



# Data synthesis for crop variety evaluation. A review

David Brown<sup>1,2</sup> · Inge Van den Bergh<sup>3</sup> · Sytze de Bruin<sup>1</sup> · Lewis Machida<sup>4</sup> · Jacob van Etten<sup>2</sup>

Accepted: 23 June 2020 / Published online: 9 July 2020  
© The Author(s) 2020

## Abstract

Crop varieties should fulfill multiple requirements, including agronomic performance and product quality. Variety evaluations depend on data generated from field trials and sensory analyses, performed with different levels of participation from farmers and consumers. Such multi-faceted variety evaluation is expensive and time-consuming; hence, any use of these data should be optimized. Data synthesis can help to take advantage of existing and new data, combining data from different sources and combining it with expert knowledge to produce new information and understanding that supports decision-making. Data synthesis for crop variety evaluation can partly build on extant experiences and methods, but it also requires methodological innovation. We review the elements required to achieve data synthesis for crop variety evaluation, including (1) data types required for crop variety evaluation, (2) main challenges in data management and integration, (3) main global initiatives aiming to solve those challenges, (4) current statistical approaches to combine data for crop variety evaluation and (5) existing data synthesis methods used in evaluation of varieties to combine different datasets from multiple data sources. We conclude that currently available methods have the potential to overcome existing barriers to data synthesis and could set in motion a virtuous cycle that will encourage researchers to share data and collaborate on data-driven research.

**Keywords** Data-driven decision-making · Meta-analysis · Multi-environment trials · On-farm trials · Data repurposing · Environmental data · Farmers' preferences

## Contents

1. Introduction
2. Data required for crop variety evaluation
  - 2.1 Agronomic performance data
  - 2.2 Environmental data
  - 2.3 Food quality and consumer preference data
3. Data management challenges
  - 3.1 Main barriers for data availability and integration
  - 3.2 Efforts to overcome data management problems
  - 3.3 Further work to improve data availability and data integration
4. Analysis of different types of data
  - 4.1 Multi-environment trial analysis
  - 4.2 Evaluation in target environments and including user requirements
  - 4.3 Multi-dimensional assessment for decision-making
5. Data synthesis approaches
  - 5.1 Rank aggregation methods
  - 5.2 Network meta-analysis
  - 5.3 Assessment of available data synthesis methods
    - 5.3.1 Partial overlapping among evaluated accessions across trials
    - 5.3.2 Measurements based on different rating scales or different methods
    - 5.3.3 Relative strengths and weaknesses of data synthesis methods
6. Conclusions and recommendations
  - 6.1 Elements for a data synthesis approach are available and aligned around ranks and reliability
  - 6.2 Data synthesis should progress from general to specific and from simple to complex

✉ David Brown  
david.brownfuentes@wur.nl

<sup>1</sup> Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands

<sup>2</sup> Bioversity International, Turrialba 30501, Costa Rica

<sup>3</sup> Bioversity International, C/O KU Leuven, W. De Croylaan 42, P.O. Box 2455, 3001 Leuven, Belgium

<sup>4</sup> Bioversity International, C/O International Institute of Tropical Agriculture (IITA), Nelson Mandela African Institute of Science and Technology, P.O. Box 447, Arusha, Tanzania

### 6.3 Use cases can spur further data sharing and model development

[Acknowledgements](#)  
[References](#)

## 1 Introduction

Farmers, especially smallholders in developing countries, are facing ever more challenging production conditions and product requirements. Extreme weather events are on the rise as one of the effects of climate change (Lesk et al. 2016; Coumou and Rahmstorf 2012). Emerging pests and diseases, as well as declining soil fertility, are also constraining farm productivity. Evolving crop production practices require the development of new genotypes that meet specific agronomic traits (Collard and Mackill 2008). Markets are also evolving, and taste preferences need to be considered if new crop varieties are to easily find their way to the consumer (Dawson and Healy 2018). Furthermore, there is growing knowledge of the different product needs and preferences relative to gender, which are influenced by their different roles in the value chain, differences in access to land and other inputs and differences in decision-making power (Christinck et al. 2017). There is also an increasing interest in more sustainable crop production systems, which would require a redesign of the whole food system and the role of players involved, including breeders (Lammerts van Bueren et al. 2018). Crop improvement aims to address the multiple challenges faced by farmers through delivering improved varieties (Malosetti et al. 2013). However, simply using the most recently released variety will not always lead to improvement, as breeding cannot address all requirements in all contexts. Decision-makers involved in crop improvement, including breeders, agronomists and farmers, evaluate multiple aspects and trade-offs relevant to the context in which they use the varieties. Crop variety evaluation is critical in decision-making in crop variety release, crop seed marketing or distribution and generating crop variety recommendations for farmers.

Crop variety evaluation is mainly conducted through field trials (Fig. 1), which are expensive and time-consuming (Lecomte et al. 2010; Tenkouano et al. 2012; Kipp et al. 2014). The limitations in resources, space, time and the required logistics in field trials also make it almost impossible to test all the varieties of interest in the same trial or in all the possible environments (Simko et al. 2012; Singh et al. 2014; Lecomte et al. 2010). Crop variety evaluation usually considers yield as the main trait while disease resistance and climate adaptation are secondary traits. Other characteristics of interest in crop variety evaluation, such as product quality and consumer preferences, are obtained through quality assessments and sensory evaluations, which are also expensive (Tomlins et al. 2004). An exception in terms of costs of data



**Fig. 1** NARITA hybrid field trials in Mbarara, Uganda, mulched with swamp grass to reduce weeds and soil moisture loss. Photo credit: Bioversity International/L. Machida

relevant to crop variety evaluation is climatic data, which acquiring costs have been decreasing due to advances in remote sensing and computational power.

Crop variety evaluation has not always kept pace with the growing complexity of agricultural production and the growing availability of data. As a data-driven type of research, crop variety evaluation can benefit from multiple revolutions occurring in several fields such as genomics, phenomics, big data and machine learning (Bolger et al. 2019; Esposito et al. 2020; van Etten et al. 2017; Tardieu et al. 2017). These revolutions are driven by increased data storage and computing capacity, the availability of sensors, improved DNA sequencing technologies and new field data collection approaches, such as high-throughput and high-precision field phenotyping and crowdsourcing (Tardieu et al. 2017; Esposito et al. 2020; Chawade et al. 2019; Reynolds et al. 2020; Van Etten et al. 2016). This has caused not only a quantitative leap in data volumes but also a shift to ‘big data’ approaches that move beyond small-sample statistics to data analysis based on machine learning (Breiman 2001; Thessen 2016; van Etten et al. 2017; Ersoz et al. 2020). While there are multiple examples of useful applications of big data analysis in agriculture (Kamilaris et al. 2017; Liakos et al. 2018), such cases are still few compared to other industries (Kamilaris et al. 2017).

