TC-EDC/Jan14/10
ORIGINAL: English
DATE: December 13, 2013

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

Geneva

ENLARGED EDITORIAL COMMITTEE

Geneva, January 8 and 9, 2014

REVISION OF DOCUMENT TGP/8: PART II: SELECTED TECHNIQUES USED IN DUS EXAMINATION, SECTION 3: METHOD OF CALCULATION OF COYU

Document prepared by the Office of the Union

Disclaimer: this document does not represent UPOV policies or guidance

1.   The purpose of this document is to report on developments concerning the method of calculation of COYU.

BACKGROUND

2.   At its twenty-sixth session held in Jeju, Republic of Korea, from September 2 to 5, 2008, the Technical Working Party on Automation and Computer Programs (TWC) considered document TWC/26/17 "Some consequences of reducing the number of plants observed in the assessment of quantitative characteristics of reference varieties[1]" and a presentation by Mr. Kristian Kristensen (Denmark), a copy of which is reproduced as document TWC/26/17 Add.

3.   Document TWC/26/17 states the following with regard to the current method of calculation of the Combined-Over-Years Uniformity Criterion (COYU):

"Conclusions

"18.   From the above it can be concluded that the variances calculated in the present system do not reflect the expected value of the true variance as they are too small, partly because the expected value of RMS [residual mean square] from the ANOVA is less than the expected value of $Var(Y_v)$ and partly because only the number of varieties used in the local adjustment influence[s] this variance (and not the total number of reference varieties). However, the present method probably adjusts for this bias by using a large t-value (by using a small α-value). Also it can be concluded that the residual mean square (RMS) may depend significantly on the number of observations recorded as the component of RMS that depends on the number of observations (degrees of freedom) was not a negligible part."

4.   The TWC noted the following possible actions to address the bias in the present method of calculation of COYU, as identified and commented on by Mr. Kristensen:

---

[1]   The term "reference varieties" here refers to established varieties which have been included in the growing trial and which have comparable expression of the characteristics under investigation.

(i)   Ignore the biases
        (comment: the test will most probably be too liberal);

(ii)   Correct only for the bias introduced by the smaller sample sizes
        (comment: the test will be too liberal, but will be comparable to those in the past);

(iii)   Correct only for the present bias
        (comment: the test will be conservative, but not comparable to the past);

(iv)   Correct for all biases
        (comment: there will be no biases, but the tests will not be comparable to the past).

5.   The TWC agreed that Denmark and the United Kingdom should prepare a new document, including a simulation using the smoothing spline method. It was noted that that would also allow experts further time to reflect on the situation and possible ways forward.

6.   The Technical Committee (TC), at its forty-fifth session, held in Geneva from March 30 to April 1, 2009, requested the TWC to make its recommendations to the TC concerning the proposals set out in paragraph 3 of this document.

7.   The developments between 2009 and 2012 are reported in document TC/49/11 "Method of calculation of COYU", paragraphs 6 to 17 (available at: http://upov.int/meetings/en/details.jsp?meeting_id=28343).

DEVELOPMENTS IN 2013

8.   The TC, at its forty-ninth session, held in Geneva, from March 18 to 20, 2013, 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 and noted that a document on possible proposals for improvements to COYU would be prepared for the TWC session in 2013 (see document TC/49/41 "Report on the Conclusions, paragraph 113).

9.   At their sessions in 2013, the TWO, TWF, TWV and TWA considered documents TWO/46/15, TWF/44/15, TWV/47/15 and TWA/42/15, respectively.

10.   The TWO TWF, TWV and TWA noted that:

        (a)   the TC had requested 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; and

        (b)   a document on possible proposals for improvements to COYU would be prepared for the TWC session in 2013 (see documents TWO/46/29 "Report", paragraph 34, document TWF/44/31 "Report", paragraph 37, document TWV/47/34 "Report", paragraph 37, and document TWA/42/31 "Report", paragraph 37, respectively).

11.   At its thirty-first session, held in Seoul, from June 4 to 7, 2013, the TWC considered document TWC/31/15 Corr. including the proposals for improvements to COYU prepared by experts from the United Kingdom and Denmark, which is attached as Annex I to this document, and received a presentation from an expert from the United Kingdom, as contained in document TWC/31/15 Add. and reproduced as Annex II to this document.

12.   The TWC noted that the present method of calculation of COYU was overly strict due to the method of smoothing used and that very low probability levels were used in compensation (e.g. p=0.1%). The TWC agreed that the bias in the present method of calculation of COYU could be addressed by a change of smoothing method from "moving average" to "cubic smoothing splines" (see document TWC/31/32 "Report", paragraph 91).

13.   The TWC welcomed the offer by the experts from the United Kingdom to write software for the proposed COYU method in FORTRAN for integration into the DUST software and to present a demonstration version of the DUST software using the proposed COYU method at the thirty-second session of the TWC (see document TWC/31/32 "Report", paragraph 92).

14.   The TWC agreed that the probability levels to be used in the proposed COYU method should be discussed on the basis of the experience of UPOV members in using the proposed method (see document TWC/31/32 "Report", paragraph 93).

15.   The TWC agreed that a circular should be prepared by an expert from the United Kingdom and issued by the Office to the TC representatives, in order to investigate which members of the Union used the current COYU method and in which software it was used (see document TWC/31/32 "Report", paragraph 94).

16.   The TWC agreed that the document containing the proposal for an improvement to COYU should be summarized by an expert from the United Kingdom and presented to the TC at its fiftieth session and the TWP sessions to be held in 2014. The document should explain the bias of the present method that justified the proposed changes (see document TWC/31/32 "Report", paragraph 95).

17.   The TWC agreed that guidance should be developed on the minimum number of varieties that would be suitable for using the COYU method (see document TWC/31/32 "Report", paragraph 96).

18.   The TWA supported the continuation of work of the TWC to improve the COYU method and noted that the TWC would provide information on the proposed changes to the COYU method and eventual consequences in DUS examination (see document TWA/42/31 "Report", paragraph 38).

