The „Hohenheim-Gülzow-method“

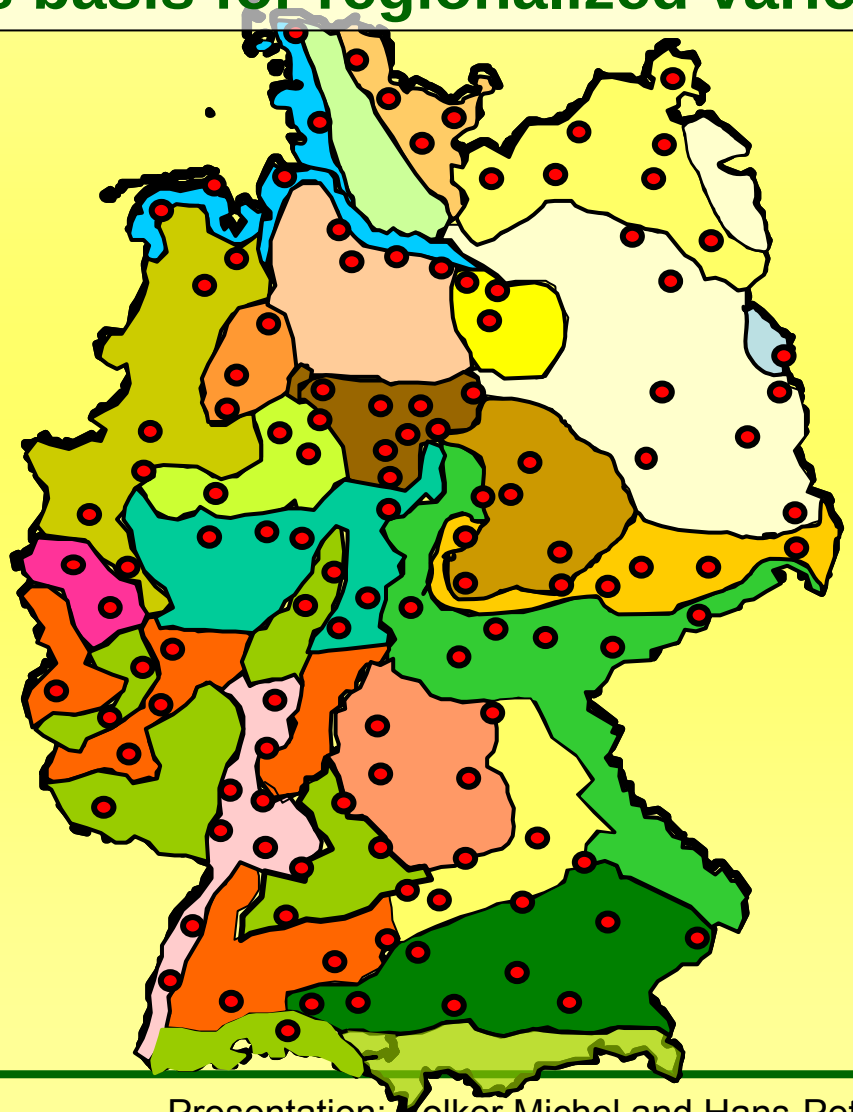
for analysis of series of trials
as basic procedure with PIAFStat and SAS
in a regionalized field trial system

Volker Michel, Andrea Zenk (Gülzow)

Jens Möhring, Hans-Peter Piepho (Stuttgart-Hohenheim)

Rudolf Graf (Freising)

Post-registration State Variety Trials
in Germany (*VCU testing*)
as basis for regionalized variety recommendations



- locations of trials with wheat in the
  - federal states
  - **growing-regions**

Incentive to improve analysis systems in competitive variety testing

number of trials is decreasing

loss in regionalized specification of varieties ?

loss in earliness and precision of estimation ?

