E

UPOV

TWC/28/27
**ORIGINAL:** English
**DATE:** June 22, 2010

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

## TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

## Twenty-Eighth Session
## Angers, France, June 29 to July 2, 2010

# ALTERNATIVE METHODS TO COYU FOR THE ASSESSMENT OF UNIFORMITY

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

# ALTERNATIVE METHODS TO COYU FOR THE ASSESSMENT OF UNIFORMITY

Kristian Kristensen and Adrian Roberts

## Introduction

1. At its twenty-sixth session, held in Jeju, Republic of Korea, from September 2 to 5, 2008, the TWC considered document TWC/26/17 and a presentation by Mr. Kristian Kristensen (Denmark), a copy of which was reproduced as document TWC/26/17 Add..

2. 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. Ignore the biases
   (comment: the test will most probably be too liberal)
2. Correct only for the bias introduced by the smaller sample sizes
   (comment: the tests will be too liberal, but will be comparable to those in the past)
3. Correct only for the present bias
   (comment: the test will be conservative, but not comparable to the past)
4. Correct for all biases
   (comment: there will be no biases, but the tests will not be comparable to the past)

3. 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, on the basis of a presentation by Mr. Adrian Roberts (United Kingdom). A copy of the presentation was provided in document TWC/27/15 Add. The TWC agreed that a new document should be prepared by Mr. Kristensen and Mr. Roberts for consideration at its twenty-eighth session (see document TWC/27/21 “Report” paragraphs 42 to 44).

4. In this paper, the present COYU method is compared to a revised version of the regression method presented in 2009 and to some alternative methods where the estimation of relationship between log *sd* and the mean and the statistical tests are done simultaneously (one-step methods). For some of the methods, performance is examined when the number of recorded plants is smaller for reference varieties than for candidates.

## Description of methods

### *Summary*

5. The current COYU method compares the overall level of uniformity for the candidate variety, over two or three years of test, with that of the reference varieties. Here uniformity is measured by the logarithm of the within-plot standard deviation (for a characteristic). The comparison is based on a t-test for the difference between the candidate’s uniformity and the mean of the reference varieties’ uniformity. The standard error for this is based on a pooled residual mean square consisting of the variety mean square and the variety x year mean square. Since the uniformity of a variety may be related to the level of expression of a characteristic, an adjustment is made to the log standard deviations based on the mean scores and using the moving average method. (See document TGP/8 for more details.)

6. In TWC/26/17, it was noted that this method produced a criterion that was biased. In document TWC/27/15, it was postulated that this was because both the degrees of freedom and the standard error need to be modified to allow for the moving average adjustment. It was also proposed that other methods of adjustment might perform better; these included linear, quadratic and cubic smoothing spline (with low degrees of freedom).

*Two-step regression*

7. The two step regression presented in document TWC/27/15 was shown to be biased due to the method of calculation of the standard error. In this paper an unbiased estimate of the standard error was used, which for a linear adjustment is given by:

$$\sqrt{\hat{\sigma}^2\left(1+\frac{1}{r}+\frac{\left(M_c-\overline{M}\right)^2}{\sum\limits_{i\in reference}\left(M_i-\overline{M}\right)^2}\right)}$$

where $r$ is the number of reference varieties, $M_c$ is the mean score for candidate $c$, $M_i$ is the mean score for reference variety $i$, $\overline{M}$ is the mean of the reference variety scores and $\hat{\sigma}^2$ is the pooled residual mean from an analysis of variance of the adjusted log standard deviations for the reference varieties taking into account the effect of year (so $\hat{\sigma}^2$ is based on variety mean square and the variety×year mean square). However, in many cases, the variety means square can be expected to be larger than the variety x year mean square. This means that observations from the same variety are correlated, so the effective degrees of freedom is smaller than *k(r-1)* (where k is the number of years). In Table 1, results are shown both for *k(r-2)* and *r-1* degrees of freedom. The latter should provide a more realistic value when the variety mean squares are much larger than the variety x year mean square.

8. This standard error varies between candidates, with candidates further from the mean score of the reference varieties having larger standard errors. This has some similarity with the second one-step method (see below).

9. At this stage we have only looked at linear adjustments with the improved approach. It is thought that quadratic and smoothing spline adjustments should also work well but the latter may be harder to implement.

*One-step methods*

10. Several different methods were applied. The following methods were all based on methods where the relationship between log *sd* was estimated simultaneously with the estimation of variety estimates and uncertainty as described by Büchse et al. (2007):

> Method 1. Application of a fixed upper limit, above which the candidate will be declared as not uniform – assuming that the relationship between log *sd* and the mean can be estimated using linear or second degree polynomial regression. This method is called $UC_{Ancova}$ in this document. Results are reported in Table 2 for second degree polynomial regression.

