Spatial Modeling of forest (site) properties – some pratical experiences with GAM

Karl Mellert

Bayerische Landesanstalt
für Wald und Forstwirtschaft

BAYERISCHE FORSTVERWALTUNG

ZENTRUM WALD FORST HOLZ
WEIHENSTEPHAN

Three Examples with GAM (mgcv)

- **1. Site quality of forests in the Bavarian Alps**
  - Background: biomass harvesting

- **2. Site-specific tree species selection**
  - Species distribution modeling (SDM)
  - Background: Climate change

- **3. Vulnerability of forests against storm damage**
  - Background: Increased storm frequency / Climate change

Spatial modeling of forest site quality in the Bavarian Alps based on indicator values and environmental predictors

Karl Mellert & Jörg Ewald

University of Applied Sciences Weihenstephan-Triesdorf, Freising, Germany

WINALP
Waldinformationssystem Nordalpen

Europäische Territoriale Zusammenarbeit | Grenzübergreifendes Förderprogramm
INTERREG Bayern – Österreich 2007-2013
Bayern
Oberösterreich
Salzburg
Tirol
Vorarlberg
ZENTRUM WALD FORST HOLZ WEIHENSTEPHAN
HOCHSCHULE WEIHENSTEPHAN-TRIESDORF UNIVERSITY OF APPLIED SCIENCES
Technische Universität München
LWF Bayerische Landesanstalt für Wald und Forstwirtschaft
Bayerische Forschungsallianz
BAYERISCHE STAATSFORSTEN Nachhaltig Wirtschaften.
BAYERISCHE FORSTVERWALTUNG
Wald Land Salzburg
tirol Unser Land

Aim sub-project

Geo-information about nutrient availability
→ decision support to forest management

Background

→ biomass harvesting = removal of woody debris from forests

Whole tree harvesting of Norway spruce in Bavaria increases exports:
→ base cations (Ca, Mg, K) 2-fold → N, P 3.5-fold

(Ref.: Weis & Göttlein 2011)

Effects are site specific
→ Spatial information about site trophy are necessary

Approach

Site quality (SQ) assessment by plant response:

→ **Mean Ellenberg indicator values**

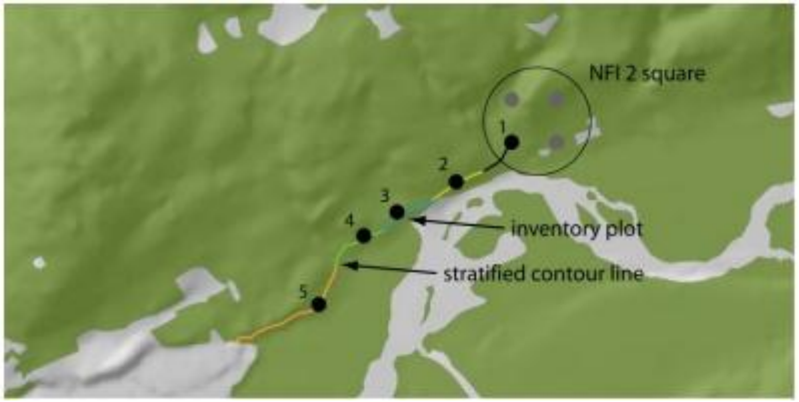
scale: optimum of a plant species along environmental gradients ranging from 1 – 9

Nutrient availability	Reaction (**mR**) = Ca Mg (K) Nutrients (**mN**) = N P K
Climate	Temperature (**mT**) = Ca Mg (K) Moisture (**mM**) = N P K

WINALPecobase ➔ >1,500 soil+relief+vegetation

Basis: Second National Forest Inventory (NFI 2)
regular 4 km x 4 km grid

Even sampling of all forest types along a contour line
maximal spread of sample plots along local + regional environmental gradients
within an inventory region



WINALPecobase → >1,500 soil+relief+vegetation

Data	Abbreviation	Description [SI unit]	Source / Reference
WINALP ecobase	mN	average indicator value for nutrients, weighted by log cover	Ellenberg et al. (2001)
	Altitude	Elevation above sea level [m]	
	TAspect	Transformed slope aspect (folded around thermal optimum)	Beers et al. (1966)
	Slope	Slope [°]	
	Sgrp	Site group	Kölling et al. (1996), Mellert & Ewald (2011)
	ORatio	Thickness of organic layer / thickness of humic topsoil	
	SoilD	Soil depth [cm]	
	Gravel	Gravel content of soil profile %	
	Clay	Clay content of soil profile %	
	AWC	Available water capacity [mm/m³]	
	DecD	Depth of decalcification [cm]	Ewald (1997)
	CPB	Chemical properties of bedrock	Kolb (2011)
GIS	Soil variables (see ecobase)		Beck et al. (2009)
	CPB	Chemical properties of bedrock	Kolb (2011)
Forest inventory	SI	Height of dominant *P. abies* individuals at age 95-105 years [dm]	

Approach

Specific aim of the subproject

Spatial prediction of mN based on abiotic environmental predictors

Model: mN = f(abiotic environment)

I. Internal validation

Multiple data splitting

II. Validation with independent forest inventory data

a) Can regionalised indicator values predict site index (SI) of Norway spruce?
b) Does regionalised mN contribute significant to the that prediction?

Model: SI = f(indicator values)

