Field Crops Research 183 (2015) 294–302

Contents lists available at ScienceDirect

Field Crops Research

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

Taylor's power law and the stability of crop yields

Thomas F. Döring[a,*], Samuel Knapp[b], Joel E. Cohen[c]

[a] Department of Agronomy and Crop Science, Albrecht-Thaer-Institute of Agricultural and Horticultural Sciences, Faculty of Life Sciences, Humboldt University Berlin, Albrecht-Thaer-Weg 5, 14195 Berlin, Germany
[b] Department of Plant Sciences, Technical University of Munich, Emil-Ramann-Straße 2, 85350, Freising-Weihenstephan, Germany
[c] Laboratory of Populations, Rockefeller University & Columbia University, 1230 York Avenue, Box 20, New York, NY 10065, USA

ARTICLE INFO

Article history:
Received 19 May 2015
Received in revised form 5 August 2015
Accepted 6 August 2015

Keywords:
Coefficient of variation
Crop
Finlay–Wilkinson regression
Stability
Taylor's power law

ABSTRACT

Taylor's power law (TPL) describes the empirical relationship $\sigma^2 = a\mu^b$ where σ^2 are sample variances and μ are sample means in subsets of data in a data set. Equivalently, TPL states that the logarithm of the sample variance is a linear function of the logarithm of the sample mean across different subsets of data. Here we show that crop yields follow this relationship in several published data sets from varied situations. We show that TPL is frequently, but not always, valid for various factors structuring the data including varieties, crop species, trial environments or countries. We propose that the residuals from the linear regression of $\log(\sigma^2)$ against $\log(\mu)$ can be used as a measure of stability, called POLAR (Power Law Residuals). We compare POLAR stability with other commonly used measures of stability, and show that POLAR stability offers an advantage over some frequently used stability measures.

1. Introduction

In 1961, the British ecologist and entomologist Roy Taylor reported a series of observations that showed surprisingly similar patterns in data from several insect groups, earthworms and other invertebrates as well as plant viruses and fish (Taylor, 1961). For all these groups, population counts or other measures of density showed a power-law relationship between the sample variance σ^2 and the sample mean μ. This relationship, $\sigma^2 = a\mu^b$, became known as Taylor's Power Law (TPL), or Taylor's Law of fluctuation scaling (Cohen, 2013), though the relationship had been published and used earlier by other researchers. Logarithmic transformation of the equation $\sigma^2 = a\mu^b$ results in a linear relationship: $\log(\sigma^2) = \log(a) + b\log(\mu)$. Taylor, who originally found this relationship for data from aerial sampling of aphids, suggested that aggregation is key to understanding the mechanisms underlying this power law and he interpreted b as an index of aggregation.

Subsequently, TPL was verified in hundreds of further species (e.g., Eisler et al., 2008), with data stemming from a wide variety of ecosystems, from bacterial cultures in a laboratory (Ramsayer et al., 2012) to forests (Cohen et al., 2012; Cohen et al., 2013). TPL was also found to be valid for a tremendous range of non-biological phenomena including traffic of Internet routers (Duch & Arenas, 2006), transactions of the New York Stock Exchange, rainfall, and the printing activity of a large university printer (Eisler et al., 2008). Because of its ubiquity, TPL has attracted wide attention from empirical researchers and theorists of multiple, often unrelated disciplines.

As data demonstrating TPL are not restricted to ecological research, the original notion suggested by Taylor and co-workers that TPL follows from explicit behavioural mechanisms such as aggregation and migration has become untenable. An alternative explanation of TPL (Anderson et al., 1982) is that TPL has stochastic roots and that behavioural mechanisms are not necessary to generate data sets exhibiting TPL. In this view, TPL can result from demographic and environmental stochasticity. TPL can indeed be generated by very different population growth models such as the Lewontin-Cohen random walk (Cohen et al., 2013) and the exponential model (Cohen, 2013). While debate on stochastic vs. behavioural explanations of TPL continues (Kendal and Jørgensen, 2011; Arruda-Neto et al., 2012), it has recently been suggested that both perspectives are valid in that "the behavioral models have the potential to explain how the interactions of individuals could generate the population-level phenomenology summarized by the stochastic models" (Cohen, 2013).

In a related area of research, H. Fairfield Smith showed linear relationships between the logarithm of plot size in so-called 'blank' field experiments (without any treatments) and the logarithm of the variance of wheat yield per plot (Smith, 1938). Smith, however, was not interested in the relationship between means and variance

---

* Corresponding author. Tel.: +4930209346467.
E-mail address: thomas.doering@agrar.hu-berlin.de (T.F. Döring).

http://dx.doi.org/10.1016/j.fcr.2015.08.005
0378-4290/

per se, but his main interest was to use this relationship to find an optimal plot size (also see Taylor et al., 1999).

As far as we are aware, TPL has not yet been applied to or tested for crop yields. In this paper we demonstrate the validity of TPL for crop yields in various contexts and on various scales. Further, we introduce a new measure of yield stability which is based on TPL. We call this new index POLAR stability (based on POwer LAw Residuals) and compare this approach with the behaviour of two commonly used stability indices, the coefficient of variation (CV) (Francis and Kannenberg, 1978) and the Finlay–Wilkinson (FW) regression slope (Finlay and Wilkinson, 1963).

Material and methods

Data sets and data filtering

Data sets were analysed to represent crop yield variation (1) at various levels including plot level (within field trials) (Seufert et al., 2012); (2) between environments (years and trial sites) (Jones et al., 2010; Vlachostergios et al., 2011); and (3) at a global level (national yield statistics from the Food and Agriculture Organization database, FAOStat) (FAOSTAT, 2014).

The dataset from Seufert et al. (2012) was collected for a meta-analysis comparing the yields of conventionally and organically managed crops. The dataset contains observations from published studies where for each observation the mean yields and corresponding standard deviations were collected. The original dataset, containing 316 paired observations of conventional and organic yields, was downloaded from the supplementary files published online with the original study.

The data were then filtered in the following way. First, all data not convertible into units of t ha^{-1} were discarded (e.g., units of boxes or bales per ha, or units of kg per plant). Data in bushels per acre were converted as 1 bu ac^{-1} = 0.06277 t ha^{-1} for maize and 0.06725 t ha^{-1} for wheat and soybean. Second, the original dataset contained some double or multiple entries where means and variances were equal among different observations within studies. For each system separately, only one of the equal observations within studies was kept. The final dataset contained 268 observations from the organic system and 228 for the conventional system, i.e., a total of 496 data points from 33 crop species. Observations from maize, wheat, tomato and soybean dominated the dataset with 128, 80, 45 and 43 observations, respectively.

The second dataset is from a study on the adaptability of different lentil genotypes to organic farming, conducted in Greece (Vlachostergios et al., 2011). Twenty genotypes were grown in five environments (three years in one location and two years in another location). In each environment (i.e., year × location combination), the lentils were grown under both organic and conventional cropping management, in separate trials, and with three replicates each. Calculations are based on the reported means over three replicates.

In the third dataset, 19 genotypes of wheat, from two quality groups (milling and feed), were grown at two organic sites (Sheepdrove and Wakelyns) and two conventional sites (Metfield and Morley) in the UK over three years (Jones et al., 2010); the published data are the means of three replicates per trial environment. For calculation of TPL, means and variances for each genotype are calculated across the environments. For the lentil and the wheat study, means and variances were calculated across (a) all organic environments ($n_{L,O} = 5$ for lentils, $n_{W,O} = 6$ for wheat); (b) all conventional environments ($n_{L,C} = 5$, $n_{W,C} = 6$); and (c) all environments ($n_{L,A} = 10$, $n_{W,A} = 12$).

