

# Auswertung von Sortenversuchen mit überlappenden Großräumen.

II. Umsetzung in SAS und PIAF.

J. Möhring

Universität Hohenheim, Fachgebiet Bioinformatik

# Inhalt

- Einführung Sortenversuche
- Modell
- Bestimmung der Gewichte
- Sortenschätzwert
- Probleme bei der Umsetzung
- Zusammenfassung

# Sortenprüfwesen in Deutschland

~ 6 Jahre Selektion durch den Züchter

Wertprüfung,  
bundesweit

Landessorten-  
versuche

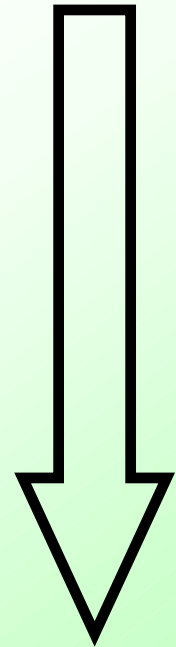
**Anmeldung**

~ 100 Sorten im 1. Jahr  
~ 25 Sorten im 3. Jahr

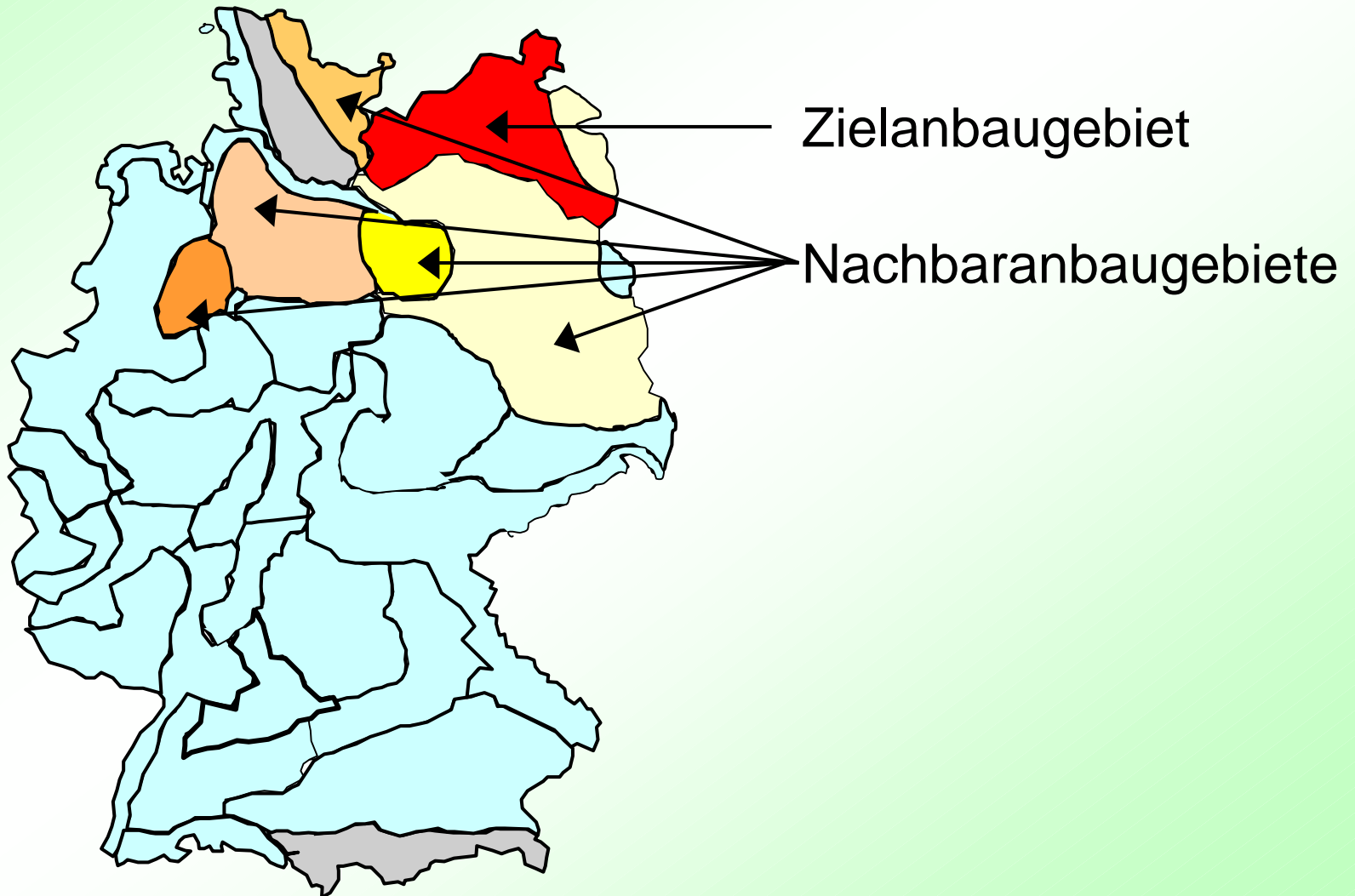
**Zulassung**

2-3 Jahre LSV

**Empfehlung**



# Beispiel Anbaugebiet 4



# Das Modell

Gemischtes Modell mit den Faktoren

- Sorte (S)
- Jahr (J)
- Region / Anbaugebiet (R)
- Ort, geschachtelt in Region (O(R))

# Das Modell

Berücksichtigung des Versuchstyps

- Hinzunahme des Faktors „Typ“ modelliert Unterschiede zwischen 2 Versuchen im selben Jahr am selben Ort, z.B. WP und LSV
- $T (J^*O(R))$  dient als Blockungsfaktor

# Das Modell

$$y_{ijrlm} = sr_{ir} + jo(r)_{jrl} + t(jor)_{jrlm} + sj_{ij} + so(r)_{irl} + sjr_{ijr} + sjo(r)_{ijrl} + e_{ijrlm}$$

*s* = Sorte

*j* = Jahr

*r* = Region

*o* = Ort

*t* = Versuchstyp (WP,LSV)

*e* = Restfehler

# Die Auswertung

1. Schritt: Berechne Sortemittelwert (LSMeans) pro Region (mit der Varianz-Kovarianz-Matrix  $\Sigma_e$ )

