E

TWC/25/14 Add.
ORIGINAL: English
DATE: February 6, 2008

UPOV

**INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS**

GENEVA

# TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

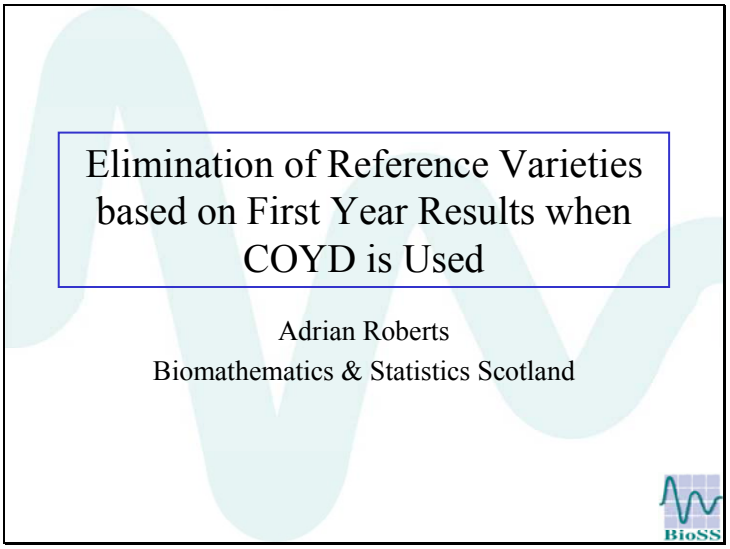
## Twenty-Fifth Session
## Sibiu, Romania, September 3 to 6, 2007

ADDENDUM TO
A RATIONALE FOR EXCLUDING VARIETIES OF COMMON KNOWLEDGE FROM THE SECOND GROWING CYCLE WHEN COYD IS USED

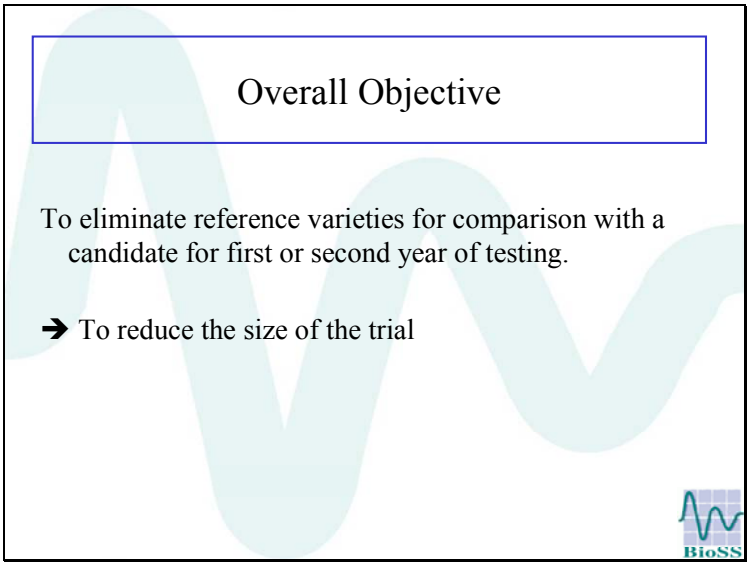
"ELIMINATION OF REFERENCE VARIETIES BASED ON FIRST YEAR RESULTS WHEN COYD IS USED"

*Document prepared by an expert from the United Kingdom*

Slide 1




Slide 2



Slide 3

## How?

Eliminate reference varieties that would almost certainly be found distinct if testing were carried out

- For first year of tests:
  - Use less official information from TQs etc
- For second year of tests:
  - Use data from first year of tests plus historical data

I will concentrate on the 2nd case in this talk

Slide 4

## Two approaches

A. Character-by-character

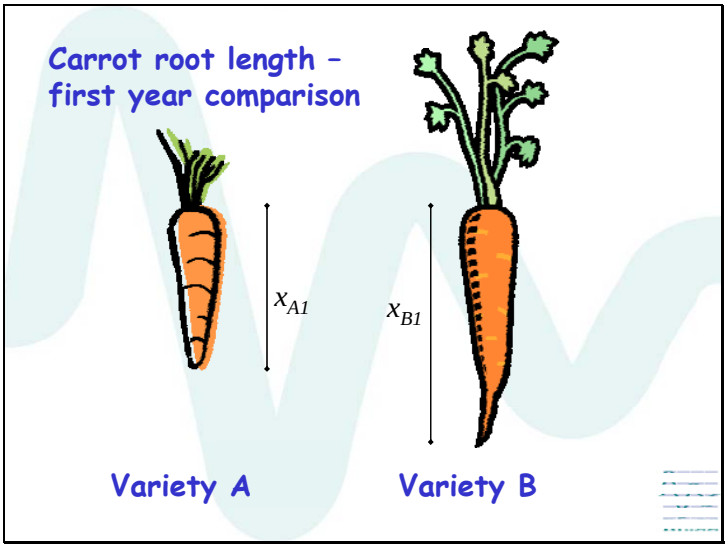
If the reference variety is likely to be found distinct from the candidate in any character then eliminate it.