The FAO dataset was downloaded in November 2014 from http://faostat3.fao.org/download/ For wheat, rice and potatoes, data were extracted for area (recorded in ha) and yield (recorded in hg ha^{-1}). These data were then filtered. First, only those countries with a complete set of the last ten years (2004–2013) were retained. Second, countries where average area between 2004 and 2013 was below 10,000 ha were excluded. In addition, to exclude data that indicated low reliability, we excluded countries in which yields were rounded to 1000 hg ha^{-1} in two or more of 10 years and countries where yields were equal in consecutive years in two or more of 10 years. For wheat, rice and potatoes, these filters resulted in final datasets of crop yields from 90, 73 and 89 countries, representing 99.3%, 97.9% and 96.8% of the global area grown with these crops in 2013, respectively.

Calculations and statistical analysis

All analyses were performed with the programme R, version 3.0.0. Means (μ) and variances (σ^2) were calculated per variety over all environments for the lentils and wheat dataset, and per country over all years for the FAO dataset. Subsequently, a linear regression was calculated for $\log_{10}$ of the variance over the $\log_{10}$ of the mean. In order to test the linear relationship between $\log(\mu)$ and $\log(\sigma^2)$ we applied a forward selection procedure by starting with a linear model and subsequently adding higher order polynomials (quadratic and cubic terms) and evaluating the model fits by the Akaike information criterion (AIC), i.e., terms were added until the AIC increased (Burnham and Anderson, 2002). Cubic terms were never significant in any of the analyses.

Residuals from the linear regression models were tested for normality, skewness (m_3) and kurtosis (m_4). Deviation from normality was tested with the Shapiro–Wilk test (Royston, 1982b). Deviation of skewness from zero and deviation of excess kurtosis ($m_4 - 3$) from zero were tested with *t*-tests on $n - 2$ degrees of freedom (Crawley, 2013); ('excess kurtosis' = 0 is equivalent to kurtosis = 3, since the normal distribution has a kurtosis of 3). In case of non-normal residuals, the variable $y = \log(\sigma^2)$ was transformed using the Box–Cox transformation ($z = (y^\lambda - 1)/\lambda$) with maximum likelihood optimisation of λ (Crawley, 2013), in order to quantitatively describe the behaviour of the residuals. Homoscedasticity was checked visually using the plot(model) function in R and with the Goldfeld–Quandt test (Goldfeld & Quandt, 1965) using the *gqtest* function in the R library *lmtest*. In addition, the relationship between $\log(\mu)$ and $\log(\sigma^2)$ was tested with Spearman's rank correlation coefficient as a non-parametric test not dependent on the assumptions of linear regression.

2. Results

Taylor's power law: testing validity for crop yields at various scales

Here we show that TPL is sometimes but not universally valid for crop yields across various scales (across environments, and at a global level) and for various factors structuring the data (genotypes, crop species, countries, years).

For the dataset from Seufert et al. (2012), based on variances of crop yields mainly within field trials, there was a highly significant ($P < 0.001$) positive correlation between $\log(\mu)$ and $\log(\sigma^2)$ (Fig. 1a), whether the organic and conventional systems were analysed separately or together. Paired observations from the two systems (organic and conventional) of both means and variances are stochastically dependent. Strictly speaking, this would affect the validity of significance tests for the regression and also tests of assumptions. However, similar results were obtained when fitting regressions within systems separately (Fig. 1a), where means and variances were stochastically independent. According to the Goldfeld–Quandt test, the data were not significantly heteroscedas-

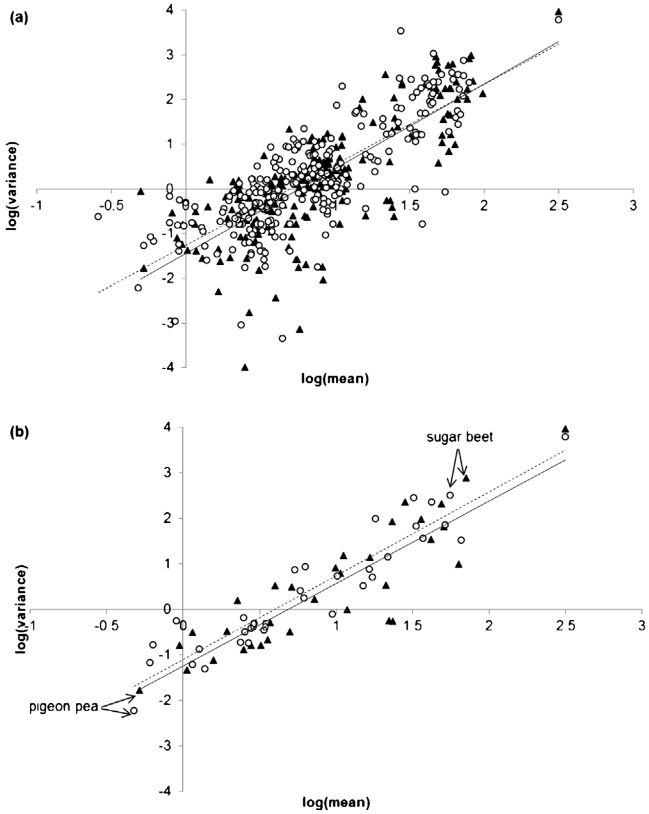

Fig. 1. Relationship between $x = \log_{10}$(spatial mean) and $y = \log_{10}$(spatial variance) for crop yields as published in a meta-analysis comparing conventional and organic yields in multiple field trials (Seufert et al., 2012), with (a) each point representing one observation from one management system and (b) each point representing one crop species from one management system. The organic system is represented by the open circles and the dashed regression line ((a): $y = a + bx$, with $a = -1.255 \pm 0.082$, $b = 1.799 \pm 0.084$ (S.E.), adj. $R^2 = 0.630$, df = 266; (b): $a = -1.103 \pm 0.129$, $b = 1.839 \pm 0.119$, adj. $R^2 = 0.882$, df = 31). Conventional systems are shown by black triangles and the continuous regression line ((a): $a = -1.447 \pm 0.105$, $b = 1.90 \pm 0.103$, adj. $R^2 = 0.601$, df = 226; (b): $a = -1.237 \pm 0.197$, $b = 1.810 \pm 0.173$, adj. $R^2 = 0.772$). The slopes of the two regression lines were not significantly different from each other in either case ((a): difference = -0.101 ± 0.132, $P = 0.445$; (b): difference = -0.030 ± 0.208, $P = 0.887$); the common slope for the two systems was $b = 1.837 \pm 0.065$ in (a); and $b = 1.817 \pm 0.103$ in (b).

tic. The slopes of $\log(\sigma^2)$ against $\log(\mu)$ for organic and conventional systems were not significantly different ($P = 0.445$). Also, separate intercepts for both systems were not significantly different ($P = 0.115$), i.e., the most parsimonious model was a common linear model for both systems.

The relationship between $\log(\mu)$ and $\log(\sigma^2)$ was not linear, contrary to TPL, as there was a highly significant ($P < 0.001$) positive quadratic term when the organic and conventional systems were analysed together. However, adding the quadratic term to the linear model reduced the residual standard deviation (RSD) from 0.743 to 0.7293, i.e., by only 1.8%; accordingly, the convexity of the resulting function was weak. The quadratic term was also significantly positive when the two farming systems were analysed separately ($P < 0.01$ for both systems), with a corresponding reduction in the RSD by 1.6% and 2.1%, for the organic and conventional system, respectively.