Specifically, crop variety evaluation has taken little advantage of the potential benefits of data synthesis. Data synthesis allows the combination of data from different sources, producing new information and knowledge to support decision-making (Pillemer and Light 1980; Pickett et al. 2007; Carpenter et al. 2009; Wyborn et al. 2018).

Interest in the value from combining and (re)using datasets in agriculture has grown, supported by open data and data sharing initiatives (Leonelli et al. 2017). As new analytical technologies and methods become available, legacy data could be reanalysed (White and van Evert 2008; Hampton et al. 2013). Data synthesis can improve the efficiency of data use in crop variety evaluation by combining and repurposing

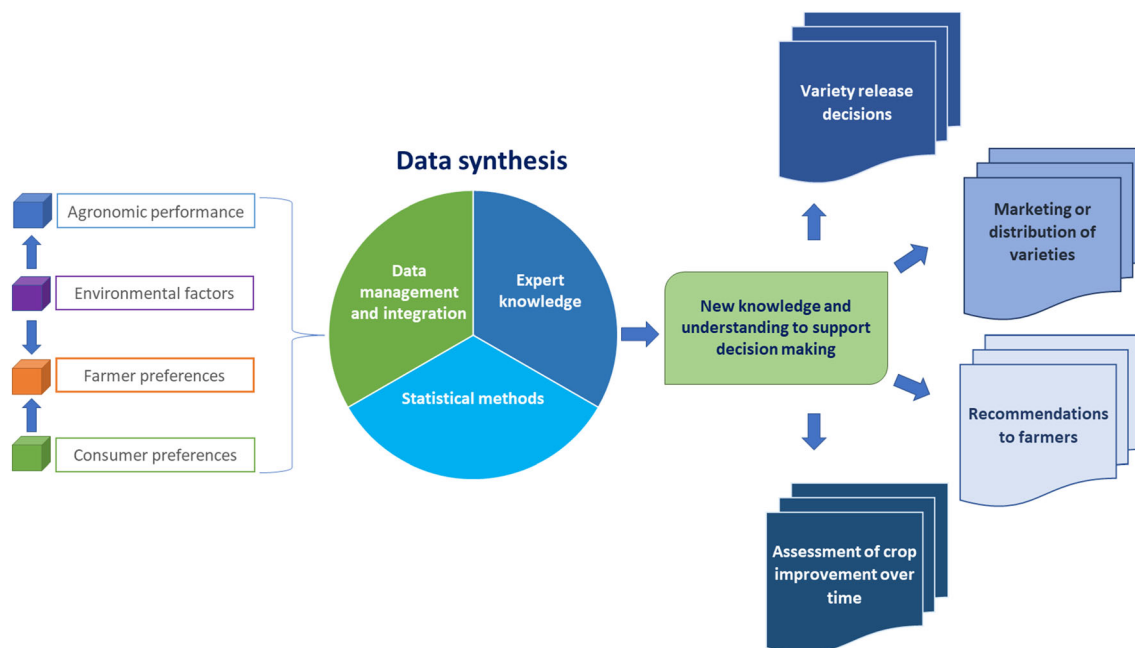
new and legacy data from field trials, environmental measurements, farmer requirements and consumer preferences (Fig. 2). In the agricultural sciences, it has so far mainly taken the form of meta-analysis (Philibert et al. 2012; Krupnik et al. 2019).

Data synthesis can play a role in different functions of crop variety evaluation. The selection of genotypes to be released as cultivars can benefit from data synthesis to assess genetic gain (progress over time) (Streck et al. 2018), to benchmark against other breeding programs, to improve accuracy through multi-season assessments and to predict performance beyond the trial environments. The latter involves analysing a combination of variety trial data and environmental data (Hyman et al. 2013). The analysis of trial data can be made more accurate when data from the last trial season is combined with historical variety performance data (Arief et al. 2015).

To release a new variety, breeders need to evaluate the proposed genotypes against existing varieties in a country or region. Data synthesis could facilitate enriching data from trials including the new varieties with data on the past performance of the older varieties to gain accuracy (Damesa et al. 2017). Seed companies need to assess variety performance to take seed production and marketing decisions. Service providers, such as agro-input suppliers, cooperatives, agricultural extension organizations and NGOs, need to make recommendations to farmers, considering the multiple dimensions of variety performance (and trade-offs between these dimensions) in different environments and under different types of crop management. Information from existing crop trials to formulate recommendations is often used for this end. Such an analysis of existing data could benefit from data synthesis if

the data that is available comes from different sources. Some service providers produce their own data about on-farm variety performance to generate recommendations or refine existing ones, which could also benefit from data synthesis to combine the new data with existing data. In many contexts, variety evaluation is done in a fragmented way (Rangarajan 2002), which can preclude centralized coordination or standardization of data collection and weakens variety evaluation as each entity assesses genotype by environment interactions in a limited set of environments. Data synthesis could help to gain a better understanding of genotype by environment interactions across space and time. Flexible data synthesis could take advantage of heterogeneous data from different actors in the seed sector and provide value to the several functions that variety evaluation plays in each step of the crop improvement cycle.

A clearer perspective on data synthesis for crop variety evaluation is needed to achieve these potentials. Here, we review the literature relevant to data synthesis for context-specific decision-making in variety management. The objective of this article is to provide an overview of the required elements, current approaches and research gaps in data synthesis for crop variety evaluation, focusing on decision-making for variety pre-release and post-release. We limit ourselves to these later stages of the breeding process, and therefore, we do not cover the genomic and high-throughput phenotyping data. Even though these types of data are clearly part of the data revolution in crop improvement, they generally concern early and intermediate stages of the breeding process. We briefly refer to high-throughput field phenotyping data, as it has the potential to support later stages of the breeding



**Fig. 2** Different elements and processes involved in data synthesis for crop variety evaluation

process. In Section 2, we discuss the types and sources of data that are required. In Section 3, we discuss how data synthesis relies on proper data management, including sharing data across different trials and the compatibility of datasets. Data synthesis requires not only combining datasets to assess variety performance but also beyond assessing average performance, a careful analysis of how different genotypes respond to diverse environments and match the preferences of farmers, consumers and other stakeholders. Therefore, in Section 4, we review how data analysis is currently dealing with the end-users, their context and what is still lacking to evaluate crop varieties through a data synthesis approach. In Section 5, we review existing data synthesis approaches used in crop improvement and assess how they can be enhanced to include use context. In Section 6, we present our conclusions and recommendations.

