Geoderma 243–244 (2015) 10–17

Contents lists available at ScienceDirect

Geoderma

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

Hypotheses presence and acceptance in seven soil science journals

Pierfrancesco Nardi [a,*], Giovanni Di Matteo [a], Alfred E. Hartemink [b]

[a] Consiglio per la Ricerca e la sperimentazione in Agricoltura - Agricultural Research Council, Research Unit for Climatology and Meteorology applied to Agriculture, Via del Caravita, 7/a 00186 Rome, Italy

[b] University of Wisconsin — Madison, Department of Soil Science, FD Hole Soils Lab, 1525 Observatory Drive, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 14 May 2014
Received in revised form 5 December 2014
Accepted 7 December 2014
Available online 20 December 2014

Keywords:
Hypotheses
Soil science
Soil journals
Hypotheses acceptance

ABSTRACT

The presence and acceptance rates of hypotheses of papers in seven major soil journals were analyzed between 2001 and 2013. The aim of the study was to quantify the testing of hypotheses in soil science and investigate how it evolved over time. The journals were *Applied Soil Ecology*, *Biology and Fertility of Soils*, *European Journal of Soil Science*, *Geoderma*, *Plant and Soil*, *Soil Biology and Biochemistry* and *Soil & Tillage Research*. In total 15,344 papers were published by the seven journals over that period. Of a sample of 620 papers, 74% tested one hypothesis, 20% tested two or more hypotheses and 6% proposed a hypothesis. In total 66% of the all tested hypotheses ($n = 783$) were accepted, and the acceptance rate for the seven journals was more or less constant over time. A single hypothesis is more likely to be accepted (75%) compared to research with multiple hypotheses (55%). Although there was some difference between journals, it was concluded that acceptance rates of hypotheses in soil science are relatively low compared to other scientific disciplines.

1. Introduction

The term hypothesis has different meanings in science. A hypothesis is viewed as the antecedent of the *if–then* statement or as a speculation (Guthery et al., 2004), as an explanation of an observe pattern (Krebs, 2000) for an observation, phenomenon, or scientific problem that can be tested by further investigation (Wolff and Krebs, 2008), or as an imaginative conjecture representing the first stage of scientific enquiry (Ayala, 2009). According to Earman and Salmon (1992), a hypothesis is a statement that is intended for evaluation in terms of its consequences. Garton et al. (2005) viewed a hypothesis as a prediction or deduction from a given theory. Ford (2000) defined hypothesis from postulate (an unexplored or a new idea) which is a data statement constructed to give a logical test. Hypotheses are classified as experimental, if they can be tested by field or laboratory experiments, or historical if they refer to past causes, unlikely to be reproduce through an experiment for currently observable phenomena. Historical hypotheses are recognized in geology (Cleland, 2001, 2002), biology (Simpson, 1967), and in soil science (Phillips, 2000). The differences in definition have led to incorrect use and a hypothesis is sometimes used as synonym of theory or theory's axioms or as synonym for postulation. In this paper, we will refer to a hypothesis as a general idea that needs to be tested and that can be confirmed or rejected.

The testing of hypotheses according to the hypothetic-deductivism scheme is considered to be one of the standard methods of reasoning in science (Ayala, 2009; Murray, 2001). According to Popper's (1959) falsificationist approach, theories can only be disproved and not proved and new knowledge arises by eliminating false theory and tests should be selected based on their capability to yield a disconfirming result instead confirming ones. The view of science as theory driven, as well as the use of hypothetic-deductive method of reasoning based on the falsification approach has been fairly well established in soil science. For example, J. von Liebig (cited in Azzone, 1991) the German soil chemist, affirmed that the natural science method is deductive and a priori which means that any experiment should be supported by a theory and therefore an experiment is significant only if it tests a theory. Prosser et al. (2007), highlighted the importance of theory for soil microbial ecology. Andrén et al. (2008) stated that the complexity and diversity of soil systems make it difficult to test hypotheses in a rigorous way and by affirming that *there must be a way to reject a hypothesis*, the authors implicitly invoked the Popper's principle of falsification. Phillips (2008) discussed how models can generate hypothesis that can be tested by field-observation and he identified testable hypothesis with falsifiable hypothesis. Bradford and Fierer (2012) attempted to apply a falsification approach toward hypotheses that are relevant in the biogeography of microbial communities (e.g. functional redundancy, similarity and equivalence hypotheses). However, the falsification approach has been challenged by Thomas Kuhn (1970), who introduced the concepts of paradigm and revolution in science and pointed out how scientists usually do not process hypotheses according to the falsification principle.

The prevalence of papers supporting the tested hypotheses among scientific disciplines has been fairly well studied (Fanelli, 2012) and the higher acceptance rate of the tested hypotheses is considered too

* Corresponding author.
E-mail addresses: pierfrancesco.nardi@entecra.it (P. Nardi), hartemink@wisc.edu (A.E. Hartemink).

http://dx.doi.org/10.1016/j.geoderma.2014.12.005

good to be true (Ioannidis, 2005). Studies with significant results were more likely to be published in journals with a high impact factor (Easterbrook et al., 1991). On the other hand, papers reporting negative findings (the tested hypothesis was not supported by evidence) are more likely published in journals with lower impact factor (Littner et al., 2005). In addition, the tendency for researchers to submit manuscripts and of editors to accept papers based on the strength and outcome of the research findings is well known (Sterling, 1959; Sterling et al., 1995; Chalmers, 1990). Therefore, the ideal of science as objective as possible, theory-driven and falsificationist in spirit, may be in contrast to the reality of science that is biased in reporting experimental results.

