Statistical upscaling of terrestrial greenhouse gas emissions

Gerard Heuvelink, Wim De Vries, Tom Hoogland, Hans Kros and Gert Jan Reinds

ALTERRA
WAGENINGEN UR

NitroEurope IP

Upscaling

- Upscaling is taken as synonymous to *aggregation*, where the interest is in obtaining an integrated value of a variable over an area or volume of a given size and/or a time interval of given length
- Example: chambers may measure emission over 30 minutes for a 0.25 m^2 surface, whereas the interest is in annual values for a large region or country
- Upscaling is based on ‘point’ observations, but perhaps auxiliary information can be used to improve accuracy

Two fundamentally different approaches to upscaling

1. Design-based

- Makes no assumption about space-time variability structure (‘model-free’)
- Does not suffer from making wrong assumptions
- Locations of observations must be selected with probability sampling, usually simple random sampling
- Simple random sampling may be replaced with more efficient designs (e.g. stratified sampling, two-phase random sampling, cluster sampling)
- Widespread misconception that design-based methods cannot be applied when there is spatial (temporal) correlation
- Problems with scarce data and preferential sampling

Two fundamentally different approaches to upscaling

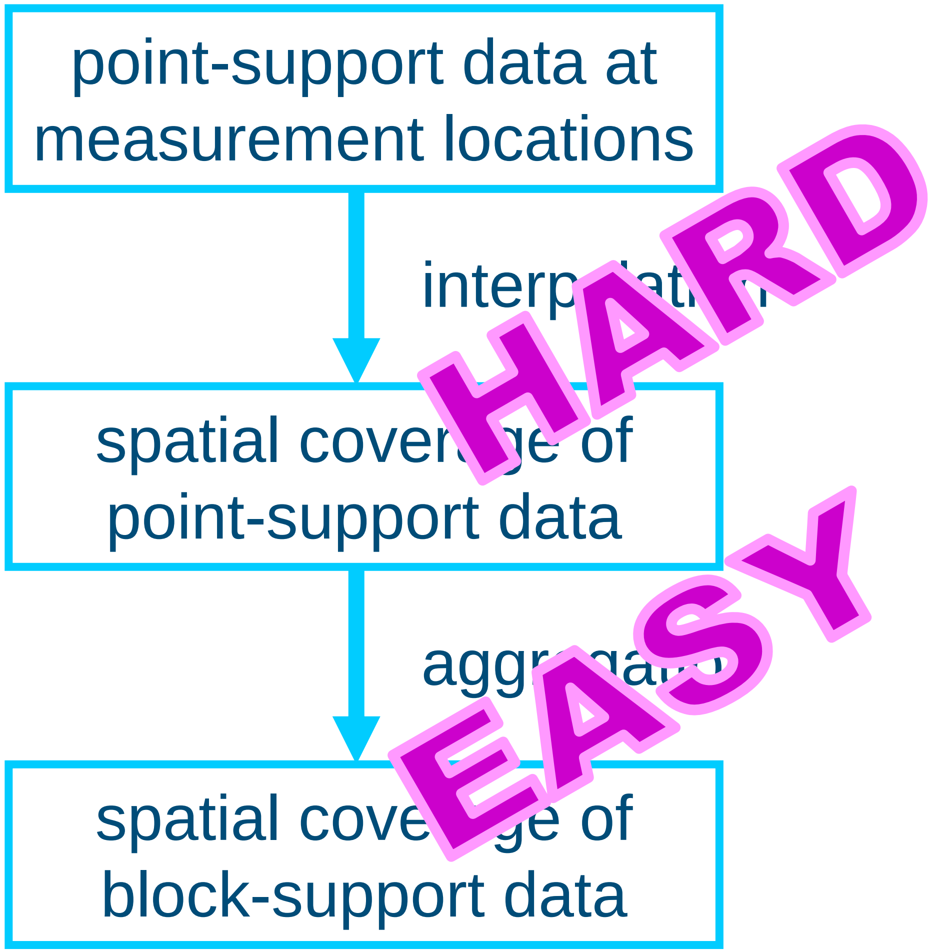
2. Model-based

- Assumes a (statistical) model that characterises the space-time behaviour of the variable of interest
- Statistical model because space-time behaviour is partially unpredictable: include stochastic term to represent uncertainty
- Model also includes a deterministic trend, ranging from an unknown constant to a complex process model such as DNDC: $Z(x,t) = m(x,t) + \varepsilon(x,t)$
- Given the model, trend and observations, estimates of upscaled variable are obtained with block-kriging
- Model-based more flexible and potentially more efficient than design-based, but makes assumptions

