European Journal of Agronomy 99 (2018) 30–36

Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

Detecting global trends of cereal yield stability by adjusting the coefficient of variation

Thomas F. Döring[a,b], Moritz Reckling[c,d,*]

[a] Department of Agronomy and Crop Science, Albrecht-Thaer-Institute of Agricultural and Horticultural Sciences, Humboldt University of Berlin, Berlin, Germany
[b] Agroecology and Organic Farming Group, Institute of Crop Science and Resource Conservation, University of Bonn, Bonn, Germany
[c] Leibniz Centre for Agricultural Landscape Research, Müncheberg, Germany
[d] Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden

ARTICLE INFO

Keywords:
Cropping systems
Crop yields
Rye
Stability index
Taylor's power law
Wheat

ABSTRACT

High stability of crop yields is a key objective in crop production and breeding, especially under the conditions of a changing climate. Reliable indices are therefore needed for quantifying yield stability. Recently it was shown that some frequently used indices of yield stability, such as the coefficient of variation (CV) may be wrongly interpreted if there is a systematic dependence of the variance σ^2 from the mean yield μ following Taylor's power law. Here we propose a method to adjust the standard CV to account for the systematic dependence of σ^2 from μ. This adjusted CV can be used as a stability index that is expressed in units that are equivalent to the standard CV, as a percentage of the mean, and can therefore be used in agronomic studies that aim to provide guidance for farmers and advisors. Applying this adjusted CV (called aCV) to FAO cereal yield data, we show that the temporal yield stability of both wheat and rye has weakly but significantly decreased over the last five decades and this trend was not picked up with the standard CV in wheat, and was more marked with the aCV than with the standard CV in rye. For the intensifying research on yield stability in agronomy, the suggested method is a novel alternative to estimate yield stability more conclusively, allowing straight-forward interpretation and providing the basis for developing cropping systems with higher yield stability in the future.

1. Introduction

High crop yield stability is an important goal shared by farmers, breeders and consumers. In the face of global change and increasing environmental variability, working towards this goal is becoming ever more imperative (Peltonen-Sainio et al., 2010; Reidsma et al., 2010). Substantial efforts are therefore dedicated to reducing variability in crop performance, e.g. through plant breeding (Mühleisen et al., 2014; Chamekh et al., 2015) and agronomic management (e.g. Smith et al., 2007). Adaptation strategies are also related to the socio-economic conditions and farm management (Reidsma et al., 2010). One of the critical issues in this endeavour is the use of appropriate measures of yield stability. Over the past few decades, scores of yield stability indices have been proposed (Eberhart and Russell, 1966; Becker, 1981; Becker and and Léon, 1988; Huehn, 1990; Eghball and Power, 1995; Piepho, 1998; Dehghani et al., 2008). An extensive literature deals with the comparison of various stability indices (e.g. Becker and and Léon, 1988; Crossa, 1988; Ferreira et al., 2006). Often, however, the interpretation of the results from stability analyses is not easy. This is partly because different stability indices may lead to contrasting conclusions as they reflect different concepts of stability (Dehghani et al., 2008). In addition, the relatively high complexity of the calculations involved in quantifying stability makes it difficult to separate 'true' effects from mathematical artefacts. In agronomic and ecological research, a popular index of yield stability is the coefficient of variation (CV) (Francis and Kannenberg, 1978; Küchenmeister et al., 2012; Ray et al., 2015; Di Matteo et al., 2016). The CV is defined as the standard deviation σ divided by the mean μ, and is expressed as percentage of the mean: $CV = \sigma/\mu \cdot 100\%$.