In the oats yield data only, the TPL regression had a significantly negative slope. The oats data originated from three studies, with estimates of means and variances being based on 2, 3 and 10 observations, respectively. When the dataset was restricted to studies for which the means and variances were based on 10 or more observations, the slope of the TPL regression was positive ($b = 2.78 \pm 1.92$, df = 4, Adj. $R^2 = 0.1779$, $P = 0.223$).

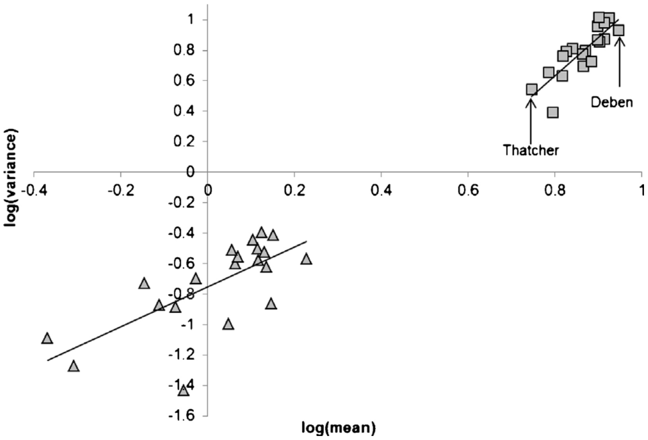

Fig. 2. Relationship between $x = \log_{10}$(mean) and $y = \log_{10}$(variance) for (a) grain yields of 20 lentil genotypes (grey triangles) (Vlachostergios et al., 2011); and (b) grain yields of 19 wheat genotypes (grey squares) (Jones et al., 2010). In both datasets, each point represents one genotype in environments that varied in space, time, and management. Statistics of regression lines are in Table 1.

The relationship between $\log(\mu)$ and $\log(\sigma^2)$ for individual crop species with ten or more observations per system in the filtered dataset was much less robust than when all crops were analysed together (Table 1). For some crop species, namely maize, soybean, barley and spinach, the regression slope was not significantly different from zero. However, when individual observations were averaged per crop within the systems, resulting in 66 observations in total, the regression across all crop species of $\log(\sigma^2)$ against $\log(\mu)$ was highly significant (Table 1, Fig. 1b). For this analysis across crop species, there was no significant deviation from linearity regardless of whether the organic and conventional systems were analysed separately or together. Generally, high significance and low S.E. values for the regression slopes were mainly observed when the range of values spanned by $\log(\mu)$ was large (Table 1).

For the lentil dataset (Vlachostergios et al., 2011), there was a highly significant linear relationship between $\log(\mu)$ and $\log(\sigma^2)$ (Fig. 2, $P < 0.001$, df = 18) and no significant deviation from linearity ($P = 0.692$, df = 17). When the means and variances were calculated within management systems, the relationship between $\log(\mu)$ and $\log(\sigma^2)$ was evident for the conventional, but not the organic systems (Table 1). For neither of the two systems considered separately was there any significant quadratic term. However, significant heteroscedasticity ($P < 0.05$) was found for the conventional lentil data subset.

For the wheat dataset from Jones et al. (2010), variances and means were calculated across environments for each variety. There was a highly significant linear relationship between $\log(\mu)$ and $\log(\sigma^2)$ (Fig. 2). The wheat dataset showed a significant linear relationship between $\log(\mu)$ and $\log(\sigma^2)$ in the organic, but not in the conventional, systems (Table 1). The Goldfeld–Quandt test did not detect significant heteroscedasticity in any subset of the data from Jones et al. (2010).

The non-significant TPL slope for the conventionally grown wheat yields from Jones et al. (2010) was robust: omitting any one of the six conventional environments from the dataset always resulted in non-significant TPL slopes; the slopes had a negative sign in five out of six cases and in the remaining case (when Metfield data from the second year were omitted), the slope was close to zero ($b = 0.09 \pm 1.14$). In comparison, the organic wheat yields from Jones et al. (2010) retained positive TPL slopes with omission of individual environments (analyses not shown). In addition, the TPL slopes of the conventional wheat remained non-significant when both groups of wheat cultivars (milling cultivars and feed cultivars, see original study) were analysed separately, whereas

Table 1
Linear regression parameters for $y = a + bx$ between $x = \log_{10}(\mu)$ and $y = \log_{10}(\sigma^2)$ for yield data (μ = mean; σ^2 = variance; unit of original data: t ha^{-1}) from various sources; Org: organic; Con: conventional; the range Q is the difference between largest and smallest value of $\log(\mu)$, i.e. $Q = \max(\log(\mu)) - \min(\log(\mu)) = \log(\max(\mu)/\min(\mu))$; (*)$P < 0.1$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ns: not significant; NA: not available; §Significance level, linear term.

Source dataset	Crops	μ and σ^2 calculated across	df	Farming system#	Min(μ) (t ha^{-1})	Max(μ) (t ha^{-1})	Range Q	a	SE(a)	b	SE(b)	Adj. R^2	Sign.§
Seufert et al. (2012)	Maize	Environments	126	Both	0.50	15.99	1.50	0.027	0.215	0.248	0.251	−0.000	ns
	Wheat		78	Both	0.58	7.40	1.10	−1.195	0.219	1.002	0.409	0.060	*
	Tomato		43	Both	2.10	97.79	1.67	−0.497	0.622	1.328	0.373	0.209	***
	Soybean		41	Both	1.10	3.67	0.52	−0.910	0.364	0.601	0.830	−0.011	ns
	Barley		26	Both	1.63	6.41	0.59	−0.643	0.445	0.292	0.741	−0.032	ns
	Spinach		16	Both	7.20	23.60	0.52	−0.245	0.644	0.129	0.641	−0.060	ns
	Oats		11	Both	1.96	5.83	0.47	1.312	0.520	−4.026	1.016	0.551	**
	Apple		11	Both	9.40	48.24	0.71	0.089	0.745	1.414	0.554	0.315	*
	Crop averages	Crop species	64	Both	0.48	316.80	2.82	−1.164	0.115	1.820	0.110	0.826	***
Vlachostergios et al. (2011)	Lentil	Environments	18	Org	0.41	1.56	0.59	−0.790	0.060	0.535	0.384	0.047	ns
			18	Con	0.45	1.81	0.60	−0.791	0.056	1.953	0.339	0.630	***
			18	Both	0.43	1.69	0.60	−0.754	0.048	1.304	0.311	0.494	***
Jones et al. (2010)	Wheat	Environments	17	Org	4.47	6.42	0.16	−0.692	0.300	1.531	0.415	0.412	**
			17	Con	6.67	11.25	0.23	0.520	1.010	−0.574	1.039	−0.040	ns
			17	Both	5.57	8.83	0.20	−1.375	0.352	2.512	0.407	0.675	***
		Genotypes	10	Both	2.54	10.58	0.62	−1.637	0.436	1.939	0.510	0.550	**
FAOSTAT (2014)	Wheat	Countries	88	NA	0.68	8.78	1.11	−1.224	0.083	0.781	0.160	0.204	***
	Rice		71	NA	1.17	9.72	0.92	−1.300	0.184	0.728	0.311	0.058	*
	Potatoes		87	NA	2.84	46.79	1.22	−1.383	0.300	1.363	0.235	0.271	***

the organic wheat data retained significantly positive TPL slopes for both milling cultivars ($b = 1.33 \pm 0.56$, df = 9, $P = 0.04$) and feed cultivars ($b = 2.53 \pm 0.92$, df = 5, $P = 0.04$).

