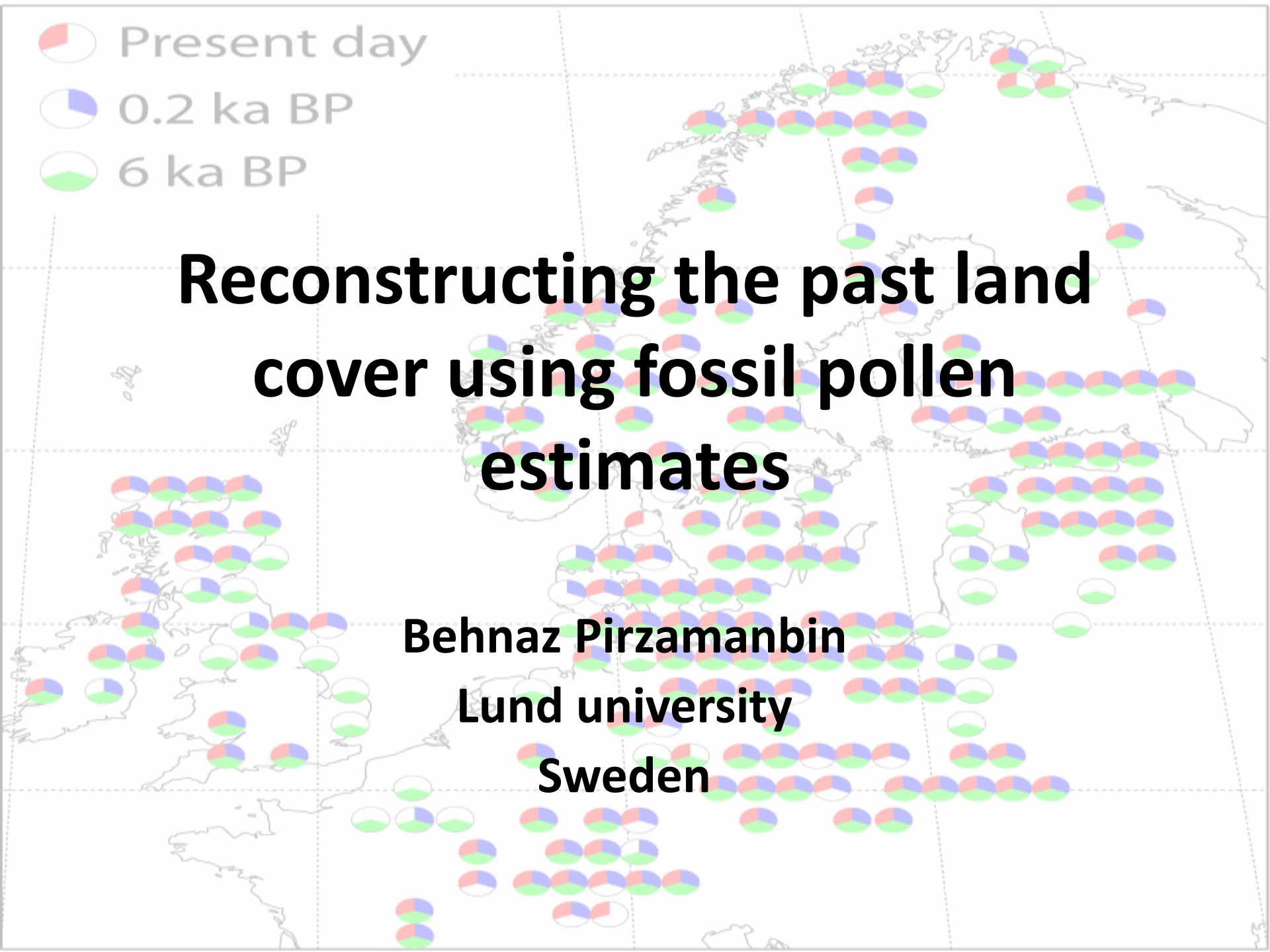


Reconstructing the past land cover using fossil pollen estimates

Behnaz Pirzamanbin
Lund university
Sweden

Collaborators:

Marie-Jose Gaillard
Department of Biology and Environmental Sciences, Linnaeus University, Sweden

Johan Lindström
Centre for Mathematical Sciences, Lund University, Sweden

Mathematical Statistics

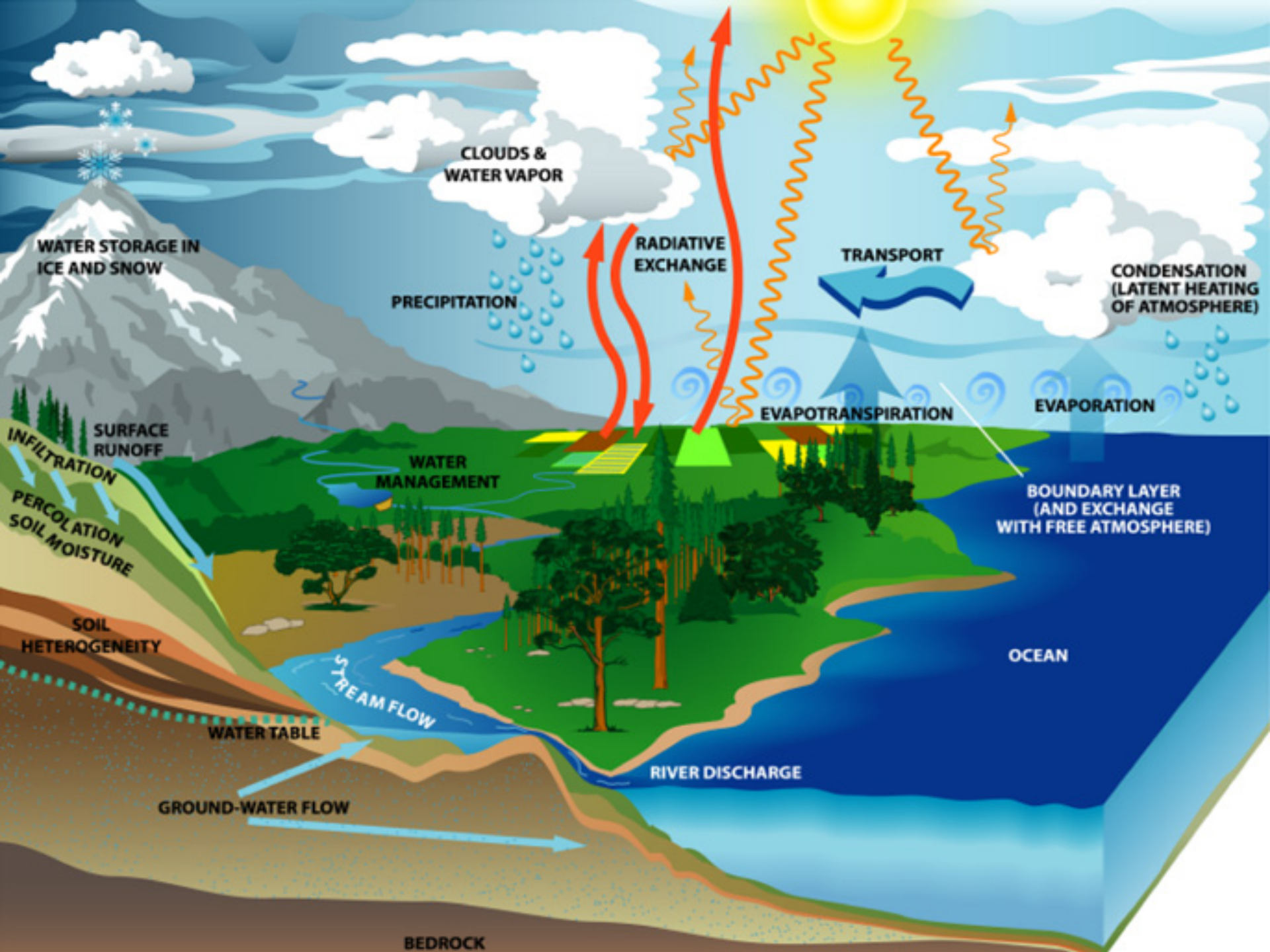
Benjamin Smith
Department of Physical Geography and Ecosystems Analysis, Lund University, Sweden

