



TWC/31/15 Corr.  
ORIGINAL: English  
DATE: May 21, 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**

METHOD OF CALCULATION OF COYU

*Document prepared by experts from the United Kingdom and Denmark*

BACKGROUND

1. 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<sup>1</sup>" and a presentation by Mr. Kristian Kristensen (Denmark), a copy of which was reproduced as document TWC/26/17 Add.

2. 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  $Var(Y_i)$  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  $\alpha$ -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."

3. 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:

- (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).

---

<sup>1</sup> 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.

4. 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 this would also allow experts further time to reflect on the situation and possible ways forward.

5. The Technical Committee (TC), at its forty-fifth session, held in Geneva from March 30 to April 1, 2009, noted the discussions concerning the current method of calculation of COYU, as set out above, and agreed that the Technical Working Parties (TWP) should be informed about those discussions at their sessions in 2009. The TC requested the TWC to make its recommendations to the TC concerning the proposals set out in paragraph 3 of this document.

#### Developments in 2009

6. At its twenty-seventh session, held in Alexandria, Virginia, United States of America, from June 16 to 19, 2009, the TWC considered document TWC/27/15 "Potential approaches to improving COYU" prepared by experts from Denmark and the United Kingdom on the basis of a presentation by Mr. Adrian Roberts (United Kingdom). The TWC agreed that it would be important to evaluate the range of circumstances that needed to be accommodated and that a new document should be prepared for its twenty-eighth session by experts from Denmark and the United Kingdom.

#### Developments in 2010

7. The TC at its forty-sixth session held in Geneva from March 22 to 24, 2010 considered document TC/46/11 "Method of Calculation of COYU". It noted the developments concerning the method of calculation of COYU as set out in document TC/46/11, paragraphs 7 to 11, and requested the TWC to make proposals to address the bias in the present method of calculation of COYU. The TC noted the observation at the twenty-seventh session of the TWC, that the way COYU made the calculations at the moment was acceptable, but that it was nevertheless desirable to find a solution.

8. The TWC, at its twenty-eighth session held in Angers, France, from June 29 to July 2, 2010, considered document TWC/28/27 "Alternative Methods to COYU for the Assessment of Uniformity", presented by Mr. Kristian Kristensen (Denmark). Mr. Kristensen proposed to carry out a survey to obtain data on the relationship between uniformity and expression of characteristics for different crops in order to determine if linear or quadratic adjustments would be suitable to correct the biases. He would then consider the implementation of the improved method. The TWC noted that experts from Germany, the Netherlands, Poland and United Kingdom would send information of averages and standard deviations to Mr. Kristensen for analysis and encouraged other experts to send such information to Mr. Kristensen (see document TWC/28/3 "Report" paragraphs 49 and 50).

#### Developments in 2011

9. The TC, at its forty-seventh session held in Geneva from April 4 to 6, 2011, noted the developments concerning the method of calculation of COYU as set out in this document, paragraphs 8 and 9, and requested the TWC to continue its work with the aim of developing recommendations to the TC (see document TC/47/26 "Report on the Conclusions", paragraph 88).

10. The TWC, at its twenty-ninth session held in Geneva, Switzerland, from June 7 to June 10, 2011, took note of the information contained in document TWC/29/10 "Method of Calculation of COYU" (see document TWC/29/31 "Report", paragraph 60).

11. The TWC also received a presentation by Mr. Kristian Kristensen (Denmark) based on document TWC/29/22 "Analysis of the Relation Between Log SD and Mean of Varieties", prepared by experts from Denmark and the United Kingdom. Document TWC/29/22 states as follows:

##### "Introduction

"1. At the twenty-eighth session on the Technical Working Party on Automation and Computer Programs, held in Angers, France, from June 29 to July 2, 2010, the bias in the present COYU method was discussed (document TWC/28/27; also previously documents TWC/26/17 and TWC/27/15). One of the possible approaches to overcome the bias in the present method was to use a linear and quadratic adjustment instead of the moving average method. However, it was questioned whether such an adjustment would be appropriate in all cases. It was decided to carry out a survey on the relationship between log SD and the mean of varieties in order to see if this could be modelled sufficiently well using a

linear or quadratic regression. In such case it would be appropriate to introduce a new method for COYU based on a linear and quadratic effect adjustment of the log SD.

