E

TWC/31/15 Add.
ORIGINAL: English
DATE: May 29, 2013

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

Geneva

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Thirty-First Session
Seoul, Republic of Korea, June 4 to 7, 2013

ADDENDUM

METHOD OF CALCULATION OF COYU

Document prepared by experts from the United Kingdom and Denmark

The Annex to this document contains a copy of a presentation on method of calculation of COYU that will be made at the Technical Working Party on Automation and Computer Programs (TWC), at its thirty-first session.

[Annex follows]

Method of Calculation of COYU

Adrian Roberts & Kristian Kristensen

BioSS

FACULTY OF AGRICULTURAL SCIENCES
AARHUS UNIVERSITY

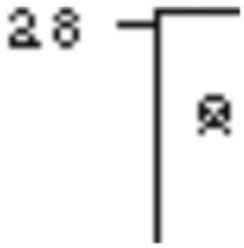
Background

- COYU is established method for assessing uniformity for measured characteristics (MS)
- TWC/26/17 "Some Consequences of Reducing the Number of Plants Observed in the Assessment of Quantitative Characteristics of Reference Varieties"
  - Demonstrated that current COYU is too lax
  - Fails more varieties than should

Background

Problem is due to the method of adjustment used

- Variability of measurements for a characteristic can depend on the level of expression
- COYU uses a 9-point moving-average adjustment method to adjust log (SD+1) based on the mean



3

Background

- In TWC/27/15, TWC/28/27 & TWC/29/10 alternative methods of adjustment were considered, including:
  - Linear regression
  - Quadratic regression
  - Cubic smoothing splines
- It was concluded that smoothing splines might provide the best approach

4

Background

- At the 29th meeting of the TWC, the TWC agreed that a new document based on the cubic spline model should be prepared for the next session of the TWC
- At the 48th meeting of the TC, TC agreed to request the TWC to continue its work with the aim of developing recommendations to the TC concerning the proposals to address the bias in the present method of calculation of COYU

In this paper

- Proposal for an improved method of calculation of COYU
  - Minimal bias
  - Uses cubic smoothing splines
  - Programmed initially in R
- Performance checked by simulation and demonstrated on real data
- The way forward?

COYU in brief

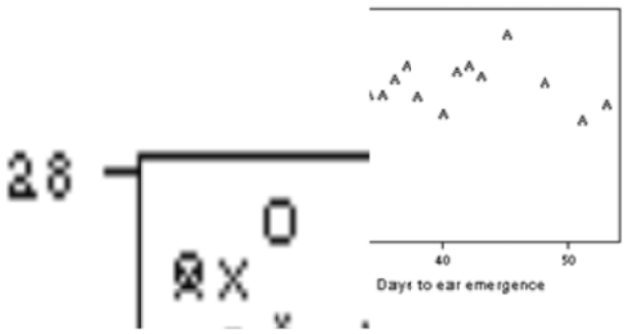
1. Calculation of within-plot SDs for each variety in each year.
2. Transformation of SDs by adding 1 and converting to natural logarithms.
3. Estimation of the relationship between the SD and mean in each year. The method used is based on moving averages of the log SDs of reference varieties ordered by their means.
4. Adjustments of log SDs of candidate and reference varieties based on the estimated relationships between SD and mean in each year.
5. Averaging of adjusted log SDs over years.
6. Calculation of the maximum allowable SD (the uniformity criterion). This uses an estimate of the variability in the uniformity of reference varieties derived from analysis of variance of the variety-by-year table of adjusted log SDs.
7. Comparison of the adjusted log SDs of candidate varieties with the maximum allowable SD.

Moving average adjustment

Order Y_i=log(SD_i+1) by X_i value to get $Y_{(i)}$

Trend value, T_i, is mean of 9 trend values $T_{(i-4)}$ to $T_{(i+4)}$

Adjusted value for $i = Y_i - T_i + \bar{Y}$



Splines

- Moving average is an example of ***smoothing***
- Another method is ***cubic smoothing splines***
- Idea is to fit a smooth curve to a set of points
- Degree of smoothing can be controlled
  - Less smooth – fits closer to the points but is more wiggly
  - Moving average controlled by number of points used for means (9 for COYU)
  - Smoothing splines controlled by a smoothing factor
    - equivalent to degrees of freedom used
    - Smoother ↔ fewer degrees of freedom

