UPOV

International Union for the Protection of New Varieties of Plants

E

Technical Working Party on Automation and Computer Programs **TWC/35/6 Add.**

Thirty-Fifth Session
Buenos Aires, Argentina, November 14 to 17, 2017

Original: English
Date: November 7, 2017

ADDENDUM TO METHOD OF CALCULATION OF COYU: PRACTICAL EXERCISE, PROBABILITY LEVELS, EXTRAPOLATION & SOFTWARE

Document prepared by an expert from the United Kingdom

Disclaimer: this document does not represent UPOV policies or guidance

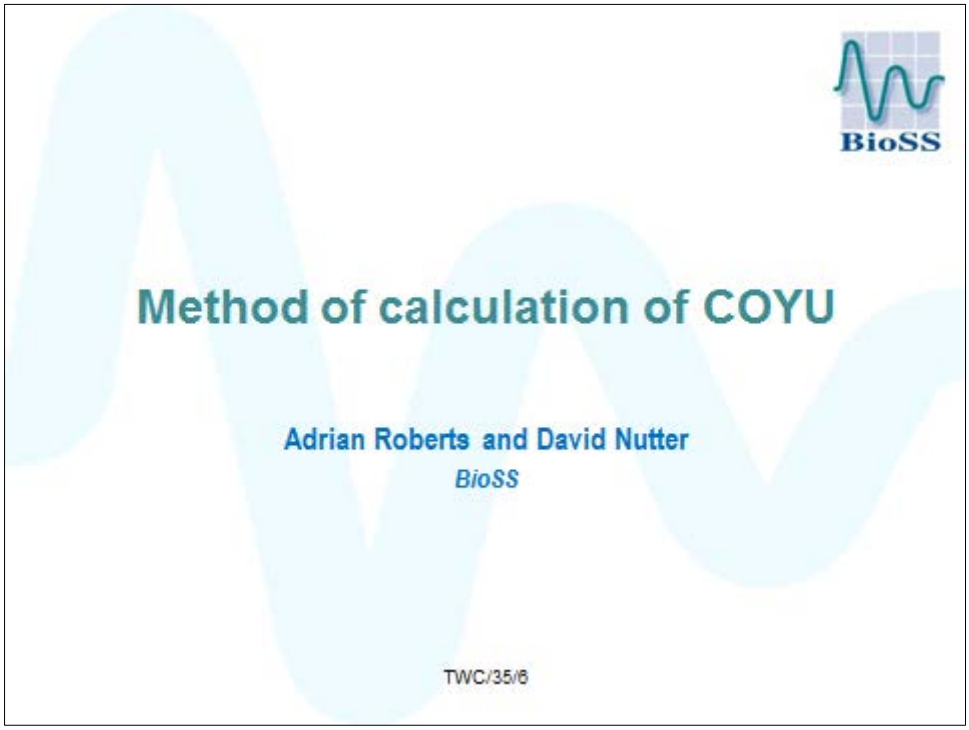
The Annex to this document contains a copy of a presentation on "Method of calculation of COYU: practical exercise, probability levels, extrapolation & software", prepared by an expert from the United Kingdom, to be made at the thirty-fifth session of the Technical Working Party on Automation and Computer Programs (TWC).