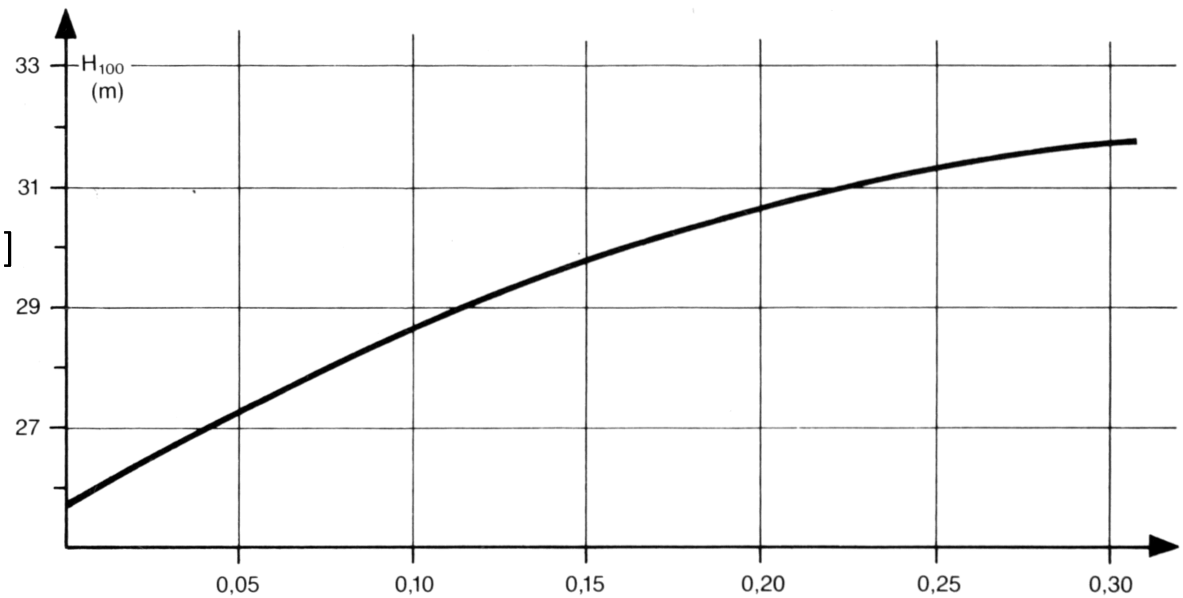
Approach

Plant response as an indicator for site trophy / productivity

Site index (SI) = Height of a tree species at reference age

Norway spruce

Site index [m]



Ca content topsoil [%]

Ref.: Nebe & Hofmann 1985

II. Validation with independent forest inventory data

SI = f(mN, mR, mM, mT)

LWF

Spatial modelling: Spatial autocorrelation of residuals?

Consequences of spatial autocorrelation (SAC)

1. Type I errors may be inflated → significance levels too optimistic
2. Model selection and parameter estimation may be biased

Techniques to integrate “space”

Random effect in a mixed model → Gamm

```
mN ~ s(x1) … +s(xi)
random=list(Group=~1)
correlation=corExp(form= ~ East + North)
```

Spatial effect in Gam → GamSE

```
mN ~ s(East,North) + s(x1)… +s(xi)
```

Methodological Questions

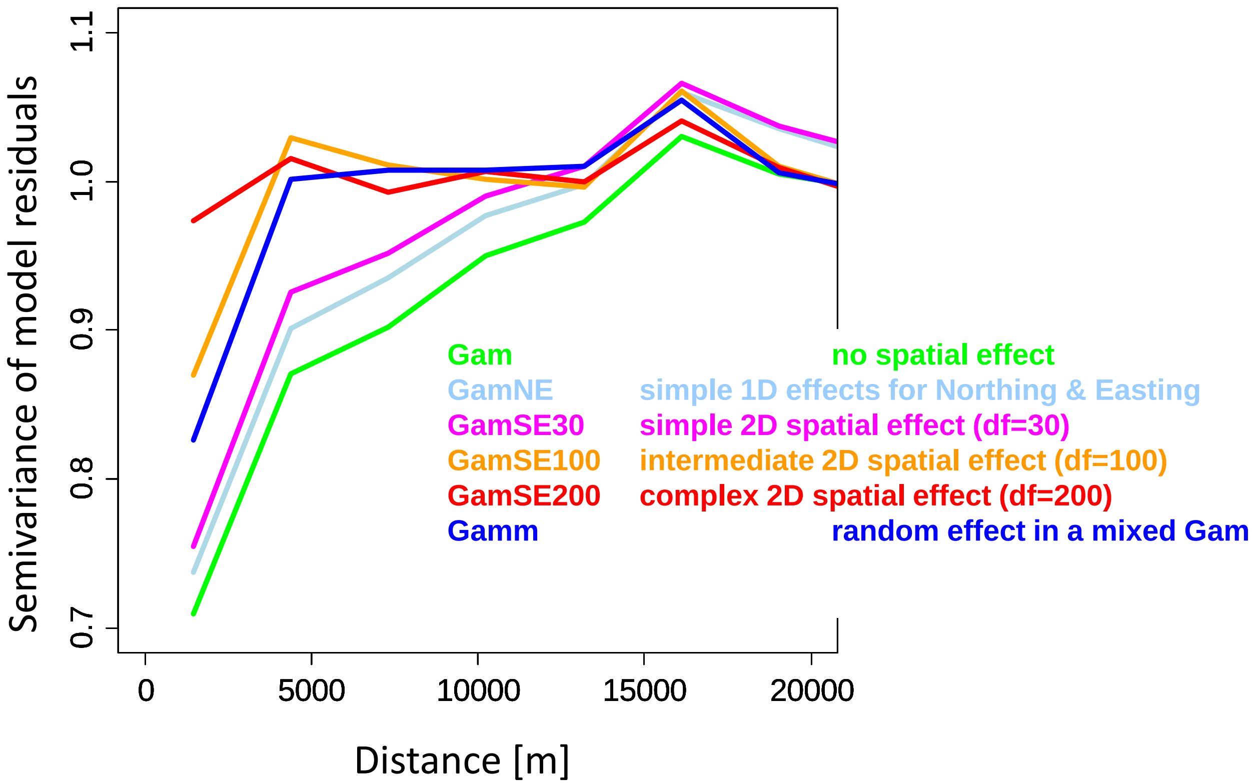
Technical questions

1. Can spatial autocorrelation of model residuals (SAC) be removed?
2. Are there severe influences of spatial autocorrelation on significance levels and parameter estimations of models?

