Field Crops Research 161 (2014) 128–136

Contents lists available at ScienceDirect

Field Crops Research

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

Any chance to evaluate in vivo field methods using standard protocols?

R. Confalonieri[a], C. Francone[a,*], M.E. Chiodini[a], E. Cantaluppi[b], L. Caravati[b], V. Colombi[b], D. Fantini[b], I. Ghiglieno[b], C. Gilardelli[a], E. Guffanti[b], M. Inversini[b], L. Paleari[a], G.G. Pochettino[b], S. Bocchi[a], S. Bregaglio[a], G. Cappelli[a], P. Dominoni[a], N. Frasso[a], T. Stella[a], M. Acutis[a]

[a] Università degli Studi di Milano, Department of Agricultural and Environmental Sciences—Production, Landscape, Agroenergy, Cassandra lab, Via Celoria 2, I-20133 Milan, Italy
[b] Università degli Studi di Milano, Students of the Cropping Systems Ms Course, Via Celoria 2, I-20133 Milan, Italy

ARTICLE INFO

Article history:
Received 24 April 2013
Received in revised form 28 February 2014
Accepted 2 March 2014
Available online 27 March 2014

Keywords:
Accuracy
ISO 5725
Leaf color chart
Precision
Rice
Chlorophyll meter

ABSTRACT

The lack of standardized information on the evaluation of in vivo field methods is an important source of uncertainty in the interpretation of field data. The same words *precision* and *accuracy* can be frequently found in the agronomic and ecological literature, although often used without a real attempt to give these terms rigorous and shared meanings. On the contrary, standard protocols for determining accuracy and precision of analytical methods were successfully proposed in the last two decades and are now routinely used, especially within the chemical community. A first attempt to compile a standard guideline for in vivo field methods, derived by adapting the ISO 5725 protocol for the validation of analytical methods, is here presented. The concepts of levels, reference material, and inter-laboratory test derived from the protocol are redefined, and the underlying assumptions behind the adaptation of the ISO norm are introduced and discussed. Applicability and effectiveness of the proposed procedure are shown by means of a case study where the accuracy – i.e., trueness and precision, the latter composed by repeatability and reproducibility – of two diagnostic methods for indirect estimates of plant nitrogen nutritional status (chlorophyll meter and leaf color chart) was determined. The chlorophyll meter was more precise than leaf color chart, with precision value – expressed as relative standard deviations – lower than 6%. On the other hand, trueness indices showed better performances for leaf color chart, thus demonstrating the suitability of this method for supporting low-income farmers in managing topdressing fertilization, although at the price of performing a large number of reading replicates. However, these results are not aimed at drawing conclusions on techniques for supporting fertilization: the one presented is indeed just a case study used to assess the possibility of adopting the proposed procedure, as well as to highlight potential limits for its application. In this regard, the identification of reference values – needed for trueness quantification – is surely the most delicate issue, since the absence of conventional true values leads to the need of finding the most suitable solution according to the specific variable investigated and to the specific contexts in which the method under evaluation is applied. Hence, in light of both the encouraging results and the underlined limits, we just aim here at opening a discussion on the need for standardizing approaches and terminology for the evaluation of indirect field methods.

1. Introduction

Methods for in vivo data collection in field campaigns are often evaluated without providing readers with crucial information related to the methods themselves, such as accuracy, precision, trueness, etc. (Balasubramanian et al., 1999; Hyer and Goetz Scott, 2004). In other cases, the lack of clarity in the exposition of results may lead to a mis-conception or mis-use of these terms, that are associated to qualitative or quantitative concepts without any standardization, leading to uncertainties in the interpretation of the information coming from these evaluations and to difficulties in the comparison of results achieved by different authors with the

* Corresponding author. Tel.: +39 02 50316578; fax: +39 02 50316575.
E-mail address: caterina.francone@unimi.it (C. Francone).