ALTERRA
WAGENINGEN UR

NitroEurope IP

Model-based upscaling: two main steps



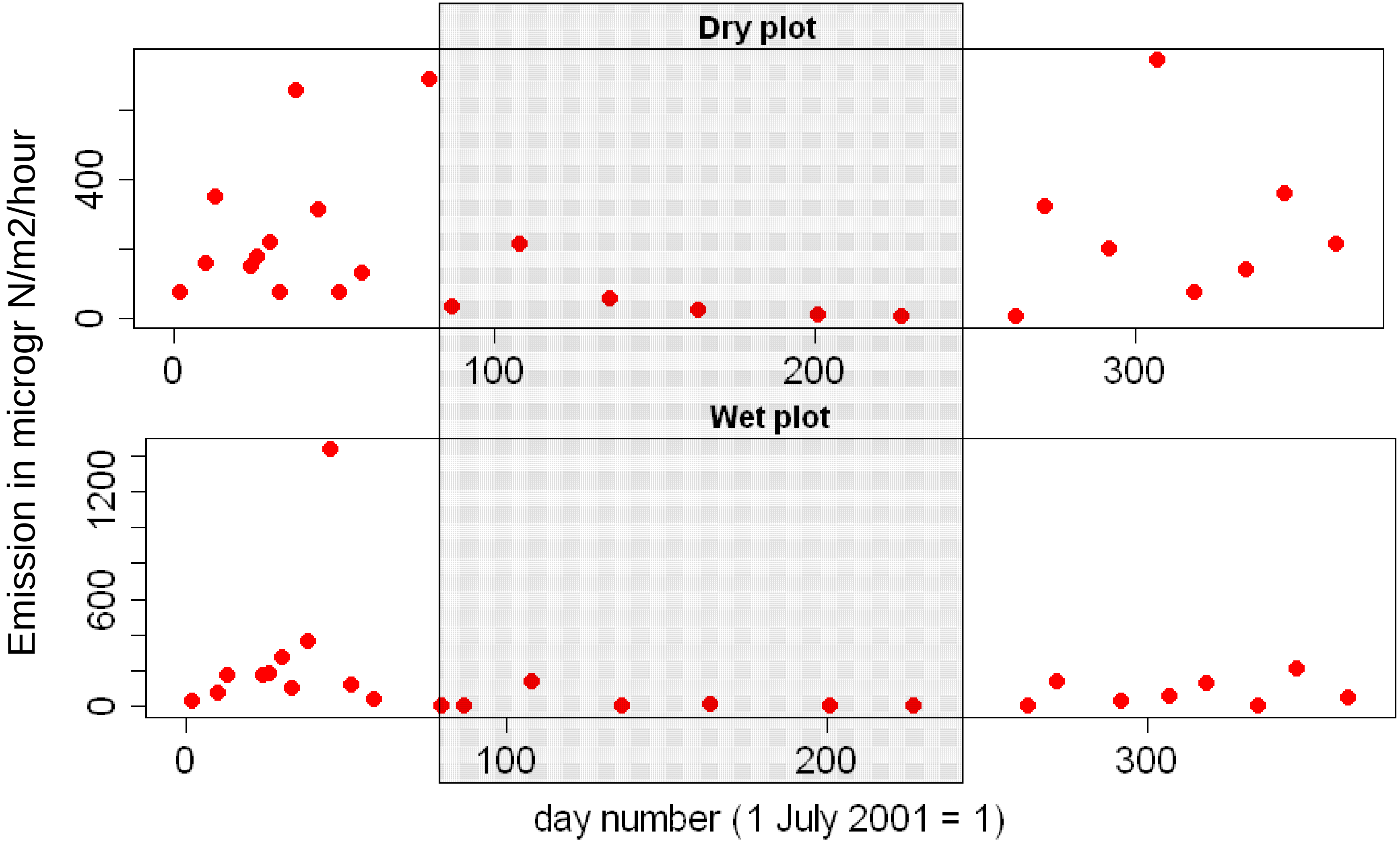
ALTERRA
WAGENINGEN UR

NitroEurope IP

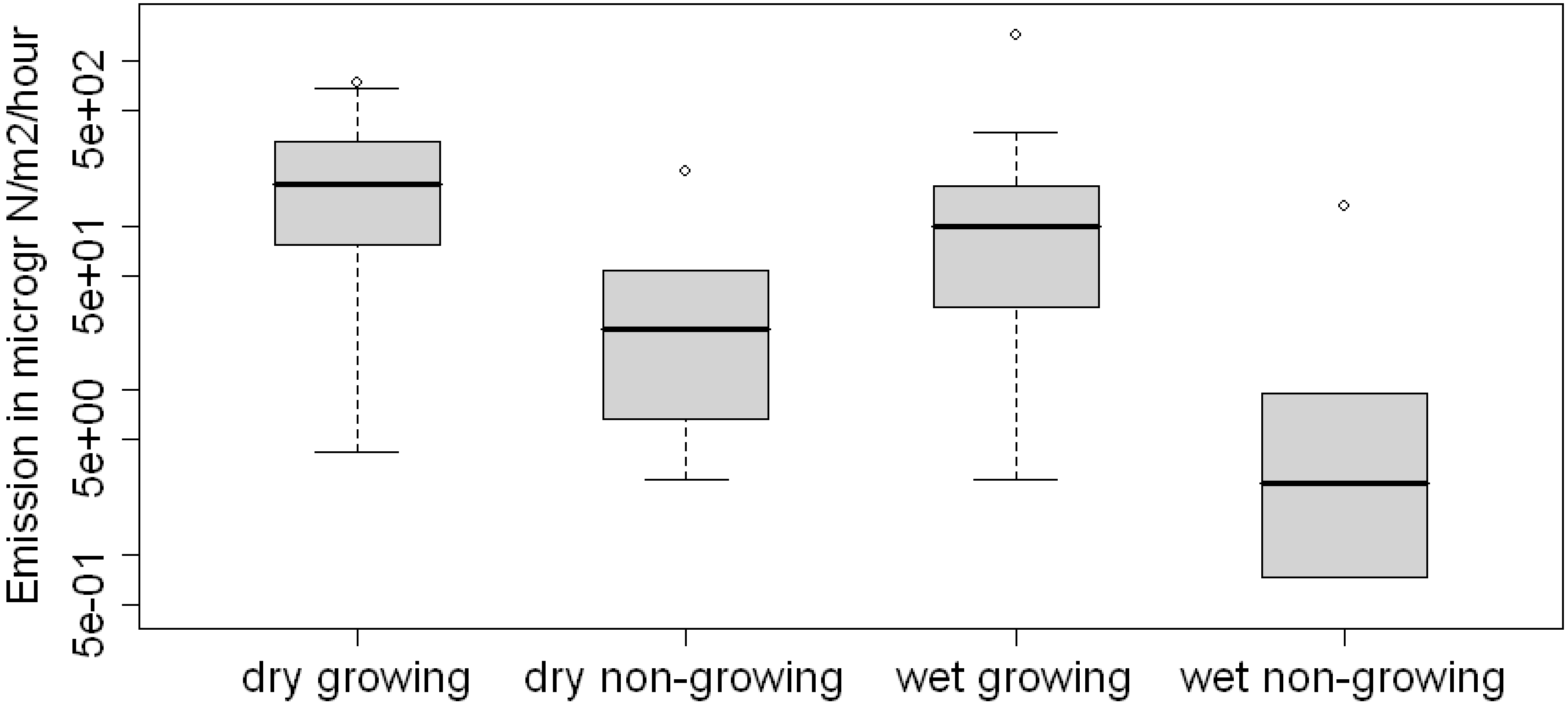
Example design-based upscaling

- Aggregation over time only: from half-hour chamber measurements to annual average (1 July 2001 to 30 June 2002)
- N_2O measured at 26 times for two grassland parcels (dry and wet) in Western Dutch peat soil area
- Assume stratified random sampling with two strata: growing season (1 March to 30 September) and non-growing season
- Higher sampling density in growing season

Measurements over time