Deviation from linearity was significant for the organic systems (significant quadratic terms, $P < 0.05$) but not the conventional systems, nor when both systems were considered together. When variances and means were calculated across varieties within each of the 12 environments, there was again a significant linear relationship.

For all three crop species from the FAOStat yield dataset, there were significant linear relationships between $\log(\mu)$ and $\log(\sigma^2)$ (Fig. 3, Table 1). Deviations from linearity of the regression lines of $\log(\sigma^2)$ against $\log(\mu)$ were not significant in any of the three crop species. For the wheat data, but not for the rice and potato data, the Goldfeld–Quandt test showed significant heteroscedasticity ($P < 0.05$).

Consequences of Taylor's Power Law for some commonly used yield stability measures

In the FW regression approach to yield stability analyses (Finlay and Wilkinson, 1963), sometimes also called 'ecoregression' (Meinel et al., 1997), yields of individual genotypes are displayed in a scatter diagram against the environmental means, which are the means over a set of genotypes within each environment. These environmental means are used as a surrogate environmental index for the yielding ability or quality of the respective environment. If a genotype (G) shows no interaction between genotype performance and environment, the slope β_G of the regression line of genotype G yields against the environmental means is 1. Genotypes that show relatively better performance in poor environments (defined as environments with lower mean yields) have slopes $\beta_G < 1$, whereas genotypes with disproportionately higher yields in good environments (defined as environments with higher mean yields) have slopes $\beta_G > 1$.

What are the consequences if, as our previous results indicate, variances scale with environmental means in accordance with TPL? For example gain yield data of genotypes of wheat from Jones et al. (2010) resulted in $\log(\sigma^2) = -1.637 + 1.939 \times \log(\mu)$. Consequently, the confidence interval for yields of individual genotypes or individual varieties becomes broader as the environmental mean increases (Fig. 4). These increasing confidence interval could arise

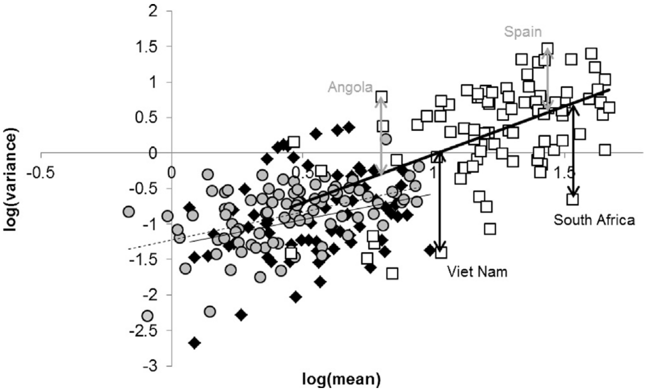

Fig. 3. Relationship between $x = \log_{10}$(temporal mean) and $y = \log_{10}$(temporal variance) for national yield data of wheat (grey circles, dashed line), rice (black diamonds, continuous line) and potatoes (white squares, bold line). Data are from FAOSTAT for 2004–2013. Each point represents one country; mean and variances are over years. For statistics of regression lines see Table 1. Low residual potato yield variability (i.e. high temporal yield stability) is indicated for two examples, Viet Nam and South Africa (black arrows, negative residuals from the regression line, i.e., POLAR < 0). Angola and Spain show positive residuals (POLAR > 0), i.e. low yield stability (grey arrows).

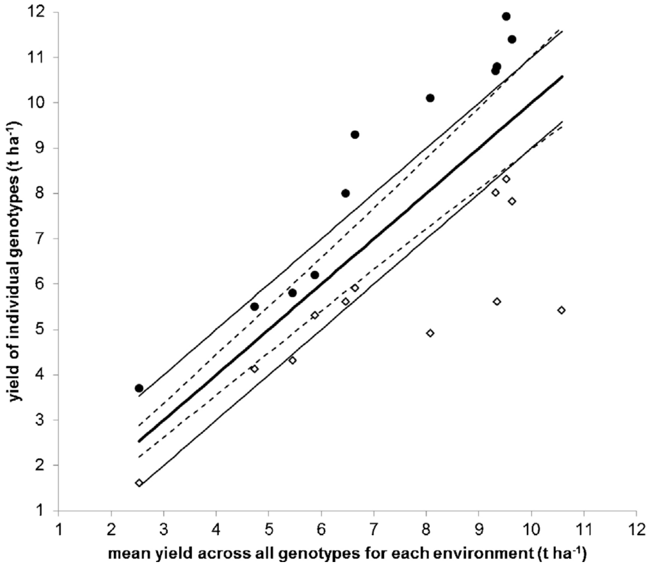

Fig. 4. Schematic representation of FW regression, based on grain yield data of wheat from Jones et al. (2010). Yields of individual genotypes are plotted against the environmental means, exemplified by the genotype with the highest and lowest mean yield, Deben (black circles) and Thatcher (white diamonds) respectively. Values on the *x*-axis are environmental means, i.e. yields are averaged across all genotypes in each environment. Around the bold line ($y = x$), the two fine lines ($y = x + 1$ and $y = x - 1$) show hypothetical genotypes for which there is no interaction between genotype (G) and environment (E). Traditional stability analyses interpret slopes β_G of genotype-specific lines deviating from $\beta_G = 1$ as an indication of G × E interactions. However, if variances scale with the environmental means, as predicted by Taylor's power law (TPL) and as demonstrated for this wheat dataset, the confidence interval for values of individual genotypes, indicated by the dashed lines, becomes broader with an increasing environmental mean. Both axes are on the original scale of measurement, not on the logarithmic scale as in TPL.

if genotypes with below-average yields have FW regression slopes of $\beta_G < 1$ while genotypes with above-average yields have slopes $\beta_G > 1$. Alternatively, increasing confidence intervals could arise if $\beta_G = 1$ for all genotypes, and the absolute value of residuals d_G from the FW regression lines (i.e., the difference between the yield of an individual genotype and the yield predicted by the regression across environmental means) increased with increasing environmental mean. The variance of the residuals d_G is often used as a further stability measure along with the FW regression slope β_G (Eberhart and Russell, 1966).

The wheat data support the hypothesis of increasing FW regression slope β_G with increasing individual mean yield for genotype *G*. Values of the genotype-specific slopes β_G calculated for the 19 varieties were highly positively correlated with mean yields of the varieties (Adj. $R^2 = 0.716$, df = 17, $P < 0.001$), while the variances $({\sigma^2}_d)_G$ of the residuals were negatively, but less strongly, correlated with the mean (Adj. $R^2 = 0.250$, df = 17, $P < 0.05$).