## 2 Data required for crop variety evaluation

In this section, we describe the data types required by a data synthesis approach for crop variety evaluation. Field trial data are important to analyse the phenotypic response of a given genotype, to the environmental characteristics of the testing location and, in some cases, to management practices. Not only yield but also product quality is considered in variety evaluation. The evaluation of crop varieties also involves data about the preferences of farmers obtained from participatory and on-farm trials, and consumers, obtained from sensory evaluations.

### 2.1 Agronomic performance data

Agronomic performance data are collected from field trials, which can be set up in several ways depending on the context and purpose. A rough classification of contexts includes (1) public international breeding programs (e.g. breeding programs within the CGIAR), (2) private breeding programs at commercial seed companies and (3) agricultural research at national or regional level, conducted by National Agricultural Research Systems, often in partnership with International Agricultural Research Organizations.

Field trials of breeding programs are usually known as performance trials or yield trials, given the importance of yield as the main trait (Acquaah 2012). There are two main types of yield trials: (1) breeder trials and (2) official trials (Acquaah 2012). Breeder trials aim to assess the performance of a set of genotypes to decide which ones should be released as cultivars (Priyadarshan 2019). An official variety trial is part of the variety release and registration process, which varies among countries, but in most of the cases, it is conducted by an independent body, such as an official seed agency or under the

jurisdiction of a variety release committee. Depending on the stage of the breeding process, the breeder's trials can be divided into preliminary yield trials (PYTs) and advanced yield trials (AYTs) (Priyadarshan 2019). A PYT often concerns many genotypes (and few replications), whereas an AYT evaluates a small number of genotypes (selected from the PYT), with more replications over different environments, and during several years (Priyadarshan 2019). In this review, we are focusing on data generated from AYT. Crop variety trials can also be established to test improved varieties to be recommended to farmers (Yan 2014b). Both for breeding and to generate variety recommendations, field trials aim to evaluate varieties in different environments, where the environment is considered a combination of location and season (Acquaah 2012). For this purpose, crop variety trials can be established mainly in three different levels of combination of location and season: (1) a single location in a single season, (2) multiple locations in a single season and (3) multiple locations in multiple seasons (Yan 2014a). Crop variety trials can also be established in a centralized or decentralized system. The centralized approach involves on-station trials; it is the conventional approach for many crops and contexts. Decentralized methods include establishment of trials at farm locations, with different levels of participation from farmers. As not only biophysical factors determine the suitability of a variety, it has been suggested that the concept of environment should be extended to include the socioeconomic context of the target location (Desclaux et al. 2008). Participatory plant breeding and participatory varietal selection methods aim to better consider farmers' preferences and context in order to increase the adoption of improved varieties (Weltzien and Christinck 2017; Ceccarelli and Grando 2007). These approaches often include participatory on-farm trials, which may produce insights that are complementary to insights derived from conventional trials (Coe 2007; Coe 2002; Atlin et al. 2001). Although farmers' participation is commonly associated with on-farm trials, farmers can also be involved in on-station trials. Participatory varietal selection (PVS) can be done through mother-baby trials, where the mother trial includes the full set of testing genotypes and the baby trial only includes a subset of test genotypes alongside the control genotype (Virk et al. 2009; Snapp 2002). On-farm trials may produce unbalanced data, due to differences in the particular conditions of farmers' fields and the limited availability of seeds of the new varieties (Virk et al. 2009).

Data collected from field trial evaluations typically include trial design, the trial location, the date of establishment, trial management, evaluated genotypes and observations of the target traits (e.g. yield). Observations of the target trait can be either measured or estimated and should be ideally referenced to the observation date and to the phenological stage of the plant during observation (Billiau et al. 2012; White et al. 2013; Germeier and Unger 2019). For instance, the second



phase of the International Musa Testing Program (IMTP) used the following attributes (Orjeda 2000): genotype, time from planting to shooting (days), time from shooting to harvest (days), time from planting to harvest (days), height of the pseudostem at shooting (cm), height of the following sucker at harvest (cm), bunch weight (cm), number of hands per bunch (hands), total number of fingers per bunch (fingers), average fruit weight (g) and leaf emission rate.

Technological innovations allowed the development of new methodologies for collecting data from field trials. These include high-throughput field phenotyping methods supported by satellite imagery or data from unmanned aerial vehicles (UAVs) and proximal phenotyping (Chawade et al. 2019).

Given the multiple context and evaluation objectives, each organization conducting crop variety trials may use its own experimental design and employ different methods and technologies for collecting, storing and publishing and/or sharing data. In Section 3, we review potential obstacles for data integration resulting from these differences. Furthermore, the diversity of goals and evaluation methodologies also produce different approaches to analyse collected data. We review these in Section 4.

## 2.2 Environmental data

To analyse the phenotypic response of genotypes to the testing environment, a fundamental step is to characterize the environment. The environment is the first source of yield variability in plant breeding trials (Chenu 2015). Hence, environmental data are required to characterize the trial location and to understand its influence on the performance of tested genotypes in that particular location. It is known that the use of environmental data as model covariates analysing multi-location trial data improves the degree of accuracy in the prediction of genotype performance (Piepho 2000; Piepho et al. 1998). A recent study by van Etten et al. (2019) demonstrated an improvement in variety recommendations using seasonal climate data as model covariates. Xu (2016) proposed to consider all environmental factors affecting growth and production of plants through an approach called ‘envirotyping’. Environmental data can be collected at trial sites directly. But even if environmental data were not collected during trials, geolocating trial sites allows enriching the dataset with existing environmental data. Adding environmental data to legacy trial data allows comparisons among trials conducted at different locations. Some types of environmental data are increasingly available through open and public repositories (Hyman et al. 2013). For instance, data on rainfall, temperature, elevation and soils are openly and freely available from open and public databases such as Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) (Funk et al. 2015), MODIS Land Surface Temperature (Wan et al. 2015), Hole-filled SRTM for the globe version 4 (Jarvis et al. 2008) and SoilGrids (Hengl et al. 2017). The European Centre for Medium-Range Weather

Forecasts (ECMWF), through the Copernicus Climate Change Service (C3S), provides a comprehensive collection of climatic datasets, including the recently deployed ‘Agrometeorological indicators from 1979 to 2018 derived from reanalysis’, known as AgERA5. Available climatic data can be used to calculate climatic indices (Table 1), which were proven to be useful as model covariates in the analysis of crop variety trials (van Etten et al. 2019; Kehel et al. 2016). Even though climatic data opens a wide range of possibilities for crop variety evaluation, the resolution of available data has to be carefully considered (Parkes et al. 2019).