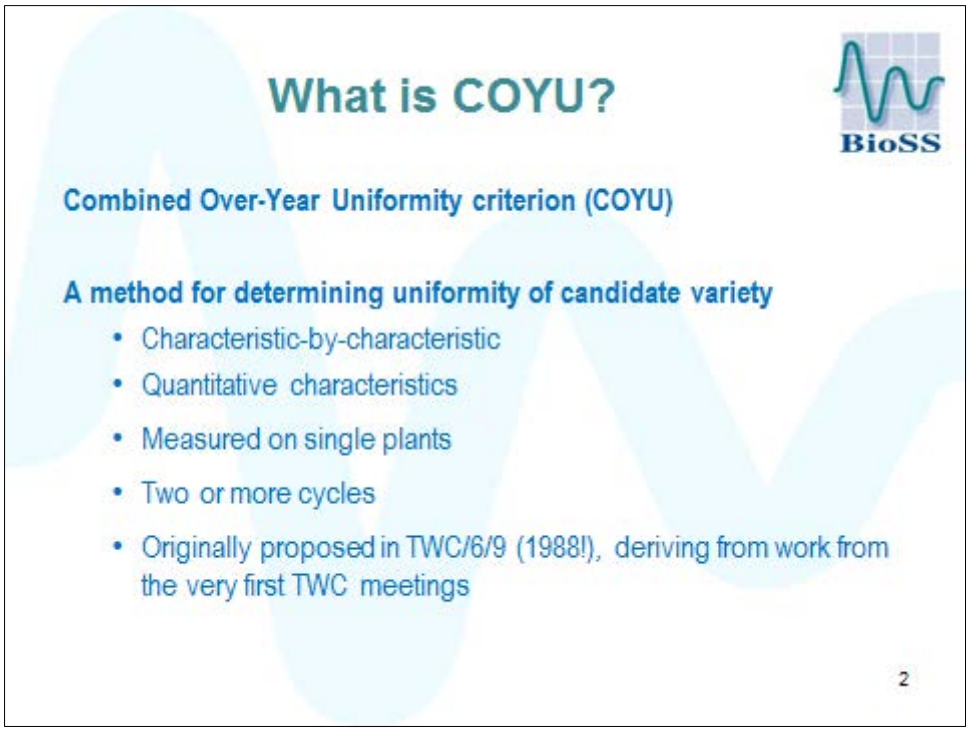
[Annex follows]

ANNEX

METHOD OF CALCULATION OF COYU: PRACTICAL EXERCISE, PROBABILITY LEVELS, EXTRAPOLATION & SOFTWARE

Presentation prepared by an expert from the United Kingdom





COYU key concepts

BioSS

Compares uniformity with similar varieties

Measures uniformity through standard deviation (SD) of measurements within plots

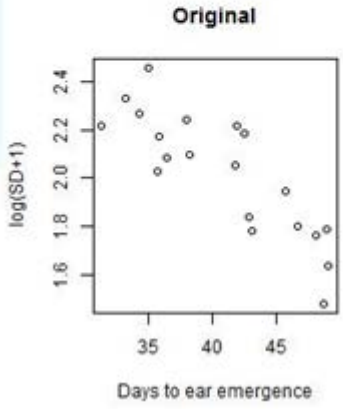
- Log (SD+1)

Adjust for any relationship between variability (SD) and level of expression

3

Relationship between uniformity & mean

BioSS

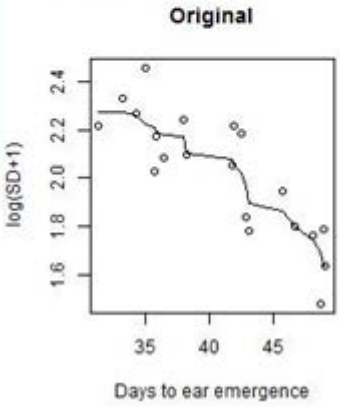


4

Relationship between uniformity & mean

BioSS

Moving average adjustment

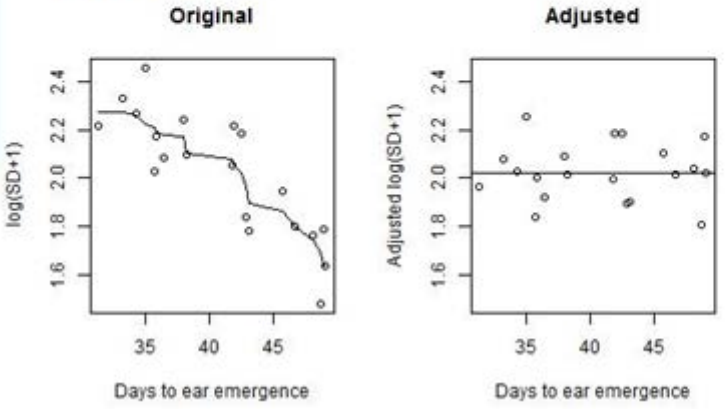


5

Relationship between uniformity & mean

BioSS

Moving average adjustment



6

Relationship between uniformity & mean

Currently adjustment uses moving average method

Found that current COYU method tends to find more varieties non-uniform than desirable

7

More recently

New method proposed and tested

- Uses spline method instead of moving average
- Software developed in R and DUST

Practical Exercise

- Data from Finland, France, Kenya, United Kingdom
- Reported on previously - TWC/33/16