Shinya Sugita
Institute of Ecology, Tallinn University, Estonia

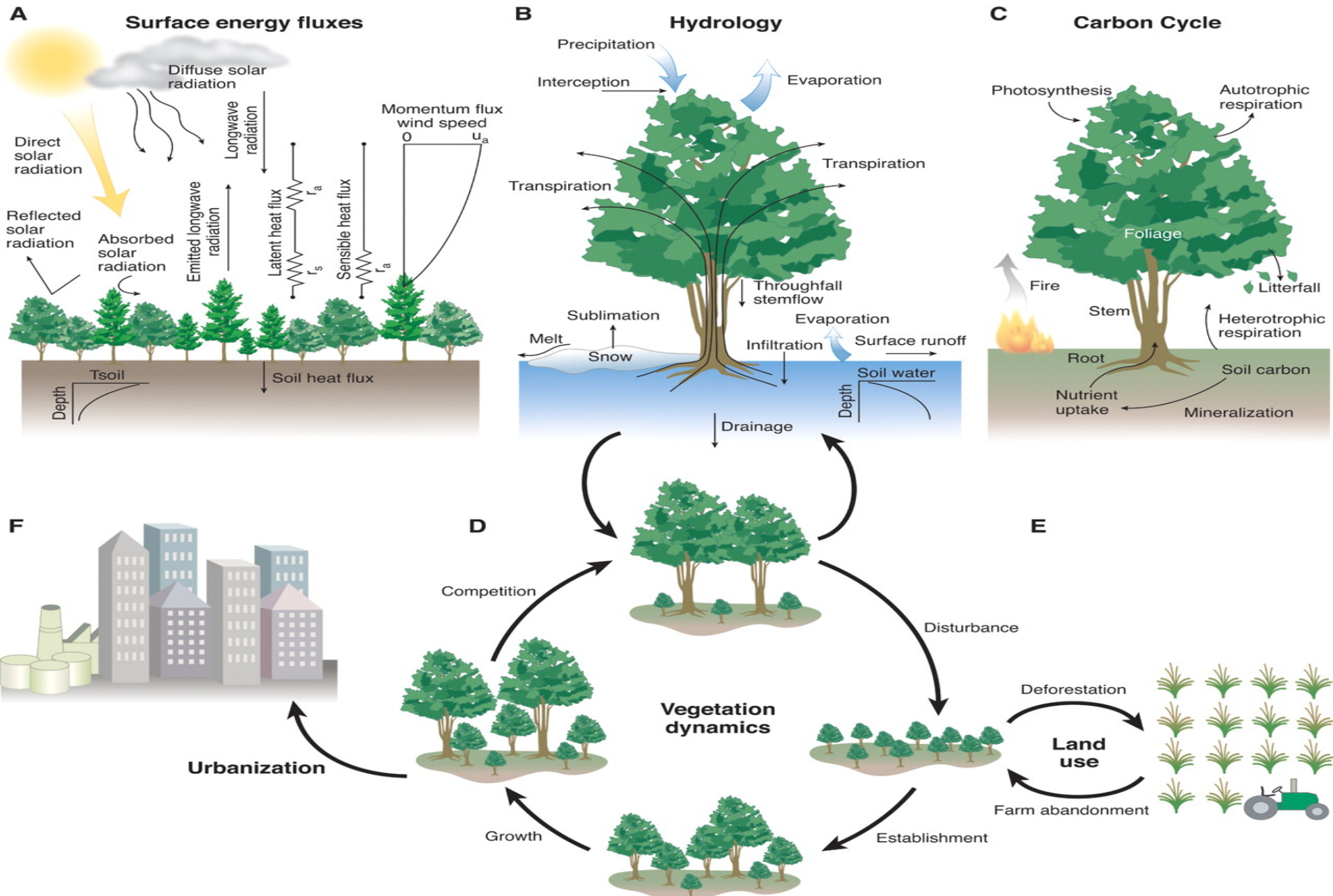
Palaeoecology

Anneli Poska
Department of Physical Geography and Ecosystems Analysis, Lund University, Sweden

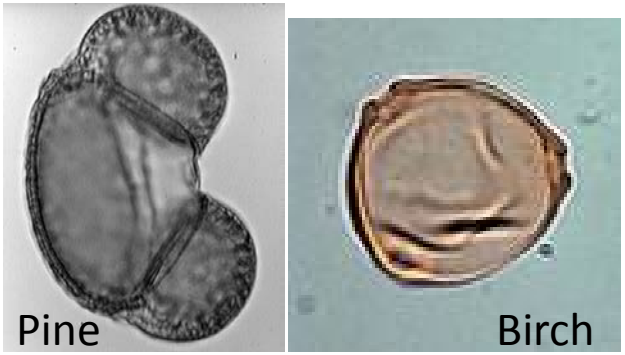
Climate modeling

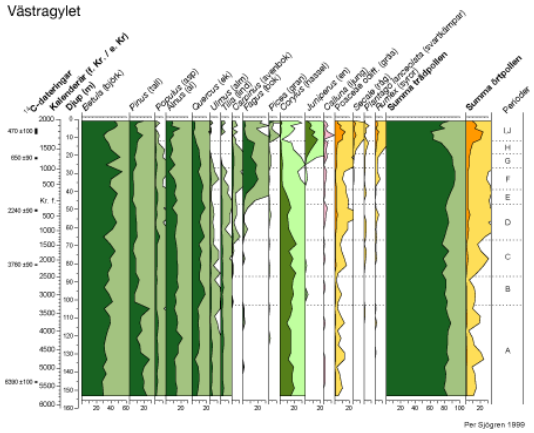

CLOUDS &
WATER VAPOR
WATER STORAGE IN
ICE AND SNOW
PRECIPITATION
RADIATIVE
EXCHANGE
TRANSPORT
CONDENSATION
(LATENT HEATING
OF ATMOSPHERE)
EVAPOTRANSPIRATION
EVAPORATION
SURFACE
RUNOFF
INFILTRATION
PERCOLATION
SOIL MOISTURE
WATER
MANAGEMENT
BOUNDARY LAYER
(AND EXCHANGE
WITH FREE ATMOSPHERE)
SOIL
HETEROGENEITY
STREAM FLOW
OCEAN
WATER TABLE
RIVER DISCHARGE
GROUND-WATER FLOW
BEDROCK

Vegetation (Land-cover) is a inherent part of climate system.