A further commonly used measure of yield stability is the coefficient of variation, $CV = \sigma/\mu$, where σ is the sample standard deviation and μ the sample mean. This stability index is popular with ecologists (Tilman et al., 1998; Tilman, 1999; Anderson et al., 2013) and agronomic researchers and is frequently calculated along with other stability measures (Francis and Kannenberg, 1978; Smith et al., 2007; Dehghani et al., 2008). However, if TPL holds, as it does for (some) crop yield data, then the CV may change non-linearly with increasing mean because $CV = \mu^{b/2-1} g^{a/2}$, 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. Consequently, if $b > 2$ (as for conventional and organic wheat yields combined, in the data of Jones et al., 2010; Table 1), CV increases with increasing μ; if $b < 2$ (as in all remaining cases in Table 1), CV decreases with increasing μ; and if $b = 2$, CV is constant and independent of μ. The FAO wheat data analysed in Table 1 have $b = 0.781 < 2$, hence CV decreases with increasing mean yield μ. The non-linear decrease of CV with increasing mean is also evident for the independent FAO dataset of the wheat yields in 1994–2003 ($n = 90$ countries) (Fig. 5).

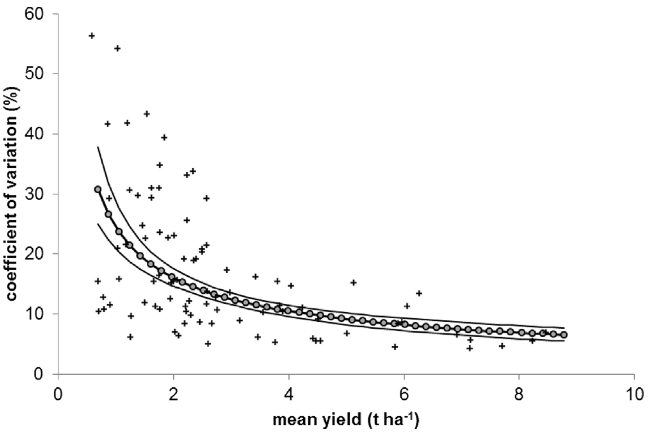

Fig. 5. Coefficient of variation (%) depending on mean yield for national yield data of wheat from 2003–2014 (grey circles, based on TPL regression, including confidence interval), and from 1994–2003 (crosses); data from FAOSTAT (2014). Both axes are on the original scale of measurement, not on the logarithmic scale as in TPL.

Characterisation and interpretation of residuals from TPL regression

For the data from Seufert et al. (2012), residuals from the linear TPL regression were non-normal, with significantly negative skewness and significantly positive excess kurtosis (Table 2). Negative skewness of residuals was also found for the lentil dataset (Vlachostergios et al., 2011), the wheat dataset (Jones et al., 2010), and for wheat and potatoes of the FAO dataset. Deviation from normality of the residuals was significant in half of the cases, but excess kurtosis was mostly not significantly different from zero. For the data from Seufert et al. (2012), where individual means and variances were based on different numbers *n* of measurements, residuals from the linear TPL were not significantly correlated with *n* (adj. $R^2 < 0$, $P > 0.5$, df = 378). Non-normality of residuals is a violation of the assumptions of linear regression. However, Spearman's rank correlation coefficient as a non-parametric way to measure the strength of the relationship between $\log(\mu)$ and $\log(\sigma^2)$ showed significant positive relationships between the two variables in all cases (Table 2) in which linear regression had been found to be significant. Further, with Box–Cox transformation of the data, normality of residuals could be obtained in most but not all cases.

The residuals from the linear TPL regression can be interpreted as a measure of yield stability (Fig. 3); positive residuals indicate that the $\log(\sigma^2)$ is larger than expected for a given $\log(\mu)$, and negative residuals indicate that the $\log(\sigma^2)$ is smaller than expected, i.e. that the yield is less variable. We call these residuals POLAR (POwer LAw Residuals). POLAR stability is measured in units of log(variance), not variance per se, because POLAR are not residuals on the variance scale from the power law, but residuals from the linear regression of log (variance) as a function of log (mean).

There was no significant correlation between these POLAR stability values and the original yield means μ for any of the data analysed (Adj. $R^2 < 0.01$ and $P > 0.6$ for all three crop species in the FAO dataset, as well as for the datasets from Jones et al. (2010) and Vlachostergios et al. (2011); for data from Seufert et al. (2012), Adj. $R^2 = 0.003$; $P = 0.129$).

In comparison to the results obtained with non-transformed data, Box–Cox data transformation for the normalisation of residuals changed the stability ranking of individual entries, though

Table 2

Characterisation of the residuals from the TL regression; #Org: organic; Con: conventional; deviation from normality was tested with the Shapiro–Wilk test (Royston, 1982a); (*)$P<0.1$; *$P<0.05$; **$P<0.01$; ***$P<0.001$; ns: not significant; NA: not available; heteroscedasticity was tested with the Goldfeld–Quandt test; Spearman's ρ is the rank correlation coefficient for the relationship between log(mean) and log(variance); λ is the Box–Cox transformation parameter for the transformation $z=(y^{\lambda}-1)/\lambda$, with $y=\log(\text{variance})$; when Box–Cox transformation was not possible because of negative values on the logarithmic scale, original data (means and standard deviations) were first multiplied by 10 (data from Jones et al. (2010)) or by 100 (data from Vlachostergios et al. (2011) and from FAOSTAT (2014)) prior to the Box–Cox transformation.

			Original data								Transformed data	
Source dataset	Crops	Farming system#	Skewness		Excess kurtosis		Deviation from normality	Heteroscedasticity	Spearman's ρ		λ	Deviation from normality
Seufert et al. (2012)	Various	Org & Con	−0.95	***	2.39	***	***	ns	0.770	***	1.24	***
Vlachostergios et al. (2011)	Lentil	Org	−0.79	(*)	−0.97	ns	**	ns	0.386	ns	5.45	*
		Con	−0.57	ns	−0.11	ns	ns	*	0.774	***	–	–
		Both	−1.39	*	1.42	ns	**	ns	0.671	**	6.01	ns
Jones et al. (2010)	Wheat	Org	−0.46	ns	−0.96	ns	ns	ns	0.712	***	–	–
		Con	−1.47	**	2.17	*	**	ns	−0.252	ns	4.13	ns
		Both	−0.73	ns	−0.10	ns	ns	ns	0.868	***	–	–
FAOSTAT (2014)	Wheat	NA	−0.49	*	0.11	ns	ns	*	0.431	***	–	–
	Rice	NA	0.13	ns	−0.12	ns	ns	ns	0.252	*	–	–
	Potatoes	NA	−0.57	*	−0.25	ns	*	ns	0.439	***	2.77	ns

these changes were small. In the organic subset from Vlachostergios et al. (2011), 12 out of 20 lentil genotypes retained their POLAR stability ranking, and 6 changed by only one rank; in the joint dataset (organic and conventional), 4 genotypes did not change their ranking, whereas 10 changed by one rank. Similarly, for the conventional wheat data from Jones et al. (2010), 14 wheat genotypes retained their POLAR stability ranking. In the potato data from FAO, 64 of 89 countries changed POLAR stability ranking by three or fewer ranks following data transformation.

3. Discussion

Validity of Taylor's power law for crop yields

For several diverse datasets of crop yields, $\log(\sigma^2)$ increased with increasing $\log(\mu)$; in several cases, this relationship was linear. Given the wide applicability of TPL (Eisler et al., 2008), this finding is not surprising. However, it raises the question why TPL has not been shown for crop yields before (to our knowledge).