Box plots show differences between plots and season



ALTERRA WAGENINGEN UR

NitroEurope IP

Statistical inference

$$\mu = \frac{T_g}{T}\mu_g + \frac{T_n}{T}\mu_n$$

$$\hat{\sigma}_g^2 = s_g^2 = \frac{1}{n_g - 1}\sum_{i=1}^{n_g}(x_{gi} - \bar{x}_g)^2$$

$$\hat{\mu}_g = \bar{x}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} x_{gi}$$

$$\hat{\mu} = \frac{T_g}{T}\hat{\mu}_g + \frac{T_n}{T}\hat{\mu}_n$$

$$\sigma(\mu_g - \hat{\mu}_g) = \frac{\sigma_g}{\sqrt{n_g}}$$

$$\sigma(\mu - \hat{\mu}) = \sqrt{\left(\frac{T_g}{T}\right)^2 \frac{\sigma_g^2}{n_g} + \left(\frac{T_n}{T}\right)^2 \frac{\sigma_n^2}{n_n}}$$

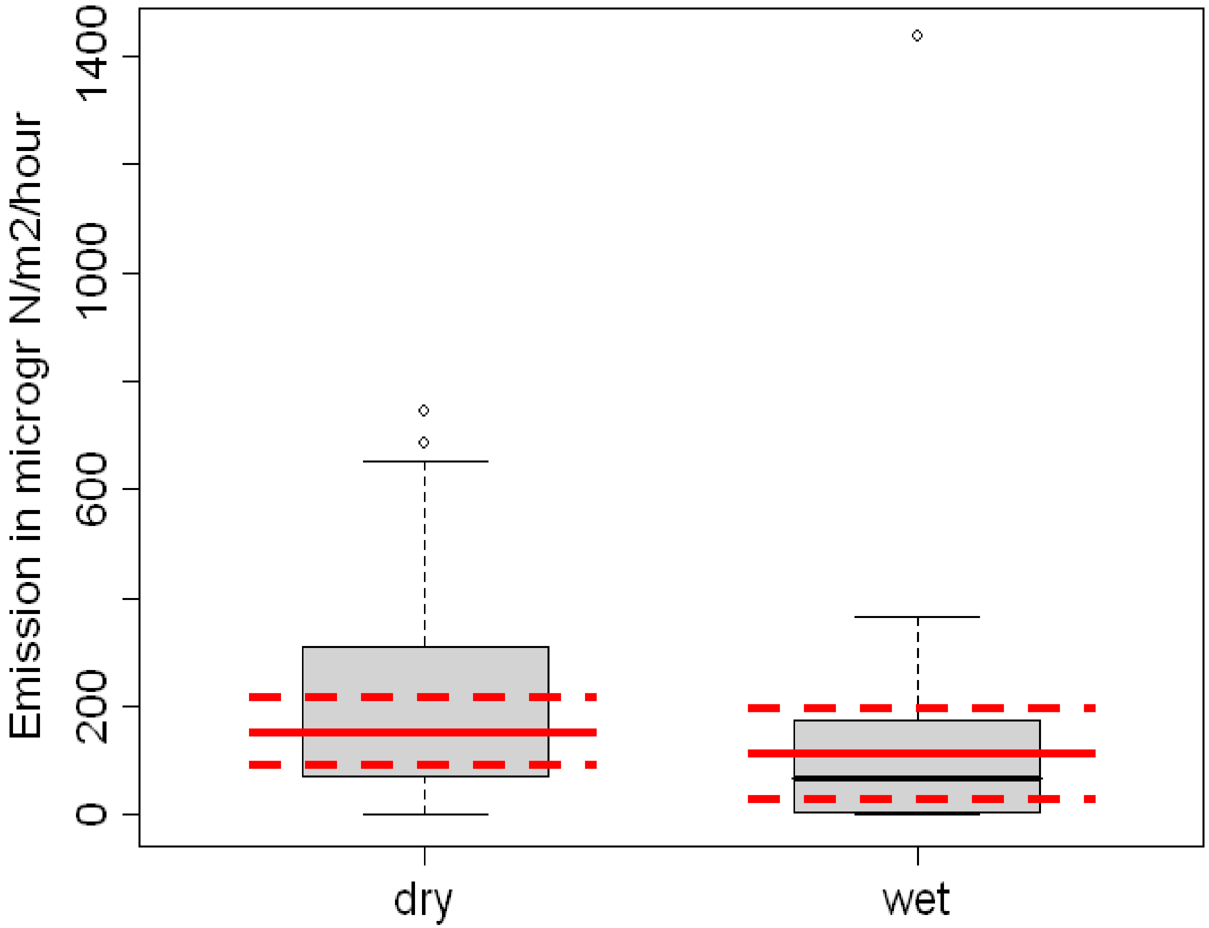
Results (recall n=26)