GAM variants

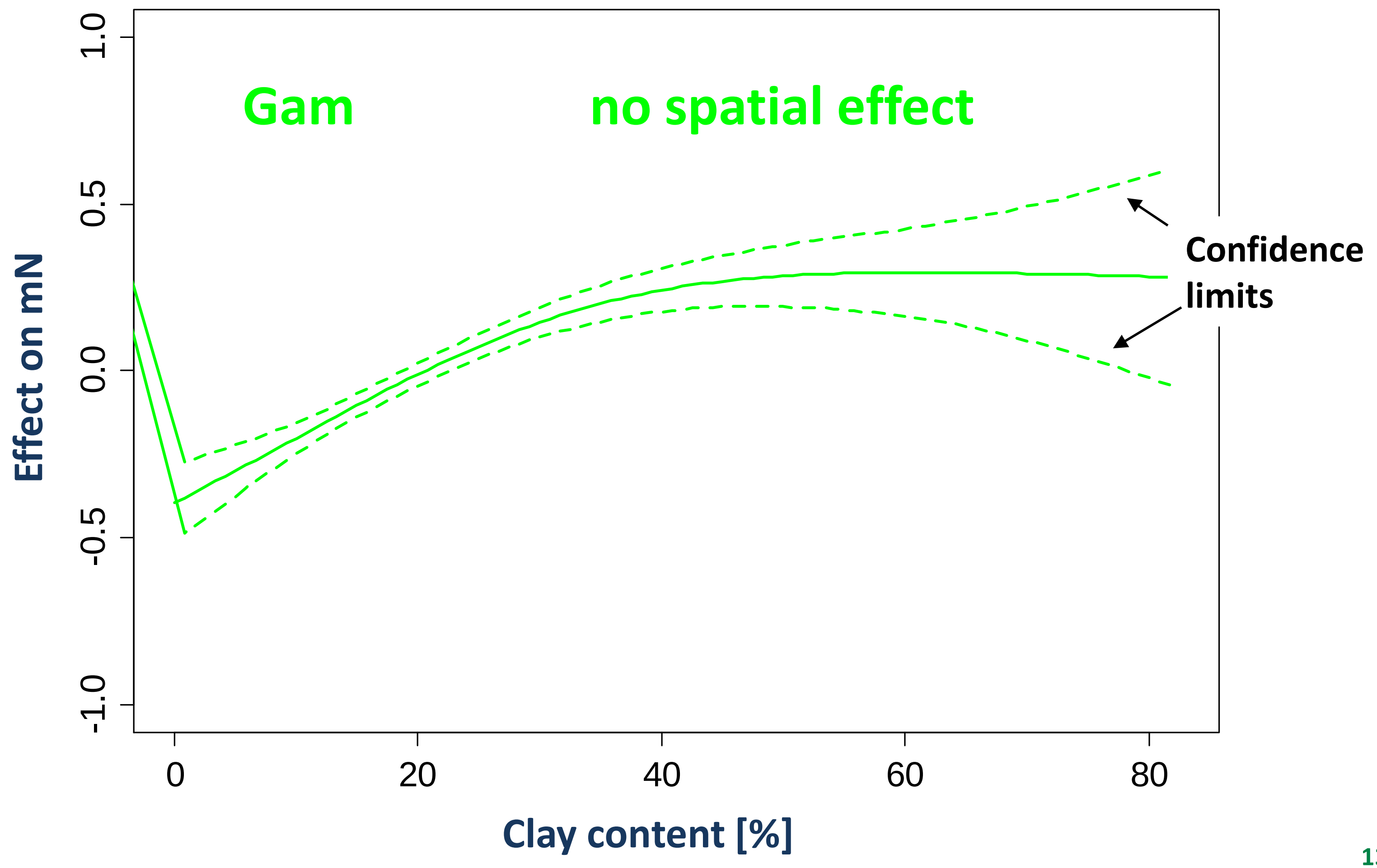
Gam	**no spatial effect**
GamNE	**simple 1D effects for Northing & Easting**
GamSE30	**simple 2D spatial effect (df=30)**
GamSE100	**intermediate 2D spatial effect (df=100)**
GamSE200	**complex 2D spatial effect (df=200)**
Gamm	**random effect in a mixed Gam**