TWC/35/6

Activities in 2016-7

BioSS

New data received from Denmark and Slovakia

Identification of suitable probability levels

Thoughts on extrapolation

Update on software

TWC/35/6

Practical Exercise

BioSS

Country	Participant	Software/Data	Crops
Denmark	Erik Lawaetz	Data	Oilseed rape
Finland	Sami Markannen	DUST	Timothy, meadow fescue, tall fescue, Canarian reed grass, red clover, white clover, turnip rape
France	Christophe Chevalier	R	Fescue
Kenya	Abraham Lagat	R	Wheat
Slovakia	Ľubomír Bašta	Data	Red fescue
United Kingdom	Sally Watson	DUST	Perennial ryegrass
United Kingdom	Haidee Philpott	DUST	Oilseed rape
United Kingdom	Tom Christie	DUST	Field pea

(Germany – currently use SAS for COYD and COYU)

TWC/35/6

Practical Exercise

BioSS

Country	Participant	Software/Data	Crops
Denmark	Erik Lawaetz	Data	Oilseed rape
Finland	Sami Markannen	DUST	Timothy, meadow fescue, tall fescue, Canarian reed grass, red clover, white clover, turnip rape
France	Christophe Chevalier	R	Fescue
Kenya	Abraham Lagat	R	Wheat
Slovakia	Ľubomír Bašta	Data	Red fescue
United Kingdom	Sally Watson	DUST	Perennial ryegrass
United Kingdom	Haidee Philpott	DUST	Oilseed rape
United Kingdom	Tom Christie	DUST	Field pea

(Germany – currently use SAS for COYD and COYU)

THANKS !

TWC/35/6

Country	Crop	Number of sets of years	Number of years for each set	Probability level for COYU	Number of characters	Overall number of candidates
Denmark	Oilseed rape	3	2	0.001	15	570
Finland	Timothy	1	2	0.001	6	3
Finland	Timothy	2	3	0.001	1-7	6
Finland	Meadow fescue	1	2	0.001	6	2
Finland	Meadow fescue	1	3	0.001	6	2
Finland	Tall fescue	1	2	0.001	6	1
Finland	Canarian reed grass	1	3	0.001	8	1
Finland	Red clover	2	2	0.001	6	1
Finland	Red clover	2	3	0.001	7	1
Finland	White clover	1	2	0.001	9	1
Finland	White clover	1	3	0.001	9	1
Finland	Turnip rape	1	2	0.001	8	3
Finland	Turnip rape	1	3	0.001	8	1
France	Fescue	1	2	0.001	11	4
Kenya	Wheat	1	2	?	3	2
Slovakia	Red fescue	9	2	0.001	4-7	57
United Kingdom	Perennial ryegrass	2	3	0.001	30	46
United Kingdom	Winter oilseed rape	1	2	0.001	12	128
United Kingdom	Pea	5	2	0.001	13-19	47

Identifying suitable probability levels

Probability level needs to be set for decision-making with COYU

Typically 0.001 is used with current method for final decisions

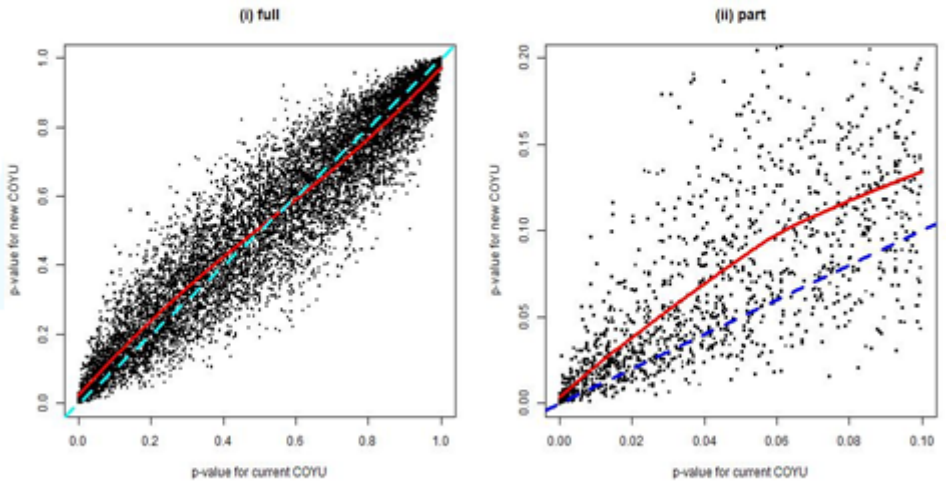
Larger probability used ⇨ fewer varieties found uniform