One plausible reason is that TPL tends to become more visible when the range of $\log(\mu)$ is large. Taylor et al. (1988a,b) recommended that "the range of values of values of $\log(\mu)$ in the regression form of the power law... should be as large as practicable." (also see Taylor et al., 1998a,b.) Though not all datasets analysed here conform to this suggestion (e.g., wheat from organic environments in Jones et al., 2010), in general, the data with the largest overall range (from Seufert et al. (2012)) point into this direction (Table 1). If a larger range does render TPL more likely to become evident in crop data, then the chance of observing TPL will be low when comparing treatments or genotypes with similar means. Small ranges of means in crop yield data, and associated low detectability of relationships between means and variances may also be the reason why logarithmic transformation of yield data, though suggested early (Finlay & Wilkinson, 1963) and also used more recently (Piepho et al., 2014), is currently not common. In datasets with a relationship of means and variances, the assumption of homogeneity of variances may be achieved through a log-transformation. However, when crop yields are log-transformed TPL is likely to disappear e.g., when the original wheat data from Jones et al. (2010) were log-transformed prior to the calculation of means and variances, the relationship of means and variances was non-significant (Spearman's rank correlation: $P=0.90$).

A second reason why TPL has not been shown in crop yields so far may be that many measurements of yield are required to do so. For example, the unexpected negativity of the TPL slope for oats yield data reported in Table 1 disappeared when data were excluded from the analysis that were based on small numbers of yield measurements for each estimate of a mean and a variance. Taylor et al. (1988a,b) suggested that when testing TPL, at least 15 observations (here, measurements of yield) should be available for each estimate of a mean and a variance, and at least 5 pairs (μ, σ^2) should be available for the regression of $\log(\sigma^2)$ against $\log(\mu)$. Taylor et al. (1988a,b) provided little statistical justification for these recommendations. Most estimates of means and variances reported in this paper are based on fewer than 15 measurements of yield.

Further, Taylor et al. (1998a,b) suggested that TPL is most reliable when sampling is conducted consistently, i.e. when the same statistical population is being sampled. The study by Fairfield Smith (1938) (and previous work cited therein) indicated that plot size is a major factor determining variance of yield. Consistent sampling, e.g. in terms of identical plot size in all environments, is likely to be the case with the data from Vlachostergios et al. (2011) and from Jones et al. (2010), but within the data from Seufert et al. (2012)

and within the FAO data, compilations are based on different data collection methods. Despite this, TPL fits well for both the Seufert et al. data and the FAO data.

In view of the limitations of sample size and sampling consistency, we regard the present results as (possibly) the first, exploratory word rather than the last, definitive word on crop yields and TPL. Further research will need to investigate how TPL in crop yield data depends on numbers and groupings of observations, as well as on methods of crop yield measurement.

A further important point highlighted by our analysis is the non-normality of residuals for TPL log–log regression of some of the crop yield datasets (Table 2). When the residuals are not normal, or when any other assumption made in linear regression is violated, we can still fit the model to the data by least squares, but the confidence intervals of intercept and slope and the significance level assigned to the linear correlation are no longer valid. (Ordinary least squares regression also assumes that the values of the x-coordinate, here $\log(\mu)$, are known with no sampling or other error. This assumption is violated in all of our data, though the effect of this violation is likely to be small.) However, a non-parametric test that does not rest on the assumptions made for linear regression (Spearman's rank correlation, Table 2) reveals significant relationships between $\log(\mu)$ and $\log(\sigma^2)$ in all cases where ordinary linear regression had shown significance. More caution is due when inferences are made based on estimates of variance (see below).

We regard TPL (and the associated analysis of skewness and kurtosis of residuals) mainly as a useful description of the relationship between mean and variance in crop yield data. The approach presented here provides opportunities of investigating effects of agricultural system design, e.g. comparisons of organic and conventional cropping management. Specifically, the analysis of the wheat data from Jones et al. (2010) suggests that under high-input (conventional) conditions, cultivars with high mean yields have lower than expected yield variability, since the associated TPL slope is not significantly positive. In contrast, the same varieties tested under organic conditions show a significantly positive TPL slope. Although these results need to be interpreted with caution because only two sites per management system were included in the field trials, the findings are plausible and in accordance with existing literature on the interactions between wheat genotypes and cropping system (Murphy et al., 2007). In particular, our results indicate that wheat breeding may have succeeded in combining high mean yields with low yield variability, but that these gains depend on the yield stabilizing effects of high inputs typically found in conventional crop management. However, why similar results were not found in the lentil dataset is open to speculation.

Consequences of Taylor's power law for yield stability measurement

When TPL holds for crop yield data, as is often the case according to our results, some commonly used yield stability measures, such as the FW regression slope β_G, the variance of the residuals, d_G, from the FW regression slope and the coefficient of variation (CV), need to be interpreted with caution.

The FW regression slopes β_G (or alternatively, the absolute residuals from the FW regression) are likely to increase with increasing means (Fig. 4) when the TPL slope b is positive. In most cases, $b > 0$. In Table 1, $b < 0$ only for oats in the data of Seufert et al. (2012) and conventional wheat in the data of Jones et al. (2010). According to our analysis, an FW regression slope $\beta_G \neq 1$ does not necessarily result from G × E interactions. If genotypes i, j have different mean yields ($\mu_i \neq \mu_j$), then TPL with slope $b \neq 0$ is incompatible with all genotypes showing the same yield stability as measured by β_G or absolute residuals from the FW regression. Either the FW regression slopes of the two genotypes must differ to satisfy TPL, or the absolute residuals from FW regression must differ between the two genotypes.

This is also evident when considering some known biometric relationships. According to Becker (1981), the variances $\sigma^2{}_G$ of individual genotypes G across environments ('places') p are a function of the FW regression slope β_G and the variance $(\sigma^2{}_d)_G$ of the deviations d_G from the FW regression, namely, $\sigma^2{}_G = \sigma^2{}_p + 2(\beta_G - 1)\sigma^2{}_p + (\beta_G - 1)^2\sigma^2{}_p + (\sigma^2{}_d)_G = \sigma^2{}_p\beta^2{}_G + (\sigma^2{}_d)_G$. Here $\sigma^2{}_p$ is the variance of the effects of the places p, and $\sigma^2{}_p$ is taken to be the same for all genotypes. Thus, if $\sigma^2{}_G$ increases with increasing μ_G (as in TPL), then the means μ_G must correlate positively with the FW regression slopes β_G or the deviations $(\sigma^2{}_d)_G$ or both. As a consequence, differences in FW regression slopes or absolute residuals from the FW regression among genotypes may not always result from breeding, e.g. selection for stability, but may just be typical distributions of yields that result in variances scaling with the means according to TPL.

Further, Fig. 5 showed that the CV can decrease systematically with increasing mean yield in line with earlier work on wheat yield data (Taylor et al., 1999). We can specify that CV systematically decreases with increasing mean in crop yield data following TPL with $b < 2$, and that CV increases with increasing mean yield if $b > 2$. If CV is used as an index of yield stability, genotypes (or any treatments) with higher means will automatically have greater stability as assessed by smaller CV (if $b < 2$) or smaller stability as assessed by large CV (if $b > 2$). However, it is often desirable to separate yield stability from mean yields in order to detect biological synergies and trade-offs between mean performance and stability of performance.

As the ratio of standard deviation to mean, the CV is designed to make standard deviations more directly comparable across different means. However, when TPL holds with $0 < b < 2$, the CV 'overcompensates' the scaling effect of the mean, so that higher means systematically lead simultaneously to higher variances and lower CVs. Accordingly, in a study comparing several indices of yield stability with a principal component analysis (Dehghani et al., 2008), the CV lay half-way between the mean and various other indices of yield stability.