The number of soil science publications is steadily increasing and so is the impact of most journals (Hartemink, 2001; Minasny et al., 2010). In part this reflects the vibrance of the discipline (Hartemink and McBratney, 2008), in part it is due to the “publish or perish” culture that seems to influence universities and research centers across the world. The rise in soil science publications has been fairly well quantified but the analysis of the testing of hypothesis remains to be investigated.

Here, we have researched the testing of hypotheses in soil science papers to address the following questions: (1) is the testing of hypothesis dominated by confirmation?, (2) are there differences in terms of hypothesis testing outcome between soil science and other scientific disciplines?, (3) can testing one or multiple hypotheses affects the hypothesis testing outcome?, and (4) how the testing of scientific hypotheses evolved with time?

This research quantified the testing of hypotheses in soil science and investigates how it evolved over the time. A survey was conducted for seven major soil journals (*Applied Soil Ecology*, *Biology and Fertility of Soils*, *European Journal of Soil Science*, *Geoderma*, *Plant and Soil*, *Soil Biology and Biochemistry*, *Soil & Tillage Research*). In total 655 papers were reviewed over the period 2001–2013. A systematic analysis of soil hypotheses in soil science has not been done before.

2. Data and analysis

2.1. Data collection

The survey covered a period of 13 years from 2001 to 2013 and was based on seven soil journals that are representative of different subdisciplines: *Applied Soil Ecology*, *Biology and Fertility of Soils*, *European Journal of Soil Science*, *Geoderma*, *Plant and Soil*, *Soil Biology and Biochemistry* and *Soil & Tillage Research*. The survey was conducted using the Elsevier's Scopus database and was performed in two steps. Firstly, a total of 15,344 articles were identified using the journal ISSN (International Standard Serial Number) codes. We then identified hypothesis papers using the following search criteria: ISSN (journal) AND TITLE-ABS-KEY (hypothesis) OR hypotheses OR hypothesize OR hypothesise OR hypothesized OR hypothesised AND DOCTYPE (ar).

In total 969 hypothesis papers were found.

To determine the sample size of papers to be analyzed we used the equation (Cochran, 1963):

$$n_0 = \frac{Z^2 * p * q}{e^2} \tag{1}$$

where n_0 = sample size. Z^2 is the abscissa of the normal curve that cuts off an area α at the tails and was equal to 1.96. The e is the accepted sampling error and was equal to 0.05, whereas p is the estimated proportion of an attribute in the population (e.g. accepted and rejected hypotheses). As we had no a priori knowledge of these proportions among the seven journals, we fixed p = 0.5 assuming therefore the maximum variability. However, as the seven journals and related papers corresponded to seven different finite populations, we used:

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}} \tag{2}$$

for finite population correction adjustment where n is the adjusted sample size and N is the population size. That is: 83 for *Applied Soil Ecology*, 54 for *Biology and Fertility of Soils*, 51 for *European Journal of Soil Science*, 95 for *Geoderma*, 270 for *Plant and Soil*, 343 for *Soil Biology and Biochemistry* and 73 for *Soil & Tillage Research*. In total, 655 papers containing at least one hypothesis were analyzed.

By reading the abstract or the full text, they were classified as follows:

1. papers that tested one hypothesis (single hypothesis papers);
2. papers that tested two or more hypotheses (multiple hypotheses papers);
3. papers that proposed a hypothesis (hypothesis formulation papers).

In the first two types of papers, the tested hypotheses were classified as accepted *or* rejected. A hypothesis was considered accepted according to what authors declared in the abstract or discussion sections. For example, sentences like “our *data support the hypothesis*…” or “*our hypothesis is accepted*” or “*we verify our hypothesis*” or “our hypothesis is confirmed…”. On the contrary, when authors clearly stated that they did not find any support for the stated hypothesis like “*data do not support our first hypothesis*” or “*we rejected our hypothesis that*…” or “*our hypothesis cannot be confirmed*” a hypothesis was considered rejected.

In case of single hypothesis papers, the number of papers corresponds to the number of hypothesis (one paper = one hypothesis). For papers with multiple hypotheses the number of hypotheses is higher than that of papers. To avoid bias that could alter proportions between hypotheses accepted and rejected, all tested hypotheses (single or multiple, n = 783) were considered. From the 655 randomly selected papers, 35 were excluded because of the lack of clarity of the outcomes and all analyses were done on the remaining 620 papers.

We considered only research papers while short papers, reviews, letters and other types were excluded.

2.2. Data analysis

We have listed the number of hypotheses as counts and percentages. To test the association between dependent (acceptance and rejection of the tested hypothesis) and independent variables (e.g. single hypothesis or multiple hypotheses papers, journals) the chi-square test (χ^2) was used. In the case of n by n (n > 2) contingency tables when the chi-square was significant, an analysis of adjusted residuals was performed to determine which cells were the major contributors to the chi-square significance (Sheskin, 2000). By definition, a residual is the difference in the observed frequency and the expected frequency. Adjusted residuals are then calculated as follows: standardized residuals / estimated standard error. Adjusted residuals are approximately normally distributed with a mean of 0 and a standard deviation of 1. Adjusted residuals with an absolute value that is equal to or greater than the tabled critical two-tailed 0.05 ($Z_{0.05} = 1.96$) or 0.01 ($Z_{0.01} = 2.58$) values are significant at 0.05 and 0.01 levels respectively. The sign of the adjusted residual indicates whether the observed frequency of the cell is above (+) or below (−) the expected frequency (Sheskin, 2000).

In addition to the chi-square test, that is dependent of sample size, we incorporate a sample size independent measure like odds. For the two by two contingency tables, odds, odds ratios (OR) and a 95% confidence interval were calculated. The odds of something happening are how many times more likely the event happen are to something not happening. In particular, odds were calculated by dividing the frequency of