Method 2. Comparing each new candidate with the mean of the reference varieties using a t-test that takes into account the actual adjustment – assuming that the relationship between log *sd* and the mean can be estimated using linear or second degree polynomial regression. This method is called $t_{Ancova}$ in this document. Results are reported in Table 2 for second degree polynomial regression.

11. In Methods 1 and 2, the comparisons were based on a mixed model for the reference varieties with a fixed effect of year and mean within each year together with a random effect of varieties and variety x year (residual variation). Following Büchse et al. (2007) the model for the reference varieties may be written as:

$$\log sd_{yv} = \mu + \alpha_y + f_y(M_{yv}) + D_v + E_{yv} \quad \text{with} \quad f_y(M_{yv}) = \beta_y M_{yv} + \gamma_y M_{yv}^2$$

where

$\log sd_{yv}$ is the logarithm of the pooled standard deviation for variety $v$ in year $y$

$M_{yv}$ is the mean of the observations for variety $v$ in year $y$ with $M_{yv}$ centered within each year.

$D_v$ and $E_{yv}$ are random effects, assumed to be independent normally distributed with mean 0 and variances $\sigma_D^2$ and $\sigma_E^2$, respectively

$\mu, \alpha_y, \beta_y, \gamma_y, \sigma_D^2$ and $\sigma_E^2$ are parameters to be estimated

12. For *k* years the mean of all reference varieties vas calculated as $\log sd_{ref} = \frac{1}{k}\sum_y^k \{\hat{\mu} + \hat{a}_y\}$

13. For each candidate, *c*, an adjusted log *sd* was calculated as

$$\widehat{\log sd_c} = \frac{1}{k}\sum_y^k \{\hat{\mu} + \hat{\alpha}_y + \hat{\delta}_c - \hat{f}_y(M_{yc})\}$$

14. For Method 1 the log sd for each variety was compared to an upper limit calculated in a similar way as for the present COY-U method:

$$UC_{Ancova} = \widehat{\log sd_{ref}} + t_{.95,r-3}\sqrt{\frac{1}{r}\hat{\sigma}_D^2 + \frac{1}{kr}\hat{\sigma}_E^2)}$$

15. For Method 2, the log *sd* for each variety was compared to the mean of all reference varieties using a t-test taking into account the actual adjustment for the candidate in question.

16. For both methods the statistical analyses were carried out using the SAS procedure Mixed (SAS Institute 2008). Both the reference and candidates was analyzed in one step, but the estimation process was separated by introduction of a dummy variable with the value zero for reference varieties and one for candidates as proposed by Büchse et al. (2007). This means that a common effect of year is estimated, while random effects are estimated for the reference varieties, fixed effects are estimated for candidates and that the relationship between log sd and mean are based on the reference varieties alone. The centering of means is also based only on the reference varieties and this centering is also then applied to the candidates.

Method 3. Application of an upper limit based on a quantile of the distribution of the log *sd* for reference varieties – assuming that the relationship between log *sd* and the mean can be estimated using a second degree polynomial regression. Here the distribution of reference varieties can be carried out in two different methods as suggested by Büsche et al. (2007):

(a). The variety effects are assumed to be random: this method is called “95% quantile with means, BLUP” in this document. The results are reported in Table 3 for second degree polynomial regression.

(b) The variety effects are assumed to be fixed: this method is called “95% quantile with means, BLUE” in this document. The results are reported in Table 3 for second degree polynomial regression.

Method 4. Application of an upper limit for the upper confidence limit on a quantile of distribution of the log *sd* for reference verities – assuming that the relationship between log *sd* and the mean can be estimated using a second degree polynomial regression. Again the distribution of reference varieties can be carried out in two different methods as for method 3

(a) The variety effects are assumed to be a random effect: this method is called “95% Cl of 95% quantile with means, BLUP” in this document. The results are reported in Table 3 for second degree polynomial regression.

(b) The variety effects are assumed to be a random effect: this method is called “95% Cl of 95% quantile with means, BLUE” in this document. The results are reported in Table 3 for second degree polynomial regression.

*One-step methods with different number of recorded plants for candidates and reference varieties*

17. In order to check whether the methods could be used for the testing of uniformity when a reduced number of plants is recorded for the reference varieties, some of the methods were also carried out for the same cases but with only 10 plants per replicate for reference varieties (but still 20 plants per replicate for candidates).