1. Can spatial autocorrelation of model residuals (SAC) be removed?

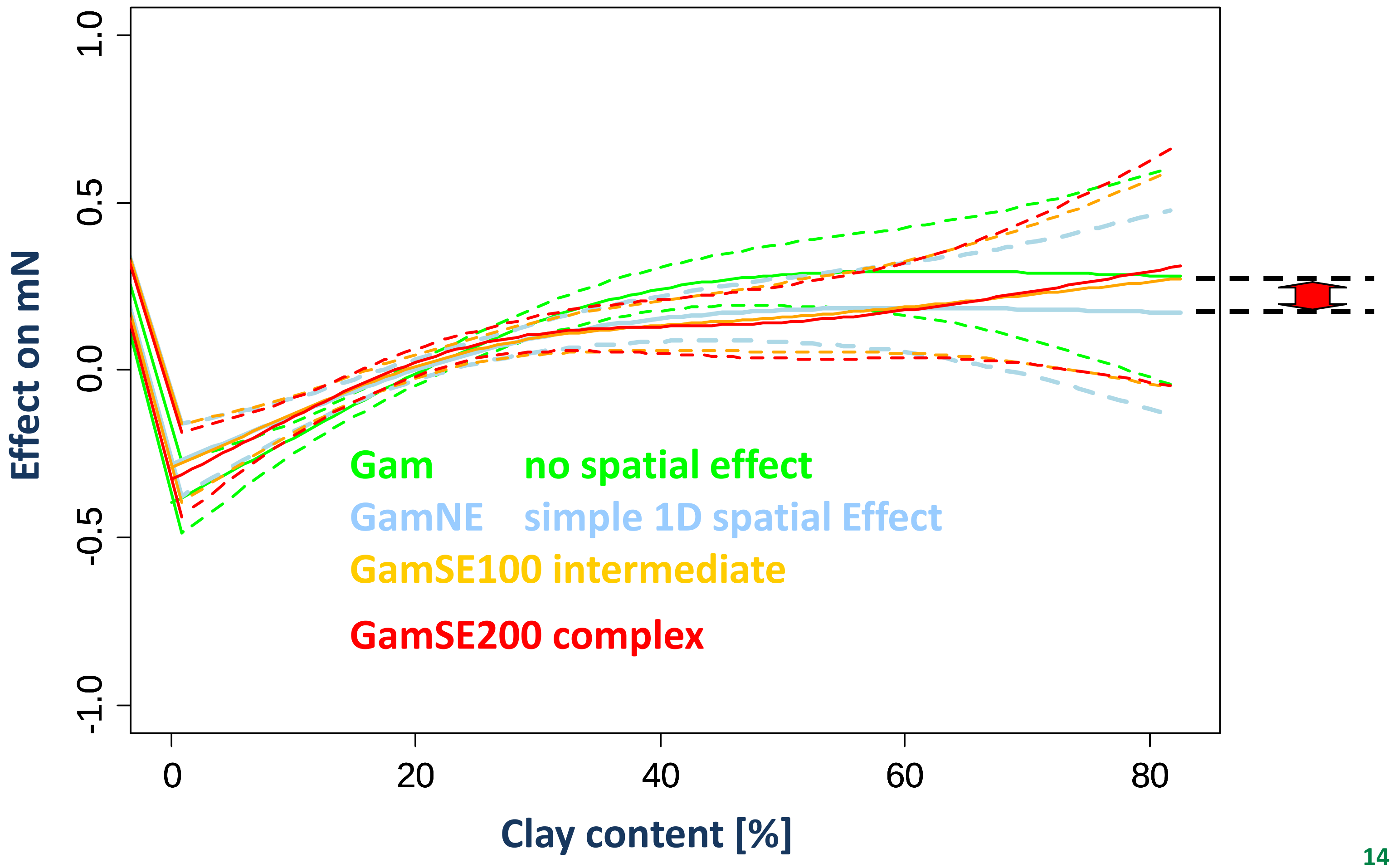


WINALP
Waldinformationssystem Nordalpen

2. Are there severe influences of spatial autocorrelation on significance levels and parameter estimations of models?



2. Are there severe influences of spatial autocorrelation on significance levels and parameter estimations of models?



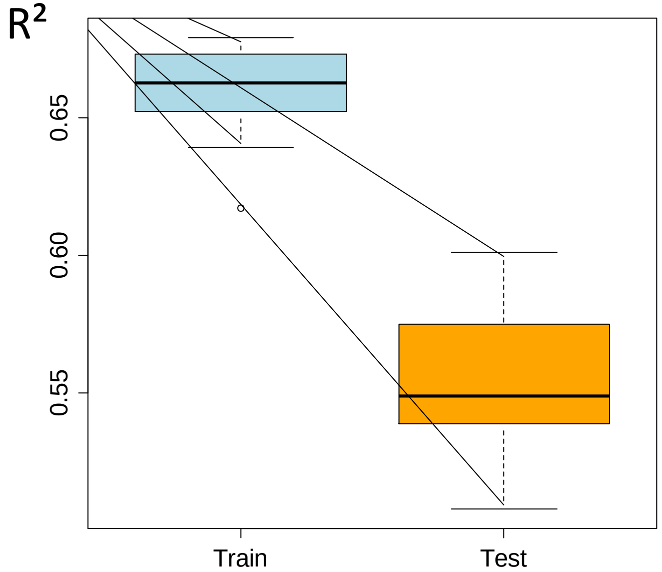
„Model of the choice“

GamSE200 ➔ complex spatial effect (df=131)

explained deviance = 0.55
baseline Model without SE explained deviance = 0.3

I. Internal validation

Model: mN = f(abiotic environment)



R^2 ➔ observed vs. predicted mN

Train ➔ $R^2 = 0.66$

Test ➔ $R^2 = 0.55$

II. Validation with independent forest inventory data

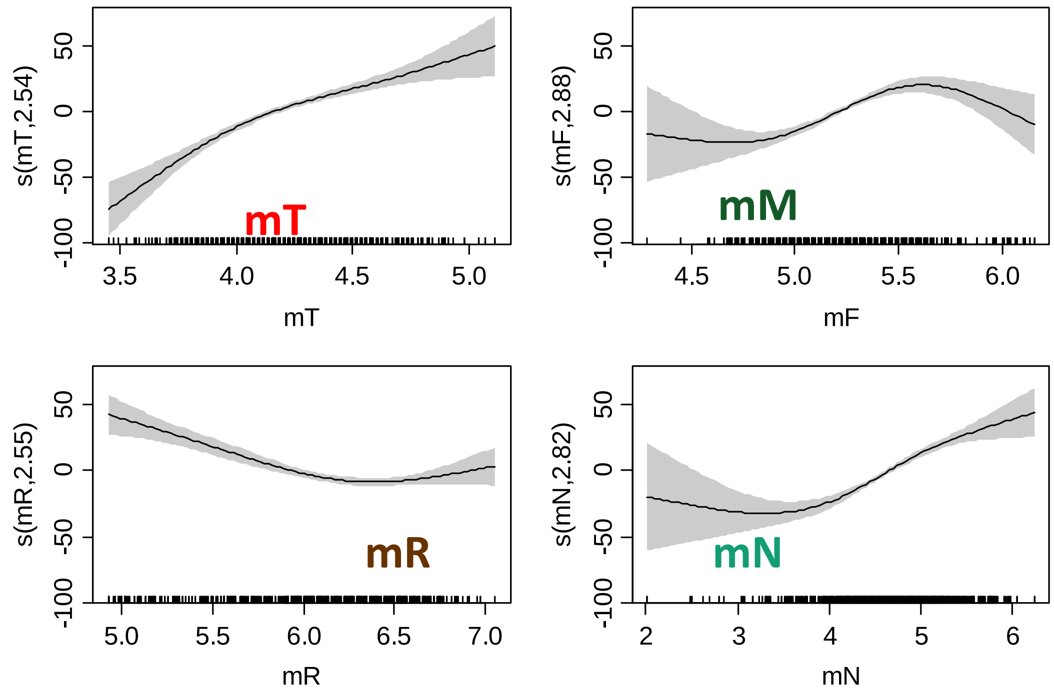
II. a) Can regionalised indicator values predict SI of Norway spruce?

Model: SI = f(indicator values)

Expl. deviance = 31%

Climate predictors

Nutrient predictors



II. b) Contribution of mN?

LWF

Answers to technical questions

1. Can spatial autocorrelation of model residuals (SAC) be removed?

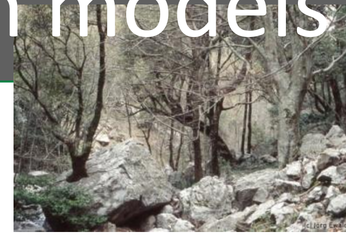
1.1 SAC could only be removed by a high order SE (df=131)
1.2 “Space” as a random effect in a GAMM could reduce but not remove SAC

2. Are there severe influences of spatial autocorrelation on significance levels and parameter estimations of models?

➔ SAC appeared to be of minor relevance for parameter (effect) estimation and significance levels

➔ Stratified large sample

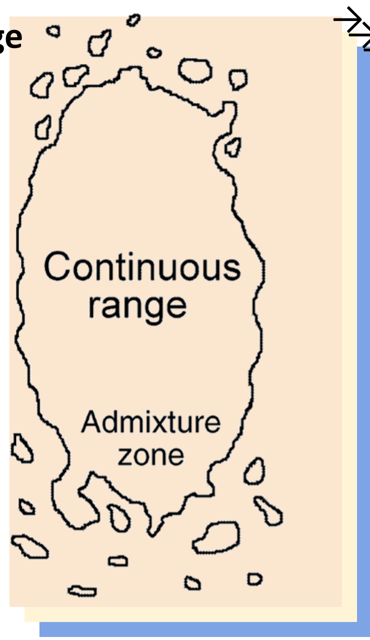
Species distribution models (SDM)