"2. *The following data were received:*

- Data on *Lolium perenne* (perennial ryegrass) from Germany, the Netherlands and the United Kingdom through the years 1993-2002. Part of the data has previous been used in TWC/28/31 "A study on Grass Reference Collections in Different Locations"
- Data on *Brassica napus* L. *oleifera* (spring oil seed rape) from Denmark through the years 1997-2005
- Data on *Pisum sativum* (field pea) from Denmark through the years 1997-2005"

[...]

#### "Discussion and conclusions

"18. For most cases a model with a linear and quadratic effect described the relation between Log SD and the mean sufficiently well. In many of the cases where the preferred model included a significant cubic term, this seemed to be caused by a few unusual varieties. The 3<sup>rd</sup> degree polynomial fit seemed to be strongly influenced by extreme standard deviations particularly for varieties with either low or high means. In such cases it might be questioned whether a model that displaying such sensitivity should be used or whether such extreme varieties should be left out of the model fitting. The cubic spline applied seemed to be less influenced by extreme observations and seemed to describe the relationship at least as well as the 2<sup>nd</sup> and 3<sup>rd</sup> degree polynomial. In most cases the cubic spline (with 4 degrees of freedom set) was in most cases located between the 2<sup>nd</sup> and 3<sup>rd</sup> degree polynomial. However the cubic spline would be a little more difficult to implement than the polynomial regressions and some technical challenges remain."

12. The TWC agreed that a new document based on the cubic spline model should be prepared for the next session of the TWC.

#### Developments in 2012

13. The TC, at its forty-eighth session, held in Geneva from March 26 to 28, 2012 noted the latest developments concerning the method of calculation of COYU, as set out in paragraphs 10 to 13 of document TC/48/11 "Method of Calculation of COYU" (see document TC/48/22 "Report on the Conclusions", paragraph 126).

14. The 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 (see document TC/48/22 "Report on the Conclusions", paragraphs 126 to 127).

15. In that regard, the Office of the Union has been informed by Mr. Adrian Roberts (United Kingdom) that a document on possible proposals for improvements to COYU could not be prepared for consideration by the TWC at its thirtieth session, to be held in Chisinau, Republic of Moldova, from June 26 to 29, 2012, and could be prepared for the TWC session in 2013.

### PROPOSALS FOR IMPROVEMENTS TO COYU

#### Introduction

16. 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.

17. 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.

18. 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.

19. 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.

20. 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?

21. Smoothing is a commonly used procedure for fitting a relationship when the form of the relationship is unknown. This illustrated in Figure 1.

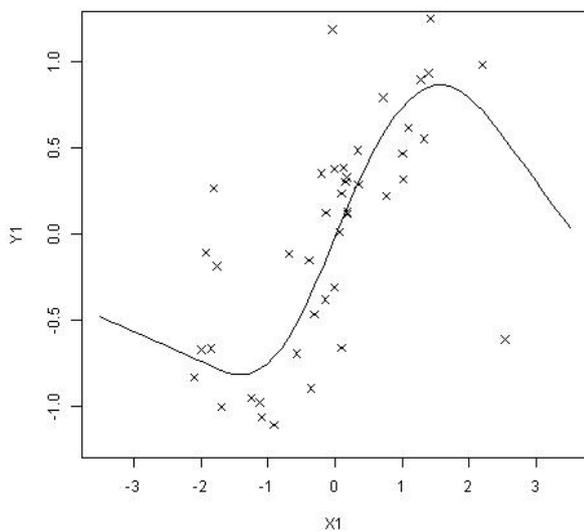


Figure 1: Example of cubic smoothing spline (with 4 degrees of freedom) fitted to simulated data. Observations are represented by "x" and the smooth fit is represented by the line.

22. 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.

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

### Why cubic smoothing splines?

24. 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

25. 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

26. 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.

27. 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.