18. For Method 1, the upper limit above which the candidate should be rejected was calculated in two different ways – using either a separate estimate of the residual variance (variety-by-year interaction) for the candidates, or a theoretical adjustment for degrees of freedom of the estimate for reference varieties:

$$1:\ UC_{Ancova} = \widehat{\log sd_{ref}} + t_{.95,r-3}\sqrt{\frac{1}{kr}(\hat{\sigma}_D^2 + \hat{\sigma}_E^2) + \frac{1}{k}(\hat{\sigma}_D^2 + \hat{\sigma}_C^2)}$$

$$2:\ UC_{Ancova} = \widehat{\log sd_{ref}} + t_{.95,r-3}\sqrt{\frac{1}{kr}(\hat{\sigma}_D^2 + \hat{\sigma}_E^2) + \frac{1}{k}(\hat{\sigma}_D^2 + \hat{\sigma}_E^2 + \frac{0.5}{f_r} - \frac{0.5}{f_c})}$$

where $\hat{\sigma}_C^2$ is the estimate of residual variance (variety $\times$ year effect) for candidates. $f_r$ and $f_c$ are the degrees of freedom for the pooled variance of a reference variaty and candidate, respectively. All other terms are as defined previously.

*Smoothing spline adjustment*

19. Cubic splines should in principle provide a more flexible basis for adjustment than linear or quadratic regression. However, implementation has proved to be technically challenging and limited progress has been made to date. A complete one-step approach similar to Methods 1 and 2 above may be possible in GenStat's REML procedure, but the degree of smoothing would be estimated from the data and so we would not be easily able to constrain the degrees of freedom for the curves. A slightly simpler version has been applied for one set of simulation results only; this carries out the estimation process on the reference varieties only and then applies the estimates to the candidates (using REML and VPREDICT in GenStat). This is most similar to Method 2.

20. Equally, a two-stage approach might be possible without recourse to more sophisticated software, but calculating the standard error is not as straightforward as for the linear or quadratic adjustments.

## Simulation of data

21. In order to compare the different methods the methods were each applied to simulated data. Eight sets of simulated data were used for all methods. The eight sets were obtained using the combinations of the following 3 parameters:

1 - Number of reference varieties: $r$=10 or $r$=50
2 - Interaction between year and variety: $\sigma_{VY}^2$=0 or $\sigma_{VY}^2$=100
*3 - sd*'s dependence on variety mean: $\beta$=0, $\sigma_V^2$=0 or $\beta$=0.1, $\sigma_V^2$=100

22. In all cases we simulated the data for 3 years, using complete block design with 3 blocks each with 20 recorded plants, $r$ reference varieties, $c$=10 candidate varieties. For each case 1000 datasets were simulated according to the following model:

$$X_{yvbp} = \mu + A_y + B_v + C_{yv} + D_{yb} + E_{yvb} + F_{yvbp}$$

where

$X_{yvbp}$ is the recorded value for plant $p$ of variety $v$ in blok $b$ in year $y$

$\mu$ is the average value of the characteristic, here $\mu = 200$

$A_y$ is the effect of year $y$, here $A_y = 0$

$B_v$ is the effect of variety $v$, here $B_v \sim N(0,\sigma_V^2)$, $\sigma_V^2 = 0$ or 125

$C_{yv}$ is the interaction effect of year and variety, here $C_{yv} \sim N(0,\sigma_{YV}^2)$, $\sigma_{YV}^2 = 0$ or 100

$D_{yb}$ is the effect of block $b$ in year $y$, here $D_{yb} = 0$

$E_{yvb}$ is the effect of the plot with variety $v$ in block $b$ in year $y$, here $E_{yvb} = 0$

$F_{yvbp}$ is the effct of plant $p$ of variety $v$ in blok $b$ in year $y$, here $F_{yvbp} \sim N(0,\sigma^2)$, here $\sigma = \sqrt{200} + \beta B_v$

with $\beta = 0$ when $\sigma_V^2$=0 or $\beta$=0.1 when $\sigma_V^2 = 125$

$y = 1,2,3;\ v = 1,2,3,\ldots,r+c;\ b = 1,2,3;\ p = 1,2,3,\ldots,20;\ r = 10$ or $50;\ c = 10$

23. From these simulated data the mean and the log of the pooled standard deviation were calculated for each combination of year and variety as:

$$\bar{X}_{yv} = \frac{1}{3\times 20}\sum_{b=1}^{3}\sum_{p=1}^{20} X_{yvbp} \qquad \text{and} \qquad \log sd_{yv} = \log\left\{\sqrt{\frac{1}{3}\sum_{b=1}^{3}\left(\frac{1}{19}\sum_{p=1}^{20}(X_{yvbp}-\bar{X}_{yv})^2\right)}\right\}$$

Result

*Two-step methods*

24. For comparison the result of simulation using no adjustment, the present moving average adjustment and the improved linear adjustment are shown in Table 1.