RECENT DEVELOPMENTS

19.   On November 4, 2013, the Office of the Union issued Circular E_13/268 to the designated persons of the members of the Union in the TC, inviting them to provide information on which members of the Union used the current COYU method and in which software it was used (see paragraph 15 of this document).

20.   Annex III to this document contains the questionnaire and the results of the survey.

21.   The proposal for an improvement to COYU prepared by an expert from the United Kingdom for presentation to the TC at its fiftieth session is attached as Annex IV to this document (see paragraph 16 of this document).

> *22.   The TC-EDC is invited to note the information in this document to be presented to the TC and propose any improvements to the document in that regard.*

[Annexes follow]

ANNEX I

PROPOSALS FOR IMPROVEMENTS TO COYU

Prepared by experts from the United Kingdom and Denmark
and presented at the thirty-first session of the TWC, held in Seoul, from June 4 to 7, 2013

Introduction

We report on progress in the development of an improved version of COYU. In particular, we have investigated the performance and practicality of an approach using cubic smoothing splines.

The existing COYU procedure is described in TGP/8/1 Part II. 9. Briefly it compares the uniformity of candidate varieties to that of reference varieties. Uniformity is represented by the standard deviation (SD) of the measurements on individual plants within a plot. The SDs are transformed by natural logarithms after adding 1. Often there is a relationship between variability of measurement and the level of expression of the character. The COYU method uses a moving-average method to estimate and adjust for any such relationship. As revealed in the previous papers described above, this method of adjustment produces an inherent bias in the COYU thresholds; in practice this is compensated for by using smaller p-values than usual.

In the previous papers, we have considered different methods of adjusting for the relationship between variability and level of expression. These included linear regression, quadratic regression and smoothing splines. In TWC/29/22, we showed that smoothing splines performed best at fitting real data. The cubic smoothing spline method fitted data at least as well as these methods without being so sensitive to unusual observations.

In previous work, we noted that it is preferable to calculate individual COYU threshold values for each candidate. This is because, whatever the adjustment method, there is more confidence in the fit of the curve for varieties with average levels of expression than those with more extreme levels. We have found previously that use of a single threshold for all candidate tends to lead to more varieties being rejected than desired with a given probability level, particularly when there are few reference varieties.

We have thus pursued further the idea of replacing the moving-average adjustment in COYU by one based on a cubic spline. We have implemented a revised COYU method in R (a free and powerful statistical programming package) and tested it using simulated and real data sets. We have considered issues in respect of implementation, with some initial thoughts on software and probability levels.

What is smoothing?

Smoothing is a commonly used procedure for fitting a relationship when the form of the relationship is unknown. This illustrated in Figure 1.
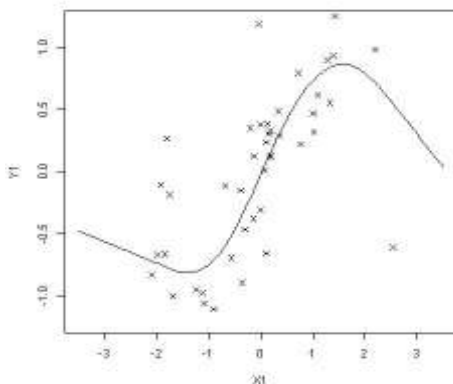
Figure 1: Example of cubic smoothing spline (with 4 degrees of freedom) fitted to simulated data. Observations are represented by "x" and the smooth fit is represented by the line. There are many different methods of smoothing, include the moving-average method found in the current COYU. Whereas for linear or quadratic regression, a particular form of curve is fitted to the whole data set,

with a smoothing method the fit at a certain point depends more on the observations that are around that point. Usually the degree of smoothing can be controlled through a parameter. Note that smoother fits correspond to use of few degrees of freedom.

Smoothing methods are described at length in several text books, including Hastie and Tibshirani (1990) and Hastie *et al* (2001).

Why cubic smoothing splines?

As mentioned above, there are many smoothing methods. However one that is commonly used is known as the cubic smoothing spline method. This is described in 5.4 of Hastie *et al* (2001). Its derivation is quite mathematical so we will not reproduce that here. However cubic smoothing splines have some useful properties. They have the following advantages that lead to their selection here for use with COYU:

- Flexibility.
- The degree of smoothness can be controlled directly through the effective degrees of freedom.
- The method uses natural splines (see 5.2.1 of Hastie *et al* (2001)), which have the benefit that the behaviour at the extremes of the data is reasonable compared to some other smoothing methods. In fact here the fit is linear.
- The method is well known and well described, facilitating implementation in different software packages.
- FORTRAN code is available for cubic smoothing splines, making it easier to implement in DUST.

Details on methodology and implementation in R

Functions and procedures for cubic smoothing splines are readily available in various software packages, including:
- SAS – using PROC GAM
- R – various functions available including smooth spline, gam in the gam library, gam in the mgcv library and sreg in the fields library
- GenStat – using the REG directive with the S function.
- FORTRAN

However in the most part, these do not give access to standard errors for the fit of new observations (as opposed to those used to fit the curve). So we have developed the methodology for this below, allowing straightforward implementation, at least in R and FORTRAN.

As indicated in document TWC/28/27 and Büsche *et al* (2007), an ideal approach to COYU might be to carry out a one-step approach. In these two papers, a mixed model was used. However this would introduce extra complexity, making the method harder to implement. Instead, we note that a model with a different smooth curve for each year can equivalently be fitted by fitting curves to the data sets for each year separately (see Hastie and Tibshirani, 1990, section 9.5.2; we have checked this for linear regression). This simplifies the programming considerably.

In smoothing, we adopt the following model for the relationship between a response variable, **y**, (in our case log(SD+1)) and an explanatory variable, **x**, (in our case the trial mean measurement for each variety):

$$y = f(x) + \varepsilon \qquad\qquad (1)$$

where *f* is a smooth function and ε is an error (independent and identically normally distributed, with variance σ^2).