Varianzen lassen sich mittels Estimate in MIXED berechnen (Sorte\*Region fest).

$$\text{cov}(A, B) = \frac{\text{var}(A - B) + \text{Var}(A) + \text{Var}(B)}{-2}$$

$\Sigma_e$  = umweltbedingte Varianz-Kovarianz-Matrix



# Umweltbedingte Varianz-Kovarianz-Matrix

Betrachte Sorte als zufällig:

$$\text{var}(\hat{\boldsymbol{\eta}}_i) = \boldsymbol{\Sigma}_g + \boldsymbol{\Sigma}_e$$

$\boldsymbol{\Sigma}_g$  = genetische Varianz-Kovarianz-Matrix

$\boldsymbol{\Sigma}_e$  = umweltbedingte Varianz-Kovarianz-Matrix

$\hat{\boldsymbol{\eta}}_i$  =  $(R \times 1)$ -Vektor der Schätzwert für den Erwartungswert der Sorte  $i$  im  $r$ -ten Anbaugebiet

# Genetisch bedingte Varianz-Kovarianz-Matrix

Sorte\*Region zufällig:

$$\begin{aligned}\Sigma_g &= \mathbf{I}\sigma_{sr}^2 + \mathbf{J}\sigma_s^2 \\ \Sigma_g &= \mathbf{I}\sigma_{sr_i}^2 + \mathbf{J}\sigma_s^2 \quad \dots\end{aligned}$$

**I** = Einheitsmatrix

**J** = (n×n)-Matrix mit allen Elementen gleich 1

$\sigma_s^2$  = Sortenvarianz

$\sigma_{sr}^2$  = Sorte\*Region-Varianz

# Optimale Gewichte

$$\mathbf{w}' = \mathbf{u}'_r \boldsymbol{\Sigma}_g (\boldsymbol{\Sigma}_g + \boldsymbol{\Sigma}_e)^{-1}$$

$$(\mathbf{w} \lambda)' = \mathbf{u}'_r \begin{pmatrix} \boldsymbol{\Sigma}_g & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_g + \boldsymbol{\Sigma}_e & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix}^{-1}$$