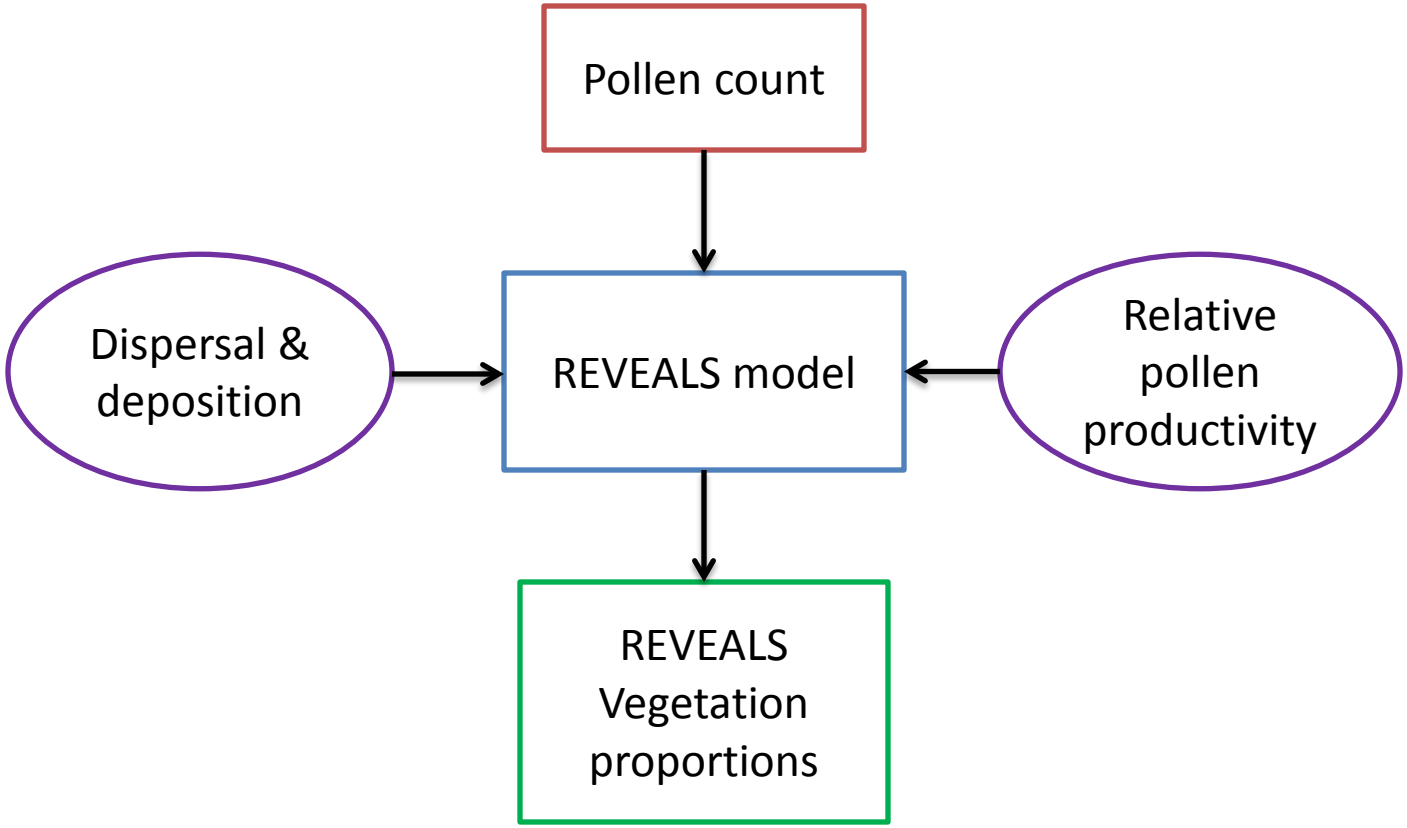
Fossil pollen has been extensively used to estimate past vegetation in regional and global scales.





REVEALS model

(Regional Estimates of VEgetation Abundance from Large Sites)
(Sugita S. 2007*)



*) Sugita S. 2007 .Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition. The Holocene 17 (2), 229-241.

Pollen data:

- 25 taxa
- 10 plant functional types (PFT)
- 3 groups, land-cover types (LCT)

PFT	Description	LCT
TBE1	Shade-tolerant-boreal (Picea)	Coniferous forest
TBE2	Shade-tolerant-temperate (Abies)	
IBE	Shade-intolerant-boreal (Pinus)	
TSE	Tall shrub evergreen trees(Juniperus)	
IBS	Shade-intolerant–boreal (Alnus)	Broadleaved forest
TBS	Shade-tolerant-temperate (Carpinus)	
TSD	Tall shrub summergreen trees(Salix)	
LSE	Low evergreen shrub (Calluna)	Unforested land
GL	Grassland – all herbs (Filipendula)	
AL	Agricultural land- cereals (Cereals)	

behnaz@maths.lth.se

Goals

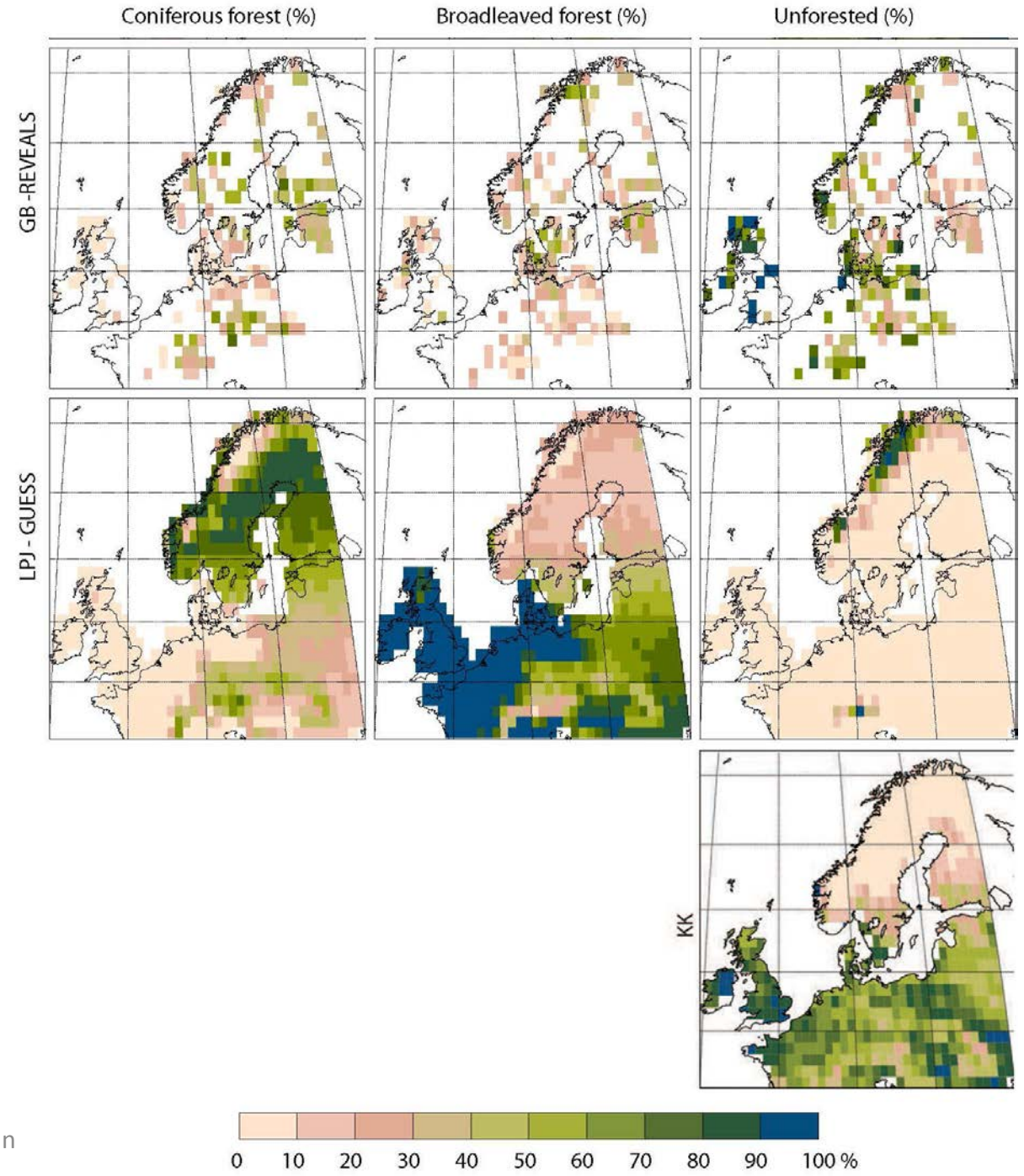
- To construct a model that explains Reveals data for present day time window,
  - **Present day (1850- 1950),**