Main possibilty: enlarge the data basis

In the past the analysis systems of German states involved only:

- State Variety Trials of each individual state
- only 2 or 3 years
- only the balanced data - subset

in the interest of earliness and precision it is better to enlarge the dataset by using all available regional data →

Enlarge the data basis

data basis can be enlarged by:

- using trials for registration by the Federal Office or from other pre-registration trial systems
- using more than three years
- using trials from neighbouring regions

but: such long-term series of trials in variety testing involve many complex and complicated system-inherent characteristics →

The H-G-method comprises the following features

a) **unbalancedness**

b) regional neighbourhood and similarity

c) heterogeneous standard errors

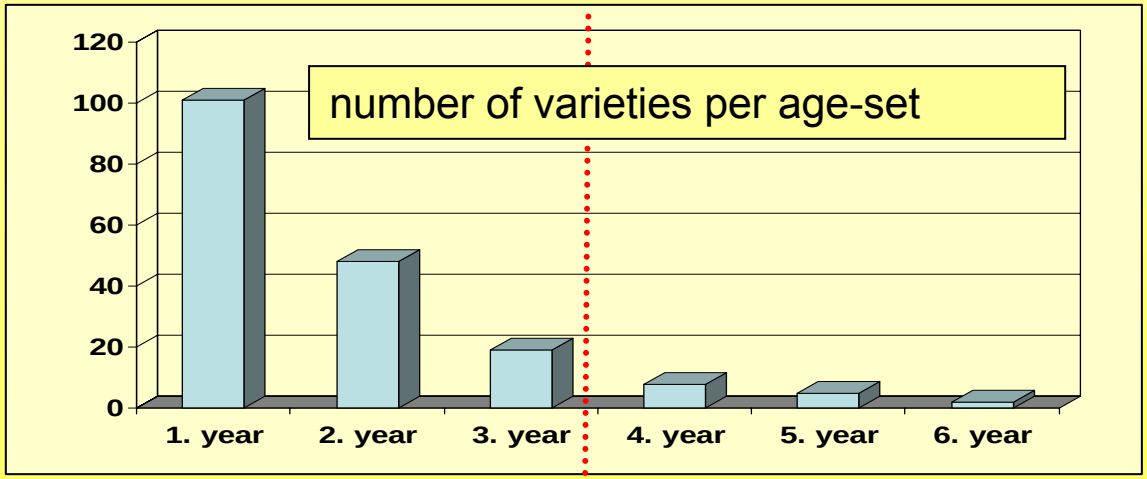
d) non-additive data behavior

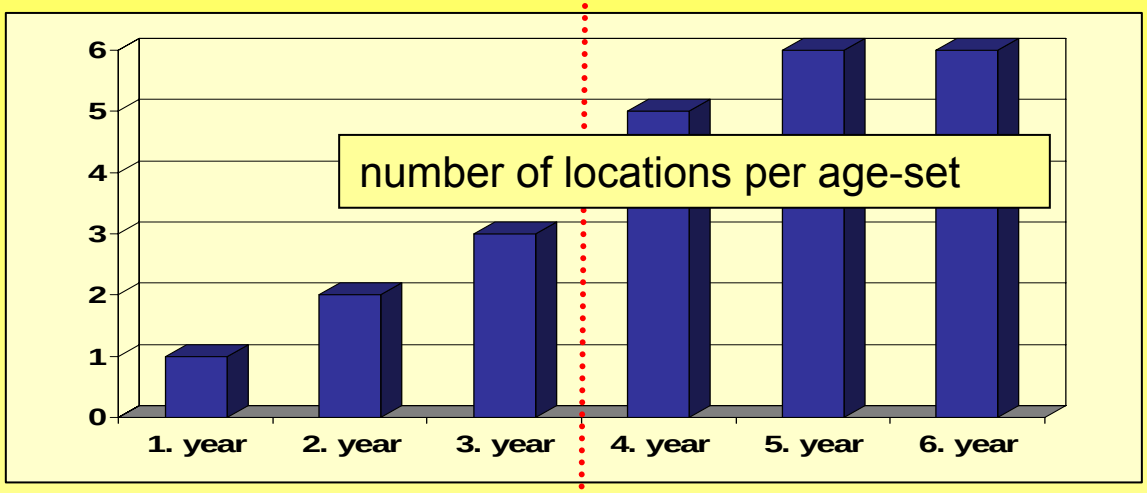
a) Unbalancedness

- combined with variety-environment-interactions -

year	Jahr	1995															1997															1998										1999					
trial system type	Typ	LSV						BSV			WP2		WP1				LSV					BSV	WP2		WP1						LSV					BSV	WP2		WP1			LSV				
location	Ort	1	5	7	10	11	12	1	3	10	5	9	2	6	8	9	1	5	10	11	12	13	1	1	5	9	2	6	8	9	1	3	5	10	12	1	5	9	8	9	1	4	5	10	12