## 2.3 Food quality and consumer preference data

Sensory and nutritional evaluation has received more attention in recent decades, countering the narrow focus of crop improvement on yield, disease resistance and uniformity (Folta and Klee 2016). At present, consumer markets are evolving, with consumers seeking additional product qualities such as nutritional and sensorial characteristics (Folta and Klee 2016). Food quality involves both objective and subjective analyses, involving measurements of contents, texture as well as sensory analyses. Sensory evaluation is formally defined as ‘a scientific discipline used to evoke, measure, analyze, and interpret reactions to those characteristics of foods and materials as they are perceived by the senses of sight, smell, taste, touch, and hearing’ (Anonymous 1975; Stone et al. 2012). Consumer preference data are obtained from sensory evaluations by panels of regular or specialized consumers, with different methods, such as descriptive analysis or rapid sensory evaluations (Dawson and Healy 2018). Sensory and hedonic (i.e.

**Table 1** Temperature and precipitation indices commonly used as covariates in crop variety trial analyses

Environmental index	Unit
Maximum daytime temperature	°C
Minimum daytime temperature	°C
Maximum nighttime temperature	°C
Minimum nighttime temperature	°C
Mean difference between daily maximum temperature and daily minimum temperature	°C
Number of days with maximum temperature > 30 °C	Days
Number of nights with maximum temperature > 25 °C	Days
Maximum length of consecutive days with precipitation < 1 mm	Days
Maximum length of consecutive days with precipitation ≥ 1 mm	Days
Number of days with precipitation > 5 mm	Days
Number of days with precipitation > 10 mm	Days
Maximum 1-day precipitation	mm
Maximum 5-day precipitation	mm

Adapted from van Etten et al. (2019) and Kehel et al. (2016)

related to pleasant or unpleasant) experiences cannot be measured directly and should be inferred from descriptive or numerical representations (hedonic scales) of subjects' responses (Lim 2011). There are four main types of scales used in hedonic scaling, which are presented in Table 2.

The 9-point hedonic scale developed by Peryam and Girardot (1952) is the most widely used method for scaling consumer preference and food acceptability (Lim 2011). It is composed of the following values and their correspondent description: 9, like extremely; 8, like very much; 7, like moderately; 6, like slightly; 5, neither like nor dislike; 4, dislike slightly; 3, dislike moderately; 2, dislike very much; and 1, dislike extremely (Peryam and Girardot 1952). More recently, other scaling methods have been proposed, such as the labelled affective magnitude (LAM) scale (Schutz and Cardello 2001) and labelled hedonic scale (LHS) (Lim et al. 2009). This diversity of measurement scales can pose a challenge to combine data from different sources, such as different laboratories testing food quality, or a sensory evaluation with farmers testing different varieties as part of a breeding process.

Agronomic performance, food quality and preference data are still expensive and complex to acquire and manage. In contrast, weather and soil data are increasingly available at significantly reduced costs. More effort is required to improve the efficiency in data use in the evaluation of crop varieties. The higher availability of weather and soil data can motivate and increase data reuse, repurposing legacy crop variety trial data by adding environmental data to extract new insights.

### 3 Data management challenges

As a data-driven research, data synthesis requires availability of the data to be reused. It also requires careful data management to integrate data of heterogeneous nature from different sources and formats, as described in Section 2. Here, we discuss the challenges in data management that are relevant

to data synthesis for crop variety evaluation, the main efforts to address these problems and further research needs.

#### 3.1 Main barriers for data availability and integration

Data should be available to be integrated and then synthesized using formal statistical techniques. However, data are still rarely shared for reuse, especially in the case of raw data from variety trials. Diekmann (2012) and Williams (2012) found that researchers are often unwilling to share raw data out of concern about data being taken out of context, which could lead to incorrect results and misinterpretation. Authors may also oppose freely releasing data that cost them substantial work and resources, whereas they may be more willing to share data with colleagues (Diekmann 2012).

In addition to cultural constraints, technical challenges arise for data sharing among both individuals and institutions. It has been a common practice for researchers and research centres to develop their own system for storing data, mainly because they do not trust global repositories with which they do not have a direct relationship and for which long-term support may not be guaranteed (Leonelli et al. 2017). This has resulted in a myriad of individual databases that are neither open nor compatible among research centres. This not only inhibits collaboration among scientists but also promotes duplication of efforts and increases costs, a luxury that the scientific community cannot afford in times of scarce economic resources for agricultural research.

In cases when data are available, data integration sometimes encounters problems due to lack of standardization in terms of syntax, semantics and structure. Crop variety trial datasets are often very heterogeneous in terms of quantity, quality, types and formats (Hyman et al. 2017; Leonelli et al. 2017). Individual trial designs and observational methods vary according to the specific purpose of trials (Section 2). This lack of standardization of crop variety trial data makes it difficult to compare results between trials and to reuse datasets with traditional tools (Rijgersberg and Top

**Table 2** Types of hedonic scales

Scale	Basic empirical operation	Number usage	Permissible statistics	Example scale
Nominal	Determination of equality (categorization)	As labels	Non-parametric: number of cases; mode	1 (good) 2 (bad)
Ordinal	Determination of greater or less	To recognize the rank order	Non-parametric: median; percentiles	Rank rating
Interval	Determination of equality of intervals or differences	To represent degrees of differences	Parametric: mean; standard deviation	Equal intervals
Ratio	Determination of equality of ratios	To represent relative proportions	Parametric: log mean; standard deviation	Labelled affective magnitude (LAM) scale; labelled hedonic scale (LHS)

Adapted from Stevens (1946) and Lim (2011)

2000; Leonelli et al. 2017). Combining crop trial datasets often presents the following problems: (1) incomplete or in-existent overlap among evaluated accessions across trials, (2) measurements based on different rating scales and (3) the use of different methods for observing the same trait (Simko and Pechenick 2010). Methods to solve the problem of measurement in different rating exist such as the threshold model (Hartung and Piepho 2005), but it does not solve the problem of partial overlap among tested varieties in the different trials. In Section 5, we explore and evaluate how different data synthesis methods deal with this kind of problems.

The dearth of relevant data in the public domain limits the possibilities of data synthesis, as it provides a large initial cost of assembling, cleaning and reformatting data. For individual data synthesis efforts, this initial investment may be relatively very high, even though it could be worthwhile if data can be repurposed more than once. Furthermore, practices limiting data reuse go against the aim of science of building universal knowledge, in which public funds play a fundamental role.