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

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

where  $f$  is a smooth function and  $\varepsilon$  is an error (independent and identically normally distributed, with variance  $\sigma^2$ ).

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

$$\hat{f} = N(N^T N + \lambda \Omega_N)^{-1} N^T y = S_\lambda y \quad (2)$$

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

30. 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:

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

31. For new observations (i.e. for candidate varieties), the prediction is formulated as follows:

$$\mathbf{n}_0 N^- S_\lambda y \quad (3)$$

where  $\mathbf{n}_0$  is the projected basis vector for the new observation and superscript  $-$  denotes a generalized inverse.

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

- Classical:  $(\mathbf{n}_0 N^- S_\lambda S_\lambda^T (N^-)^T \mathbf{n}_0^T + 1) \sigma^2$ . (4a)
- Bayesian:  $(\mathbf{n}_0 N^- S_\lambda (N^-)^T \mathbf{n}_0^T + 1) \sigma^2$ . (4b)

33. 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.

34. 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 [ $\rightarrow \log SD_{ij}$ ]	
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 [ $\rightarrow \log \overline{SD}_{ij}$ ]	predict.smooth.spline
7	For each year, calculate the mean of the logSDs over the reference varieties only [ $\rightarrow \overline{\log SD}$ ]	
8	For each year, calculate the adjusted logSDs: $\overline{\log SD} + \log SD_{ij} - \log \overline{SD}_{ij}$ [ $\rightarrow adj \log SD_{ij}$ ]	
9	For each year, calculate the basis matrix for the reference varieties – this needs the smoothing parameter $\lambda$ 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 $\lambda$ and knots from step5 [ $\rightarrow N_0$ ]	ns function from splines library
11	For each year (and each candidate variety), calculate a variance factor ( $N_0 N^{-1} S_\lambda S_\lambda^T (N^-)^T N_0^T$ for classical or $N_0 N^{-1} S_\lambda (N^-)^T N_0^T$ 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 $\frac{\sum_j SS_j}{d_r}$ [ $\rightarrow \sigma^2$ ]	
14	Take the mean of the $adj \log SD_{ij}$ values over years and varieties for the reference varieties only [ $\rightarrow \overline{adj \log SD}$ ]	
15	For the candidate varieties, calculate the mean variance factors over years [ $\rightarrow \overline{f}_i$ ]	
16	For each candidate variety, take the mean of the $adj \log SD_{ij}$ values over years [ $\rightarrow \overline{adj \log SD}_i$ ]	
17	For each candidate, calculate the COYU threshold: $\overline{adj \log SD} + t_{\alpha, d_r} \sqrt{\frac{\sigma^2}{k} (1 + \overline{f}_i)}$ where $t_{\alpha, d_r}$ is the $100(1-\alpha)$ percentile of the Student t-distribution with $d_r$ degrees of freedom [ $\rightarrow UC_i$ ]	
18	Compare the $\overline{adj \log SD}_i$ 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

35. 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.

36. 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.

37. 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

38. 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

39. 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.

40. 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  $\alpha=0.05$ ) – simulated data has a linear relationship*

Set No	Assumptions in simulations			Method				
	No of reference varieties, $n_r$	Variety, $\sigma_v^2$ / Slope	Interac- tion, $\sigma_{yv}^2$	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

41. 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.

42. 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.