Rather than compare decisions made with different probability levels, better to compare probability values directly

- Each candidate has a probability value from COYU calculation
- When smaller than set probability level ⇨ non-uniform

Probability level

Example: Denmark oilseed rape (all candidates and years)



Probability level

BioSS

Approx equivalent probability levels to current COYU at 0.001

- Log-log additive models – see text for detail

Data set	Number of cases	Equivalent p-value	95% confidence interval	
			lower	upper
Overall	1756	0.00288	0.00258	0.00322
Denmark	1269	0.00284	0.00255	0.00317
Finland	15	0.00629	0.00450	0.00879
France	5	0.00293	0.00167	0.00514
Kenya	2	0.00372	0.00154	0.00898
Slovakia	51	0.00223	0.00182	0.00274
GB oilseed	137	0.00339	0.00293	0.00394
GB pea	92	0.00256	0.00217	0.00303
GB ryegrass	185	0.00273	0.00237	0.00315

Probabi 0.003?

BioSS

Approx equivalent probability levels to current COYU at 0.001

- Log-log additive models – see text for detail

Data set	Number of cases	Equivalent p-value	95% confidence interval	
			lower	upper
Overall	1756	0.00288	0.00258	0.00322
Denmark	1269	0.00284	0.00255	0.00317
Finland	15	0.00629	0.00450	0.00879
France	5	0.00293	0.00167	0.00514
Kenya	2	0.00372	0.00154	0.00898
Slovakia	51	0.00223	0.00182	0.00274
GB oilseed	137	0.00339	0.00293	0.00394
GB pea	92	0.00256	0.00217	0.00303
GB ryegrass	185	0.00273	0.00237	0.00315

Probability level

BioSS

Approx equivalent probability levels to current COYU at 0.01

- Early decisions to accept eg at year 2 of 3

Data set	Number of cases	Equivalent p-value	95% confidence interval	
			lower	upper
Overall	1674	0.0188	0.0179	0.0198
Denmark	1200	0.0186	0.0177	0.0196
Finland	15	0.0410	0.0302	0.0557
France	4	0.0171	0.0095	0.0309
Kenya	2	0.0245	0.0106	0.0566
Slovakia	47	0.0141	0.0119	0.0169
GB oilseed	133	0.0226	0.0203	0.0251
GB pea	90	0.0171	0.0150	0.0195
GB ryegrass	183	0.0180	0.0164	0.0198

Probabi 0.02?

BioSS

Approx equivalent probability levels to current COYU at 0.01

- Early decisions to accept eg at year 2 of 3

Data set	Number of cases	Equivalent p-value	95% confidence interval	
			lower	upper
Overall	1674	0.0188	0.0179	0.0198
Denmark	1200	0.0186	0.0177	0.0196
Finland	15	0.0410	0.0302	0.0557
France	4	0.0171	0.0095	0.0309
Kenya	2	0.0245	0.0106	0.0566
Slovakia	47	0.0141	0.0119	0.0169
GB oilseed	133	0.0226	0.0203	0.0251
GB pea	90	0.0171	0.0150	0.0195
GB ryegrass	183	0.0180	0.0164	0.0198

Extrapolation

Extrapolation is when the candidate has a mean value outside those of the reference varieties

- In any cycle
- Important because COYU is based on the relationship between mean and variability (uniformity)
  - For both current and proposed methods of COYU
- Need to agree what to do in cases of extrapolation

TWC/35/6

Defining Extrapolation?