SDM calibration

Geographic space

Leading Edge



Rear Edge

Environmental data

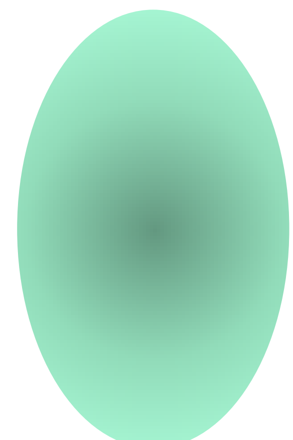
Model

SDM

Niche model

Environmental space

Response curves
Tolerance ranges



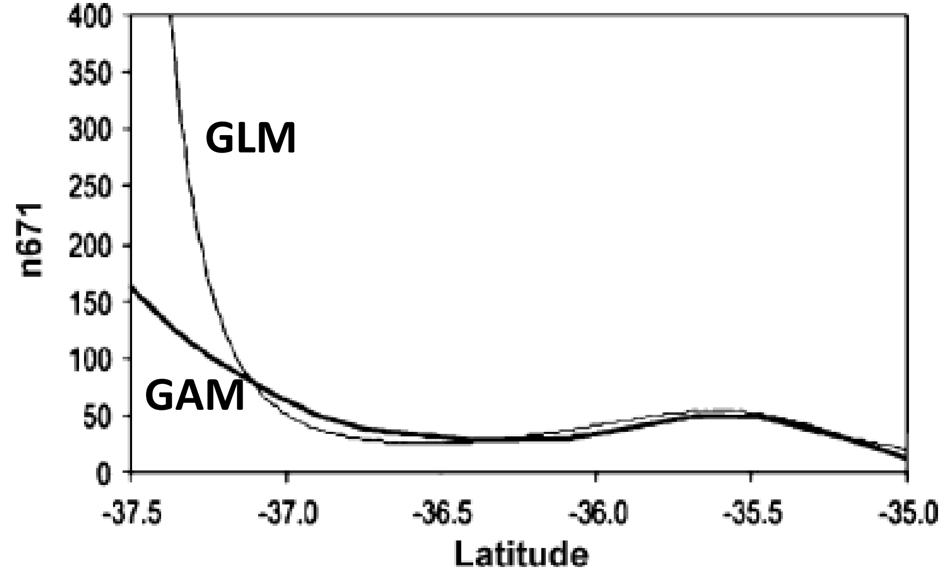
Niche

Hampe u. Petit (2005) Ecology Letters *8: 461–467*

Species-environment relationship ➔ site-specific tree species selection

Why GAMs are so popular in SDM?

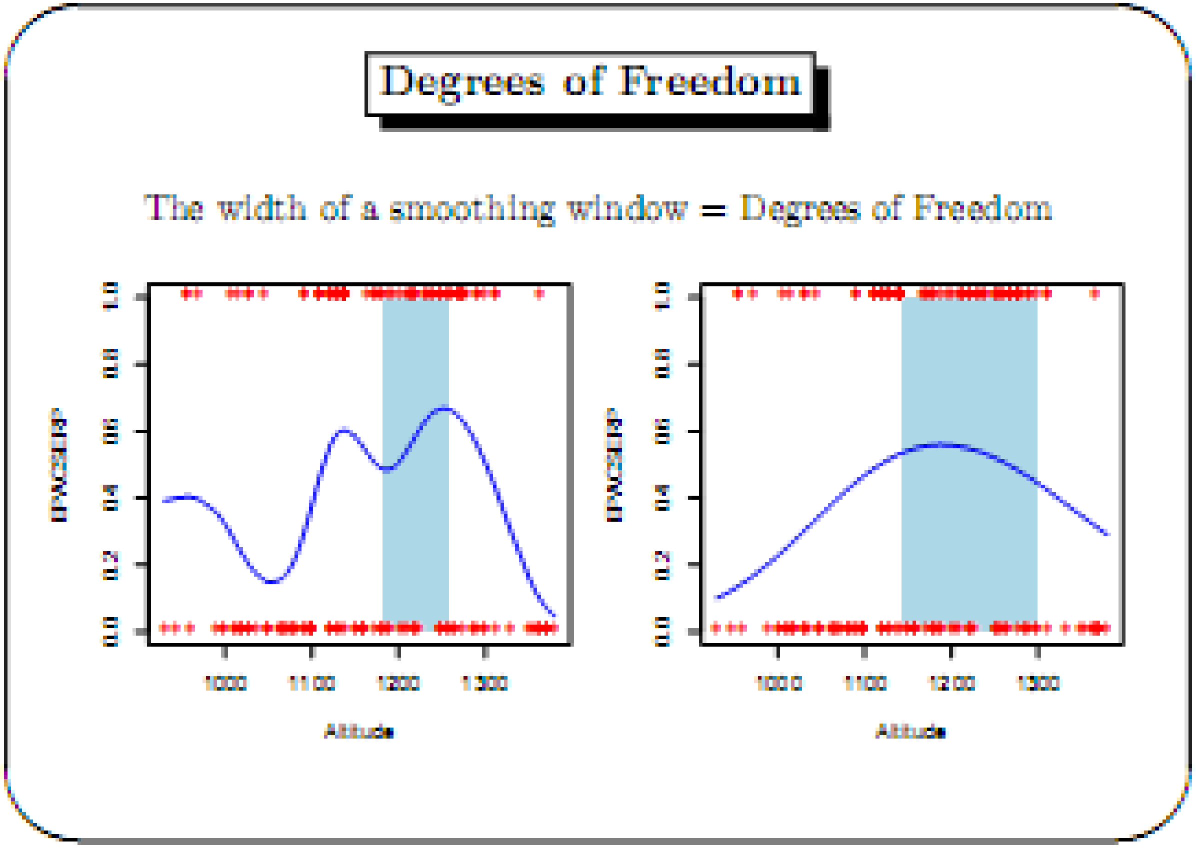
GAM are more flexible than GLMs
- Skewed and multimodal responses are possible
- At the edge GAMs provide more robust responses



Austin et al. 2006

Why GAMs are so popular in SDM?

Easy handling, e.g. complexity of response curves can be controlled by df



Oksanen 2003

Spatial SDMs using spatial effects

Comparison of different GAM implementations with SE

- GAM (mgcv)
- GamBoost (mboost, Hothorn et al.)