$$\hat{\mu}_{dry} = 168.6$$

$$\sigma(\mu_{dry} - \hat{\mu}_{dry}) = 31.8$$

$$\hat{\mu}_{wet} = 112.5$$

$$\sigma(\mu_{wet} - \hat{\mu}_{wet}) = 40.6$$



ALTERRA WAGENINGEN UR

NitroEurope IP

Example model-based upscaling

- N_2O emission in natural areas in Europe, data from (after screening 115 observations remain):
  - NOFRETETE
  - Stehfest and Bouwman
  - Denier van der Gon

- Candidate predictors (auxiliary information incorporated in trend):
  - Climate (precipitation, nr frost days, temperature)
  - Soil (pH, organic carbon, texture, CN ratio, bulk density)
  - N deposition
  - Vegetation type (coniferous, deciduous, grass&heath)

- Use regression kriging

Regression-kriging

target variable = f(explanatory variables) + stochastic residual

possibly spatially autocorrelated

Example:

$$\log(N_2O) = \beta_0 + \beta_1 \cdot f(\mathrm{Prec}) + \beta_2 \cdot \log(N_{dep}) +$$
$$\beta_3 \cdot pH_{soil} + \beta_4 \cdot OrgC_{soil} + \beta_5 \cdot \mathrm{vegtype} + \beta_6 \cdot nr_{frost\,days} + \varepsilon$$

Steps in regression kriging upscaling

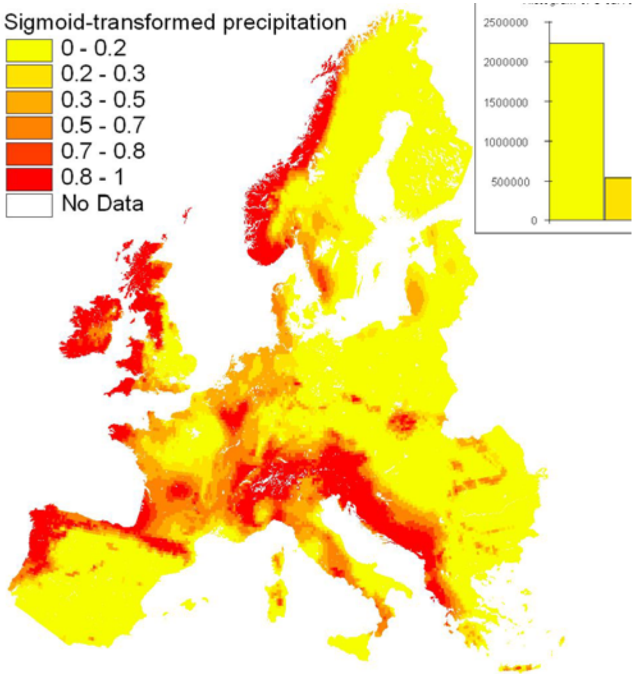
- Build and fit regression model using emission and predictor data (auxiliary information)
- Run regression model for the whole of (natural) Europe
- Compute regression residuals at measurement locations
- Estimate spatial correlation structure of residuals
- Interpolate residuals using kriging
- Add interpolated residual to regression model output
- Slightly better approach is to integrate estimation of regression coefficients and kriging of residuals (WLS instead of OLS)
- Aggregate resulting map to desired support (e.g. compute average emission over regions or nations)

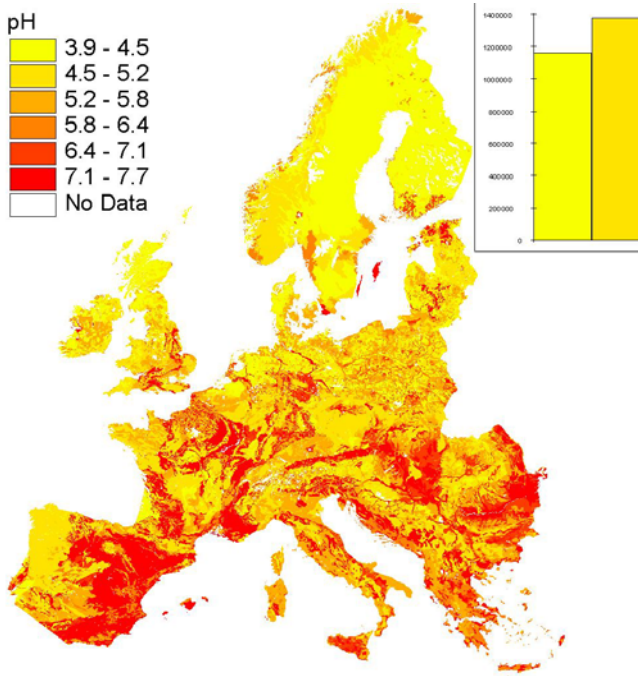
Regression model and parameter estimates based on 115 observations across Europe