### 3.2 Efforts to overcome data management problems

Several international agricultural research organizations have made efforts to address the barriers of poor data availability and data format incompatibilities (McLaren et al. 2005; Ritchie 1995; Germeier and Unger 2019). Several global funders of agricultural research are increasingly seeking mechanisms to guarantee that research investments generate benefits for smallholder farmers in developing countries (Dalrymple 2008). Global open data and sharing initiatives aim to facilitate data accessibility, allowing the generation of new knowledge (Wilkinson et al. 2016). With valuable contributions from diverse partner organizations, CGIAR centres have been developing information systems and platforms, aiming to integrate heterogenous data sources to support crop improvement research (McLaren et al. 2005; Hyman et al. 2017). Examples of this type of systems are the International Crop Information System (ICIS) and its derivatives, the International Rice Information System (IRIS) and the International Maize Information System (McLaren et al. 2005; Shrestha et al. 2010). The recently created CGIAR ‘Platform for Big Data in Agriculture’ aims to materialize the potential of big data-related methods and technologies to improve agricultural production. Outside the CGIAR system, agricultural researchers and organizations are also endeavouring to construct better ecosystems of data and methods. Table 3 contains a compilation of the main international initiatives on standardization and data sharing. Data standardization and sharing systems include online databases such as AgTrials, YamBase, CassavaBase and MusaBase, which all implement ontologies to standardize vocabularies and terminologies (see Table 3). Ontologies

formally define the relationships among concepts within a given domain (Matteis et al. 2013). Similar approaches have also been proposed by other authors. For instance, Rijgersberg and Top (2000) proposed data model templates—a generalization of data models—to achieve a balance between standardization and flexibility. See Spyns et al. (2002) for an overview of specific differences between data models and ontologies. Germeier and Unger (2019) applied a modelling approach that goes further than data models, considering also statistical models in the implementation of a phenotyping information system. Efforts to standardize phenotyping data include the Minimal Information About Plant Phenotyping Experiment (MIAPPE) (Krajewski et al. 2015; Ćwiek-Kupczyńska et al. 2016; Papoutsoglou et al. 2020). Efforts to standardize data from field experiments include the International Consortium for Agricultural Systems Applications (ICASA) standard, initially developed by the International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) and updated by ICASA (White et al. 2013).

The FAIR (findability, accessibility, interoperability and reusability) guiding principles are intended for producing and publishing data, aiming to facilitate and enable data sharing and reuse (Wilkinson et al. 2016). To achieve these principles, a set of mechanisms is considered. These include unique and persistent identifiers, a standardized communications protocol and the use of domain-specific standards for both data and meta-data.

Despite these efforts, recent literature indicates that there are still serious challenges in implementing data standardization and sharing. About 85% of the more than 35,000 records in the AgTrials database contain only meta-data; hence, those interested in the underlying data should contact the original data provider (Hyman et al. 2017). Problems persist on both the supply and demand side. It has been found that researchers are often reluctant to use data produced by others because there is no guarantee about the quality (Diekmann 2012). Data standards are now available (Table 3), but it is still difficult to persuade the agricultural research community to adopt the suggested standards. The lack of flexibility to adapt to scientific progress is indeed one of the arguments stated against standards (Germeier and Unger 2019). For both new and legacy data, standardization requires considerable efforts, which will not immediately pay off, hindering its implementation. In addition, the efforts required for processing datasets with the accompanying meta-data to facilitate open access and use by others are often not acknowledged (Diekmann 2012; White and van Evert 2008). From 112 surveyed users of the AgTrials platform, 34% considered that their data are incorrectly organized to be shared publicly (Hyman et al. 2017). In the cases where data are shared, it is mostly not through global repositories but rather as supplementary data on the journal

**Table 3** Main international initiatives to increase data sharing and standardization in agriculture

Initiative	URL
AgTrials	<a href="http://agtrials.org">http://agtrials.org</a>
AgroPortal	<a href="http://agroportal.lirmm.fr">http://agroportal.lirmm.fr</a>
Breeding Application Programming Interface (BrAPI)	<a href="https://brapi.org">https://brapi.org</a>
Breeding Management System	<a href="https://bmspro.io">https://bmspro.io</a>
CassavaBase	<a href="https://www.cassavabase.org">https://www.cassavabase.org</a>
Crop Ontology	<a href="http://www.cropontology.org">http://www.cropontology.org</a>
GARDIAN	<a href="http://gardian.bigdata.cgiar.org">http://gardian.bigdata.cgiar.org</a>
Global Open Data for Agriculture and Nutrition (GODAN)	<a href="https://www.godan.info">https://www.godan.info</a>
Global Trial Data Management System	<a href="https://research.cip.cgiar.org/gtdms">https://research.cip.cgiar.org/gtdms</a>
Integrated Breeding Platform	<a href="https://www.integratedbreeding.net">https://www.integratedbreeding.net</a>
MIAPPE	<a href="https://www.miappe.org">https://www.miappe.org</a>
MusaBase	<a href="https://musabase.org">https://musabase.org</a>
Sol Genomics Network	<a href="https://solgenomics.net">https://solgenomics.net</a>
YamBase	<a href="https://yambase.org">https://yambase.org</a>

website where associated articles were published (Williams 2012). Sharing data as supplementary materials on a journal website would be adequate if journals followed commonly agreed guidelines, such as the FAIR principles. Although there has been an increase in data shared by researchers during the recent years, there is still a lack of awareness and hence compliance with FAIR principles (Mark et al. 2018).

### 3.3 Further work to improve data availability and data integration

We identified barriers for data availability and data integration. The main efforts to stimulate data open access and repurposing have focused on compliance with standards and data sharing as a goal. Given the modest progress so far, we suggest that this focus be complemented with efforts to make data sharing more appealing, by the stimulation of data demand for data synthesis. This might set in motion a virtuous cycle of collaboration around data synthesis, providing clear incentives in the form of authorships on joint publications and citations to datasets (Fig. 3).

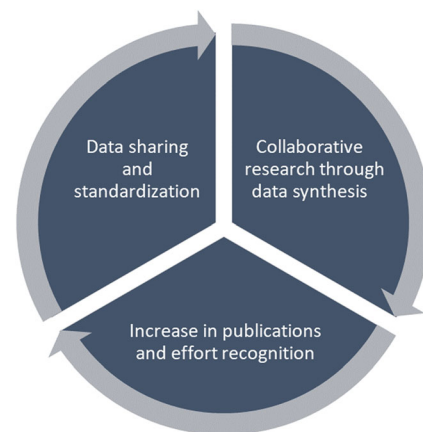
Therefore, we posit a need for simple methods that can deal with highly heterogeneous datasets to start to show the potential benefits of data synthesis (see Section 5). Ecology research is a concrete example of how increasing the number of datasets, publications and collaboration among larger groups of scientists through meta-analysis can result in even larger collaborative initiatives which enhance the scope and potential impact of research (Cadotte et al. 2012). Data journals may help to boost this kind of data-centred collaboration (Candela et al. 2015). This could provide further motivation and feedback to continue the consensus building processes around data management.

## 4 Analysis of different types of data

Data synthesis requires the combination and analysis of different types of data as described in Section 2. Here, we review current approaches to combine those types of data and research gaps to achieve data synthesis for crop variety evaluation.