For cubic smoothing splines, it can be shown (see Hastie *et al* 2001, 5.4.1) that the fitted smooth curve is given by:

$$\hat{f} = N(N^T N + \lambda \Omega_{\mathrm{N}})^{-1} N^T y = S_\lambda y \qquad\qquad (2)$$

where λ is a parameter controlling the degree of smoothing, *N* is a natural spline basis based on knots at each of the observations **x** and S_λ is known as the smoother matrix. Note that the effective number of degrees of freedom is given by $\mathrm{trace}(S_\lambda)$ (the sum of the diagonal elements of the smoother matrix).

For each of the observations that are used to fit the smooth (these would be for reference varieties), standard errors can be calculated for the corresponding point of $\hat{f}$. There are two distinct formulations:

a) Classical, given by the diagonal element corresponding to the observation of $S_\lambda S_\lambda{}^T \sigma^2$.

b) Bayesian (Wahba, 1983), given by the diagonal element corresponding to the observation of $S_\lambda \sigma^2$.

These two formulations are discussed in section 3.8.1 of Hastie and Tibshirani (1990). They note that little difference can be found in practice between these two. However we find in practice that, although the standard errors are very similar throughout most of the range of the observations, they start to differ for observations at the outer limits of the range. For extrapolation (see below), they can be very different.

For new observations (i.e. for candidate varieties), the <u>prediction</u> is formulated as follows:

$$\boldsymbol{n_0} N^- S_\lambda \boldsymbol{y} \qquad\qquad\qquad (3)$$

where $\boldsymbol{n_o}$ is the projected basis vector for the new observation and superscript $^-$ denotes a generalized inverse.

For a new observation (i.e. for candidate varieties), the standard error for the <u>prediction</u> are formulated as follows:

a) Classical: $\left(\boldsymbol{n_0} N^- S_\lambda S_\lambda{}^T (N^-)^T \boldsymbol{n_0}^T + 1\right)\sigma^2$. (4a)

b) Bayesian: $(\boldsymbol{n_0} N^- S_\lambda (N^-)^T \boldsymbol{n_0}^T + 1)\sigma^2$. (4b)

Based on the above, we lay out below a basic algorithm for our proposal for an improved COYU procedure. The right hand column indicates the R functions that could be used.

We recognize that some of the calculations might be done in a more computationally efficient manner than indicated in the formulae here. The sparse nature of the matrices involved is likely to help. In particular, the generalized inverse used may mean that data sets with many reference varieties run slowly. One way to reduce the computational cost in such circumstances is to use fewer knots.

Table 1: Algorithm for COYU using cubic smoothing splines

Step	Process	R
1	Calculate within-plot standard deviations and means	
2	Average the within-plot standard deviations $[\rightarrow SD_{ij}]$ and means $[\rightarrow M_{ij}]$ over the plots in a trial to give one for each year (j) and variety (i) combination	
3	Transform the SD_{ij} using the natural logarithm after adding 1 $[\rightarrow logSD_{ij}]$	
4	Divide the data set into two: one for the reference varieties and one for the candidate varieties	
5	For each year, fit a smoothing spline with set degrees of freedom (d=3 or 4) to the reference variety data set – save the smoothing parameter $[\rightarrow \lambda)]$ the set of knots, and the sums of squares of the residuals $[\rightarrow SS_j]$	smooth.spline(x=M,Y=logSD, all.knots = TRUE, df = d)
6	For each year, use this fitted spline to predict the logSDs for both the reference and candidate varieties $[\rightarrow \widehat{logSD}_{ij}]$	predict.smooth.spline
7	For each year, calculate the mean of the logSDs over the reference varieties only $[\rightarrow \overline{logSD_{.j}}]$	
8	For each year, calculate the adjusted logSDs: $\overline{logSD_{.}} + logSD_{ij} - \widehat{logSD}_{ij} [\rightarrow adjlogSD_{ij}]$	
9	For each year, calculate the basis matrix for the reference varieties – this needs the smoothing parameter λ and knots from step5 $[\rightarrow N]$	ns function from splines library
10	For each year, calculate the basis matrix for the candidate varieties – this needs the smoothing parameter λ and knots from step5 $[\rightarrow N_0]$	ns function from splines library
11	For each year (and each candidate variety), calculate a variance factor $(\boldsymbol{N_0}N^- S_\lambda S_\lambda{}^T (N^-)^T \boldsymbol{N_0}{}^T$ for classical or $\boldsymbol{N_0}N^- S_\lambda (N^-)^T \boldsymbol{N_0}{}^T$ for Bayesian) $[\rightarrow f_{ij}]$	ginv function from MASS library
12	Calculate the overall residual degrees of freedom $k(n_r - d)$, where k is the number of years and n_r is the number of reference varieties $[\rightarrow d_r]$	
13	Calculate the estimate of residual error $\frac{\Sigma_j SS_j}{d_r} [\rightarrow \sigma^2]$	
14	Take the mean of the $adjlogSD_{ij}$ values over years and varieties for the reference varieties only $[\rightarrow \overline{adjlogSD_{..}}]$	
15	For the candidate varieties, calculate the mean variance factors over years $[\rightarrow \bar{f}_{i.}]$	
16	For each candidate variety, take the mean of the $adjlogSD_{ij}$ values over years $[\rightarrow \overline{adjlogSD_{i.}}]$	
17	For each candidate, calculate the COYU threshold: $$\overline{adjlogSD_{..}} + t_{\alpha,d_r}\sqrt{\frac{\sigma^2}{k}(1 + f_{i.})}$$ where t_{α,d_r} is the 100(1-α) percentile of the Student t-distribution with d_r degrees of freedom $[\rightarrow UC_i]$	
18	Compare the $\overline{adjlogSD_{i.}}$ values for each candidate with the corresponding threshold UC_i. Candidates with values higher than the threshold fail under the COYU criterion.	

Choice of degrees of freedom

We can set the degree of smoothness of the cubic smoothing spline by setting the effective degrees of freedom for the curve. We need sufficient degrees of freedom to give the flexibility to fit non-linear relationships but not so many that an overly-complicated relationship is fitted that isn't supported by the data. In particular, if there are few reference varieties then fewer degrees of freedom would certainly be better.

In principle, the smoothness of the fitted curve could be determined by the data itself, through e.g. cross-validation. However this has risks particularly with smaller data sets and when the user is unlikely to review the results.