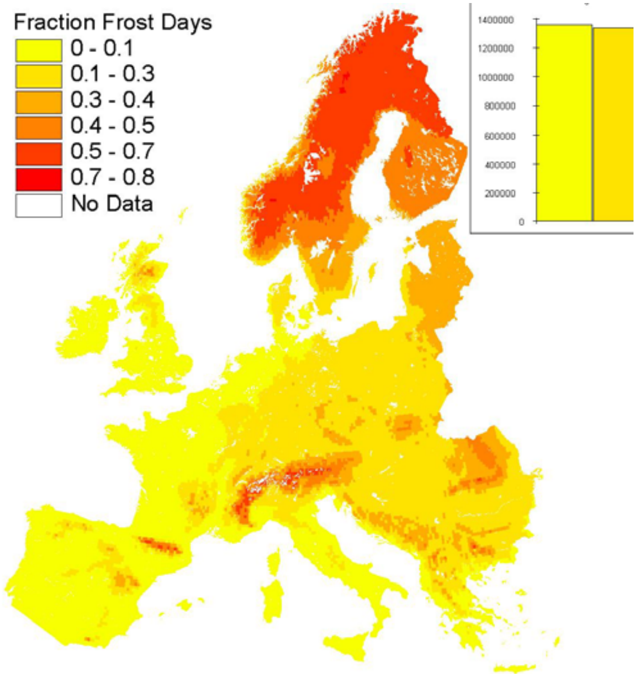
$$\log(N_2O) = \beta_0 + \beta_1 \cdot \mathrm{Sigmoid(Prec)} + \beta_2 \cdot \log(N_{dep}) + \beta_3 \cdot pH_{soil} + $$
$$+ \beta_4 \cdot \mathrm{OrgC}_{soil} + \beta_5 \cdot \mathrm{Indicator(deciduous)} + \beta_6 \cdot \frac{\text{nr frost days}}{365} + \varepsilon$$

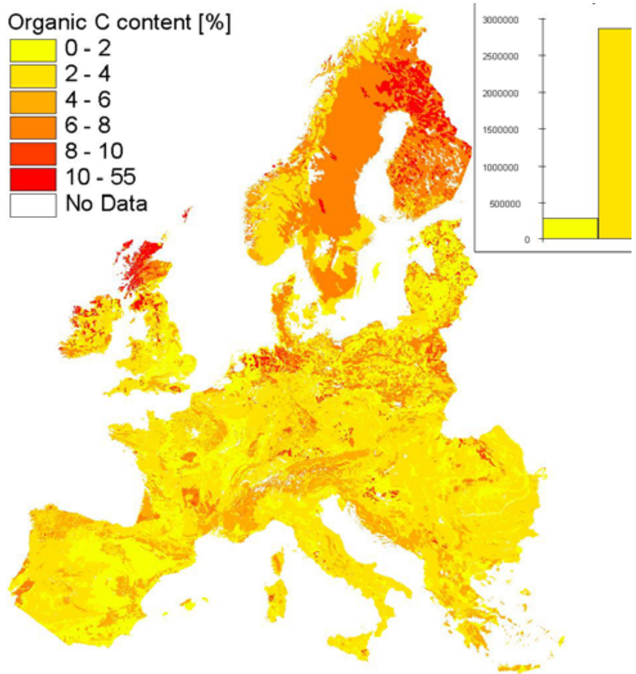
β_0	β_1	β_2	β_3	β_4	β_5	β_6
0.170	–0.246	0.214	–0.082	0.010	0.287	0.725