variaty	Sorte																																												
1		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
2		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
3											x	x					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
4											x	x					x	x	x	x	x	x									x	x	x	x	x	x	x	x	x	x	x	x	x	x
5											x	x					x	x	x	x	x	x									x	x	x	x	x	x	x	x	x	x	x	x	x	x
6											x	x					x	x	x	x	x	x									x	x	x	x	x	x	x	x	x	x	x	x	x	x
7											x	x					x	x	x	x	x	x									x	x	x	x	x	x	x	x	x	x	x	x	x	x
8											x	x					x	x	x	x	x	x									x	x	x	x	x	x	x	x	x	x	x	x	x	x
9																	x	x	x	x	x	x									x	x	x	x	x	x	x	x	x	x	x	x	x	x
10													x	x	x	x							x	x							x	x	x	x	x	x	x	x	x	x	x	x	x	x
11													x	x	x	x							x	x							x	x	x	x	x	x	x	x	x	x	x	x	x	x
12													x	x	x	x							x	x							x	x	x	x	x	x	x	x	x	x	x	x	x	x
13													x	x	x	x							x	x							x	x	x	x	x	x	x	x	x	x	x	x	x	x
14													x	x	x	x							x	x							x	x	x	x	x	x	x	x	x	x	x	x	x	x
15													x	x	x	x							x	x																	x	x	x	x	x
16																									x	x									x						x	x	x	x	x
17																									x	x									x						x	x	x	x	x
*18 -224		12	12	12	12	12	12	17	17	17	20	20	46	46	46	46	7	7	7	7	7	6	16	16	17	17	53	53	53	53	3	4	3	3	3	19	22	22	54	54	0	3	0	0	0