Consider each cycle separately

Direct method:

- If inside range of reference varieties ⇨ 0
- Otherwise, difference in level for candidate compared to the most similar reference variety divided by range of reference varieties

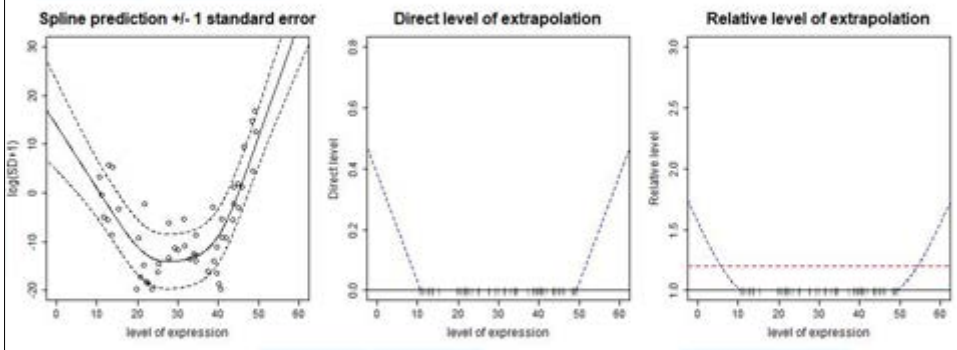
$$\frac{\min(x - x_{min}, x - x_{max})}{(x_{max} - x_{min})}$$

Relative method:

- If inside range of reference varieties ⇨ 0
- Otherwise, ratio of prediction standard error for candidate over that for the nearest comparable variety

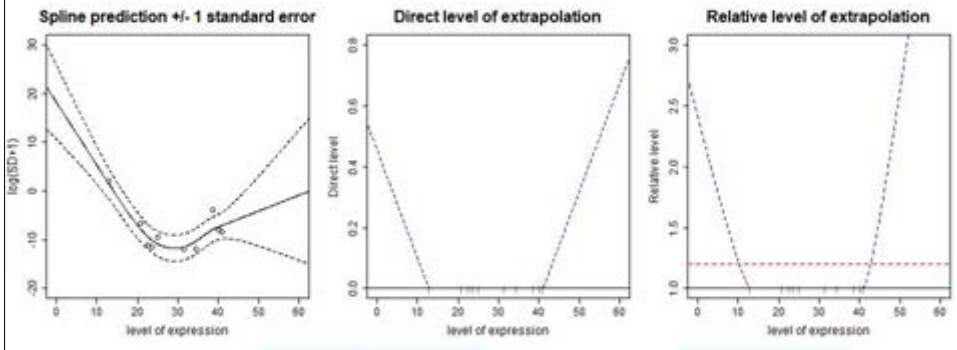
TWC/35/6

Illustration
50 reference varieties



TWC/35/6

Illustration
10 reference varieties



TWC/35/6

Defining Extrapolation

BioSS

Consider each cycle separately

Direct method:

- Does not depend on number of observations
- Does not relate to effect of extrapolation on COYU

Relative method:

- Incorporates effect of extrapolation on COYU
- More observations, safer to go further away from reference varieties

TWC/35/6

Proposal

BioSS

Use relative method

When relative level of extrapolation exceeds a threshold, expert judgement required

Threshold? 120%?

TWC/35/6

Extrapolation (Direct)

BioSS

Country	Data set	Number of cases	Frequency of extrapolation	Cases > 10% extrapolation	Cases > 20% extrapolation
Denmark	Oilseed rape	11,910	2%	0.8%	0.3%
France	Fescue	36	0%	n/a	n/a
Kenya	Wheat	6	50%	n/a	n/a
Finland	Various	137	19%	n/a	n/a
GB	Perennial ryegrass	1,381	13%	7%	4%
GB	Winter oilseed rape	1,536	1%	0.2%	0%
GB	Pea	698	8%	4%	2%
Slovakia	Red fescue	738	20%	15%	7%

Software

BioSS

R: package –source code also available on GitHub

DUST module – accesses R package

Software

BioSS

To do:

- Improve installation with DUST
- Improve error messages
- Ensuring that problematic data sets can be dealt with appropriately
- Produce extrapolation flags according to approach agreed by TWC
- Ensuring that the algorithm works well for unbalanced data (for cyclic planting).

TWC/35/6

Conclusions

BioSS

Higher probability levels required to most closely match decisions for current COYU

- Practical exercise suggests:
  - probability levels 0.003 to match 0.001 for current COYU
  - probability levels 0.02 to match 0.01 for current COYU

Extrapolation

- TWC needs to agree approach for identifying cases of problematic extrapolation
- Cases of major extrapolation should be considered by crop expert

Software

- Areas for improvement identified

TWC/35/6

[End of Annex and of document]