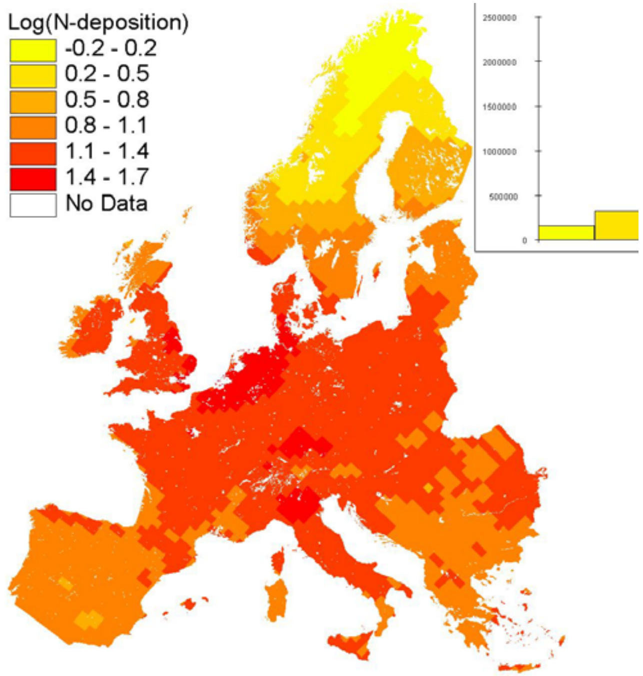
Predictor maps



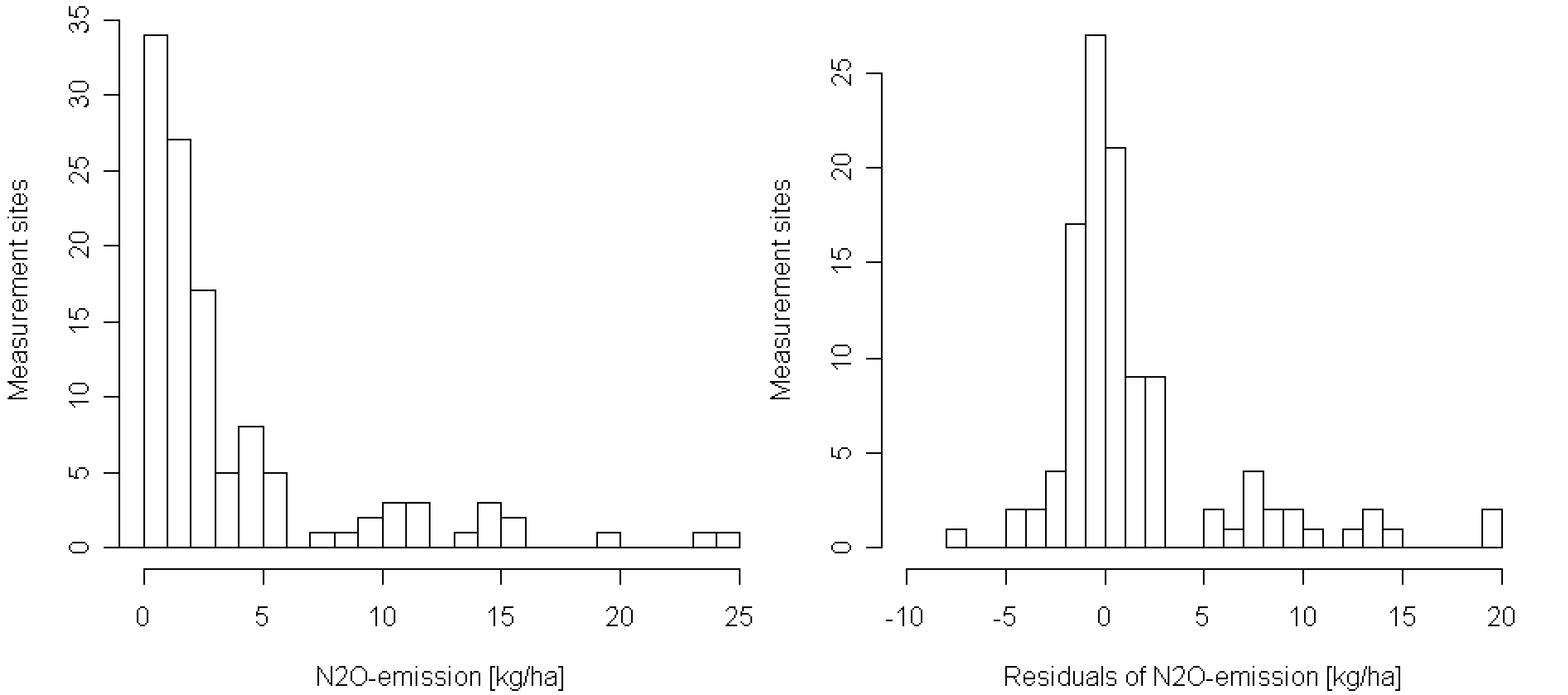








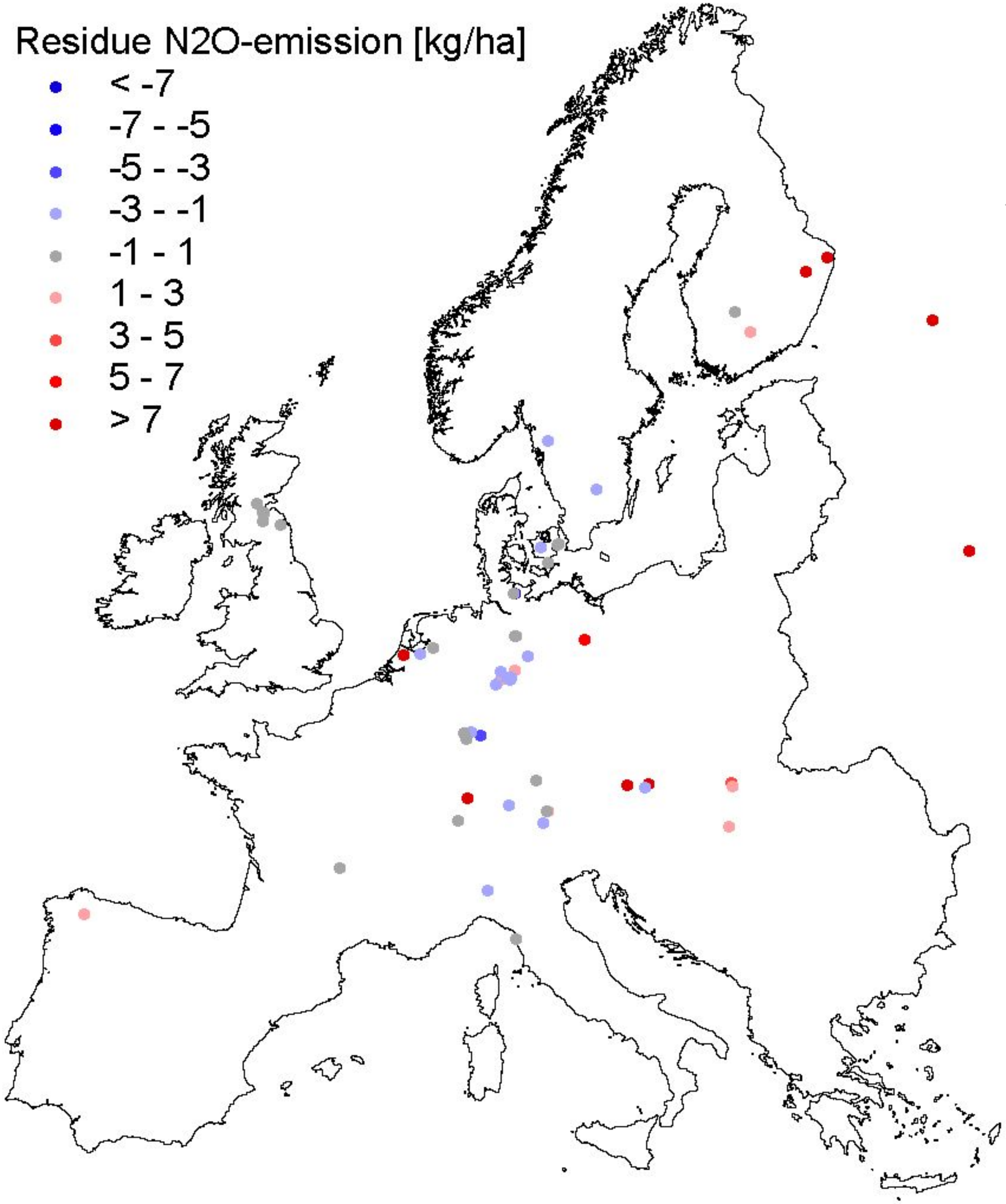
Regression explains little variation (R^2=0.20)