a) Unbalancedness





before ← **→ after**

REGISTRATION

selection process

different period of trial-process per variety

only few data per region in the early trial phase

even these data are useful to assess the interactions with years

The method comprises the following features

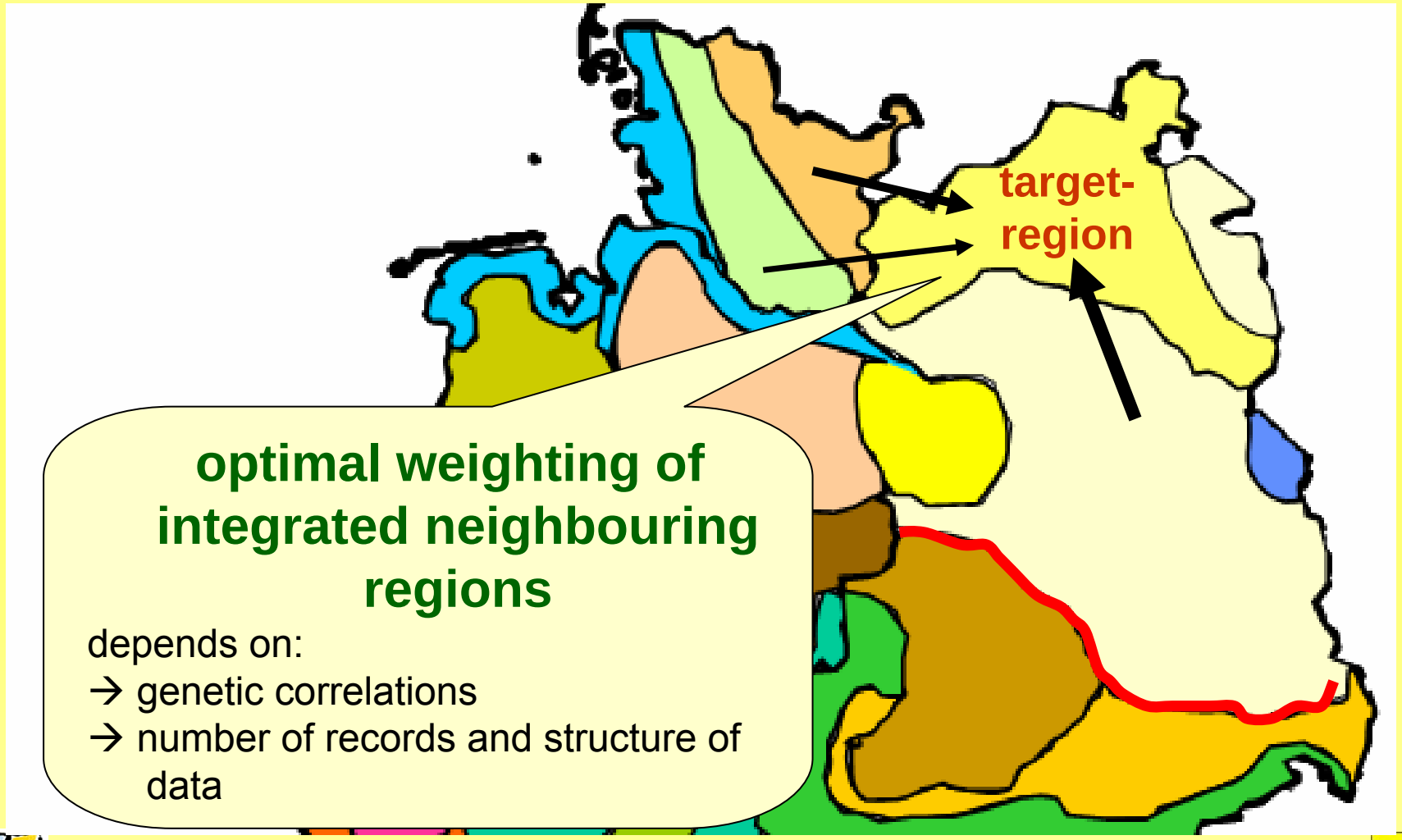
a) unbalancedness

b) regional neighbourhood and similarity

c) heterogeneous standard errors

d) non-additive data behavior

b) Regional neighbourhood and similarity



optimal weighting of integrated neighbouring regions

depends on:

→ genetic correlations

→ number of records and structure of data

The method comprises the following features

a) unbalancedness

b) regional neighbourhood and similarity

c) heterogeneous standard errors

d) non-additive data behavior

The method comprises the following features

a) unbalancedness

b) regional neighbourhood and similarity

c) heterogeneous standard errors

d) non-additive data behavior

Statistical features of the H-G-method

a) unbalanced data
b) regional neighbourhood and similarity
c) heterogeneous standard errors
→ **two-stage analysis**
→ **multi-environmental mixed model**

d) non-additive data behavior
→ **Box-Cox transformation**

Two-stage analysis (weighting of trials)

Stage 1: Compute adjusted means and standard errors (s.e.) per trial

Stage 2: Fit mixed model across series, fixing error variance at $(\text{s.e.})^2$ from Stage 1:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e} \quad \text{with} \quad \boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{R}),$$

where $\boldsymbol{R}$ is diagonal with diagonal elements equal to $(\text{s.e.})^2$

Box-Cox transformation of data

$$z = \begin{cases} \left(y^{\phi} - 1\right)/\phi & \text{for } \phi \neq 0 \\ \log(y) & \text{for } \phi = 0 \end{cases}$$

where ϕ is a transformation parameter.