Recently it has been shown that under certain conditions, the unguarded interpretation of the CV of crop yield data may be misleading (Döring et al., 2015). In particular, it was demonstrated that crop yield data, especially when it spans over a large range, may often follow a power-law relationship between the sample variance σ^2 and the sample mean μ, and in this case the CV tends to typically decrease with increasing mean, according to the data published so far. The power-law relationship, $\sigma^2 = A\mu^b$, is known as Taylor's Power Law (TPL). Logarithmic transformation of TPL results in a linear relationship, expressed as the equation $\log(\sigma^2) = a + b \log(\mu)$ with $a = \log(A)$. The relationship was first extensively described by the British ecologist and

* Corresponding author at: Leibniz Centre for Agricultural Landscape Research, Eberswalder Str. 84, 15374, Müncheberg, Germany.
E-mail address: moritz.reckling@zalf.de (M. Reckling).

https://doi.org/10.1016/j.eja.2018.06.007
Received 22 November 2017; Received in revised form 8 June 2018; Accepted 11 June 2018

entomologist Roy Taylor (Taylor, 1961) and has been detected in hundreds of data sets from population ecology (Cohen et al., 2012, 2013) and multiple other contexts and sciences (Duch and Arenas, 2006; Eisler et al., 2008). Because TPL has also been found in crop yield data (Döring et al., 2015), caution is needed when interpreting the CV. In particular, if TPL holds, then CV = $\mu^{b/2-1} g^{a/2}$, where a and b are the regression parameters (intercept and slope) of the TPL log-log regression and g is the basis of the logarithm. Thus, the CV may change in a non-linear manner with increasing mean, unless $b = 2$. For most crop yield data sets analysed for the presence of TPL-like relationships between mean and variance, the slope b has been found to be < 2 (Döring et al, 2015). In these cases, the CV systematically decreases with increasing μ in a nonlinear way.

Stability parameters need to show independence from the mean, though. Otherwise it would not be possible to differentiate true effects of stability from changes in the mean; information on stability per se would be confounded by information contained in the mean. Specifically, when using the CV, large means will often (if $b < 2$) be automatically associated with low CVs, just because of a mathematical artefact, rather than biologically or agronomically meaningful mechanisms. To solve this problem, it is necessary to account for systematic dependence of the variance from the mean. One method is the stability index POLAR, which calculates the residuals from the linear regression of $\log(\sigma^2)$ against $\log(\mu)$ (Döring et al., 2015). The new method presented here is an adjusted coefficient of variation (aCV) that removes systematic dependence of the standard CV from the mean. In contrast to the units of POLAR stability that are the logarithm of squared yield units, the aCV is expressed in percentages of the mean and thereby facilitates application in agronomic studies that aim to provide guidance to farmers and advisors. In particular, the aCV enables users to interpret results on yield stability more easily and thereby to make decisions on adjusting management.

The objective of this article is to demonstrate a novel method to quantify yield stability by adjusting the standard CV such that dependence from the mean yield is removed. We apply the novel stability index that we call 'scale-adjusted coefficient of variation' using a publicly available data set of cereal yields, obtained from the Food and Agriculture Organisation (FAO) statistics website, to examine global trends of cereal yield stability.

2. The adjusted coefficient of variation

Here we show how the scale-adjusted coefficient of variation (aCV) is calculated in four steps:

2.1. First step

Means ($\hat{\mu}$) and variances ($\hat{\sigma}^2$) are calculated for a subset of data. This creates a list of means $\hat{\mu}_i$ paired with variances $\hat{\sigma}^2{}_i$, that is, each pair (with index i) consists of a mean and a variance. Following TPL, a linear regression is calculated for $\log_{10}$ of the variance over the $\log_{10}$ of the mean. With $v_i = \log(\hat{\sigma}^2{}_i)$ and $m_i = \log(\hat{\mu}_i)$, the linear regression is $v = a + bm$.