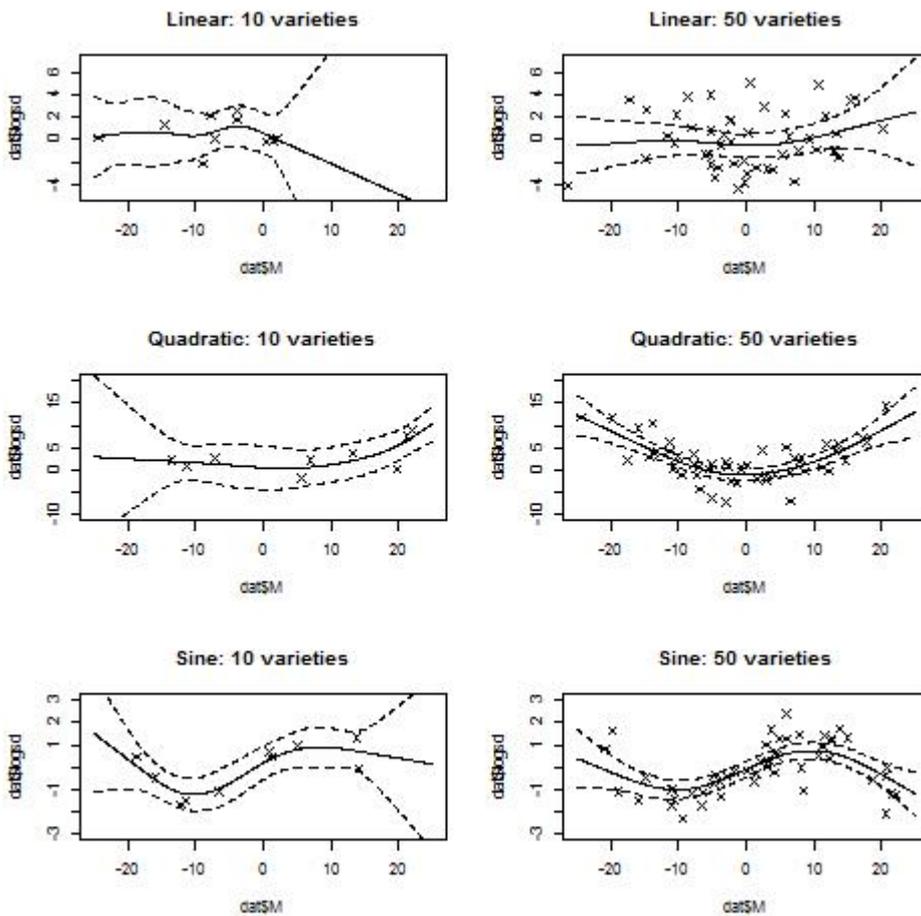


Figure 2: Examples of one-year simulated data sets with different forms of relationship. A cubic smoothing spline (with 4 degrees of freedom) is fitted to each (solid line). The dashed lines represent the pointwise 95% confidence interval for the fit (using the Bayesian formulation).

Table 3: Proportion of candidates above the COYU threshold using linear and spline (3 degrees of freedom) methods of adjustment (probability level  $\alpha=0.05$ ) – 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  $\alpha=0.05$ ) – 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

43. 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

44. 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).

45. 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).