ϕ= 1: Do not transform $\Rightarrow$ Additive effects

ϕ= 0: Log-transform $\Rightarrow$ Multiplicative effects

Log(a*b) = log(a) + log(b)

The multi-environment mixed model

$V + L \bullet Y \bullet T \; : \; V \bullet L + V \bullet Y + V \bullet L \bullet Y + V \bullet L \bullet Y \bullet T$,

where

V = variety
L = location
Y = year
T = trial (may be several per L $\times$ Y combination)

Effect L•Y•T fixed
$\Rightarrow$ no recovery of inter-trial information

Extension to sub-divided target region

Replace L by

$R/L = R + R \bullet L$,

where

R = region
L = location, nested within region

Variety effect, nested in regions

Variety effect, nested in regions: V•R

Effects for variety correlated between regions:

$$\text{var}\begin{pmatrix} V \bullet R1 \\ V \bullet R2 \\ V \bullet R3 \\ V \bullet R4 \end{pmatrix} = \Sigma_g$$

Different structures for Σ_g, e.g. compound symmetry, factor-analytic, heterogeneous

Compute adjusted means for V*R

(1) Compute adjusted means for V•R

$\Rightarrow y_{ir}$ for variety *i* regions *r* = 1, 2, ….

(2) Combine across regions as weighted mean

Yield estimate: $E_i = w_{i1} \times y_{i1} + w_{i2} \times y_{i2} + \ldots.$

Compute adjusted means for V*R

Objective: Minimize

"Error" = "Estimate – true value (V•R)"

$\Rightarrow$ Mean squared error (MSE)

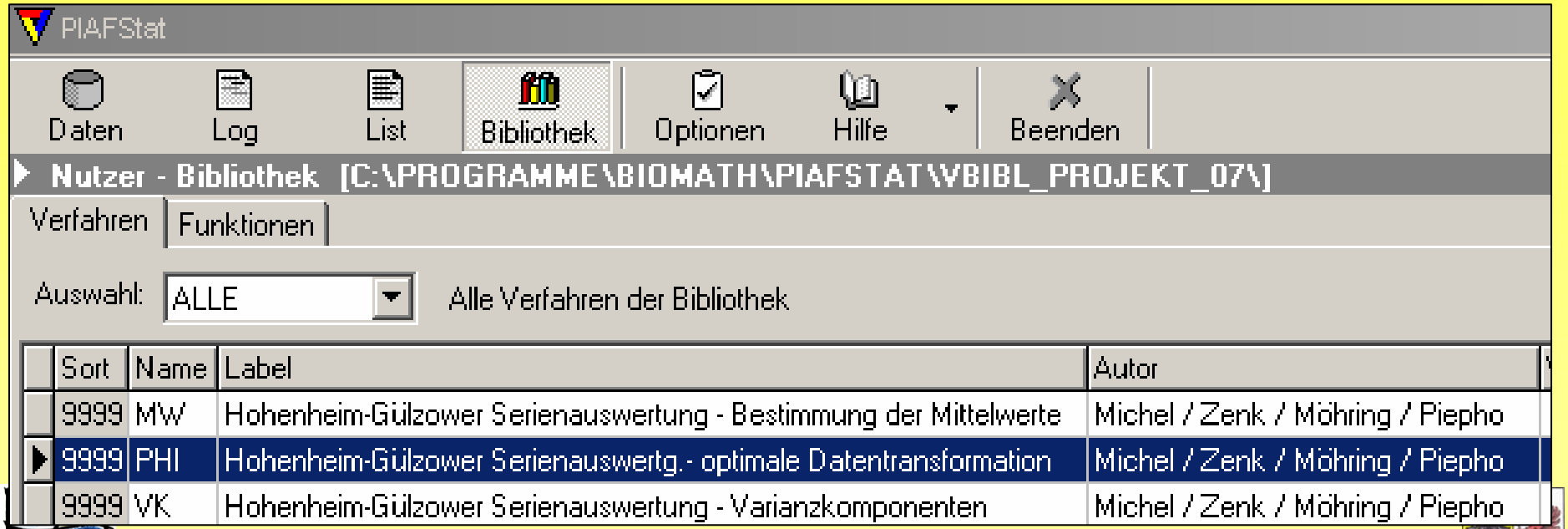
Two variants:

(1) $\text{Sum}_i(w_{ir}) = 1$
(2) No constraint on w_{ir} ($\Rightarrow$ BLUP)