Thus, both CV and FW regression stability are related to the mean of the treatments if TPL holds. Even if TPL is not strictly valid e.g. because of non-linearity, as long as variances are positively correlated with the means as shown by Spearman's rank correlation (Table 2), both CV and FW regression stability are still related to treatment means. These facts affect attempts to test the hypothesis that there is a trade-off between yield and stability as the observed relation is probably more due to statistical relationships than the actual behaviour of the treatments under test. For example, in plant breeding and variety evaluation, a stability index strongly influenced by the mean would not allow selection of truly stable genotypes. This problem is particularly relevant when means differ widely (Fig. 5), e.g. when comparing the yields of different cropping systems or crop species. For these comparisons, a number of approaches to measure stability have been developed (Piepho, 1998). In practice, however, the CV has often been used (Rao and Willey, 1980; Mesterházy, 1995; Berzsenyi et al., 2000; Smith et al., 2007; Küchenmeister et al., 2012), especially in agronomic research, though mostly in conjunction with other stability parameters.

POLAR as a new stability index

Residuals from the TPL linear regression (POwer LAw Residuals or POLAR) can be interpreted as a measure of yield stability, with lower stability (relative to all samples with that mean yield) indicated by more positive POLAR values, and higher stability (relative to all samples with that mean yield) indicated by more negative POLAR values.

In our data, there was no significant correlation between POLAR stability and mean yields, so POLAR was independent of the mean. Whether POLAR stability will be independent of the mean yields in general depends on the error structure of the relationship between $\log(\sigma^2)$ and $\log(\mu)$. If the residuals on the $\log(\sigma^2)$ scale are homoscedastic, then POLAR stability and mean yields will be independent, fulfilling an important desideratum for stability measures.

Further, the POLAR stability index is easy to calculate e.g., with commonly available spreadsheet programmes, and easy to interpret mathematically and graphically (Fig. 3). If POLAR values are available from different groups within the dataset, significance of differences in POLAR stability between these groups can be tested. Such tests may need to be non-parametric in case of non-normal distributions of POLAR values, as characterised in Table 2. Conventional statistical tests that assume residuals are normally distributed may not be appropriate in all cases. Further, POLAR can be regarded as a special case of a more general approach of stability analysis. This approach first builds a model to describe the relationship between mean and variance and estimates variances for each mean; it then calculates the differences of observed and estimated variance (i.e. the residuals) as a measure of stability.

Over the past few decades, multiple methods have been developed to measure yield stability (Becker, 1981; Becker and Léon, 1988; Gauch, 1992; Eghball and Power, 1995; Piepho, 1998). These measures have advantages and disadvantages. So does POLAR stability. Which genotype is most stable strongly depends on the stability index used (e.g., Dehghani et al., 2008). Research on yield stability has therefore often attempted to demonstrate relationships among different stability measures (Becker, 1981). We suggest that future studies of the consistency of stability measures in ranking genotypes, crop species or agronomic treatments should include POLAR stability among other stability measures, because it takes account of the dependence of the variance on the mean.

A characteristic of TPL is the robustness across datasets uniformly collected, in particular with respect to the slope b. Thus, the TPL response itself may differ among different crop varieties. In this case, changing the ensemble (i.e., subset) of varieties that is tested for stability across a number of environments may change the ensemble of associated TPL gradients. In the realistic case that two subsets A and B of varieties are tested, with some varieties being shared among the two subsets, the common gradient b_A for the subset A may be different from the common gradient b_B for subset B ($b_A \neq b_B$). This will quite likely change the POLAR response of individual varieties, drawing into question its general robustness, since each application may deliver a different result. One way to increase the robustness is to make sure that a sufficiently great number of observations is included. Similar problems are observed for other stability measures (e.g., FW regression slope β); however, more extensive datasets need to be analysed using TPL to elucidate how far POLAR is, or is not, robust against changes of ensemble and variations in the number of observations.

4. Conclusions

High stability of crop yield is an important goal in agricultural production (Becker & Léon, 1988; Piepho, 1998; Annicchiarico, 2009). To generate reliable income each year, farmers seek varieties or crops with low fluctuations in yields over time at the farm level. Since the early days of professional plant breeding, researchers and breeders have sought to develop crop cultivars with high yield stability over time and in multiple locations, so that the genotype can be successfully grown over large areas and in many different environments (Mühleisen et al., 2014; Chamekh et al., 2015; Ingvordsen et al., 2015). At geographic scales larger than an individual farm, temporal yield stability is critical for ensuring food security and reducing the risk of food shortages (McKersie, 2015). Research has developed several indices to characterise yield stability in various contexts and data structures.

Taylor's Power Law suggests a new stability index that can deal with one form of the dependence of the variance on the mean across large ranges of mean values, under assumptions spelled out above. This index (POLAR stability) is easy to calculate and interpret. Further, as shown for the dataset from Seufert et al. (2012), it is applicable to historical data where only variances and means are known, but not the original values. TPL highlights the need to be cautious in interpreting stability measures of crop yields when means vary over large ranges, because some measures of yield fluctuations may be biased when TPL holds. In cases when different entities (such as cropping systems or crop species) with large differences in mean yields are to be compared in terms of their yield stability, it is therefore essential to check for systematic relationships between means and variances; ideally, enough data points should be collected to determine TPL slopes.

We have shown that TPL is widespread but not universal in crop yield data. From the discovery of TPL until to today, the potential underlying mechanisms that lead to the TPL relationship between means and variances have been the focus of numerous studies. It is clear that as in other datasets, the original proposition of aggregation behaviour as the driving force generating TPL is not valid for crop yields. Recently it has been shown that relatively simple, non-biological mechanisms could explain some (but not all) occurrences of TPL in biology (Cohen and Xu, 2015; Xiao et al., 2015). It remains to be investigated in which cases deviations from TPL in crop data result from the size and nature of the data, and in which, alternatively, systematic factors explain these deviations from TPL. While it is too early to spell out implications of TPL for agronomy in detail, benefits can be expected for both crop science and the study of TPL across the sciences from the wide application of the TPL approach to crop yield data.

Acknowledgments

We thank Frank Ellmer and Charlotte Kling for helpful comments during the preparation of this paper. JEC acknowledges with thanks the support of U.S. National Science Foundation grant DMS-1225529 and the assistance of Priscilla K. Rogerson.

References

Anderson, R., Gordon, D., Crawley, M., Hassell, M., 1982. Variability in the abundance of animal and plant species. Nature 296, 245–248.
Anderson, S.C., Cooper, A.B., Dulvy, N.K., 2013. Ecological prophets: quantifying metapopulation portfolio effects. Methods Ecol. Evol. 4, 971–981.
Annicchiarico, P., 2009. Coping with and exploiting genotype × environment interactions. In: Ceccarelli, S., Guimarães, E.P., Weltzien, E. (Eds.), Plant Breeding and Farmer Participation. Food and Agricultural Organization, Rome, pp. 519–564.
Arruda-Neto, J.D.T., Bittencourt-Oliveira, M.C., Castro, A.C., Rodrigues, T.E., Harari, J., Mesa, J., Genofre, G.C., 2012. Global warming and the power-laws of ecology. Atmos. Clim. Sci. 2, 8–13.
Becker, H.C., 1981. Correlations among some statistical measures of phenotypic stability. Euphytica 30, 835–840.
Becker, H.C., Léon, J., 1988. Stability analysis in plant breeding. plant breeding 101, 1–23.
Berzsenyi, Z., Győrffy, B., Lap, D., 2000. Effect of crop rotation and fertilisation on maize and wheat yields and yield stability in a long-term experiment. Eur. J. Agron. 13, 225–244.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference—A Practical Information-Theoretic Approach. Springer, New York.
Chamekh, Z., Karmous, C., Ayadi, S., Sahli, A., Hammami, Z., Fraj, M.B., Benaissa, N., Trifa, Y., Slim-Amara, H., 2015. Stability analysis of yield component traits in 25 durum wheat (*Triticum durum* Desf.) genotypes under contrasting irrigation water salinity. Agric. Water Manag. 152, 1–6.