Splines

- Why use cubic splines?
  - Mathematical properties well established
  - Can be used within regression
  - Widely used
- Fixed degrees of freedom
  - Facilitates full automation
  - Low number should fit most realistic relationships (TWC/29/10)
- COYU thresholds for adjusted log SDs
  - Vary according to mean – larger at the extremes of expression
  - Current COYU method uses constant threshold (part of its problem)
  - issue re extrapolation

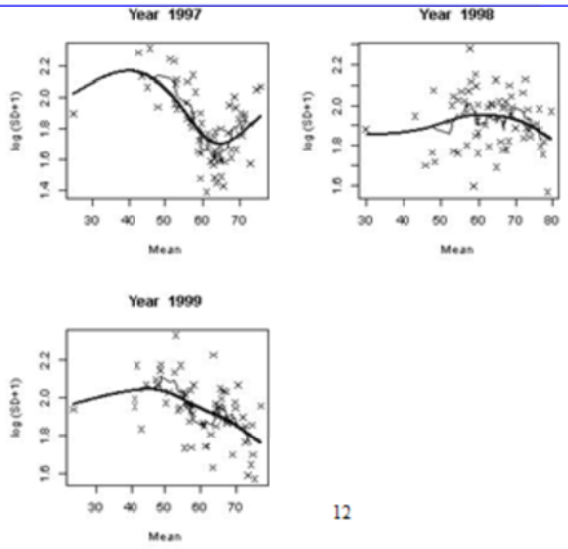
Example on real data

Lolium perenne

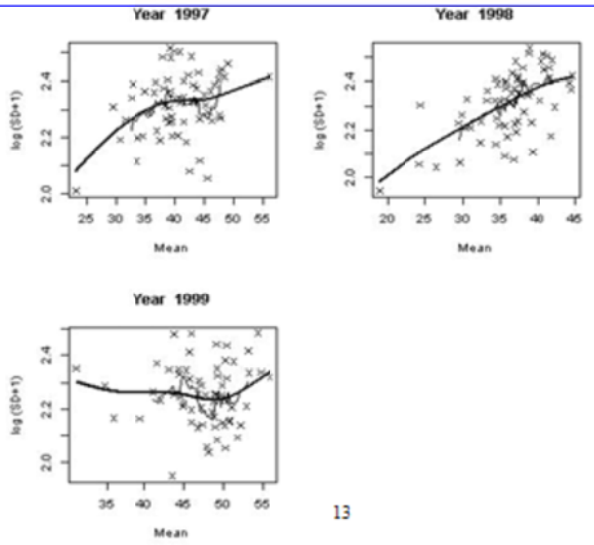
- Three years
- 63 reference varieties; two candidate varieties
- Characteristic 8:
  - Time of inflorescence emergence in 2nd year
- Characteristic 9:
  - Plant: natural height at inflorescence emergence

Thanks to the UK Herbage DUS centre

Characteristic 8



Characteristic 9



13

31/15, Seoul 2013

Candidate	Characteristic 8 A	Characteristic 8 B	Characteristic 9 A	Characteristic 9 B
Mean	48.36	67.71	45.83	42.41
logSD	2.03	1.97	2.34	2.27
Existing COYU				
Adjusted logSD	1.90	1.99	2.32	2.25
Threshold with α=0.001	2.13	2.13	2.49	2.49
Uniform with α=0.001?	Yes	Yes	Yes	Yes
COYU with Spline (4 df)				
Adjusted logSD	1.90	2.01	2.31	2.26
Threshold with α=0.05	2.03	2.03	2.40	2.40
Uniform with α=0.05?	Yes	Yes	Yes	Yes
Threshold with α=0.01	2.09	2.09	2.45	2.45
Uniform with α=0.01?	Yes	Yes	Yes	Yes
p-value	0.438	0.071	0.392	0.699

14

TWC/31/15, Seoul 2013

Simulations

We have carried out extensive simulations to test the performance of alternative methods:

- In previous papers and this paper, we looked at eight scenarios varying:
  - the number of reference varieties (10 or 50)
  - the absence/presence of a linear relationship between SD and mean
  - The size of the variety-by-year interaction
- Also in this paper, we looked at different underlying relationships between logSD and mean
  - Linear
  - Quadratic
  - Sine function
- Looked at rejection rate (coverage) compared to acceptance probability