Integration in the software-concept of PIAF

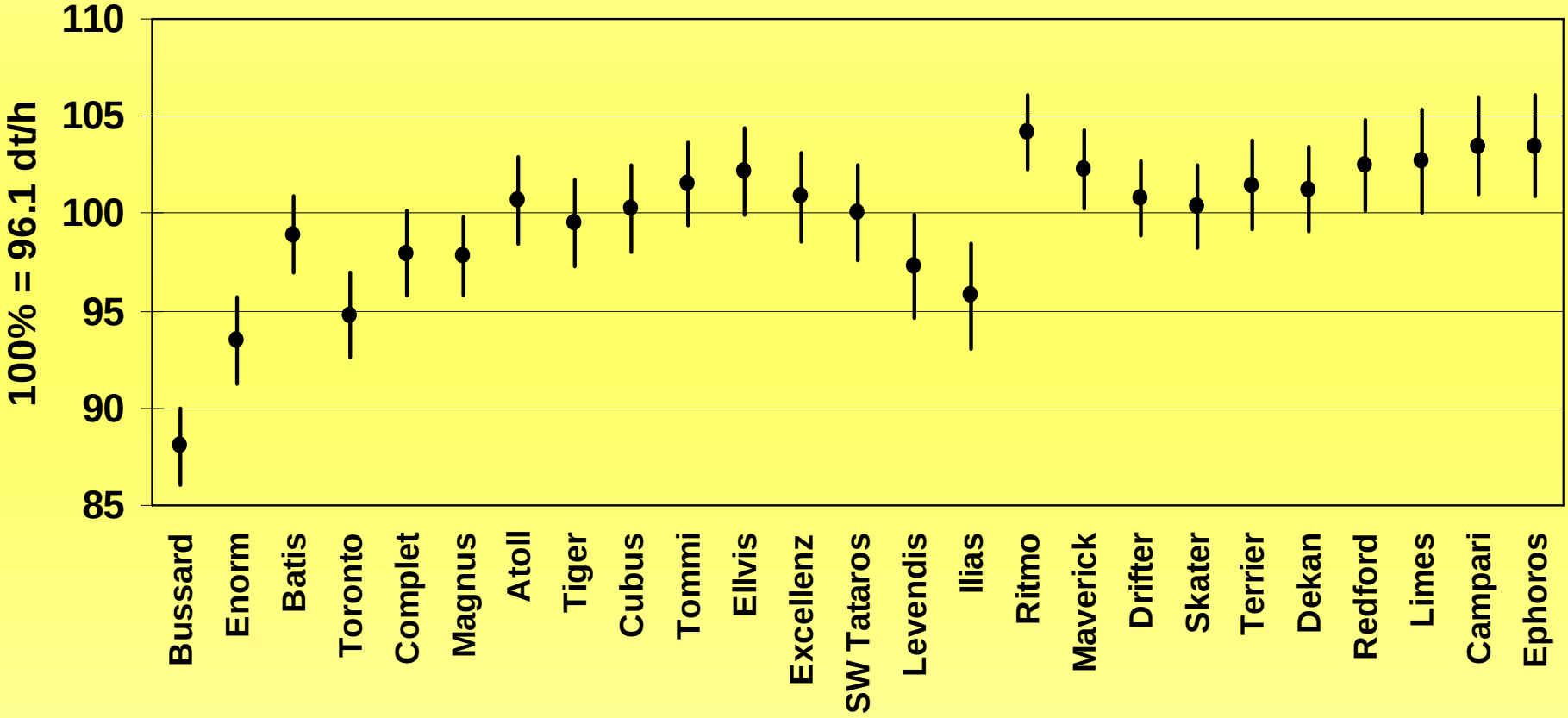
PIAFStat:

controls the statistical analysis;
connects PIAF and SAS by the procedure library of analysis algorithms