- To reconstruct land cover at locations with no Reveals estimates for past time windows,
  - **200 BP (1600-1850 AD), and**
  - 600 BP
  - 3000 BP
  - **6000 BP (4250-3750 BC),**

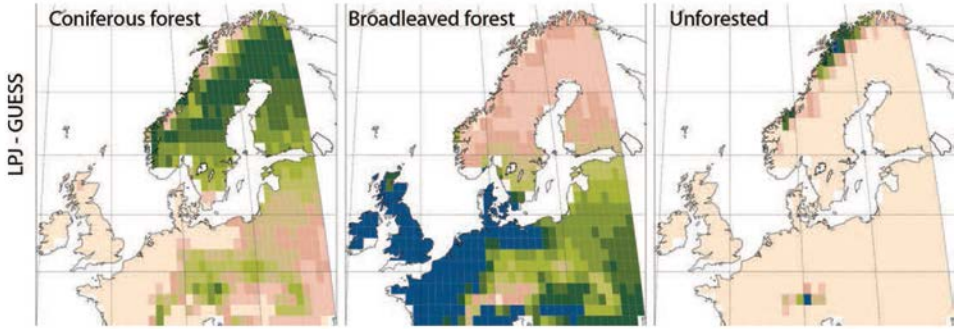
Data

- REVEALS land cover proportions,
  - Pollen based,
  - 24 taxa, 10 plant functional types (PFTs), and 3 land cover types (LCTs),
  - Present day, 200 years before present (BP), 6000 BP.
- LPJ-GUESS potential land cover proportions,
  - Vegetation model force by bio-climate variables
- KK10 estimates of human land use proportions.

Data:


Coniferous forest (%)
Broadleaved forest (%)
Unforested (%)
GB -REVEALS
LPJ - GUESS
KK
0
10
20
30
40
50
60
70
80
90
100 %

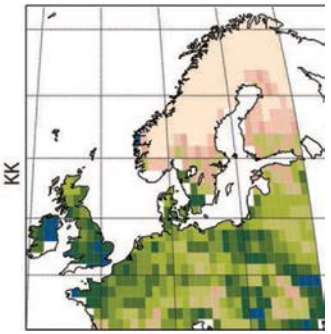
LPJ-GUESS adjusted with KK10

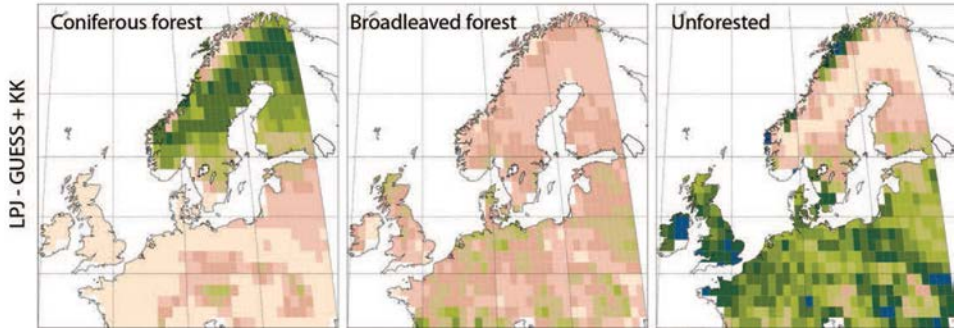


$$P(C_{adj}) = P(C) * (1 - P(HLU))$$

$$P(B_{adj}) = P(B) * (1 - P(HLU))$$

$$P(U_{adj}) = P(U) * (1 - P(HLU)) + P(HLU)$$





Compositional data:

For Land cover data set, $\{y_i \ , \ i = 1, \dots, D = 3\}$ with

$$\sum_i y_i(s) = 1, \qquad y_i \quad \in [0,1].$$

If we model y_i using linear regression

$$y_i = X\beta_i + \varepsilon_i, \qquad \hat{y}_i = X\hat{\beta}_i$$

but

$$\hat{y}_i \notin [0,1].$$

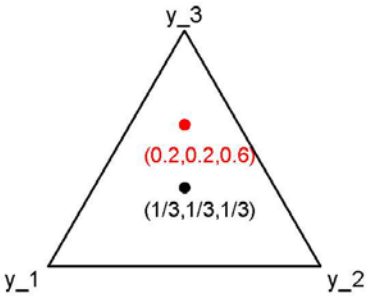
We use the log transform for compositional data, $\mathrm{j} = 1, 2$

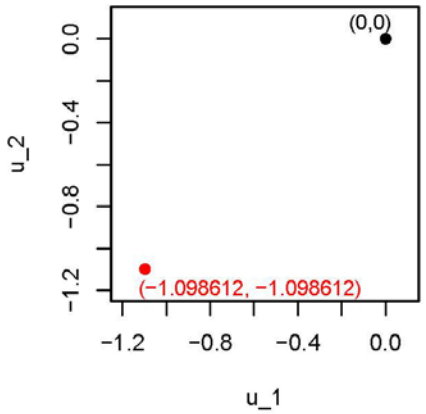
$$u_j(s) = log\frac{y_j(s)}{y_D(s)}, \qquad u_j \in (-\infty, +\infty).$$

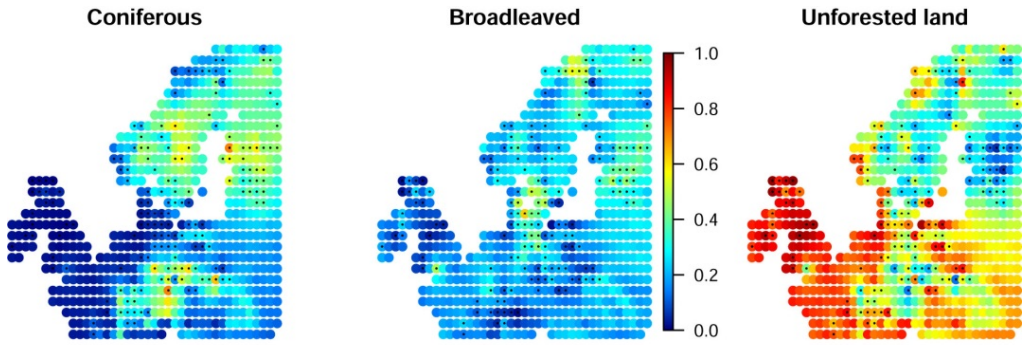
Back transformation of the data,

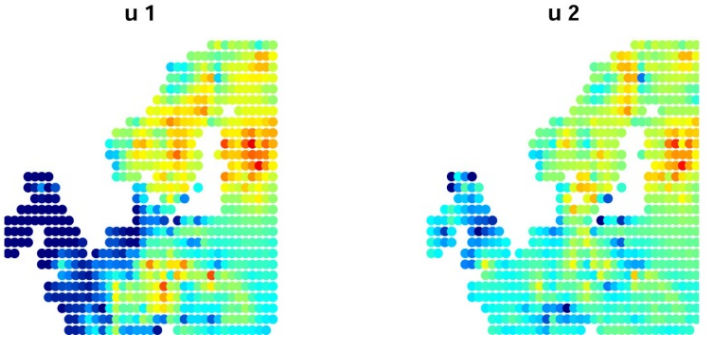
$$y_j(s) = \frac{\exp(u_j(s))}{1 + \sum_j \exp(u_j(s))}, \qquad y_D(s) = \frac{1}{1 + \sum_j \exp(u_j(s))},$$