B. Multivariate/Index

Construct an index to compare the two varieties over a number of characters

I follow approach A here, attempting to link it with the method used for final distinctness assessment – specifically COYD.

Slide 5




Slide 6

How different should $x_{A1}$ and $x_{B1}$ be before we say that A and B are distinct after only one year of tests?

- Set a threshold based on experience?
  - ad hoc – may not properly account for yearly variation
  - May be suitable for some characters
- Set a threshold based on historical data and using statistical methodology?
  - More suitable for quantitative characters?
  - Attempting this here

BioSS

Slide 7

## The approach

To estimate the probability $p_D$ that a candidate is distinct from a reference variety on the 2-year COYD criterion using results from the first year of test.

- If this probability is high then we might declare the varieties distinct and eliminate the reference variety from further comparisons with the candidate
- Looking at it from a more practical point of view, we calculate a threshold $D_1$ to achieve a preset probability $p_D$
- Applied character by character

Slide 8

### Last year

- Showed how to calculate the threshold given various assumptions
- TWC/24/9

### This year

- Tested on a field pea example

Here I concentrate on the example – see TWC/25/14 for details on methodology

Slide 9

## The COYD criterion

Varieties *A* and *B* , year *j* (1 or 2), $x_{ij}$ measurement in character of interest.

$$d_j = x_{Aj} - x_{Bj}$$

Slide 10

## The COYD criterion

COYD criterion says *A* and *B* distinct if:

$$\left| \frac{d_1 + d_2}{2 s_{12}} \right| \geq t_{1-p/2, \nu_{12}}$$

where $s_{12}$ is estimated residual std dev for COYD analysis, p is COYD significance level (e.g. 2%), $\nu_{12}$ is degrees of freedom.

Note $d_2$ and $s_{12}$ are unknown

Slide 11

## Calculation of $p_D$

$$p_D = \Pr\left( \left| \frac{d_1 + d_2}{2s_{12}} \right| \geq t_{1-p/2,\nu_{12}} \text{ given } d_1, s \right)$$

where $d_1$ is obtained from the first year results and $s$ is the long-term residual variance from historical data

Slide 12

## Assumptions

Currently I am making the following assumptions when calculating $p_D$ or $D_1$

- Measurements $x_{ij}$ are normally distributed
- Degrees of freedom for residual in two year COYD analysis and historical data are large (so that student t distributions can be approximated by normal distributions)
- Residual variance $\sigma$ is applicable all years and varieties

Slide 13

## Calculation of threshold for $d_1$

- If we set $p_D$ then we can calculate a threshold $D_1$ for the first year difference $d_1$
- Given our assumptions:

$$D_1 \approx \Phi^{-1}(1-{}^{p}\!/_{2}).s + \Phi^{-1}(p_D).{}^{s}\!/_{\sqrt{2}}$$

- Where $\Phi^{-1}$ is the inverse of the standard normal cumulative distribution function
- Note that this includes a further minor approximation for simplification (otherwise iterations are needed)