Cohen, J.E., 2013. Taylor's power law of fluctuation scaling and the growth-rate theorem. Theor. Popul. Biol. 88, 94–100.

Cohen, J.E., Xu, M., 2015. Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling. Proc. Natl. Acad. Sci., 201503824.

Cohen, J.E., Xu, M., Schuster, W.S., 2012. Allometric scaling of population variance with mean body size is predicted from Taylor's law and density-mass allometry. Proc. Natl. Acad. Sci., 201212883.

Cohen, J.E., Xu, M., Schuster, W.S., 2013. Stochastic multiplicative population growth predicts and interprets Taylor's power law of fluctuation scaling. Proc. R. Soc. B 280, 20122955.

Crawley, M.J., 2013. The R Book, second ed. Wiley, Chichester, UK.

Dehghani, H., Sabaghpour, S.H., Sabaghnia, N., 2008. Genotype × environment interaction for grain yield of some lentil genotypes and relationship among univariate stability statistics. Span. J. Agric. Res. 6, 385–394.

Duch, J., Arenas, A., 2006. Scaling of fluctuations in traffic on complex networks. Phys. Rev. Lett. 96, 218702.

Eberhart, S.A., Russell, W.A., 1966. Stability parameters for comparing varieties. Crop Sci. 6, 37–40.

Eghball, B., Power, J.F., 1995. Fractal description of temporal yield variability of 10 crops in the United States. Agron. J. 87, 152–156.

Eisler, Z., Bartos, I., Kertesz, J., 2008. Fluctuation scaling in complex systems: Taylor's law and beyond 1. Adv. Phys. 57, 89–142.

FAOSTAT, 2014. Crop production data. In: (ed Food and Agriculture Organization of the United Nations), http://faostat3.fao.org/

Finlay, K.W., Wilkinson, G.N., 1963. The analysis of adaptation in a plant breeding programme. Aust. J. Agric. Res. 14, 742–754.

Francis, T., Kannenberg, L., 1978. Yield stability studies in short-season maize. I. A descriptive method for grouping genotypes. Can. J. Plant Sci. 58, 1029–1034.

Gauch Jr., H., 1992. Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs. Elsevier Science Publishers.

Goldfeld, S.M., Quandt, R.E., 1965. Some tests for homoskedasticity. J. Am. Stat. Assoc. 60, 539–547.

Ingvordsen, C.H., Backes, G., Lyngkjær, M.F., Peltonen-Sainio, P., Jensen, J.D., Jalli, M., Jahoor, A., Rasmussen, M., Mikkelsen, T.N., Stockmarr, A., 2015. Significant decrease in yield under future climate conditions: Stability and production of 138 spring barley accessions. Eur. J. Agron. 63, 105–113.

Jones, H., Clarke, S., Haigh, Z., Pearce, H., Wolfe, M., 2010. The effect of the year of wheat variety release on productivity and stability of performance on two organic and two non-organic farms. J. Agric. Sci. 148, 303–317.

Kendal, W.S., Jørgensen, B., 2011. Tweedie convergence: A mathematical basis for Taylor's power law, 1/*f* noise, and multifractality. Phys. Rev. E 84, 066120.

Küchenmeister, F., Küchenmeister, K., Wrage, N., Kayser, M., Isselstein, J., 2012. Yield and yield stability in mixtures of productive grassland species: Does species number or functional group composition matter? Grassland Sci. 58, 94–100.

McKersie, B., 2015. Planning for food security in a changing climate. J. Exp. Bot., 1–16, eru547.

Meinel, A., Richter, C., Bätz, G., 1997. Breeding aspects of clustering winter wheat cultivars for yield response. Plant Breed. 116, 437–441.

Mesterházy, A., 1995. Types and components of resistance to *Fusarium* head blight of wheat. Plant Breed. 114, 377–386.

Mühleisen, J., Piepho, H.-P., Maurer, H.P., Longin, C.F.H., Reif, J.C., 2014. Yield stability of hybrids versus lines in wheat, barley, and triticale. Theor. Appl. Genet. 127, 309–316.

Murphy, K.M., Campbell, K.G., Lyon, S.R., Jones, S.S., 2007. Evidence of varietal adaptation to organic farming systems. Field Crops Res. 102, 172–177.

Piepho, H.-P., Müller, B.U., Jansen, C., 2014. Analysis of a complex trait with missing data on the component traits. Commun. Biometry Crop Sci. 9, 26–40.

Piepho, H.P., 1998. Methods for comparing the yield stability of cropping systems—a review. J. Agron. Crop Sci. 180, 193–213.

Ramsayer, J., Fellous, S., Cohen, J.E., Hochberg, M.E., 2012. Taylor's law holds in experimental bacterial populations but competition does not influence the slope. Biol. Lett. 8, 316–319.

Rao, M., Willey, R., 1980. Evaluation of yield stability in intercropping: studies on sorghum/pigeonpea. Exp. Agric. 16, 105–116.

Royston, P., 1982a. Algorithm AS181: the *W* test for normality. Appl. Stat. 31, 176–180.

Royston, P., 1982b. An extension of Shapiro and Wilk's *W* test for normality to large samples. Appl. Stat. 31, 115–124.

Seufert, V., Ramankutty, N., Foley, J.A., 2012. Comparing the yields of organic and conventional agriculture. Nature 485, 229–232.

Smith, H.F., 1938. An empirical law describing heterogeneity in the yields of agricultural crops. J. Agric. Sci. 28, 1–23.

Smith, R.G., Menalled, F.D., Robertson, G.P., 2007. Temporal yield variability under conventional and alternative management systems. Agron. J. 99, 1629–1634.

Taylor, L., 1961. Aggregation, variance and the mean. Nature 189, 732–735.

Taylor, L.R., Perry, J.N., Woiwod, I.P., Taylor, R.A.J., 1988a. Specificity of the spatial power-law exponent in ecology and agriculture. Nature 332, 721–722.

Taylor, R.A.J., Lindquist, R.K., Shipp, J.L., 1998b. Variation and consistency in spatial distribution as measured by Taylor's power law. Environ. Entomol. 27, 191–201.

Taylor, S.L., Payton, M.E., Raun, W.R., WR, 1999. Relationship between mean yield, coefficient of variation, mean square error, and plot size in wheat field experiments. Commun. Soil Sci. Plant Anal. 30, 1439–1447.

Tilman, D., 1999. The ecological consequences of changes in biodiversity: a search for general principles. Ecology 80, 1455–1474.

Tilman, D., Lehman, C.L., Bristow, C.E., 1998. Diversity–stability relationships: statistical inevitability or ecological consequence? Am. Nat. 151, 277–282.

Vlachostergios, D., Lithourgidis, A., Roupakias, D., 2011. Adaptability to organic farming of lentil (*Lens culinaris* Medik.) varieties developed from conventional breeding programmes. J. Agric. Sci. 149, 85–93.

Xiao, X., Locey, K.J., White, E.P., 2015. A process-independent explanation for the general form of Taylor's law. Am. Nat.