Simulations

Conclusions from simulations and previous results:

- Spline method with 4 degrees of freedom (and Bayesian standard errors) worked best
  - 4 degrees of freedom better than 3
  - Not perfect for sinusoidal:
    - Underfits
    - But works fine with real data (see TWC/29/10)
- For more detail, see TWC/31/15

Acceptance probability

- For the existing COYU method, very small acceptance probabilities (α) are typically used
  - e.g. 0.1% or 0.2%
  - Perhaps to accommodate bias?
- For proposed method using splines, bias is greatly reduced
  - Can adopt more commonly used values e.g. 0.05 or 0.01

17

TWC/31/15, Seoul 2013

Acceptance probability

How to select α for new method?

a) For each crop (& country), match to results for existing COYU so decisions similar
   - quite a bit of work

b) Same for all crops (of a type) based on principles
   - Harmonisation
   - Potential for change in toughness of uniformity criterion

18

TWC/31/15, Seoul 2013

Software considerations

- R functions written for COYU with splines
  - R is a free but powerful statistical programming language
- It should be straightforward to write FORTRAN code
  - Code is available for the difficult bits
- FORTRAN code could then be used in DUST and potentially with SAS and other packages
- See paper for more – we are happy to discuss further

Conclusions

- Proposal for new version of COYU using splines
  - Addresses bias issue
  - Performs better than existing method
  - Should be possible to produce software (e.g. DUST)
- Details over method
  - Cubic smoothing splines (natural)
  - Method of standard error calculation → Bayesian
  - Fixed level of smoothing – 4 degrees of freedom

Issues to be dealt with

- Acceptance probability
- Extrapolation
  - Issue for both current and proposed methods
  - What if the level of expression for the candidate is outside that of the reference varieties?
- Software

Going ahead

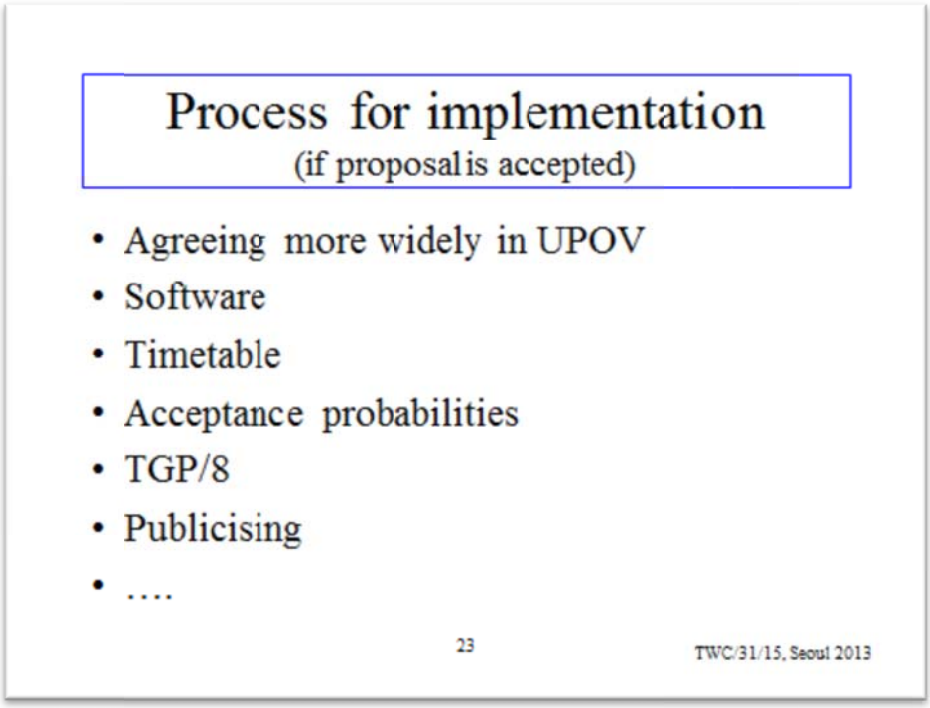
From TWC/31/15:

The TWC is invited to:

a) note the information on development of COYU provided in this document;

b) consider whether COYU method should be modified to use splines, as set out in paragraph 58 of this document; and

c) Consider the possibility to write software for COYU in FORTRAN that could then be integrated into DUST, as set out in paragraph 59 of this document.



[End of Annex and of document]