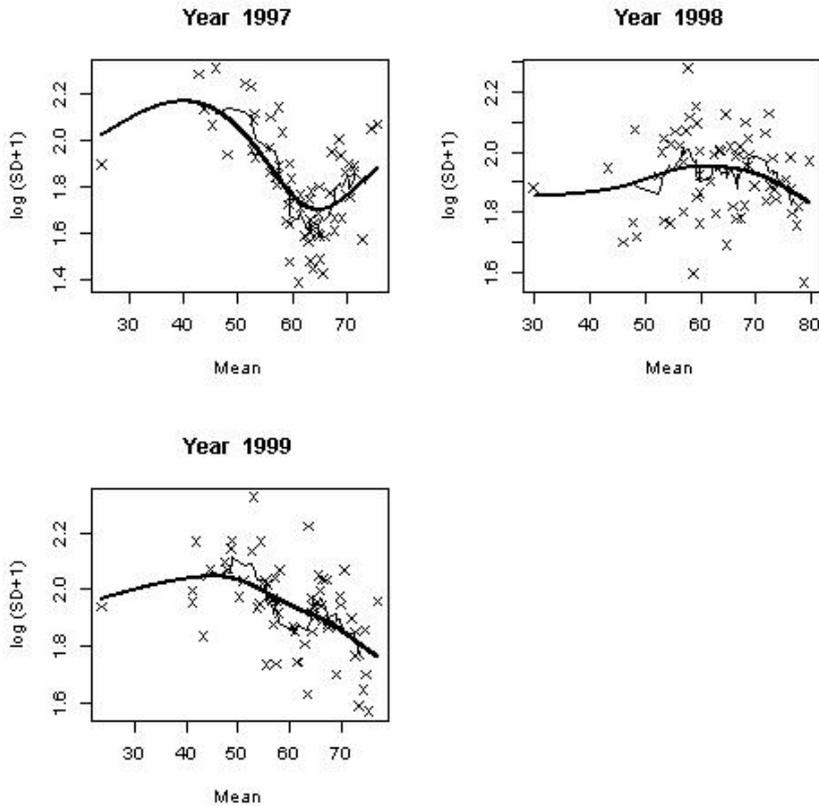


Figure 3: Relationship between  $\log SD$  and mean in each of three years for the *Lolium perenne* example with characteristic 8. A cubic smoothing spline (with 4 degrees of freedom) is fitted to each (solid line). The thinner lines represent a nine-point moving average as used in the current COYU procedure.

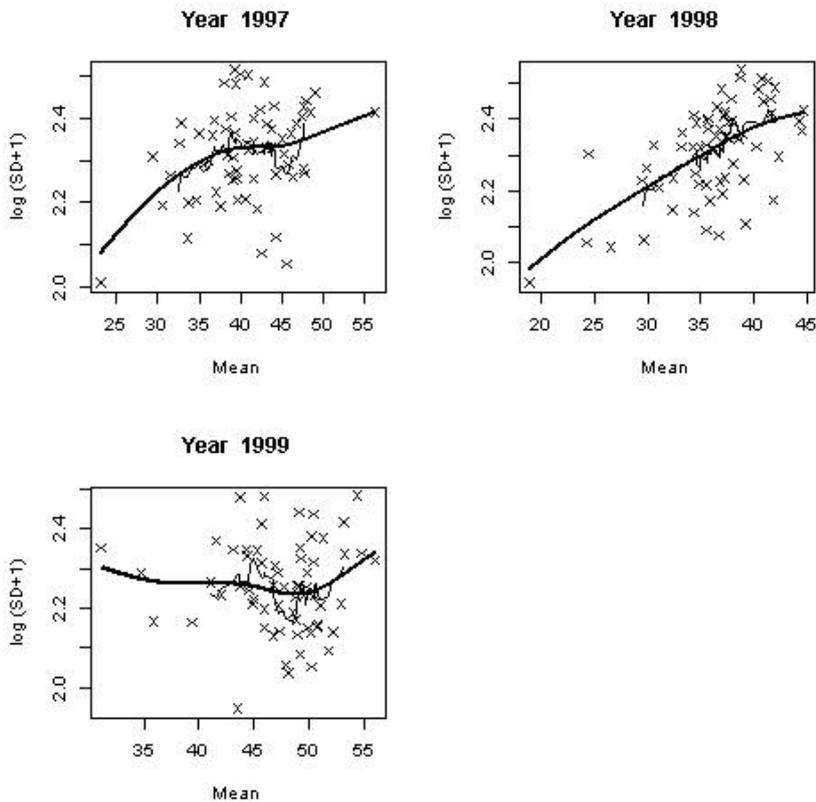


Figure 4: Relationship between  $\log SD$  and mean in each of three years for the *Lolium perenne* example with characteristic 9. A cubic smoothing spline (with 4 degrees of freedom) is fitted to each (solid line). The thinner lines represent a nine-point moving average as used in the current COYU procedure.

46. 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 ( $\alpha=0.001$ ) are higher than the proposed method, though only a little when  $\alpha$  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.

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

#### Choice of acceptance probability

47. 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  $\alpha$  (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

48. 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*

49. 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*

50. 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

51. 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

52. 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

53. 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

54. 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.

55. 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.

56. 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.

57. 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.

58. 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.

59. 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.

60. *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.

### Acknowledgements

61. The authors are grateful for funding support from the Community Plant Variety Office (European Union), Defra (United Kingdom), SASA (United Kingdom) and Department of Agroecology, University of Aarhus (Denmark). They are also grateful for advice from Zhou Fang (BioSS) on splines.

### 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.

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

Hastie T, 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.

[End of document]