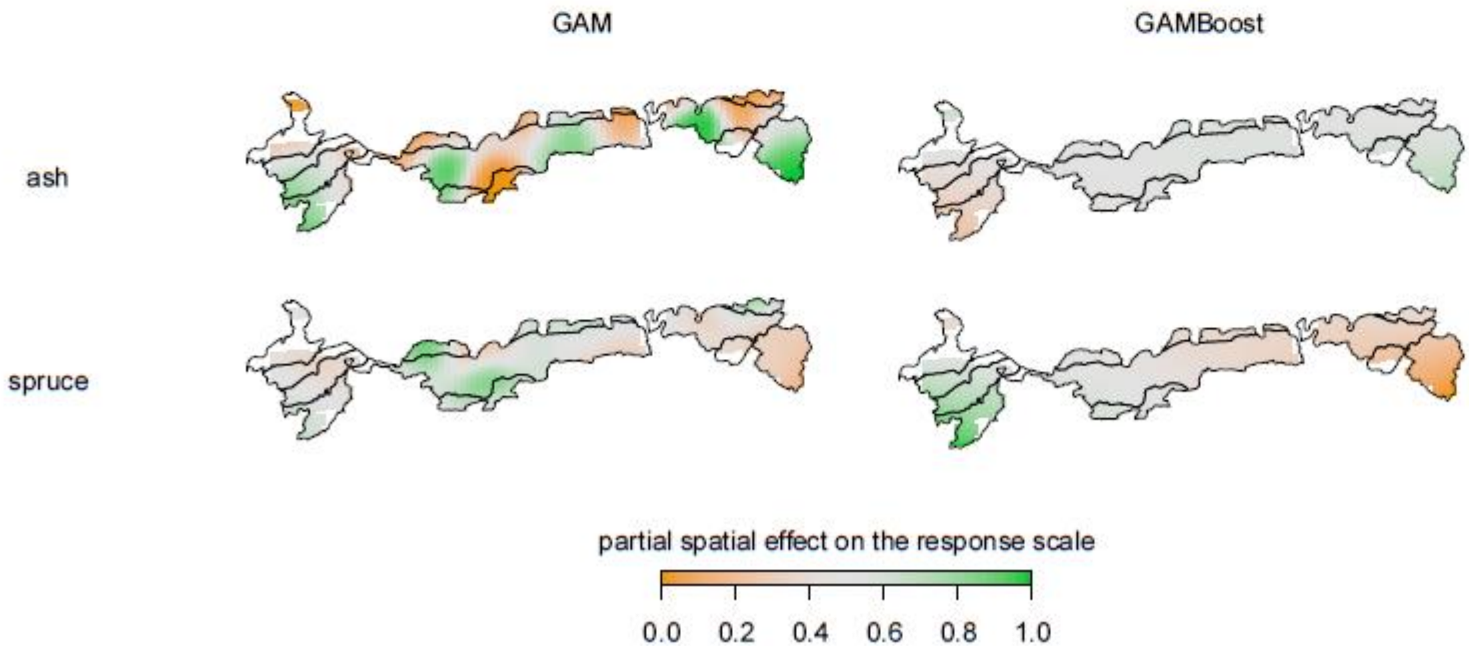
Questions

- Is the strength of SE different?
- Do spatial configurations of SE differ much?

Spatial SDMs using spatial effects

Result

- Not only the strength of SE is different, also the direction of SE may differ
- Indeed, the spatial configurations of SE differ much



Fensterer (2011): Statistical methods in niche modelling (diploma thesis, Department of statistics, LMU)

Vulnerability of forests against storm damage

Karl Mellert, Daniel Fröhlich, Lothar Zimmermann, Christoph Schulz,

Bayerische Landesanstalt für Wald und Forstwirtschaft

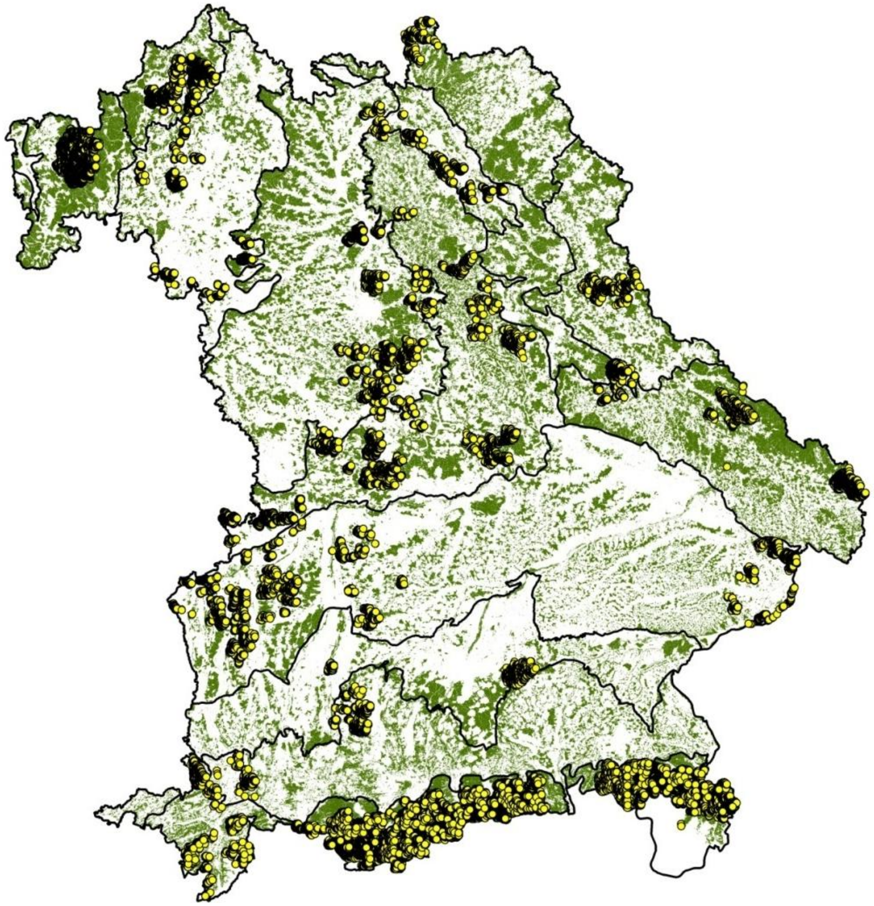
BAYERISCHE FORSTVERWALTUNG

ZENTRUM WALD FORST HOLZ WEIHENSTEPHAN

Vulnerability of forest (sites) against storm damage

Data

- 26080 plots
- Clumped distribution
- Zero-inflation: 90% of the plots without incidences