http://dx.doi.org/10.1016/j.fcr.2014.03.002
0378-4290/

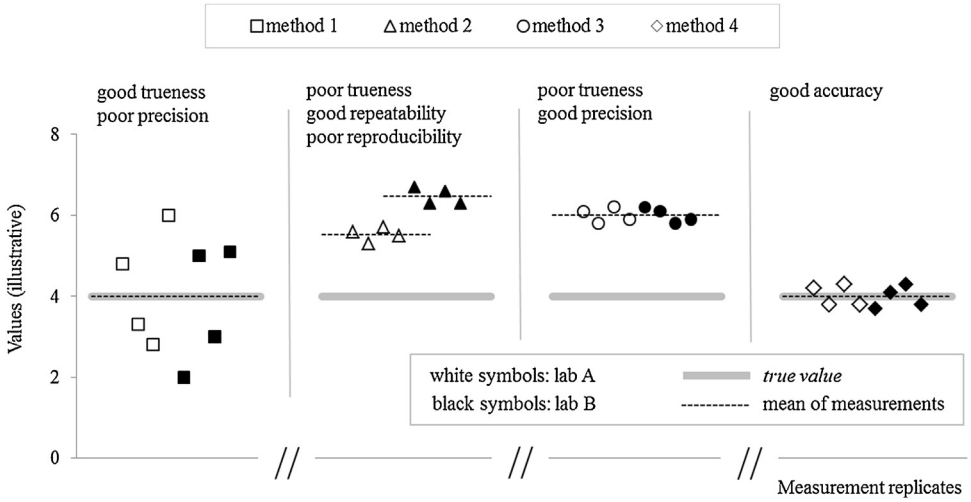

Fig. 1. *Scheme illustrating the concepts of accuracy, trueness, precision, repeatability and reproducibility.* The first method (squares) has a good trueness, since the mean of the measurement replicates from laboratories A (white) and B (black) is 4, exactly like for the true value; however, its precision is poor, since values from both the laboratories present a large variability. The second and third methods (triangles and circles, respectively) present a poor trueness, since the mean of the measurement replicates is 6; the third methods present – however – a good precision, since the variability among all measurements is small; the second method has good values for one of the component of precision (i.e., repeatability), since replicates from each laboratory are close to each other, whereas it present differences between replicates from laboratory A and those form laboratory B. The fourth method (rhombi) is accurate, since it presents good values for both trueness (mean = 4) and precision (for both repeatability and reproducibility).

same method (e.g., Garrigues et al., 2008; Peper and McPherson, 2003). One of the main reasons for this lack of standardization in the evaluation of in vivo field methods derives from the impossibility to create homogeneous reference materials with given values for the variable investigated, as for the standards used in other disciplines, like chemistry (Golubev and Fatkudinova, 2006). Another reason refers to the difficulty – for field methods – in standardizing the conditions during the evaluation (Hund et al., 2000), since some methods can be affected by factors that cannot be fully controlled (e.g., sky conditions for some instruments for leaf area index estimates), and cannot be evaluated by operators without a certain degree of subjectivity. In other disciplines, like chemistry, there is also the need to use validation procedures to certify one method (or more) among those available for a certain purpose, and to certify laboratories according to how they meet the requirements defined during the validation procedure for that method. This is because there are different reasons that push *clients* to request strict guaranties on how a method was applied to get results that can be substantial for, e.g., health-related problems, legal issues, or simply because they paid the laboratory for getting results. On the contrary, in many cases, field activities dealing with in vivo estimates of plant-related variables do not currently need neither the method nor the laboratory to be certified.

The absence of standardized information on the performances of experimental methods leads to a series of problems. One of the most important is the lack of quantitative criteria for selecting the most suitable method – among those available – for specific experimental conditions, e.g., availability of resources (money, time, people), number of entities to be monitored (e.g., plots), number of determinations during the season, etc. Another relevant problem is related to the difficulties in interpreting results obtained with a certain method, since standard information on, e.g., repeatability limits, is missing. The absence of standardized and unambiguous information on method performances could even lead to doubts on the scientific and technical bases of the method itself (Slezák and Waczulíková, 2011).

The need for a clear and rigorous definition of basic terminology used to describe method performances was already highlighted by the medical community (e.g., Menditto et al., 2007), thus showing its cross relevance to different experimental sciences. In this sense, chemistry can be considered as a reference point in evaluating method quality, reliability, and consistency. Different guidelines were indeed proposed for the evaluation of analytical methods (e.g., ISO, 1994a) and an effort to harmonize them was also made (Horwitz, 1995; Wood, 1999). Validation of a method is defined as the process by which the reliability and relevance of a procedure are established for a specific purpose (Balls et al., 1990). The final goal of the validation is to ensure that future measurements in routine analysis will be close enough to a conventional true value, i.e., to a certified reference value (Rambla-Alegre et al., 2012), or – for methods/variables for which reference materials are not available (or cannot be produced) – to the mean of specific sets of measurements (expectation) carried out under specified conditions (ISO, 2006). According to ISO (1994a), the complete validation of an analytical method requires the following metrics to be determined: accuracy (closeness of agreement between the measurement result and the true value of the measurand); linearity (proportionality of the measured to the actual quantity of the reference material); range (interval where the method is precise, accurate and linear); limit of detection (the lowest amount of analyte to be detected); limit of quantification (the lowest amount of analyte that can be measured); and robustness (closeness of results achieved under deliberately diverse laboratory conditions). Accuracy is defined as composed by trueness and precision: trueness is the degree of agreement between the mean of the measurement results and the actual (true) value; precision is instead the degree of agreement within a series of measurement replicates, and it is in turn composed by repeatability (when measurements are performed by the same operator under the same conditions in a short interval of time) and reproducibility (when measurements are performed by different laboratories) (Fig. 1). The method can be said 'fully validated' when it is assessed by collaborative trials (i.e., ring trials/tests or inter-laboratory experiments), although the cheaper 'in-house' validation (no laboratory effect is assessed) is sometimes performed (Wood, 1999).

Successful attempts to adapt the ISO 5725 (ISO, 1994a) validation protocol outside the domain of chemistry, i.e., to biological and physical methods, were carried out, by means of