In document TWC/29/22, a cubic smoothing spline with 4 degrees of freedom was found to produce reasonable fits to real data. We test the performance in simulated data below with degrees of freedom set at either 3 or 4.

Performance on simulated data sets

We have compared the performance of the spline approach with that of a linear regression using the eight sets of simulated data described in document TWC/28/27. The eight sets were obtained using the combinations of the following 3 parameters:

- Number of reference varieties: 10 or 50
- Interaction between year and variety: variance component is 0 or 100
- Slope for linear relation between SDs and mean: 0 or 0.1

The data were originally simulated at the plant within plot level with k=3 years, 3 blocks each year and 20 plants in each plot. However here we just use the variety means and SDs aggregated to the trial level. In each data set there were 10 candidate varieties – these were simulated from the same distributions as the reference varieties. For each of the eight sets, there are 500 simulated data sets.

Table 2 below compares the proportion of candidate varieties rejected using COYU with either linear regression and spline adjustment methods. The probability level adopted here was 0.05 (so an acceptance probability of 95%) so, given that the candidate varieties were simulated in the same way as the reference varieties, we would hope to achieve a 5% level of rejection. The linear regression method uses the formula $k(n_r - 2)$ for the residual degrees of freedom. For the spline method, we compare the classical and Bayesian formulation for standard error and the use of three or four degrees of freedom.

Table 2: Proportion of candidates above the COYU threshold using linear and spline methods of adjustment (probability level α=0.05) – simulated data has a linear relationship

Set No	Assumptions in simulations			Method				
	No of reference varieties, n_r	Variety, σ_v^2/ Slope	Interac-tion, σ_{yv}^2	Linear	Spline			
					Classical		Bayesian	
					3 df	4 df	3 df	4 df
1	50	0/0	0	0.044	0.046	0.048	0.045	0.046
2	10	0/0	0	0.049	0.055	0.058	0.046	0.046
3	50	125/0.1	0	0.047	0.046	0.048	0.046	0.047
4	10	125/0.1	0	0.048	0.055	0.058	0.048	0.047
5	50	0/0	100	0.045	0.046	0.047	0.045	0.045
6	10	0/0	100	0.050	0.058	0.063	0.049	0.049
7	50	125/0.1	100	0.054	0.055	0.056	0.054	0.054
8	10	125/0.1	100	0.054	0.060	0.066	0.053	0.054

The performance of the linear method and the spline method with the Bayesian standard error were very similar. Both tended to under-reject very slightly, apart from data sets 7 and 8 when they slightly over-rejected. However the match with the probability level set seems acceptable. The number of degrees of freedom makes little difference. The spline method with the classical standard errors deviated more from the target level, especially when the number of reference varieties is low.

The good performance of the linear method above might have been anticipated: the underlying relationship is linear. To provide a greater challenge, we simulated new data sets with linear, quadratic and sinusoidal relationships between the logSD and the means, with the same relationship in each year. Here we looked at data sets with either 10 or 50 reference varieties with 10 candidates tested in 3 years. Examples of each type of function are shown in Figure 2, with the splines with 4 degrees of freedom shown for data sets of 10 and 50 varieties. We ran separate sets of simulation data sets for each combination of degrees of freedom, form of function and number of reference varieties. The results for 3 degrees of freedom are shown in Table 3 and for 4 degrees of freedom in Table 4. These were based on 100,000 simulated data sets in the case of 10 reference varieties and 10,000 data sets in the case of 50 reference varieties (these ran more slowly). Note these are subject to simulation sampling error; this is why, for example, the result for linear regression with 10 reference varieties with a sinusoidal function differs slightly between the two tables.
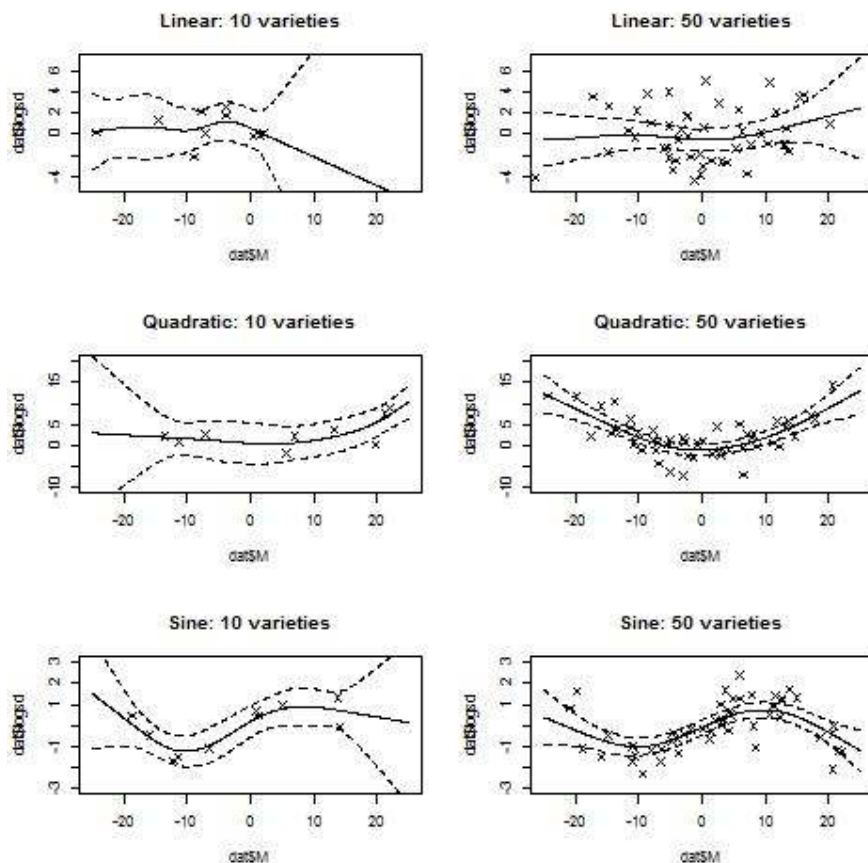

Figure 2: Examples of one-year simulated data sets with different forms of relationship. A cubic smoothing spline (with 4 degrees of freedom) is fitted to each (solid line). The dashed lines represent the pointwise 95% confidence interval for the fit (using the Bayesian formulation).