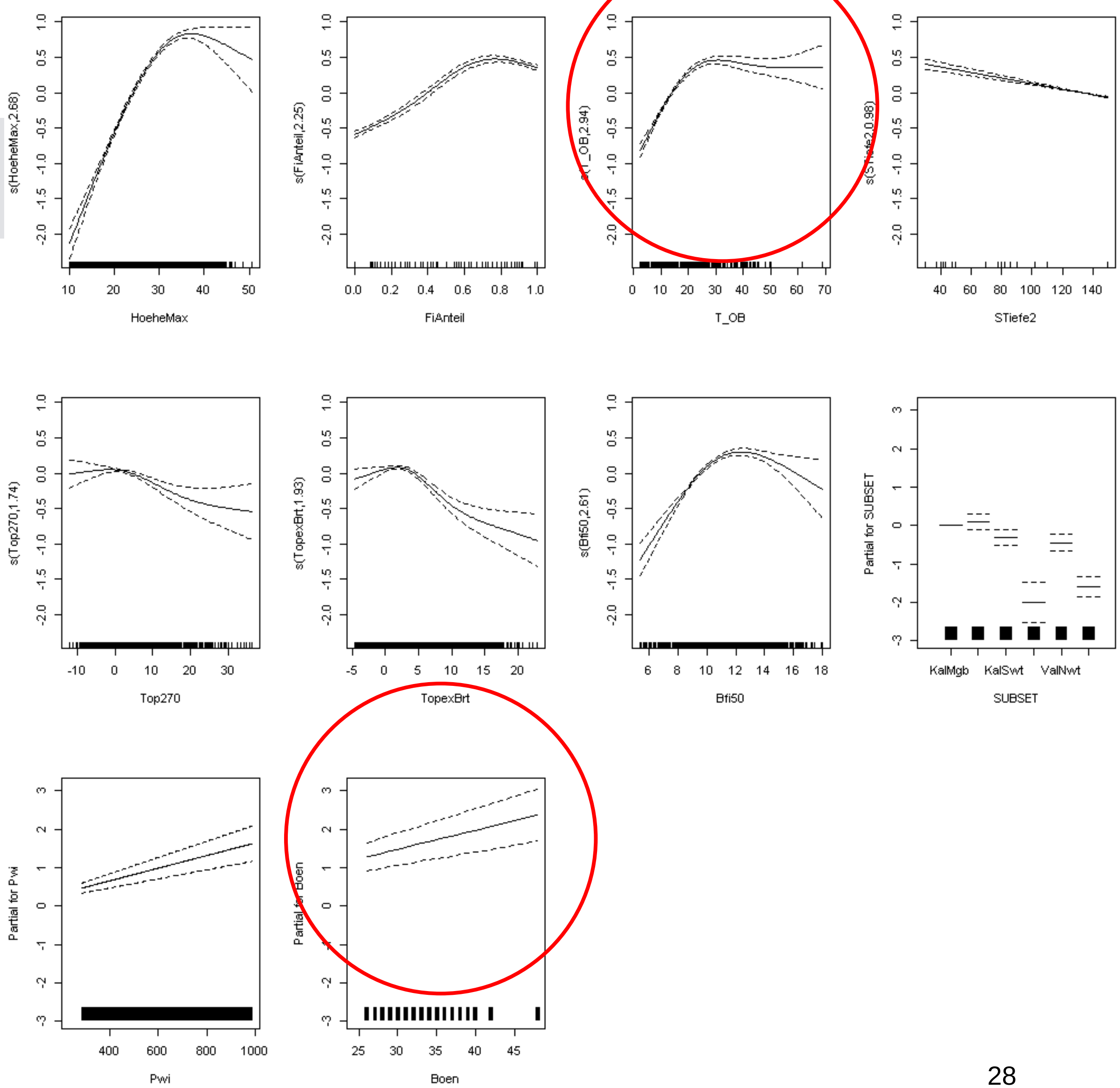
Vulnerability of forest (sites) against storm damage

- **Tested modelling approaches**
- GLM
  - [GLMM with MASS & nlme]
- GAM
  - GAM (mgcv) quasibinomial
  - [Constrained zero inflated GAM (cozigam, Liu & Chan 2010)]
  - [GAMM with mgcv & gamm4]
- Spatial effect
  - Do SE improve predicitve success of models?
  - Do SE affect shape of response of physical predictors?
  - Do SE look plausible?
  - Do SE improve transferability of models?