### 4.1 Multi-environment trial analysis

As we discussed in Section 2, the most basic evaluation of genotypes is the assessment of their agronomic performance, such as plant growth parameters, yield, plant reaction to diseases and tolerance to climate conditions (e.g. tolerance to drought, cold and flooding), among others. This type of evaluation is conducted through field trials, which can be established in different ways. Conventional trials are usually established in research stations under controlled conditions. Trials can also be established on farms and involve different



**Fig. 3** A virtuous cycle set in motion by data synthesis



levels of farmer participation, from limited participation as observers to full participation as citizen scientists (Ceccarelli et al. 2009; Ceccarelli 2012; Van Etten et al. 2016). Combinations are also possible, such as on-station trials with some level of farmer participation. Regardless of the trial design, the idea is to establish trials at different locations for several growing seasons. The combination of location and time is known as the ‘testing environment’, and the trials are known as ‘multi-environment trials’ (METs). Multi-environment trials are conducted to evaluate the suitability of crop genotypes for different agroecological conditions (van Eeuwijk et al. 2005).

Genotype  $\times$  environment ( $G \times E$ ) interaction is the relative difference in phenotypic response that a group of genotypes expresses depending on the environmental conditions (de Leon et al. 2016). Hence,  $G \times E$  assessment requires the evaluation of a minimum of two different genotypes in at least two different environments (Kang 1997). The phenotypic response of a genotype to the environment is described by a function known as the reaction norm (Bustos-Korts et al. 2019). When the reaction norm lines of evaluated genotypes in different environments are not parallel, there is presence of  $G \times E$  (Bustos-Korts et al. 2019). Especially in conventional breeding,  $G \times E$  is considered a challenge by breeders due to its implications for genotype selection (Kang and Gorman 1989). Aiming for more specific adaptation, decentralized breeding programs take advantage of  $G \times E$  instead of diminishing the effects (Ceccarelli 1989).

There are different statistical models for  $G \times E$  analysis. For an overview, we refer to recent reviews such as the work of Malosetti et al. (2013), van Eeuwijk et al. (2016) and Bustos-Korts et al. (2019). Most of the statistical models for  $G \times E$  analysis can be interpreted as (phenotypic) response functions for each genotype to environmental variables (van Eeuwijk et al. 2005). Therefore,  $G \times E$  analysis represents a combination of two of the data types presented in Section 2: the agronomic performance data and the environmental data.

Data from multi-environment trials are usually summarized in two-way tables of means, with genotypes as rows, and environments as columns (Malosetti et al. 2013). Two major groups of statistical models for the analysis of  $G \times E$  can be identified: (1) methods that only use the two-way table of means and environmental information are included only implicitly (usually as dummy variables) and (2) methods that use additional information, explicitly included as genotype and/or environment covariates (temperature, rainfall, etc.) (Malosetti et al. 2013). Examples of  $G \times E$  models from the first group include additive models (ANOVA), regression on the mean (Yates and Cochran 1938; Finlay and Wilkinson 1963), additive main effects and multiplicative interaction (AMMI) models (Gauch Jr. 1992) and the genotype + genotype  $\times$  environment (GGE) model (Yan et al. 2000). Since they only require a two-way table of means as input, these models are

considered to be good for descriptive and explorative purposes, but not for explaining  $G \times E$  (Malosetti et al. 2013). The second group of models includes factorial regression, partial least squares regression, structural equation models and mixed effect models. Factorial regression allows the use of environmental or genotypic variables as covariates to explain  $G \times E$  but has the limitation that only permits one dependent variable at a time (Vargas et al. 2007). Another limitation of factorial regression is its difficulty in dealing with multi-collinearity when several covariates are used (Vargas et al. 1999). For these cases, partial least squares regression is a more convenient approach, as it can easily handle multiple explanatory variables (Vargas et al. 1999). When the cause-effect analysis of  $G \times E$  is aimed for, partial least squares regression becomes inadequate, and methods such as structural equation modelling are more suitable (Vargas et al. 2007).

Mixed-effect models are one of the most used approaches for analysing  $G \times E$ , and they are usually implemented using either single-stage or two-stage analysis (Möhring and Piepho 2009). Single-stage models analyse data from individual plots, in which the residual effects and the  $G \times E$  effects are estimated simultaneously (Smith et al. 2005). In contrast, two-stage models include a first stage in which design features and spatial variation are modelled using data from individual trials. Next, the second stage involves fitting an overall mixed model to the genotype by environment adjusted means obtained from stage 1 (Malosetti et al. 2013; Smith et al. 2005). The analysis can be extended to more than two stages, in which case the approach is more commonly known as stage-wise analysis (Piepho et al. 2012a; Damesa et al. 2017). Although single-stage analysis is preferred from a theoretical point of view, two-stage analysis is less demanding in terms of computation requirements and provides similar results to single-stage when appropriate weights are selected (Malosetti et al. 2013). Therefore, most of  $G \times E$  models are implemented using a two-stage approach (Malosetti et al. 2013).

There are situations where non-parametric methods, such as rank-based methods, are more convenient (Elias et al. 2016). This type of non-parametric methods has been used mainly to rank genotypes at specific locations (Elias et al. 2016) and has a set of advantages that include no specific modelling assumption about the distribution of the effects and are easy to implement and interpret (Huehn 1990). Non-parametric models are considered a useful option when the interest is focused on the ranking of genotypes rather than to evaluate the level of difference on performance between genotypes (Brancourt-Hulmel et al. 1997). In the context of selection in breeding and evaluation programs, Huehn (1990) considered the rank order of genotypes to be the most important information.

Other  $G \times E$  models go beyond statistical analysis and integrate knowledge on crop physiology and expert assessments. For instance, Theobald et al. (2002) proposed

the use of a Bayesian model to incorporate expert knowledge about the analysed crop. In a bibliometric analysis, van Eeuwijk et al. (2016) identified an important growth in the application of both mixed models and crop growth models, especially after 2005. A crop growth model incorporates plant physiological aspects, along with the genotype and environment, in the analysis of interactions that produce a phenotype (van Eeuwijk et al. 2016). Furthermore, crop growth models also allow to consider the effect of cropping systems (intercropping, fertility management, etc.) on  $G \times E$  (Jeuffroy et al. 2014).

One of the goals of crop growth models for variety evaluation in multi-environment trials is to improve the characterization of the environment (Jeuffroy et al. 2014). For example, Tesfaye et al. (2016) combined geospatial analysis with crop modelling (1) to characterize a maize growing environment in Southern Africa (Malawi, Mozambique, Zambia and Zimbabwe) and (2) to evaluate the variety performance of five new drought-tolerant varieties across the aforementioned region. The environmental characterization was conducted using a standardized precipitation index, and it focused on the frequency of drought occurrences rather than drought severity (Tefaye et al. 2016). To evaluate the variety performance of new varieties, maize yields were simulated using the Crop Estimation through Resource and Environment Synthesis (CERES)-Maize model. Simulated relative yields of five drought-tolerant varieties outperformed the commercial check variety across many environments, but not in all (Tefaye et al. 2016).