Transfromation of compositional data:









Model

Land cover types $\{y_i\ ,\ i = 1, \dots, D = 3\}$, $u_j(s) = log\frac{y_i(s)}{y_D(s)}$, $j = 1, 2,$

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = A\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \varepsilon_i \qquad where \quad \varepsilon_i \in N(0, I\sigma^2)$$

- μ is mean field and modeled as a linear regression

$$\mu_i = 1.\beta_{0,i} + \sum_p B_p\beta_{p,i}$$

- B is covariate matrix

$$\mathrm{B} = \begin{bmatrix} \text{Coniferous} \\ \text{Broadleaved} \\ \text{Elevation} \end{bmatrix}^{\mathrm{T}},$$

- Z is spatial dependency field and modeled using SPDE/GMRF

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \in N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \otimes Q^{-1}\right).$$

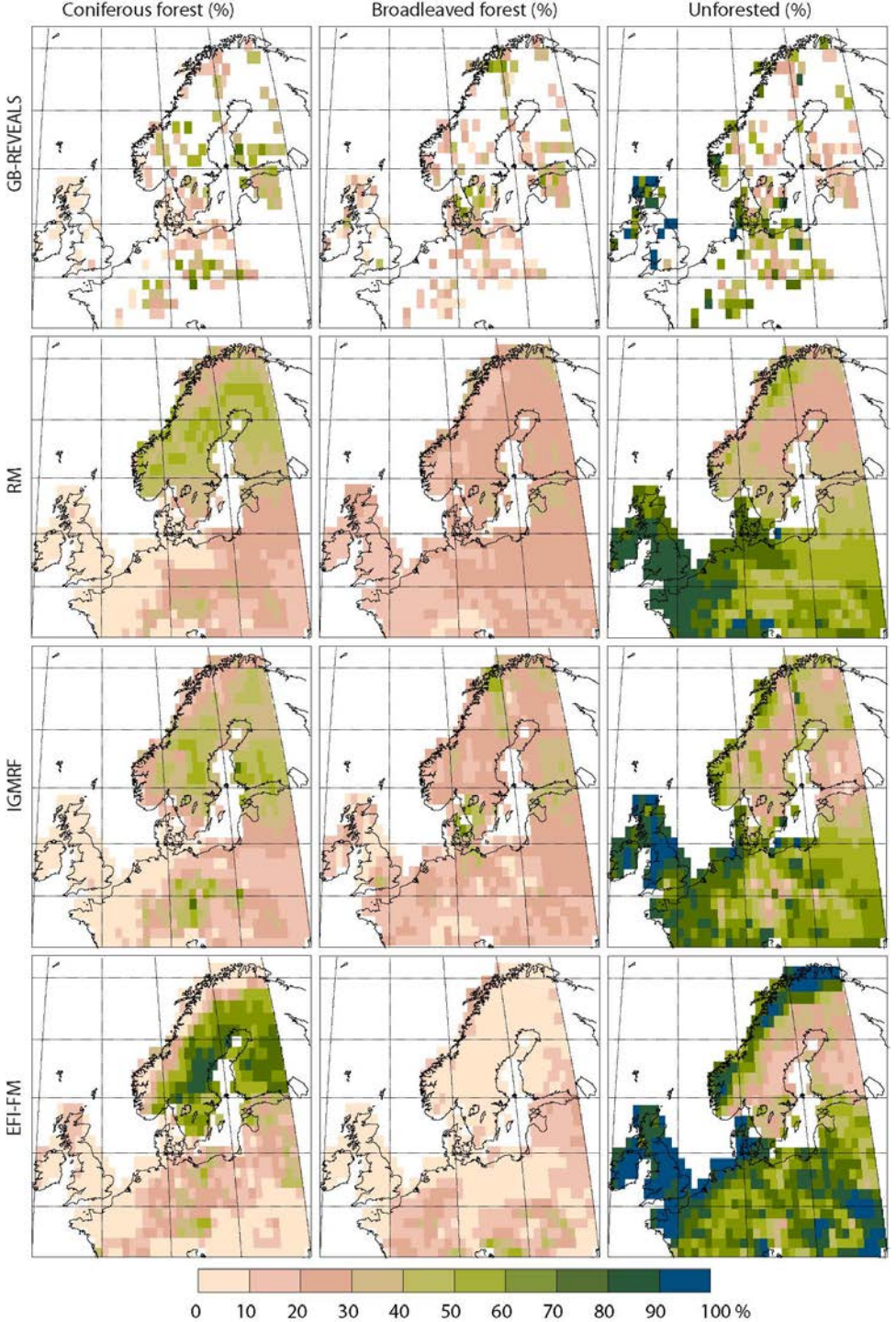
Validation of model:

To check the model performance

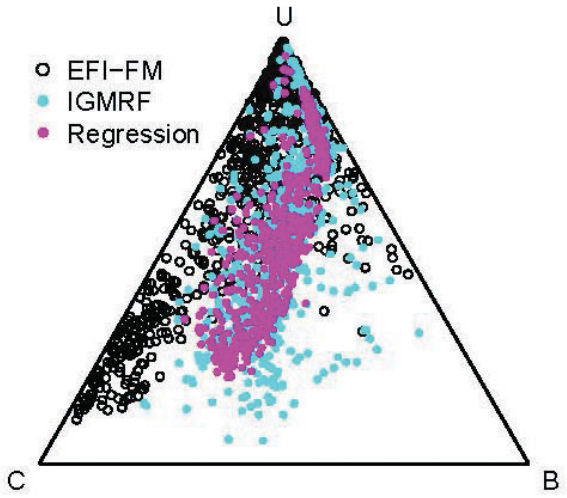
- We evaluate the model with present-day European Forest Institue Forest Map (EFI-FM) data
  - 2006 vs. 1950