GAM without SE

- GAM without SE

R-sq.(adj) = 0.0638
Deviance explained = 13.4%
GCV score = 1.0168
n = 26080

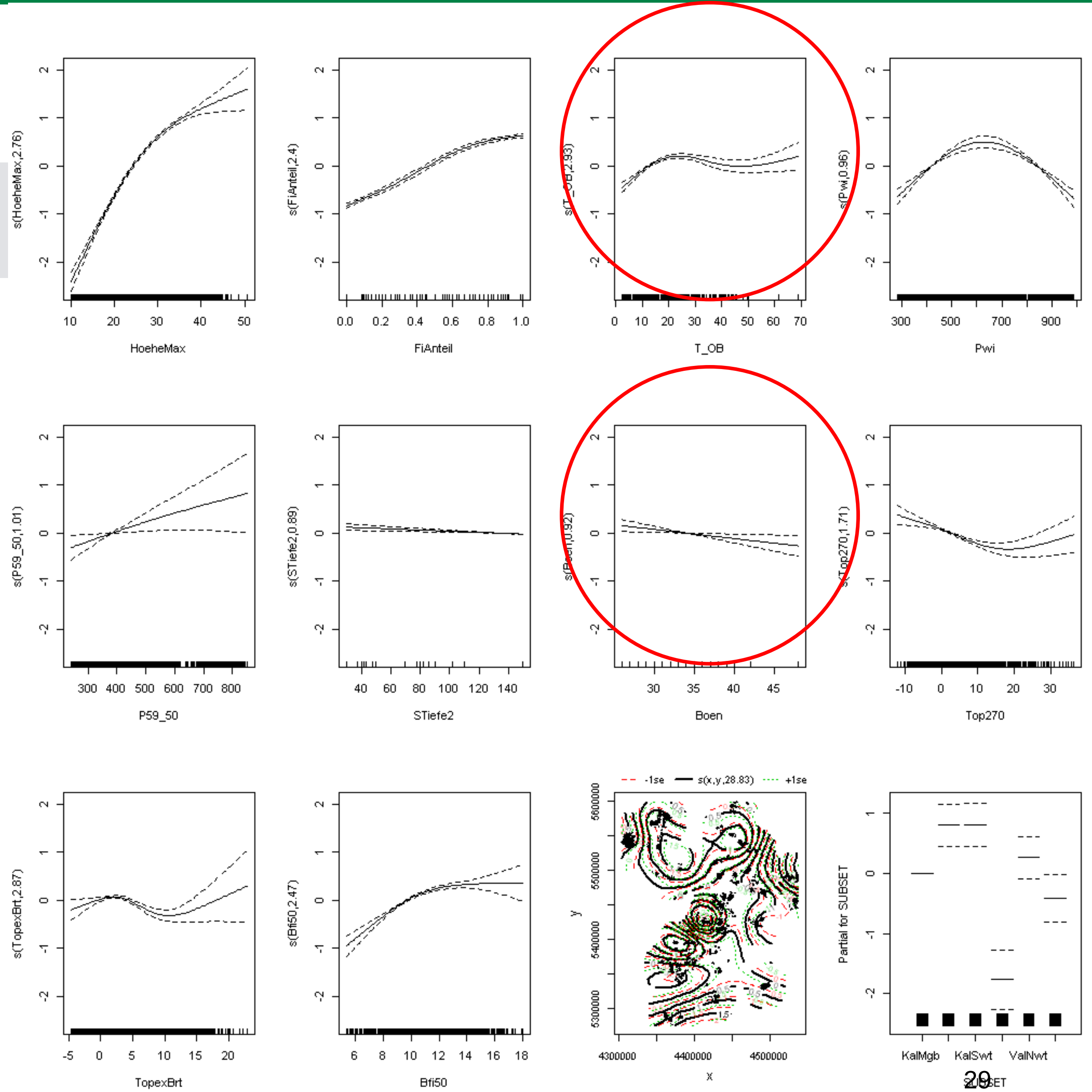


GAM with SE

- GAM with SE

R-s-sq.(adj) = 0.132
Deviance explained =   24%
GCV score = 0.89434
n = 26080

  - Higher perfomance
  - Reduction of some effects
  - Complex SE



Transferabilty GAM + SE

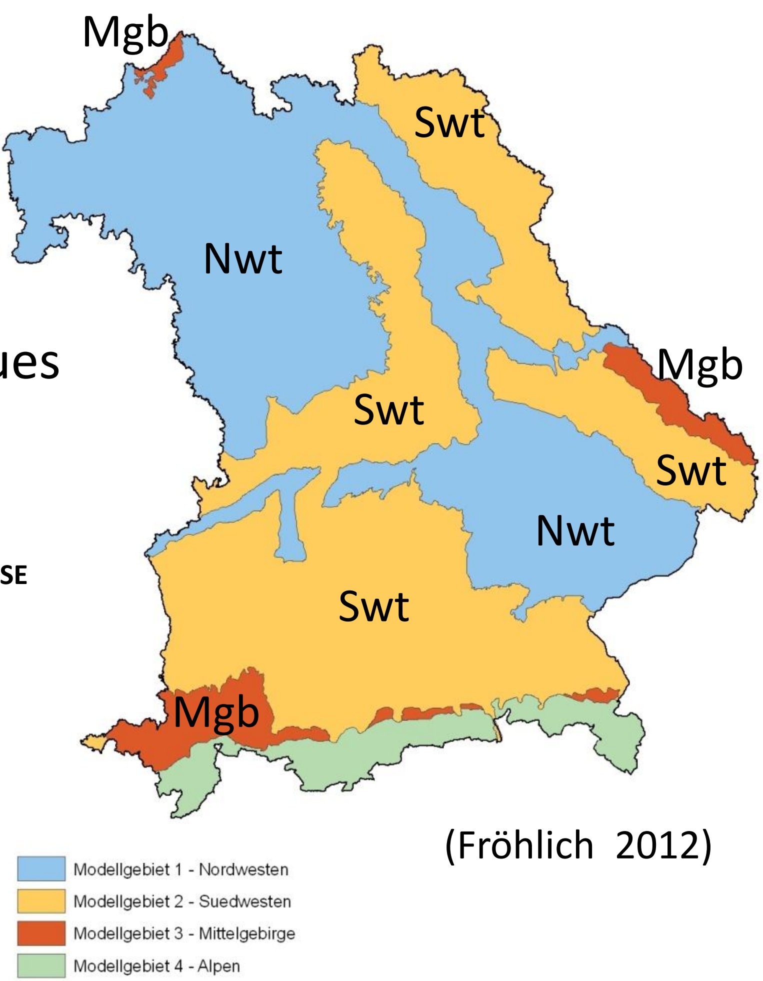
Transferability between regions

Regions: defined based on a PCA of
terrain variables and speed of wind gusts

Test:
Correlation of predicted and observed values

First row: calibrated in Mgb predicted to Nwt

Model_Data	without SE	with SE	difference	improved by SE
Mgb_Nwt	0.124	-0.061	-0.185	no
Mgb_Swt	0.085	0.068	-0.017	no
Nwt_Mgb	0.211	0.224	0.013	yes
Nwt_Swt	0.17	0.201	0.031	yes
Swt_Mgb	0.209	-0.024	-0.233	no
Swt_Nwt	0.189	0.083	-0.106	no



(Fröhlich 2012)

Vulnerability of forest (sites) against storm damage

Summary SE

- SE allways improved predictive success of models for training data
- Plausibility of SE was not easy to be validated
- SE improved transferability of models only in 1 of 3 cases

Discussion

General aspects

- Spatial effects (SE)
  - Reduce effects of physical predictors
  - High demand in df
  - Overfit data
  - Different approaches (Gam, GamBoost) result in different SE
  - SE are not allways plausible (edge!)
- Random effects (RE)
  - Did not allways help to reduce spatial autocorrelation of residuals
  - GLMM and GAMM runs often failed
    - Require enormous amounts of memory

Discussion

Questions

- Guidelines for spatial modeling with GAMs?
  - In what situations could we use SE?
  - In what situations should we definitly use GAMM?
- Spatial effects (SE)
  - How to reduce dominance of SE in models?
  - Could penalization of SE be strengthened?
- Random effects (RE)
  - Possibilities to reduce memory requirements with GAMM?
  - Could „simple random effects“ (mgcv) help in spatial modelling?
- Other questions
  - How to reduce wiggliness of response curves?

Spatial SDMs?

Most SDMs do not account for SAC

Reasons:

1. In case of low density sampling conventional models are acceptable
2. There are several methods to account for SAC (e.g. Dormann et al. 2007) but there is no unique method of the choice
3. High complexity of SAC models. Often subjective assumptions have to be made e.g. the neighbourhood size and shape
4. Application of SDM for climate change: It is not reasonable to project spatial structures of SE into the future

LWF

Spatial autocorrelation (SAC)

- Observation at nearby locations are not independent from each other

Causes for SAC (of model residuals)

biological processes are distance-related

- endogenous processes
  - → speciation
  - → extinction
  - → dispersal
  - → species interactions

(Dormann et al. 2007, Franklin 2009)

Causes for SAC (of model residuals)

Spatial processes, which affect species dispersal

2) exogenous processes (independent from 1)
   → disturbance
   → historical barriers
   → spatially structured environmental gradients

(Dormann et al. 2007, Franklin 2009)

Spatial Modelling

Causes for SAC (technical)

3) non-linear relationships between environment and species are modelled erroneously as linear

4) the statistical model fails to account for an important environmental determinant that in itself is spatially structured and thus causes spatial structuring in the response

(Dormann et al. 2007, Franklin 2009)