Table 3: Proportion of candidates above the COYU threshold using linear and spline (3 degrees of freedom) methods of adjustment (probability level α=0.05) – different forms of relationship

Relationship	No of reference varieties	Method		
		Linear	Spline	
			Classical	Bayesian
Linear	10	0.050	0.055	0.047
Quadratic	10	0.141	0.096	0.077
Sinusoidal	10	0.115	0.108	0.097
Linear	50	0.050	0.051	0.049
Quadratic	50	0.109	0.078	0.076
Sinusoidal	50	0.118	0.105	0.103

Table 4: Proportion of candidates above the COYU threshold using linear and spline (4 degrees of freedom) methods of adjustment (probability level α=0.05) – different forms of relationship

Relationship	No of reference varieties	Method		
		Linear	Spline	
			Classical	Bayesian
Linear	10	0.050	0.058	0.047
Quadratic	10	0.141	0.077	0.056
Sinusoidal	10	0.114	0.084	0.069
Linear	50	0.050	0.052	0.050
Quadratic	50	0.110	0.062	0.059
Sinusoidal	50	0.117	0.078	0.076

From this it can be seen that overall the spline method with the Bayesian standard error formulation was closest to matching the target reject rate. Unsurprisingly the version with four degrees of freedom worked better than with three degrees of freedom for non-linear relationships. Looking at the results for the sinusoidal simulations, the spline with four degrees of freedom was clearly under-fitting the sine curve, resulting in a slightly higher reject rate than desired. However results shown in TWC/29/22 demonstrate that four degrees of freedom should be adequate in practice.

Application to real data sets

We demonstrate the proposed method (with 4 degrees of freedom) on a three-year data set for *Lolium perenne* kindly supplied by the Agri-Food and Biosciences Institute, which runs the United Kingdom DUS Centre for Herbage Crops. In this data set there are 63 reference varieties and two candidate varieties tested in all three years. We look at characteristics 8 (Time of inflorescence emergence in 2nd year) and 9 (Plant: natural height at inflorescence emergence).

First we show the relationships between logSD and the means for the reference varieties in Figures 3 and 4. These plots also show the spline fit (thick line) and a moving average (thinner line).
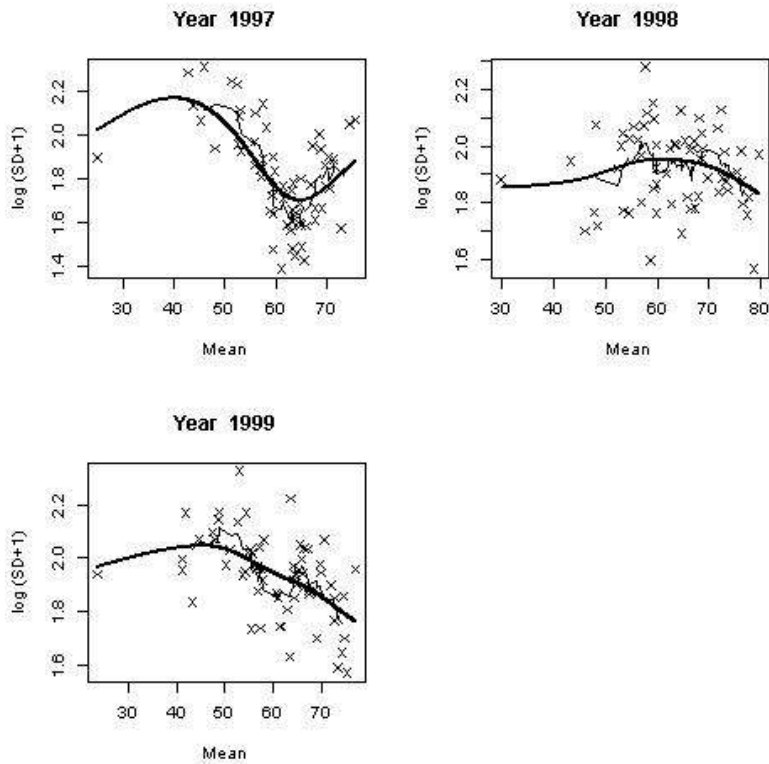
Figure 3: Relationship between logSD and mean in each of three years for the Lolium perenne example with characteristic 8. A cubic smoothing spline (with 4 degrees of freedom) is fitted to each (solid line). The thinner lines represent a nine-point moving average as used in the current COYU procedure.
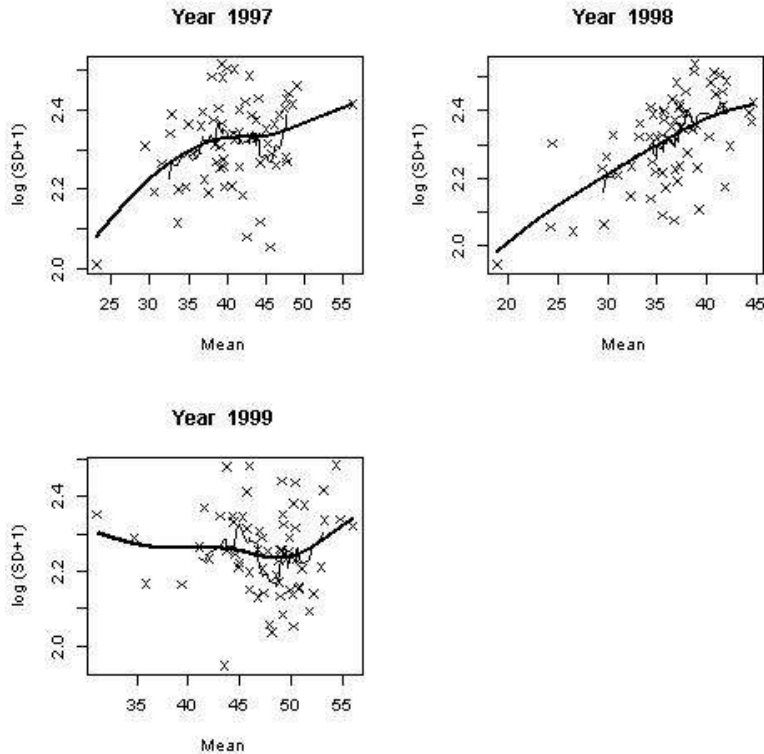

Figure 4: Relationship between logSD and mean in each of three years for the Lolium perenne example with characteristic 9. A cubic smoothing spline (with 4 degrees of freedom) is fitted to each (solid line). The thinner lines represent a nine-point moving average as used in the current COYU procedure.