$\boldsymbol{\Sigma}_g$  = genetische Varianz-Kovarianz-Matrix

$\boldsymbol{\Sigma}_e$  = umweltbedingte Varianz-Kovarianz-Matrix

$\mathbf{u}_r$  = Einheitsvektor für  $r$ -tes Anbauggebiet

$\mathbf{1}$  =  $(n \times 1)$ -Vektor mit allen Elementen gleich 1

# Schätzwert

Korrigiere um das Regionenmittel

$$sr_{ir} = \mathbf{w}' \begin{pmatrix} sr_{i1} - r_1 \\ \vdots \\ sr_{in} - r_R \end{pmatrix} + r_r$$

$sr_{ir}$  = Schätzwert für die  $i$ -te Sorte in der  $r$ -ten Region

$r_j$  = Mittelwert der  $j$ -ten Region


$\mathbf{w}$  = Vektor mit den Gewichten

# Probleme

- In welcher Software wie umsetzen?
- Varianzstruktur für  $\Sigma_g$

# Software

- Anwender sind die Ländesdienststellen
- PIAF ist allen Länderdienststellen zugänglich
- PIAF-Stat basiert auf SAS, Einbindung von SAS-Makros möglich

Erstellung von 3 Makros für  
 Varianzkomponentenschätzung, BLUP und gewichtete LSMeans

# Umsetzung in SAS

**ERROR: Out of memory.**

**NOTE: The SAS System stopped... .**

**NOTE: PROCEDURE MIXED used:**

real time	45.34 seconds
-----------	---------------

cpu time	7.20 seconds
----------	--------------

# Umsetzung in SAS

- Ausschalten der Tests für die festen Effekte
- Umwelteffekte werden fix genommen, Nichtnutzung der Interumweltinformation
- Verwenden einer subject-Variablen möglich

```
Proc mixed Data=Datei;  
class S J R O Typ;  
model y=                J*O(R) Typ*J*O(R)  
                /ddfm=residual notest;  
random int J    O(R)    J*O(R)/subject=S;  
random    R/subject=S  solution;  
run;
```



# Interumweltinformation

**Tab.1:** Vergleich der Varianzkomponenten für Umwelteffekte und Sorte\*Umwelt-Effekte bei der gemeinsamen Auswertung aller Anbaugebiete

Varianzkomponente	Gemeinsame Auswertung aller Anbaugebiete	
	Umwelteffekt	mit S gekreuzt
J	10,93	3,90
R	fest	1,32
O	28,75	2,69
JR	21,75	2,92
JO	41,82	5,94

# Probleme

- In welcher Software wie umsetzen?
- Varianzstruktur für  $\Sigma_g$

# Varianzstruktur

**Tab.2:** Vergleich verschiedener Modelle bezüglich Güte und Rechenzeit bei einem relativ **großen** Datensatz (4 Anbauggebiete, 6 Jahre, WP+LSV, Kulturart Winterweizen)

Varianzstruktur für Sorte*Region	2*Differenz der LogLikelihood zu CS	Kritischer $\chi^2$ -Wert
Compound Symmetry (CS)	0	-
Heterogene Varianz (UN(1))	0,80	7,81
Faktoranalytisch	<b>16,20</b>	12,59
Unstrukturiert	Konvergenzprobleme	15,51

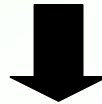
# Varianzstruktur

**Tab.1:** Vergleich verschiedener Modelle bezüglich Güte und Rechenzeit bei einem **kleineren** Datensatz (2 Anbauggebiete, 6 Jahre, WP+LSV, Kulturart Winterweizen)

Varianzstruktur für Sorte*Region	2*Differenz der LogLikelihood zu CS	Kritischer $\chi^2$ -Wert
Compound Symmetry (CS)	0	-
Heterogene Varianz (UN(1))	1,24	3,84
Faktoranalytisch	1,24	3,84
Unstrukturiert	Konvergenzprobleme	5,99

# Varianzstruktur

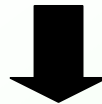
Große Datensätze



Gute Varianzkomponentenschätzungen



Komplexere Varianzstrukturen sind möglich



Eventuell bessere Modellierung der Wirklichkeit

aber

Je mehr Daten und je komplexer das Modell, desto höher die Rechenzeit

# Zusammenfassung

- Bestimmung von nicht geschrumpften Sortenschätzwerten bei Einbeziehung von verschiedenen Anbaugebieten ist möglich
- Umsetzung in SAS möglich
- Die Wahl der Varianzstruktur wird begrenzt durch Datenumfang und Rechnerressourcen