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 ($n_{\mathrm{L,O}} = 5$ for lentils, $n_{\mathrm{W,O}} = 6$ for wheat); (b) all conventional environments ($n_{\mathrm{L,C}} = 5$, $n_{\mathrm{W,C}} = 6$); and (c) all environments ($n_{\mathrm{L,A}} = 10$, $n_{\mathrm{W,A}} = 12$).

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-