The results of applying the existing and proposed versions of COYU are summarized in Table 5. It can be seen that the adjusted logSDs are similar for both methods in this small example. Candidate B is closest to failing to pass the COYU criterion with the new method, having a p-value of 0.071. The thresholds for the existing COYU method (α=0.001) are higher than the proposed method, though only a little when α is 0.05 for the new method. The setting of acceptance probabilities is discussed below.

Table 5: Summary of results of application of the existing and proposed versions of COYU.

	Characteristic 8		Characteristic 9	
Candidate	A	B	A	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

Choice of acceptance probability

Guidance on acceptance probabilities for the current version of COYU is given in TGP/8/1 Part II. 9.11. For a three-cycle testing regime, different probability levels can be set: p_{u2} for declaring a candidate as uniform after two cycles, p_{uu2} to declaring a candidate as non-uniform after two cycles and p_{u3} for the decision after three cycles. The above results seem to suggest that a reasonable α (or p_{u3}) to use in a decision taken after 3 years of test may be 0.01 as this will give a threshold that is close the one found using the existing method. However, before a final decision about the different P-values (p_{u2}, p_{uu2} and p_{u3}) is made for a particular crop, it would be best to carry out direct comparisons between the present and a new method on historical data in order to ensure that there will be a smooth transition from the present method to the new method.

Implementation in software

Although many statistical software packages do have a facility to fit cubic smoothing splines, they do not usually calculate the standard errors needed for new observations. If this calculation is not available then either a suitable powerful programming facility or the ability to interact with a FORTRAN program is required to implement COYU with splines. At this stage we have not carried out an in depth review of all software packages used by member states. Below we give some initial views on some key software options.

R

R has been used to set up and test an initial version of the improved COYU software. The "smooth.spline" function in the "stats" library and the "ns" function in the "splines" library have been used. The "gam" function in the "mgcv" library provides an alternative route.

FORTRAN

FORTRAN subroutines for the special functionality required are readily available. Indeed the authors of R functions have made available FORTRAN source code (Hastie & Tibshirani – gamfit - http://www.stanford.edu/~hastie/swData.htm ; Fields development team – css - http://www.image.ucar.edu/Software/Fields/index.shtml).

DUST

DUST has a Windows interface to FORTRAN modules. If FORTRAN code can be developed for the new COYU method, it should be straightforward to then integrate it into DUST.

GenStat

Smoothing splines are available using the REG directive (with the SSPLINE function). However this does not seem to allow prediction. There is also a facility for calculating spline bases (SPLINE procedure). Fitting of splines is also possible through the mixed model directives (VCOMPONENTS and REML) and prediction with standard errors for new observations can be done using the VPREDICT directive. However the degree of smoothing is estimated from the data rather than being fixed according to the degrees of freedom required. With some programming effort, it may be possible to alter this (essentially by fixing the variance component for the spline) but this has not been tested. In general, we have not advocated a mixed model approach to the fitting of splines because it would be difficult to implement in DUST. A more straightforward alternative for the implementation of COYU in GenStat would be to interface with a FORTRAN or R program.

SAS

In SAS/STAT, PROC TRANSREG and PROC GAM will fit splines. However we do not believe that they will directly produce standard errors for new observations. This may be possibly through coding with SAS Macro language or PROC IML but we haven't investigated this further. A more straightforward alternative would be to interface with a FORTRAN or R program.

Conclusions and outstanding issues

We have developed a new version of COYU using a spline adjustment rather than the current moving-average approach. We believe this to be an improvement on the current version.

The spline approach avoids the problem of bias exhibited by the moving-average approach yet is able to fit a non-linear relationship between variability and level of expression better than those alternatives also examined.

We think that a fixed degree of smoothing should be adopted. This avoids complexity in implementation and difficulties with choosing a level of smoothing with a small data set. We would recommend a level of smoothing equivalent to four degrees of freedom. This seems to give sufficient flexibility to fit relationships seen in practice without over-fitting. The Bayesian formulation for standard errors performs better than the classical formulation.

An issue that we have not addressed here is extrapolation. It is clearly inadvisable to adjust logSD values for a candidate whose level of expression is outwith that seen in the reference varieties. This is as true for other methods as for the spline approach, including the current COYU method. We think that a warning should appear in such cases. However at this stage we have not thought about how uniformity might be assessed when this occurs. Further consideration is required; it could be difficult to find a generally acceptable approach in such cases.

We ask the TWC to consider this paper and give guidance on whether COYU method should be modified to use splines, In that case, there needs to be an agreed process for the modification to take place.

We think that it should be relatively straightforward to write software for the method in FORTRAN that could then be integrated into DUST. It would also be straightforward to implement the method using R (a free statistical package). However, implementation in other software packages such as SAS or GenStat may be more difficult – it may be easiest simply to interface with the FORTRAN program.

References

Büsche, A.; Piepho, H.-P.; Meyer, U. (2007). Examination of statistical procedures for checking uniformity in variety trials. Biuletyn Oceny Odmian (Cultivar testing Bulletin) 32: 7-27.

HastieT,& Tibshirani R. (1990). *Generalized additive models.* Chapman and Hall.

HastieT, Tibshirani R. & Friedman J (2001). *The elements of statistical learning.* Chapter 5. Springer. A free version and updated version is also available at http://www-stat.stanford.edu/~tibs/ElemStatLearn/

Wahba G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. J. R. Statist. Soc. B 45:133-150.