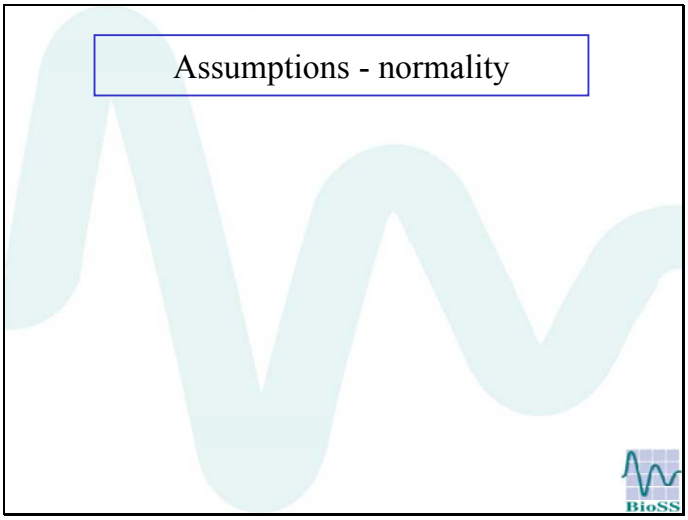
Slide 14

## Example

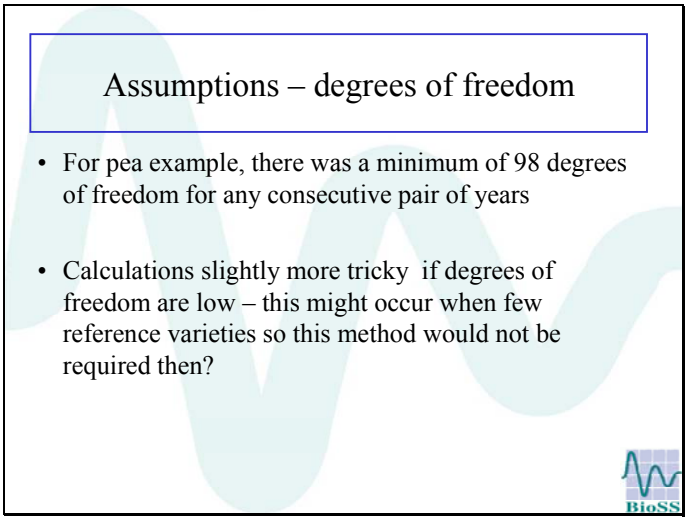
- Field pea at SASA, UK
- Stipule length
- Historical data 1995-2004
- COYD at 2%
- Range 45mm–121mm
- Long-term two-year COYD criterion is **10.64** (LSD[illegible]s 15.04mm after 1 year)

| $p_D$ | $d_1$ threshold |
|---|---|
| 99.9% | ±20.63 |
| 99% | ±18.16 |
| 98% | ±17.28 |
| 95% | ±15.95 |
| 90% | ±14.78 |
| 80% | ±13.36 |
| 50% | ±10.64 |

Slide 15




Slide 16



Slide 17

## Assumptions - consistency of GxE variance across years and varieties

In TWC/25/14, special REML analysis done for the 10 year data set to examine this

- Variances seemed reasonable consistent
- However used only varieties that were present in all years – a small fraction of total – conclusions not necessarily valid for larger data set
- Analysis of whole data set more difficult because very unbalanced – need to investigate if possible and how

Slide 18

## Application in retrospect

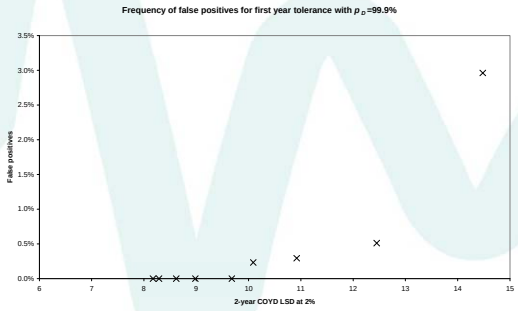
- Apply thresholds calculated on 1995-2004 data set to pairs of consecutive years: 1995-6, 1996-7 ….
- Compare candidates to reference varieties
- See how predicted COYD decision based on only first year matches two-year COYD decision

Slide 19

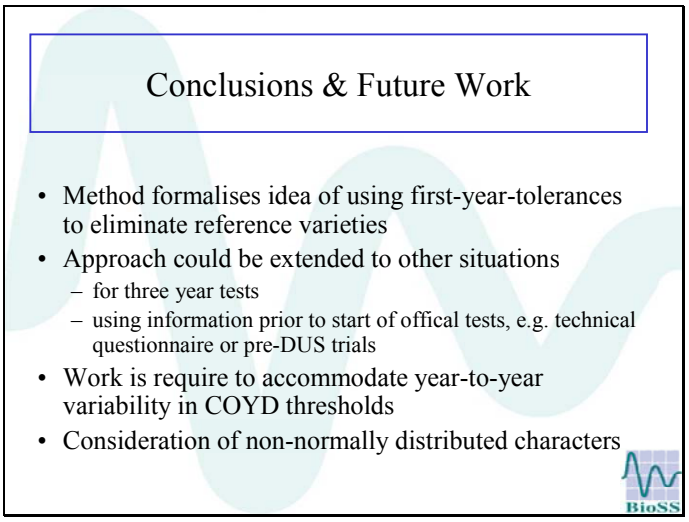
| $p_D$ | *Frequency of correct predictions of COYD decision (no. of decisions)* | |
|---|---|---|
| | *ND COYD (16751)* | *D COYD (7167)* |
| 99.9% | 99.3% | 18.5% |
| 99% | 98.3% | 28.6% |
| 98% | 97.8% | 32.8% |
| 95% | 96.7% | 39.7% |
| 90% | 95.4% | 46.5% |
| 80% | 93.1% | 54.9% |
| 50% | 84.9% | 73.3% |

Slide 20

## Year-to-year variation



Slide 21




[End of document]