25. The linear adjustment works better than the other two methods in that the proportion of significant comparisons match more closely to 0.05. Using *r-1* degrees of freedom for the linear regression method seems to work better than *k*(*r*-2) when the number of reference varieties is lower, though this method is a little conservative for all cases.

Table 1: Relative number of significant comparisons using no adjustment, the present method, the improved two-step analysis using linear regression method when using alpha=0.05

| Set No | Assumptions in simulations | | | Method | | | |
|---|---|---|---|---|---|---|---|
| | No reference varieties, r | Variety, $\sigma_V^2$/ Slope, β | Interaction, $\sigma_{YV}^2$ | No adjustment | Moving average | Linear regression | |
| | | | | | | df=*k*(*r*-2) | df=(*r*-*1*) |
| 1 | 50 | 0/0 | 0 | 0.045 | 0.111 | 0.048 | 0.046 |
| 2 | 10 | 0/0 | 0 | 0.050 | 0.121 | 0.058 | 0.047 |
| 3 | 50 | 125/0.1 | 0 | 0.111 | 0.111 | 0.048 | 0.046 |
| 4 | 10 | 125/0.1 | 0 | 0.121 | 0.119 | 0.058 | 0.047 |
| 5 | 50 | 0/0 | 100 | 0.045 | 0.117 | 0.045 | 0.044 |
| 6 | 10 | 0/0 | 100 | 0.050 | 0.123 | 0.063 | 0.051 |
| 7 | 50 | 125/0.1 | 100 | 0.093 | 0.108 | 0.047 | 0.046 |
| 8 | 10 | 125/0.1 | 100 | 0.099 | 0.116 | 0.056 | 0.046 |

*One-step methods*

26. The result of Methods 1 and 2 show that those two methods have a type I error that agrees reasonably with the nominal value although there is a tendency that the tests using the $UC_{Ancova}$ method were too liberal when the number of reference varieties was low, while the $t_{Ancova}$ were too liberal when both a slope and an interaction between year and variety existed. The slightly simpler version of $t_{Ancova}$ with a cubic spline adjustment also had type I errors that agree reasonably with the nominal value.

Table 2: Relative number of significant comparisons using second degree polynomial regression or splines to describe the relation between log *sd* and mean with a fixed upper limit or t-tests with alpha=0.05 (method 1 and 2 above)

| Set No | Assumptions in simulations | | | Method | | |
|---|---|---|---|---|---|---|
| | No reference varieties, r | Variety, $\sigma_V^2$/ Slope, $\beta$ | Interac-tion,$\sigma_{YV}^2$ | $UC_{Ancova}$ with means and means$^2$ | $t_{Ancova}$ with means and means$^2$ | Simplified $t_{Ancova}$ with spline |
| 1 | 50 | 0/0 | 0 | 0.044 | 0.052 | 0.045 |
| 2 | 10 | 0/0 | 0 | 0.065 | 0.051 | 0.054 |
| 3 | 50 | 125/0.1 | 0 | 0.050 | 0.046 | 0.047 |
| 4 | 10 | 125/0.1 | 0 | 0.064 | 0.043 | 0.049 |
| 5 | 50 | 0/0 | 100 | 0.044 | 0.053 | 0.046 |
| 6 | 10 | 0/0 | 100 | 0.071 | 0.050 | 0.054 |
| 7 | 50 | 125/0.1 | 100 | 0.053 | 0.067 | 0.057 |
| 8 | 10 | 125/0.1 | 100 | 0.068 | 0.061 | 0.061 |

27. Both the 95% quantile and the 95% confidence limit base on BLUP estimates for the reference varieties were far too liberal. The 95% quantile based on BLUE estimates yielded too many significant results in some cases – especially when the number of reference varieties was low. The 95% confidence limit for the 95% quantile seemed to be conservative in most cases, but also too liberal in one case and thus seems to depend very much on the conditions.