[Annex II follows]

Method of Calculation of COYU

Adrian Roberts & Kristian Kristensen

BioSS

FACULTY OF AGRICULTURAL SCIENCES
AARHUS UNIVERSITY

1

TWC/31/15, Seoul 2013

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 strict
  - Fails more varieties than should
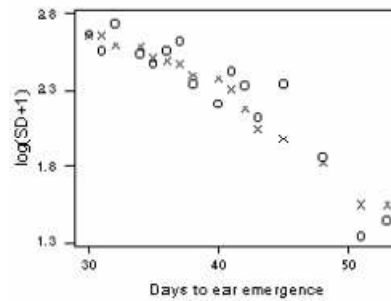 - Compensated for by using small probability level

2

TWC/31/15, Seoul 2013

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

TWC/31/15, Seoul 2013

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

TWC/31/15, Seoul 2013

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

5

TWC/31/15, Seoul 2013

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?

6

TWC/31/15, Seoul 2013

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.

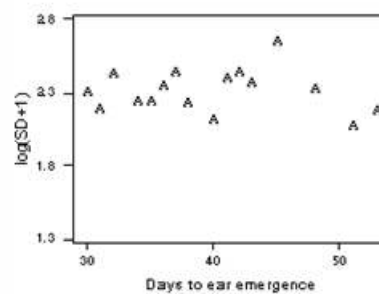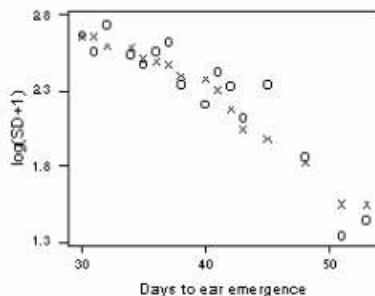7. Comparison of the adjusted log SDs of candidate varieties with the maximum allowable SD.

7

TWC/31/15, Seoul 2013

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 $= Y_i - T_i + \overline{Y}$

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

9

TWC/31/15, Seoul 2013

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
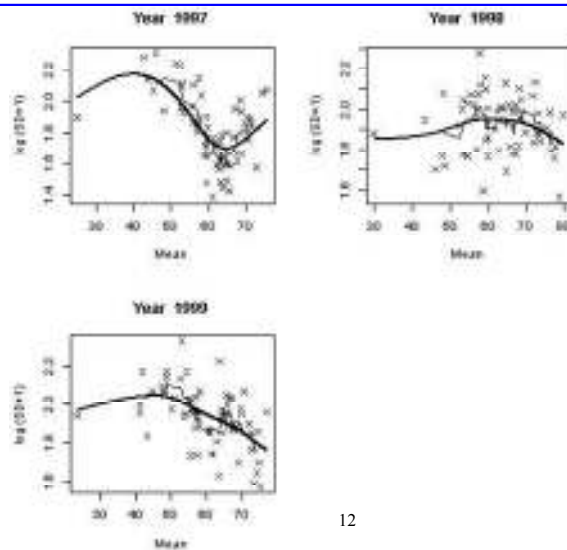
10

TWC/31/15, Seoul 2013

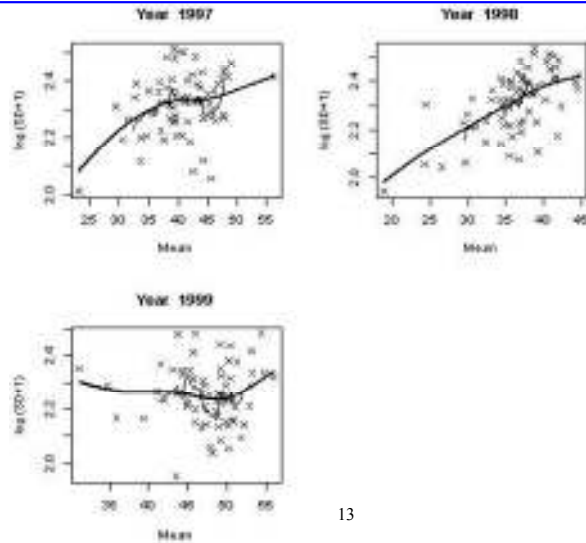# Example on real data

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

11

TWC/31/15, Seoul 2013

Characteristic 8

12

√C/31/15, Seoul 2013

Characteristic 9

	Characteristic 8		Characteristic 9	
Candidate	A	B	A	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

15

TWC/31/15, Seoul 2013

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

16

TWC/31/15, Seoul 2013

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

19

TWC/31/15, Seoul 2013

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 of method
  - Cubic smoothing splines (natural)
  - Method of standard error calculation → Bayesian
  - Fixed level of smoothing – 4 degrees of freedom

20

TWC/31/15, Seoul 2013

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

21   TWC/31/15, Seoul 2013

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.*

22   TWC/31/15, Seoul 2013

Process for implementation
(if proposal is accepted)

- Agreeing more widely in UPOV
- Software
- Timetable
- Acceptance probabilities
- TGP/8
- Publicising
- ….

23

TWC/31/15, Seoul 2013

Questionnaire concerning the Combined-Over-Years Uniformity Criterion (COYU) method

UPOV member:

Name of person answering the questionnaire:

Title:
Organization:
Contact Information:
        Address:

        Tel:
        Email:

Please answer to the following questions. You can attach a separate sheet(s) to provide more detailed information, if necessary.

1) Do you use COYU for assessment of uniformity in one or more crops?

   [   ] Yes

   [   ] No

2) If the answer to question 1 is 'yes', what software (e.g. DUSTNT) do you use for COYU?

3) If the answer to question 1 is 'yes', for what crop(s) do you use COYU?