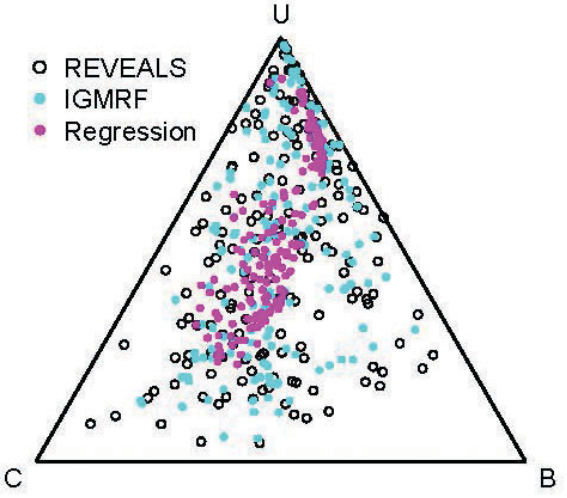
- We use 6-block cross-validation

Present-day

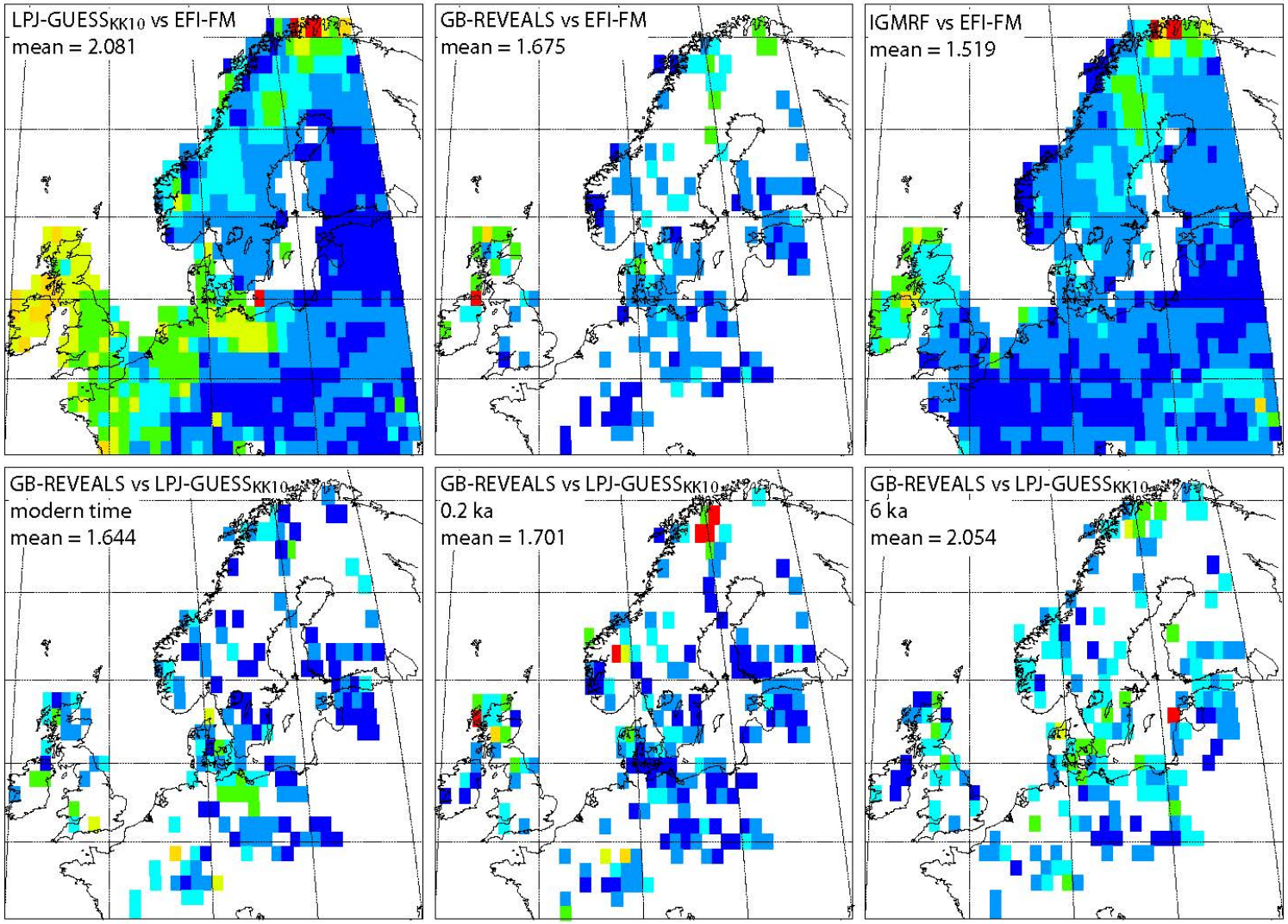

Coniferous forest (%)
Broadleaved forest (%)
Unforested (%)
GB-REVEALS
RM
IGMRF
EFI-FM
0
10
20
30
40
50
60
70
80
90
100 %