Jeuffroy et al. (2014) reviewed the use of crop growth models in variety performance prediction and concluded that although their use is increasing, they are still not mainstream for variety evaluation. For mechanistic models to have predictive power to distinguish between varieties, information is needed on the processes or the underlying genotypic factors that give rise to these differences. Some information can be derived from existing trial data through model fitting, but overfitting often occurs. Acquiring additional data to estimate crop model parameters directly is often costly or not possible retrospectively in a data synthesis context. Jeuffroy et al. (2014) argue that cost-benefit considerations to assess the value of additional information are important. On the one hand, crop growth modelling generally focuses on a narrow set of variables (mostly yield). Yield is an important input into crop variety recommendations, but other aspects cannot be ignored, including the user perspective (see Section 4.2). Their relative complexity makes their application often difficult. One possibility is to use (at least initially) very simple crop models and build up their complexity gradually (Shorter et al. 1991). Another option is to generate intermediate variables that can be used in statistical models or machine learning approaches (see Feng et al. 2019).

## 4.2 Evaluation in target environments and including user requirements

The current challenges in agricultural production are more likely to be addressed by locally adapted solutions that consider both environmental and socioeconomic information (Van Etten et al. 2016). The socioeconomic context of the target environment should be considered to match both farmer and consumer preferences. Socioeconomic data like human population, welfare and transportation infrastructure have been proposed for targeting genotypes to environments, but such recommendations mostly concerned logistics planning on germplasm deployment (Hyman et al. 2013). While this kind of socioeconomic data is indeed important, other types of data, such as consumer preferences, should also be considered. For example, Desclaux et al. (2008) proposed that, in addition to the usual biophysical and management factors, the environment should be a wider concept that also includes actors, markets, regulations and societal dynamics.

Participatory on-farm trials aim to take the variety trials closer to the target environments and user requirements (Ceccarelli and Grandó 2007). On-farm trials can provide much information, ranging from biophysical performance to economic assessment (Franzel and Coe 2002). In this type of trials, the concept of environment in a  $G \times E$  analysis is extended to include socioeconomic factors, besides the usual biophysical variables (Coe 2002). Data collected from on-farm participatory trials is often in the form of ratings or rankings, requiring different statistical models to conduct  $G \times E$  analysis (Coe 2002). For example, Coe (2002) proposes to analyse ratings using proportional odds, and rankings with the Bradley-Terry model (Bradley and Terry 1952). An extended Bradley-Terry model can incorporate covariates (Coe 2002; Ditttrich et al. 1998). As discussed before, the possibility of including covariates is especially relevant when location-specific information is to be extracted from the experimental data. It has been shown that environmental covariates can improve predictions of variety performance (Piepho 2000; Piepho et al. 1998). For data synthesis, in cases when environmental data is not collected as part of the crop trial, environmental covariates can be linked to experimental data through geolocation, as shown by Lobell et al. (2011), van Etten et al. (2019) and others.

An extended version of the Plackett-Luce model (Plackett 1975; Luce 1959) recently implemented by Turner et al. (2020) includes the use of model-based recursive partitioning (Strobl et al. 2011), allowing the incorporation of covariates for predicting rank orders. Hence, subgroups of rankings with significantly different worth parameters are identified based on covariates (Turner et al. 2020). van Etten et al. (2019) recently used this model to analyse data from on-farm participatory trials in three countries, successfully identifying

environmental covariates that consistently influence variety performance across several seasons.

### 4.3 Multi-dimensional assessment for decision-making

An overall evaluation of varieties requires joint analysis of several traits, from both biophysical and socioeconomic perspectives. Different approaches have been proposed to handle multi-criteria prioritization on ranking varieties according to different traits of interest. Abeyasekera et al. (2002) developed a methodology to combine scores and rankings assigned by farmers evaluating bean (*Phaseolus vulgaris*) varieties, using a weighted index, which allows the farmer's preference to be captured, thus combining multiple criteria. The work of Abeyasekera et al. (2002) considered the farmers' preferences in terms of not only agronomic criteria (yield, pest resistance, etc.) but also non-agronomic criteria, such as taste, marketability and cooking time. Waldman et al. (2014) used choice experiment models to estimate farmers' preferences of perennial pigeon pea. Smith and Fennessy (2011) applied the PAPRIKA (Potentially All Pairwise RanKings of all possible Alternatives) method (Hansen and Ombler 2008) to assess the relative importance of traits on the improvement of perennial pasture species. The PAPRIKA method asks participants to compare pairs of options (varieties) and select one. It assumes full transitivity to reduce the number of pairs compared. In other words, when  $A > B$  and  $B > C$ , the model assumes  $A > C$ . An alternative method for priority setting is AgroDuos (Steinke and van Etten 2017), which is similar to PAPRIKA but integrates the concept of gamification to increase participants' engagement (Deterding et al. 2011), while it does not require interactive updating of questions and can therefore be used without a digital device or Internet connection.

Farmers' comprehensive evaluations of the total value of a variety can also be derived from on-farm trials. For example, the 'tricot' (*triadic comparisons of technologies*) approach proposed by Van Etten et al. (2016) integrates farmers' feedback on variety evaluation as a ranking of varieties, based on their overall appreciation of the varieties. In the tricot approach, each farmer receives three packages of seeds, each with a different variety of the crop (Van Etten et al. 2016). Each farmer ranks the three varieties from best to worst, according to overall performance, considering traits such as pest resistance and yield (Van Etten et al. 2016). Rankings of varieties directly evaluated by farmers in on-farm trials are aggregated by rank aggregation models (see Section 5.1).