4) Additional comments (if any):

SUMMARY OF THE REPLIES TO THE QUSTIONNAIRE
CONCERNING THE COMBINED-OVER-YEARS UNIFORMITY CRITERION (COYU) METHOD

This following table summarizes the results of the survey:

Member of the Union	Use of COYU (Question 1)	Software (Question 2)	Crops COYU is used for (Question 3)
Czech Republic	Yes	DUSTNT	Oilseed rape, Fodder crops
Estonia	Yes	DUSTNT	Grasses and legumes
Finland	Yes	DUSTNT	Timothy, Meadow fescue and Tall fescue, Reed canary grass, Red clover and White clover, Turnip rape (oilseed type) and Rye
France	Yes	SAS	Forage crops, Oilseed rape, Broad bean
Japan	No	n/a	n/a
Netherlands	Yes	GenStat	Cross-pollinated crops in general, Grasses in particular
New Zealand	No	n/a	n/a
Portugal	No	n/a	n/a
Russian Federation	No	n/a	n/a
United Kingdom	Yes	DUSTNT	Festulolium, White clover, Ryegrass, Pea, Oilseed rape (winter type)

In addition to the above, the following comments were received (Question 4):

- France: for forage crops, COYU is in routine use for uniformity; for oilseed rape, COYU is used for some of the characteristics and for other characteristics, off types methods are used; for broad bean (field crop group), COYU has been used for quantitative characteristics until now, but will be abandoned to use GAIA to get a distance mixing qualitative and quantitative characteristics.

- Netherlands: Naktuinbouw is working with a full set of procedures in GenStat developed by Biometris (Wageningen University) for trial design (alpha or block design), trial analysis, COYD, COYU and 'Differ' (filter for distinctness).

- New Zealand: DUST is used for ryegrass and other forage species. New Zealand has not had sufficient variety numbers to apply COYD or COYU until this year. It is planned to undertake COYD and COYU on ryegrass and other forages measurements in January 2014. Until this time, the LSD 1% method has been used.

[Annex IV follows]

ANNEX IV

OVERVIEW OF THE PROPOSED IMPROVEMENTS TO COYU

The Combined Over-Year Uniformity method (COYU) is a method for assessing uniformity based on measured characteristics. The General Introduction (TG/1/3) states:

> "6.4.2.2.1 For measured characteristics, the acceptable level of variation for the variety should not significantly exceed the level of variation found in comparable varieties already known. UPOV has proposed several statistical methods for dealing with uniformity in measured quantitative characteristics. One method, which takes into account variations between years, is the Combined Over Years Uniformity (COYU) method."

Document TGP/10/1 "Examining Uniformity" gives a little more detail on COYU as follows:

> "5.2 Determining the acceptable level of variation
>
> 5.2.1 The comparison between a candidate variety and comparable varieties is carried out on the basis of standard deviations, calculated from individual plant observations. Comparable varieties are varieties of the same type within the same or a closely related species that have been previously examined and considered to be sufficiently uniform.
>
> 5.2.2 UPOV has proposed several statistical methods for dealing with uniformity in measured quantitative characteristics. One method, which takes into account variation between years, is the Combined Over Years Uniformity (COYU) method. The comparison between a candidate variety and comparable varieties is carried out on the basis of standard deviations, calculated from individual plant observations. This COYU procedure calculates a tolerance limit for each characteristic on the basis of varieties within the same trial with comparable expression for that characteristic."

COYU is described in much greater detail in document TGP/8 "Trial Design and Techniques Used in the Examination of Distinctness, Uniformity and Stability". In particular, software for applying COYU is available in DUST as described in document TGP/8 as follows:

> "9.9 Implementing COYU
>
> The COYU criterion can be applied using COYU module of the DUST software package
> for the statistical analysis of DUS data. This is available from Dr. Sally Watson,
> (Email: *info@afbini.gov.uk*) or from *http://www.afbini.gov.uk/dustnt.htm.*"

Over the last six years, the TWC has investigated improvements to the current COYU method. This document gives an overview of the progress made. It explains why the TWC proposes to improve the current method of COYU and how it can be done, and it discusses how the proposed improvement might be evaluated more widely. Technical detail is given in the TWC papers.

Why does COYU need improving?

The investigation by the TWC has shown that the current method tends to declare more varieties as not uniform than desirable. It is believed that smaller than customary probability levels have been widely adopted to set the COYU criterion in order to compensate for this feature. Probability levels such as 0.001 (0.1%) and 0.002 (0.2%) are typically used for COYU, whereas for COYD probability levels such as 0.01 (1%) and 0.05 (5%) are used.

However compensation using small probability is not the best way of managing the bias issue with COYU. This is an *ad-hoc* solution rather than one based on sound principles, and it is less than perfect because the actual compensation required varies from crop to crop, from characteristic-to-characteristic and from candidate-to-candidate. Of course in practice the same probability level is usually adopted over all characteristics for a crop.

The source of the bias issue is the method within COYU that is used to adjust for any relationship between levels of variability seen for a measured characteristic and the expression of that characteristic. Such relationships are quite common – see document TWC/29/22. Adjustment is required to ensure that comparisons of variability are made with "comparable varieties already known" (document TG1/3). The method of adjustment within the current COYU is known as the moving average method.

<u>The proposed improvement</u>

It was thought that an alternative method of adjustment to moving average might substantially reduce the issue of bias and allow use of more usual probability levels. The TWC investigated several likely approaches and evaluated them on simulated and real examples of data. On the basis of this, it is proposed that the moving average method be replaced by a spline method.

It was found that the spline method fits relationships between variability and level of expression seen in real examples. Further, the bias exhibited using it is very small and allows the use of more customary probability levels.

<u>The way forward</u>

Basic software for the improved COYU method has been written using R, a freeware statistical package. At the thirty-first session of the TWC, the United Kingdom agreed to add a prototype module to the DUST package in time for the thirty-second. This would allow evaluation of the new method on real examples by TWC members. On the basis of this, the setting of appropriate probability levels would be discussed.

The work has revealed an issue also present with the existing COYU method: when the expression level of a candidate falls outside that of the reference varieties, how can its uniformity be assessed? The new software will identify such cases and these can be considered as part of the TWC evaluation.

In the longer term, the new method will need to be evaluated more widely. If it is considered an improvement over the existing COYU method, then a plan for introduction of the method will be required. To aid this process, an expert of the United Kingdom has prepared a survey to investigate which members of the Union use the COYU method and the software they use for this.

[End of Annex IV and of document]