EFI-FM & REVEALS vs. IGMRF & RM

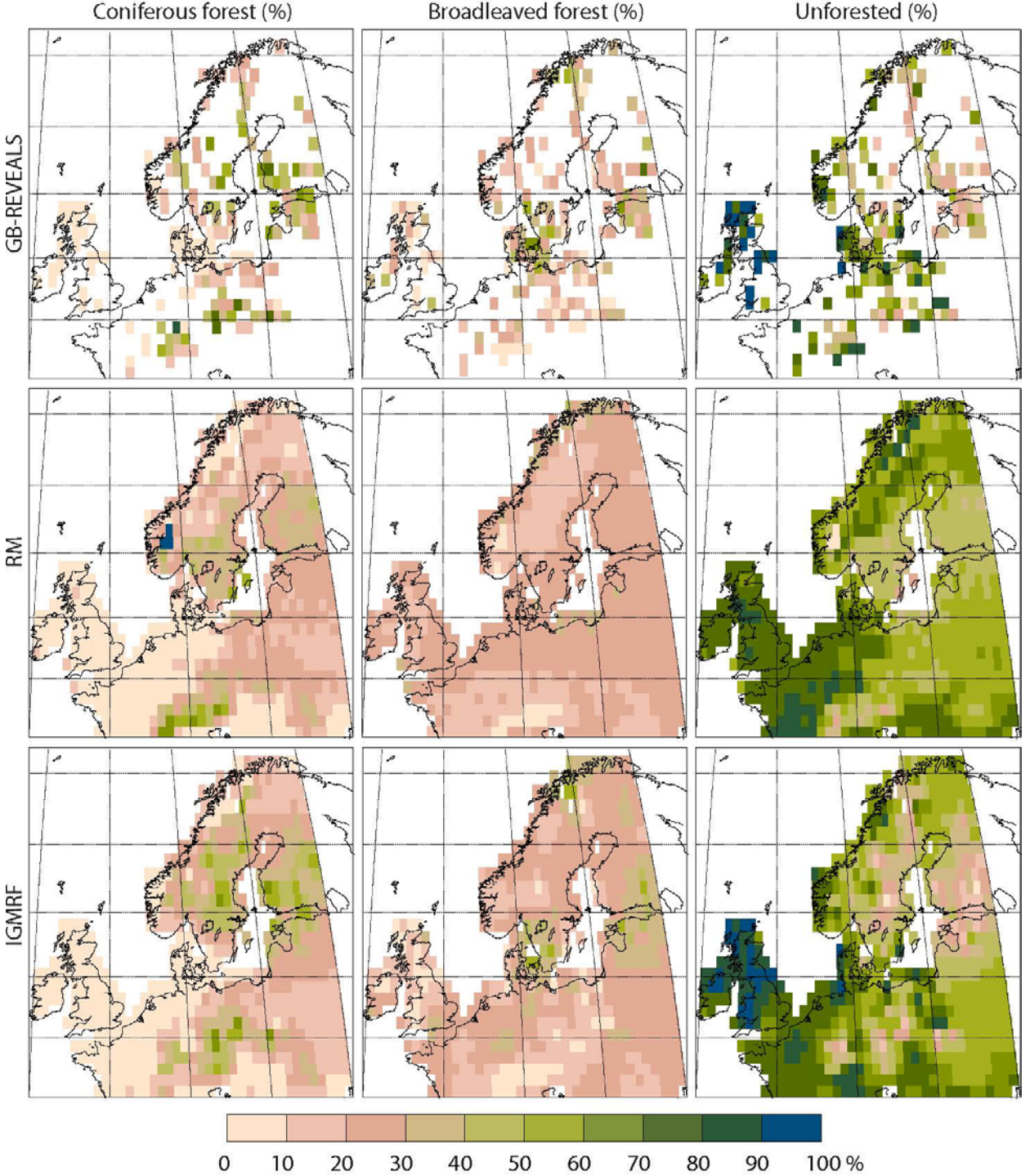




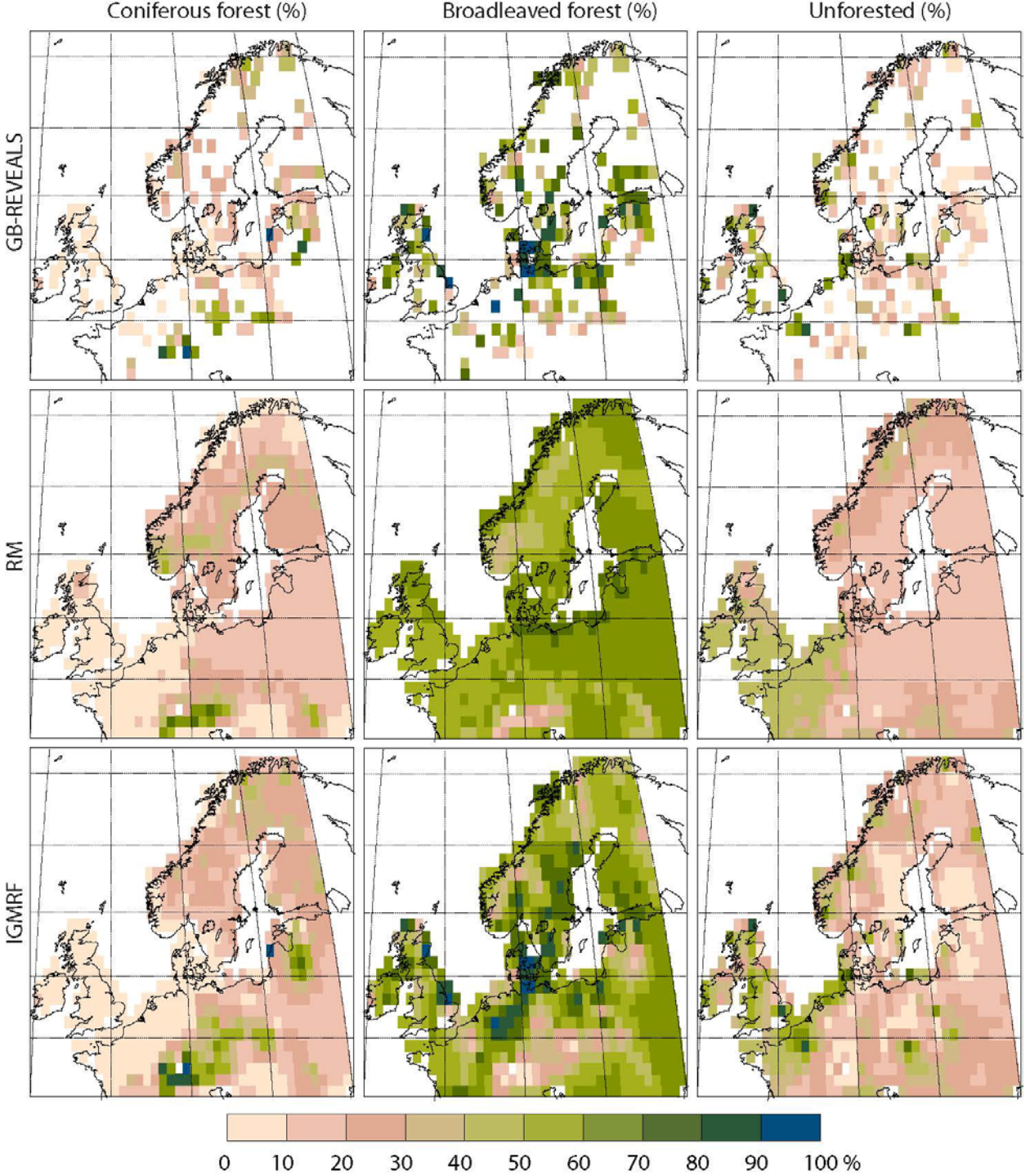
Comparison


LPJ-GUESS$_{KK10}$ vs EFI-FM
mean = 2.081
GB-REVEALS vs EFI-FM
mean = 1.675
IGMRF vs EFI-FM
mean = 1.519
GB-REVEALS vs LPJ-GUESS$_{KK10}$
modern time
mean = 1.644
GB-REVEALS vs LPJ-GUESS$_{KK10}$
0.2 ka
mean = 1.701
GB-REVEALS vs LPJ-GUESS$_{KK10}$
6 ka
mean = 2.054
0 1 2 3 4 5 6 7

200 BP


Coniferous forest (%)
Broadleaved forest (%)
Unforested (%)
GB-REVEALS
RM
IGMRF
0
10
20
30
40
50
60
70
80
90
100 %

6000 BP


Coniferous forest (%)
Broadleaved forest (%)
Unforested (%)
GB-REVEALS
RM
IGMRF
0
10
20
30
40
50
60
70
80
90
100 %

Results:

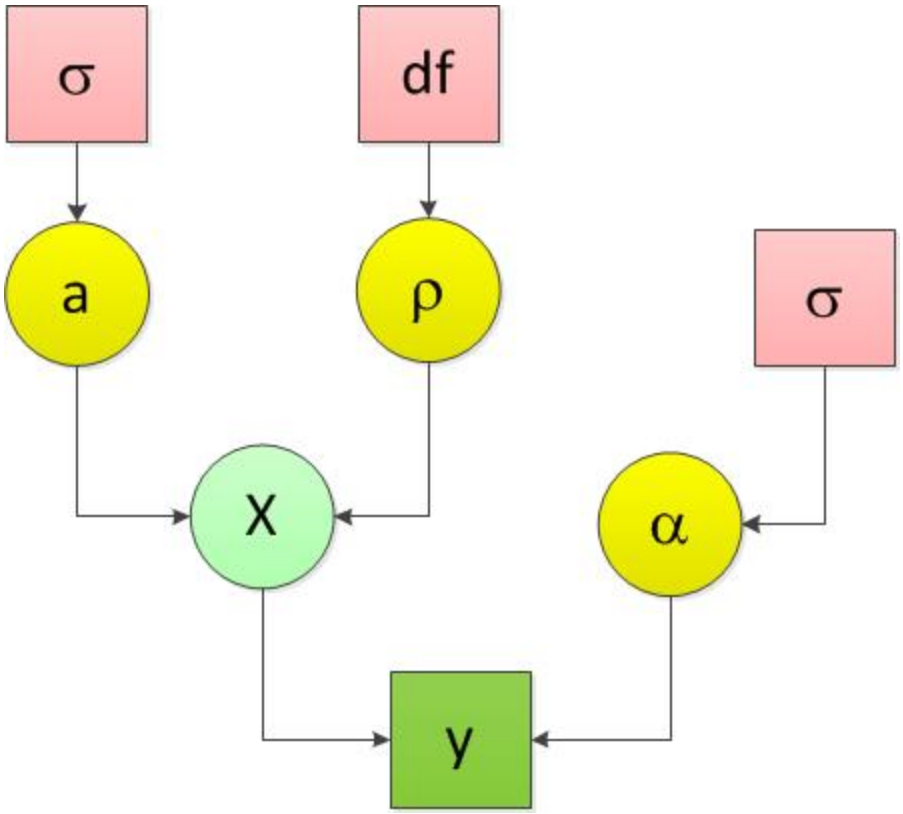
- In **general** the reconstructed maps from both models are close to EFI-FM,
- **IGMRF** reconstructed maps capture the local variability of REVEALS data,
- **RM** reconstructed maps smooth REVEALS data.

New study's goal

- Estimating the human land use,
- Investigating what LCTs is most likely to be used by humans.