2.2. Second step

The residuals u_i from this regression line, the i.e. the Power Law Residuals (POLAR), are calculated as

$$u_i = v_i - (a + bm_i) \tag{1}$$

2.3. Third step

We adjust the logarithm of the variance which is subsequently used for calculating the coefficient of variation. The adjusted logarithm of the variance $\tilde{v}_i$ is

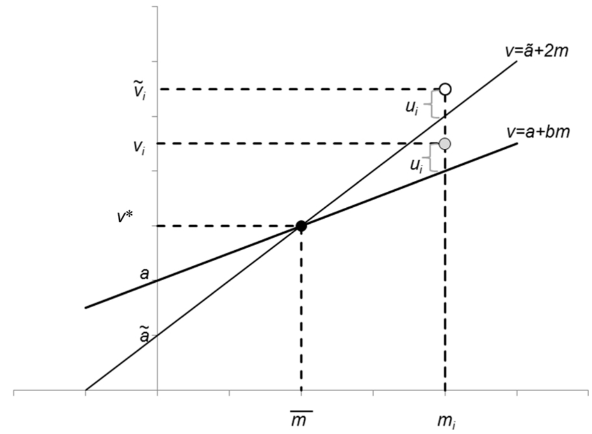

Fig. 1. Example illustrating the procedure for correcting the $v = \log(\hat{\sigma}^2)$ by setting the TPL slope to $b = 2$. The original TPL regression follows the equation $v = a + bm$, where $m = \log(\hat{\mu})$, in this case with $b < 2$ (bold line). Individual points (m;v) (e.g. the grey point) deviate from the regression by u_i. To set $b = 2$, the original regression line is rotated around the average value $\overline{m}$ over all m, resulting in the thin line represented by the equation $v = \tilde{a} + 2m$. Because $v^* = a + b\,\overline{m} = \tilde{a} + 2\,\overline{m}$, the adjusted intercept $\tilde{a}$ can be calculated as $\tilde{a} = a + (b-2)\,\overline{m}$ (see Eq. (5)). The adjusted value for $\log(\hat{\sigma}^2)$ (represented by the white point) can then be calculated by inserting m into the new $v = \tilde{a} + 2m$ and adding u_i. This results in $\tilde{v}_i = 2m_i + \tilde{a} + u_i$.

$$\tilde{v}_i = \tilde{a} + 2m_i + u_i \tag{2}$$

$$\text{With} \quad \tilde{a} = a + (b-2)\overline{m} \tag{3}$$

where $\overline{m} = \frac{1}{n}\sum m_i$.As illustrated and explained in Fig. 1, this procedure adjusts v by setting the TPL slope to $b = 2$, and rotating it around $\overline{m}$.

2.4. Fourth step

The final step is using the adjusted logarithm of the variance for calculating the adjusted coefficient of variation $\tilde{c}_i$ = aCV$_i$.

$$\tilde{c}_i = \frac{\sqrt{g^{\tilde{v}_i}}}{\hat{\mu}_i}\cdot 100\% \tag{4}$$

where g is the basis of the logarithm (which was 10 in our case). Combining and simplifying the Eqs. (1)–(4) leads to

$$\tilde{c}_i = \frac{1}{\hat{\mu}_i}\left[10^{(2-b)m_i+(b-2)\overline{m}+v_i}\right]^{0.5}\cdot 100\% \tag{5}$$

These calculations are exemplified in Table 1 for a small subset of the wheat yield data.

If TPL holds, then for a particular point i, the *anticipated* coefficient of variation $\hat{c}_i$ is

$$\hat{c}_i = \hat{\mu}_i^{\frac{b}{2}-1} g^{\frac{a}{2}} \cdot 100\% \tag{6}$$

Here, the TPL slope a and the intercept b are used to calculate the coefficient of variation that would be expected for $\hat{\mu}_i$ if there was no deviation from the TPL regression line (i.e. if u = 0). When $b < 2$, the anticipated CV decreases non-linearly with increasing mean. For adjusting the coefficient of variation we have removed the dependence of the CV from the mean by setting the slope b to 2 in Eqs. (2) and (3), so that $\hat{\mu}_i^{\frac{b}{2}-1} = \hat{\mu}_i^{0} = 1$.

In short, the adjusted coefficient of variation sets the TPL slope to 2, so that the dependence from the mean disappears.