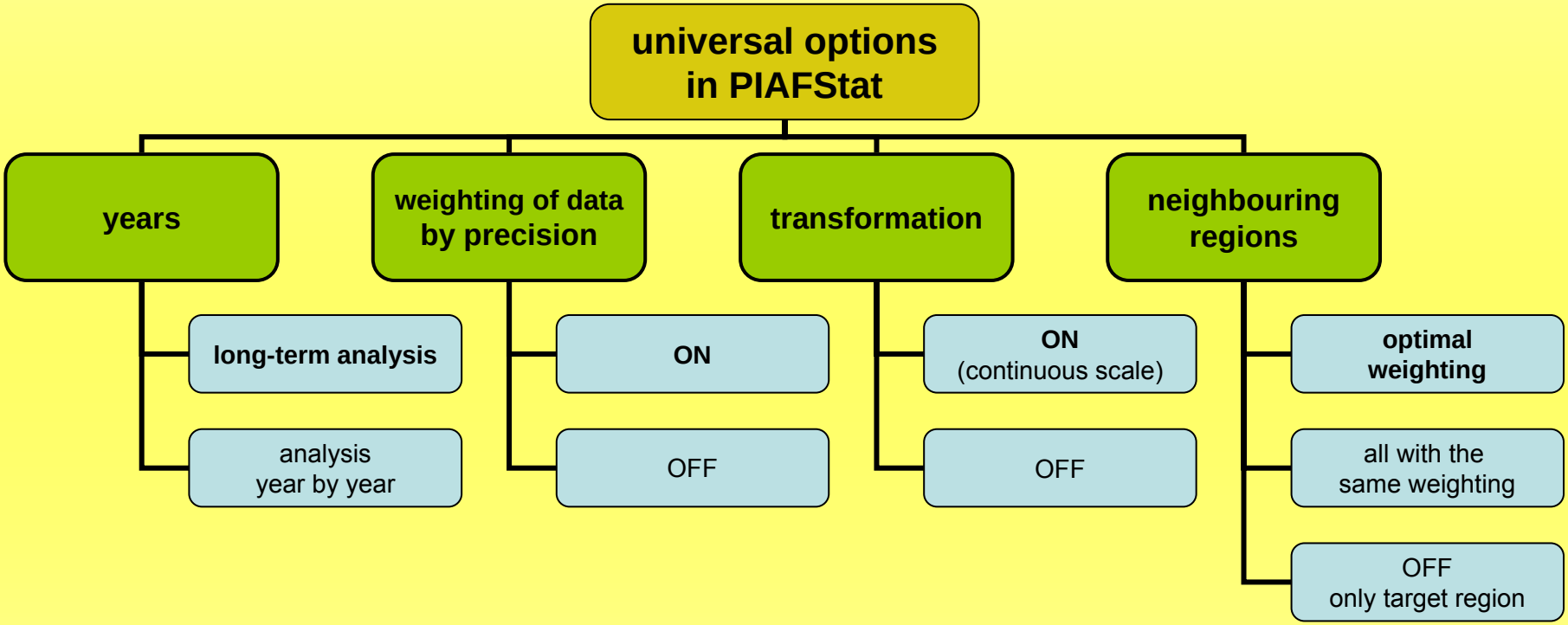
Output of PlAFStat-procedure „MW“

Estimated means and confidence intervals for paired comparisons



„Means with non-overlapping intervals are significantly different according to a t-test.“ Hochberg et al., 1982 and Piepho, 2000

„Hohenheim-Gülzow-method“ as the basic procedure in the analysis of all our series of trials



Summary (1)

- Number of trials has been reduced in German variety testing. There is an increased need to optimally exploit all information available from variety trials.
- The Hohenheim-Gülzow-method increases the earliness and precision of regional variety recommendations.
- The database can be enlarged by using unbalanced long-term series, including pre-registration trial systems and trials from neighbouring regions with optimal weighting.
- The method comprises the following features:
  - unbalancedness,
  - regional neighbourhood and similarity,
  - heterogeneous standard errors,
  - non-additive data behaviour.

Summary (2)

- Key statistical features are:
  - two-stage analysis,
  - Box-Cox transformation of data,
  - multi-environment mixed model,
  - handling of unbalancedness,
  - overlapping regionalization and optimal weighting of neighbouring regions.
- The method has been integrated in the procedure library of PIAFStat as a modul of the PIAF system.
- „Hohenheim–Gülzow-method“ is the basic procedure in the analysis of all series of trials.
- Ministers of Agriculture decided, that this method is going to be standard in the post-registration variety testing in Germany.
- In the federal states of Mecklenburg-Western Pomerania and Bavaria this has already been successfully implemented.