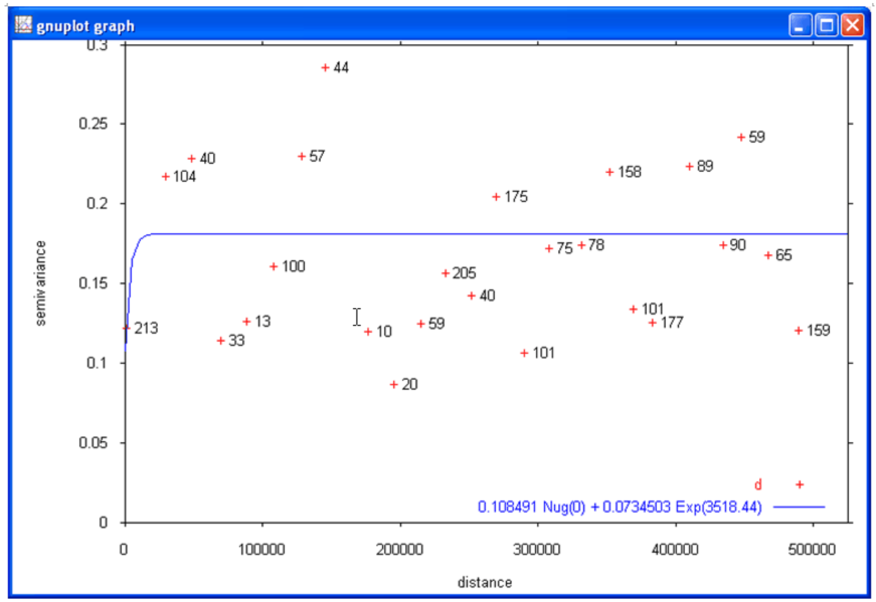


ALTERRA
WAGENINGEN UR

NitroEurope IP

Residual weakly spatially correlated

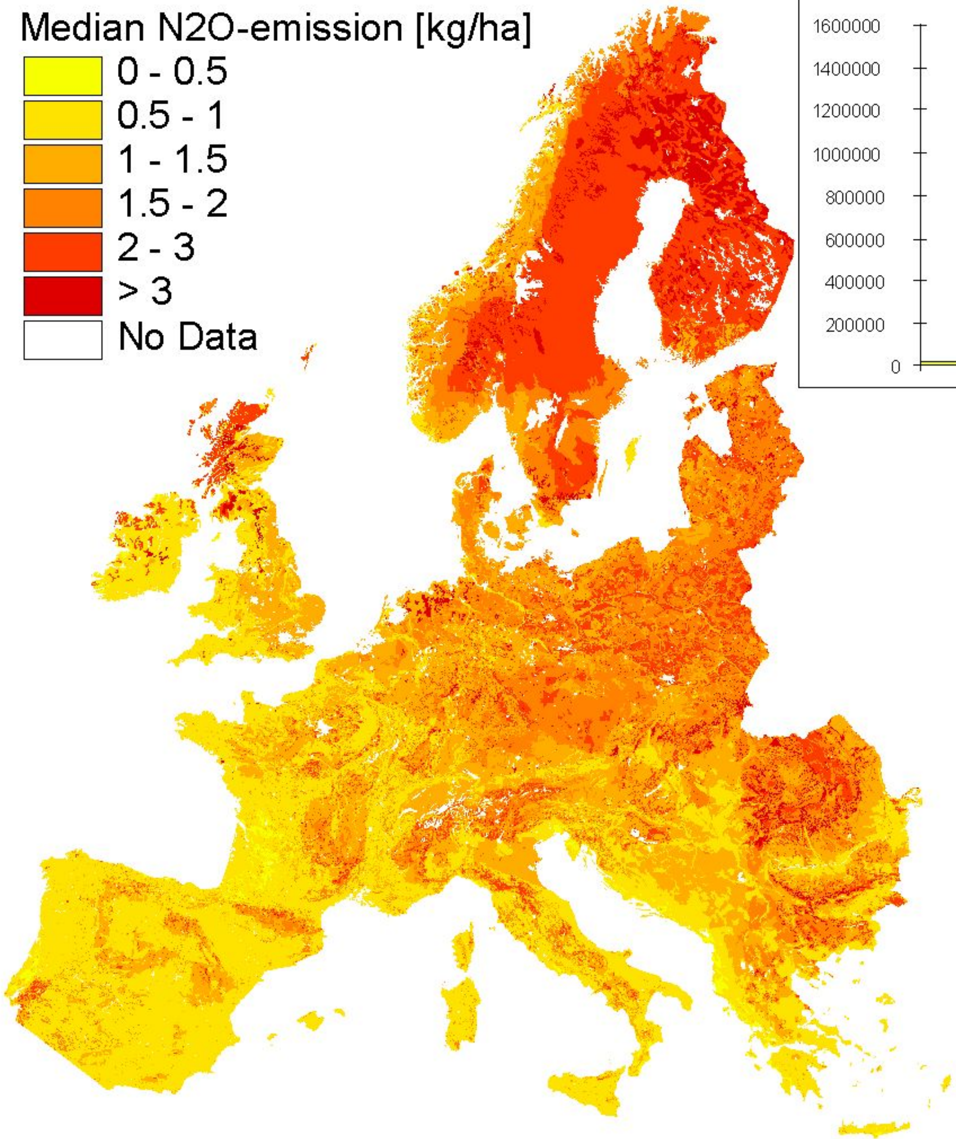


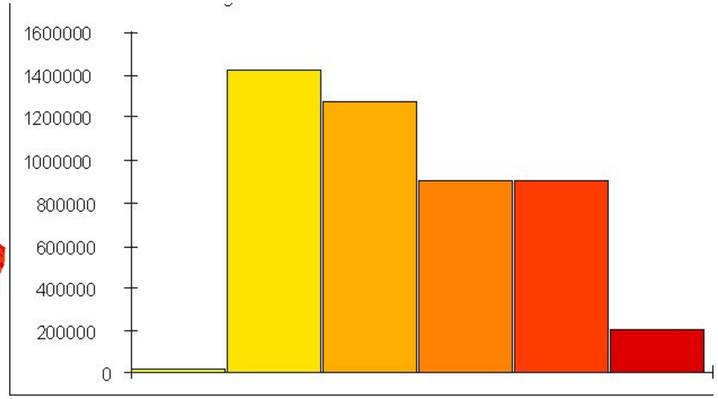


NitroEurope IP

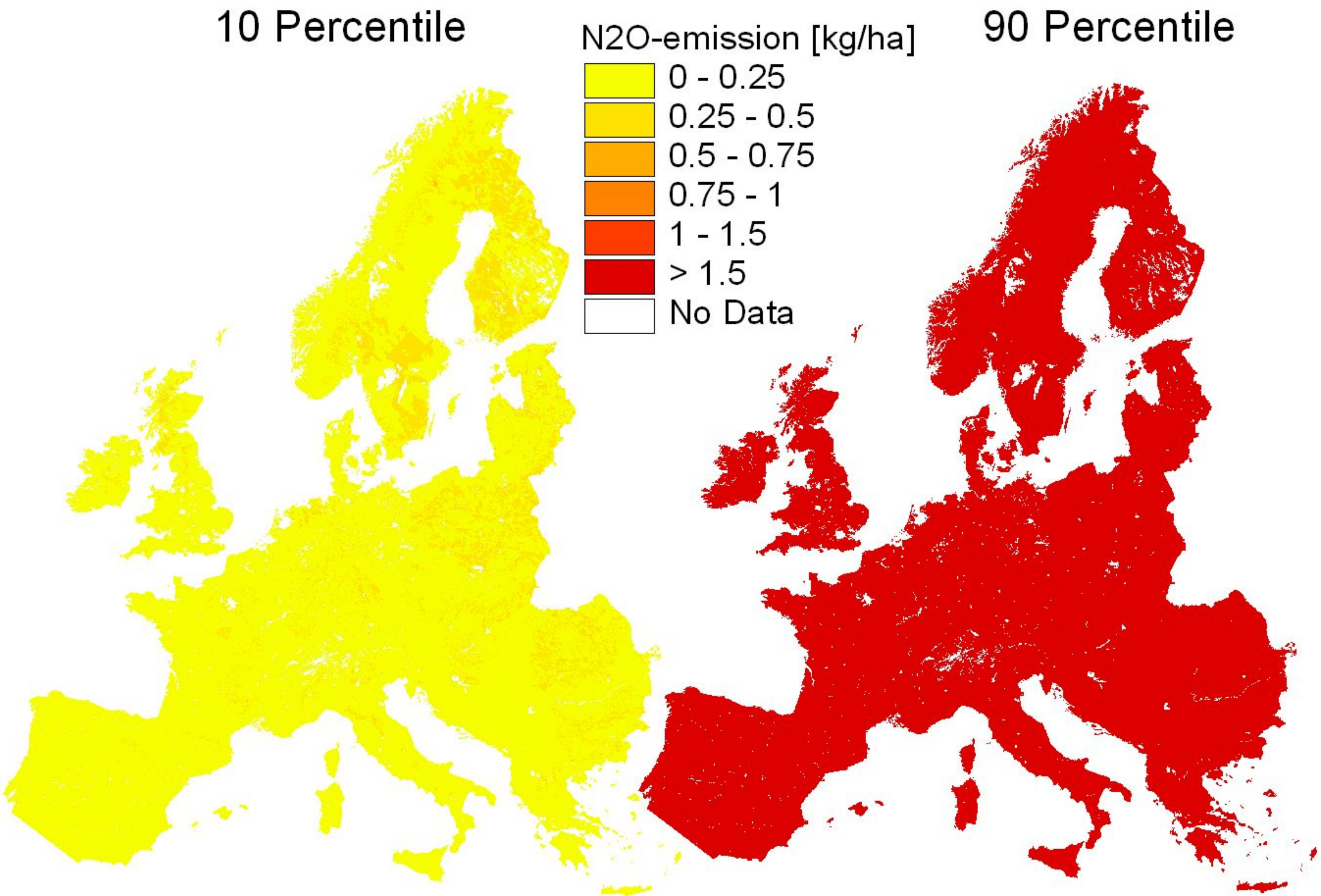
Regression kriging result: median N_2O emission

(natural areas only)





Large uncertainties (but note: point support!)



Conclusions

- Design-based upscaling attractive because it does not suffer from making wrong assumptions
- It is also suitable for validation because independence guaranteed (provided data are not used twice)
- However, measurements must be selected using probability sampling, this is rare in GHG emission research
- Time to critically evaluate measurement strategies?
- Model-based upscaling currently more suitable for GHG emission research, but model-building and data selection requires attention
- Do not expect good results with scarce and/or poor data!

Thank you