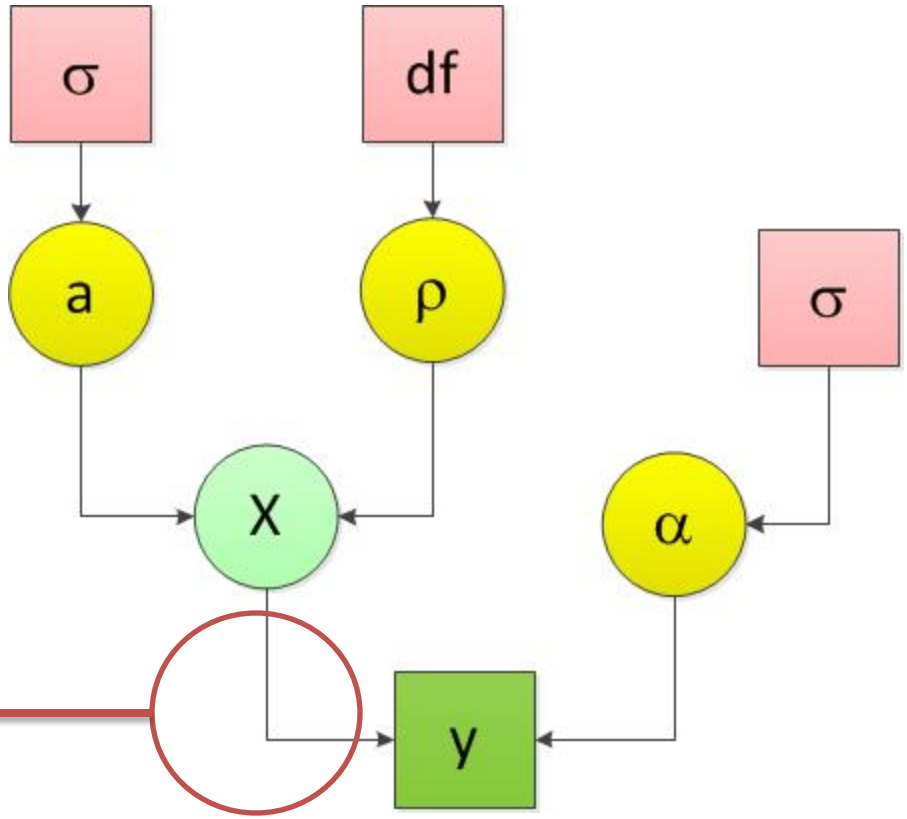
Hierarchical model:

- Data model:
  $$\mathrm{P}(y|\alpha, z(\mathrm{X}))\sim \mathrm{Dir}(\alpha.z(X))$$
- Latent field:
  $$\mathrm{P}(\mathrm{X}|\theta)\sim N(0, \rho \otimes Q^{-1}(a))$$
- Unknown parameters: α, ρ, a



Hierarchical model:

- Data model:
  $$\mathrm{P}(y|\alpha, z(\mathrm{X})) \sim \mathrm{Dir}(\alpha.z(X))$$
- Latent field:
  $$\mathrm{P}(\mathrm{X}|\theta) \sim N(0, \rho \otimes Q^{-1}(a))$$
- Unknown parameters: α, ρ, a



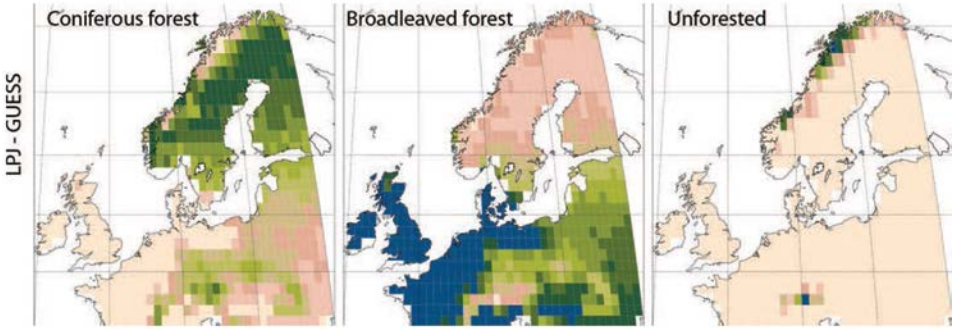
$$\begin{pmatrix} C \\ B \\ U \end{pmatrix} \sim alr(\mathrm{X}_1, \mathrm{X}_2)$$

$$HLU \sim logit(\mathrm{X}_3)$$

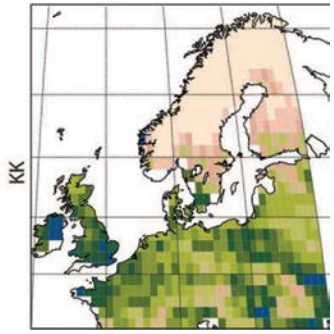
$$\begin{pmatrix} C_{adj} \\ B_{adj} \\ U_{adj} \end{pmatrix} \sim \begin{pmatrix} \mathrm{C}*(1-\mathrm{HLU}) \\ \mathrm{B}*(1-\mathrm{HLU}) \\ \mathrm{U}*(1-\mathrm{HLU})+\mathrm{HLU} \end{pmatrix}$$

Latent field: $\mathrm{P}(\mathrm{X}|\theta) \sim N(0, \rho \otimes Q^{-1}(a))$

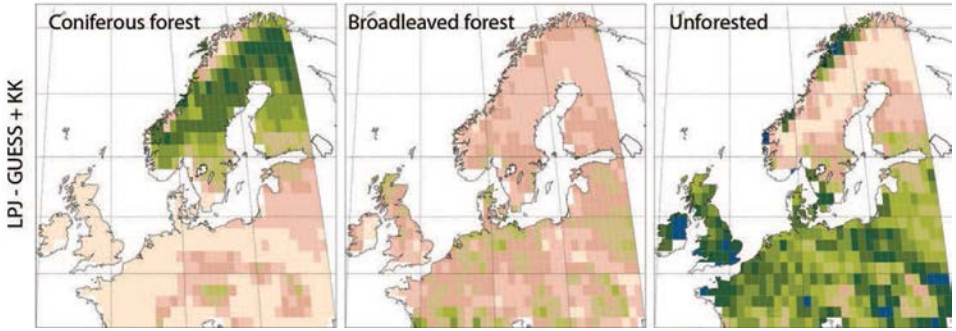
Before, we considered the transformed data directly in the models,



$$\mathrm{P}(C_{adj}) = \mathrm{P}(\mathrm{C}) * (1 - \mathrm{P}(\mathrm{HLU}))$$
$$\mathrm{P}(B_{adj}) = \mathrm{P}(\mathrm{B}) * (1 - \mathrm{P}(\mathrm{HLU}))$$
$$\mathrm{P}(U_{adj}) = \mathrm{P}(\mathrm{U}) * (1 - \mathrm{P}(\mathrm{HLU})) + \mathrm{P}(\mathrm{HLU})$$



Now, we want to model the components on the right side of the above equations,



MCMC

To fit the model and estimate the parameters we use

1. MH: $\mathrm{P}(\alpha|y, z(\mathbf{X}))$
2. MH: $\mathrm{P}(a, \rho|\mathbf{X})$
3. MH, using "Taylor expansion" or "Riemann manifold Langevin" : $\mathrm{P}(\mathbf{X}|\alpha, a, \rho, y)$

Future plan

- To reconstruct the LCTs for specific regions with higher temporal resolution, for example Baltic area, and consider the temporal dependency between the time windows.

- To reconstruct the 10 PFTs as specified in climate models, since there exist locations where some of these PFTs are nonexistence (zero proportion). This is an interesting open statistical problem!

Thanks ☺

Did you know that Badminton is

- **the world fastest racket sport &**
- **the second largest participation sport in the world (after football)**