Table 3: Relative number of significant comparisons using second degree polynomial regression to describe the relation between log *sd* and mean with 95% quantiles for BLUP or BLUE estimates or similar estimates with an upper 95% confidence limit (method 3 and 4 above)

| Set No | Assumptions in simulations | | | Method | | | |
|---|---|---|---|---|---|---|---|
| | No reference varieties, r | Variety, $\sigma_V^2$/ Slope, $\beta$ | Interac-tion,$\sigma_{YV}^2$ | 95% quantile with means and means$^2$, BLUP | 95% Cl of 95% quantile with means and means$^2$, BLUP | 95% quantile with means and means$^2$, BLUE | 95% Cl of 95% quantile with means and means$^2$, BLUE |
| 1 | 50 | 0/0 | 0 | 0.440 | 0.424 | 0.067 | 0.025 |
| 2 | 10 | 0/0 | 0 | 0.415 | 0.362 | 0.153 | 0.045 |
| 3 | 50 | 125/0.1 | 0 | 0.465 | 0.448 | 0.091 | 0.050 |
| 4 | 10 | 125/0.1 | 0 | 0.441 | 0.391 | 0.204 | 0.111 |
| 5 | 50 | 0/0 | 100 | 0.441 | 0.423 | 0.065 | 0.025 |
| 6 | 10 | 0/0 | 100 | 0.417 | 0.370 | 0.168 | 0.047 |
| 7 | 50 | 125/0.1 | 100 | 0.433 | 0.410 | 0.061 | 0.023 |
| 8 | 10 | 125/0.1 | 100 | 0.405 | 0.353 | 0.117 | 0.037 |

28. When only 10 plants per block were recorded for the reference varieties, the use of a fixed upper limit yielded too many significant results, while the t-test based on the standard deviation of the difference reflection the actual adjustment seemed to give slightly too few significant results.

Table 4: Relative number of significant comparisons using second degree polynomial regression to describe the relation between log *sd* and mean with a fixed upper limit or t-tests with alpha=0.05 for the case where only 10 plants are recorded in each block for reference varieties

| Set No | Assumptions in simulations | | | Method | | |
|---|---|---|---|---|---|---|
| | No reference varieties, r | Variety, $\sigma_V^2$/ Slope, $\beta$ | Interac-tion,$\sigma_{YV}^2$ | $UC_{Ancova}$ with means and means2 using actual residual for candidates | $UC_{Ancova}$ with means and means$^2$ using theoretical adjustment | $t_{Ancova}$ with means and means$^2$ |
| 1 | 50 | 0/0 | 0 | 0.059 | 0.054 | 0.037 |
| 2 | 10 | 0/0 | 0 | 0.074 | 0.092 | 0.033 |
| 3 | 50 | 125/0.1 | 0 | 0.075 | 0.067 | 0.032 |
| 4 | 10 | 125/0.1 | 0 | 0.093 | 0.106 | 0.032 |
| 5 | 50 | 0/0 | 100 | 0.068 | 0.062 | 0.037 |
| 6 | 10 | 0/0 | 100 | 0.102 | 0.117 | 0.039 |
| 7 | 50 | 125/0.1 | 100 | 0.087 | 0.069 | 0.049 |
| 8 | 10 | 125/0.1 | 100 | 0.099 | 0.097 | 0.043 |

## Concluding remarks

29. The one-step methods based on second degree polynomial regression and spline seemed to yield type I errors that were reasonably close to the nominal values. We did not examine the linear regression in detail, but some checks showed similar results (not shown). Because of that and because this method is a simplification of the second degree polynomial regression, it is expected that linear regression will also yield type I errors that are reasonably close to the nominal values. Also we did not examine the second degree polynomial regression and spline methods for data simulated with a second degree polynomial relationship, but we do not think that this should change the results substantially. The one-step regression method using the $t_{Ancova}$ method was also able to control the type I error when the number of recorded plants for reference varieties were reduced. The $UC_{Ancova}$ methods did not work well for that situation. The one-step methods based on quantiles were too liberal and rather unstable and, therefore, are not recommended.

30. The two step method based on linear regression yields type I errors that were reasonably close to the nominal values when the t-tests took into account the actual adjustment of the candidate and based on only r-1 degrees of freedom.

31. In principal, one-step methods should have the advantage of better statistical properties than the two-step approach. However, they rely on the use of relatively sophisticated software. The two-step approach should be easier to implement in DUST.

32. Cubic splines should, in principle, provide a more flexible basis for adjustment than linear or quadratic regression. However, implementation is more technically challenging. It would perhaps best to first evaluate whether adjustment using a second degree polynomial would provide an adequate solution for real uniformity data.

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.

SAS Institute Inc. (2008). *SAS® 9.2. Procedures Guide: Statistical Procedures*. SAS Institute Inc., Cary, NC, USA. 492 pp (online access: http://support.sas.com/documentation/cdl/en/procstat/59629/PDF/default/procstat.pdf )

[End of document]