independent variables (including spatial coordinates)	Stepwise selection of independent variables (without spatial coordinates) + kriging of residuals
METHOD	Using frequencies of clustered main map classes as independent variables (52 potential variables)	No map class clustering (200 potential independent variables)
METHOD	crossvalidation (leave-one-out)	crossvalidation loocv (only spatial part)
OUTPUT	1 final map	1 map per year
OUTPUT	All input map cells receive prediction values	No prediction for missing cells
OUTPUT	Prediction resolution 500m x 500m for cells with over 25% agriculture or nature	Prediction resolution 25m x 25m for all participating cells

Discussion

- Logically, models contain different covariables
- Which approach is 'better'? Or are they just versions?
- Methods were not aimed to be compared with eachother

- General impression: maps yield comparable pattern, though locally patterns differ
- Methods yield a different distribution when looking at classes (method A more in the centre classes, method B in lower and higher classes)

- Best fit method A explained 55% of variation
- Best fits method B explained 43-55% of variation

- However: Method A not exactly in same time frame as method B

Conclusions

- Both methods capable of estimating the measured variable
- Generally, maps give same impression, also mean value, but local differences exist
- Many choices to make, realize influence of this on modelling outcome
- For policy measures, certified laboratory samples are demanded, so method A would be presented

Thanks for the attention

Questions?