Recent work of van Etten et al. (2019) is an example of how a rank-based model was applied to consider several criteria, such as disease resistance, yield and farmer preferences into a single judgement, in combination with local environmental conditions in the analysis crop variety

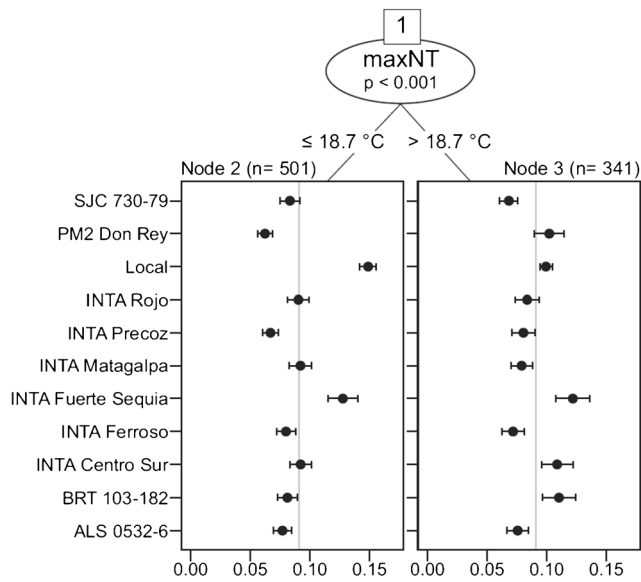
trials. The work of van Etten et al. (2019) includes three independent studies in three countries: Ethiopia, India and Nicaragua. For brevity, we focus on the case of Nicaragua, where varieties of common bean were evaluated in 842 plots. An extended version of the Plackett-Luce model, implemented in the PlackettLuce package (Turner et al. 2020), was fitted for the trial data collected from on-farm trials, which were established following the tricot approach (van Etten et al. 2019). The Plackett-Luce model estimates a worth parameter that represents the log probability of each evaluated element (a crop variety in this case) to be ranked first. Environmental conditions of the trial locations were included into the model using climatic indices (Table 1) as model covariates, through model-based recursive partitioning, implemented in the PlackettLuce package as Plackett-Luce Trees. The use of climatic variables as model covariates led to the identification of environmental factors that influenced the probability of a variety performing better than the other varieties tested in the trials (see Fig. 4 for an example).

The evaluations mentioned above focus on average performance of varieties, but other approaches focus on the variation in performance across seasons to assess farmers' risks. It has been shown that multi-environmental trial data from several seasons can be used to propose variety portfolios to reduce risk and maximize farmers' profits (Nalley et al. 2009; Nalley and Barkley 2010; Sukcharoen and Leatham 2016). These studies all focus on yield as the main evaluation criterion. Fadda and van Etten et al. (2019) proposed the adaptation of portfolio selection theory from financial asset management field, based on the portfolio management method developed by Dembo and King (1992). Instead of recommending a single variety, a portfolio of varieties is recommended based on calculations of the expected regret (Fadda and van Etten 2019). This method does not require absolute (yield) data and can also be applied to ranking data. This is interesting for progress in data synthesis, as ranking methods can play a role in combining datasets from different sources (see Section 5.1).

## 5 Data synthesis approaches

In the previous section, we reviewed methods for the analysis of different data types used for crop variety evaluation. In addition to that, data synthesis involves integration of datasets from heterogenous sources. For instance, datasets come from several research programs, each one with diverse types of data formats, measurement units and experimental designs.

Data synthesis for crop variety evaluation has followed two main lines of research: rank aggregation and network meta-analysis. In the remaining part of this section, we review relevant examples from both rank aggregation and network meta-analysis, to finally weigh their advantages, disadvantages



**Fig. 4** Plackett-Luce Tree of farmer-participatory tricot trial data in Nicaragua. The probability of each variety to perform better than the others in the trial is presented on the horizontal axis. Grey vertical line represents the average probability of better performance (1/number of evaluated varieties). From the study of van Etten et al. (2019, p. 4196, CC BY-NC-ND)

and existing gaps towards a data synthesis methodology for crop variety evaluation.

### 5.1 Rank aggregation methods

Rank aggregation methods are rank-based non-parametric statistical methods that allow for aggregation of results from individual studies to obtain one consensus ranking (Lin 2010; Yu et al. 2019). They have been applied to several fields including advertisement research, psychology, Internet search engines and biological studies (Lin 2010). Rank aggregation methods are suitable for high-level meta-analysis, where aggregation of different raw data is not feasible (Lin and Ding 2009). They also provide more statistical power than individual analyses (Simko and Pechenick 2010; Lin 2010). This coincides with one of the widely argued characteristics of meta-analysis (Cohn and Becker 2003).

Simko and Pechenick (2010) proposed to use rank aggregation methods to combine heterogeneous data from independent plant breeding trials. Simko and Linacre (2010) demonstrated how the Rasch model (Rasch 1960) can be used to combine heterogeneous data. The Rasch model is, in principle, very similar to the Luce model (Rasch 1960; Luce 1959).

Simko and Linacre (2010) presented four different real datasets as case examples; for brevity, we only focus on one of the datasets, containing data of potato chip quality evaluations. The analysis of this dataset implies two main constraints already mentioned in Section 3: (1) data measurements in different rating scales and (2) only partial overlap among tested varieties. Potato chip quality data were collected from online

databases of 10 different laboratories. As we described in Section 2, assessment of food quality and consumer preferences can be done using several rating scales. In the case of quality assessments of potato chips, Simko and Linacre (2010) explained that it is a common practice that each laboratory uses a different rating scale such as one of the following: (1) a rating scale of 5, 9 or 10 categories (the number is subjectively selected by each laboratory; lower values indicate higher quality of potato chip); (2) a measurement of the potato chip colour using specialized equipment with values ranging from 0 to 100 (higher readings indicate a lighter colour of potato chips, which is a desired trait); and (3) a percentage of chips passing a given quality test defined by the laboratory. In this example, it was not specified which rating scale was used by each laboratory in each test, but indeed different ranges of values exist across the different tests.

The data of potato chip quality assessments collected from 10 different laboratories were aggregated into one dataset. The aggregated dataset contained 63 cultivars over 157 trials, with only partial overlap among evaluated cultivars. For instance, only one cultivar was evaluated in 154 trials, while only seven cultivars were evaluated in a single trial (Simko and Linacre 2010). The resulting matrix contains 994 data points, around 10% of the expected total data points (9891) that would have resulted if all the varieties had been evaluated in all the trials (Simko and Linacre 2010). The original ordinal ratings are replaced for relative rankings (Simko and Linacre 2010). The relative rankings were used to calculate an overall performance rating, by means of an extended version of the Rasch model (Simko et al. 2012; Linacre and Wilson 1992). In this case, the extended version of the Rasch model allowed to compare 63 cultivars, even when not all were tested in the same trial.

Interestingly, rank aggregation has also found a direct application in variety trials, such as the work of van Etten et al. (2019) presented in Section 4. The successful application of rank-based methods in both trial analysis and meta-analysis shows that this is an interesting way forward in data synthesis for variety evaluation.

### 5.2 Network meta-analysis

Commonly used meta-analysis methodologies, especially in the medical sciences, are often based on pairwise comparison of treatments, usually in the form of an intervention against a control or placebo (Lumley 2002; Tonin et al. 2017). Network meta-analysis (Lumley 2002) allows the comparison of multiple treatments, even when some of them have never been compared directly in trials (Tonin et al. 2017). Although network meta-analysis is commonly used in medical sciences, it has also been used recently in other fields such as plant pathology (Madden et al. 2016). Network meta-analysis is also known by several other terms, such as ‘multiple treatments

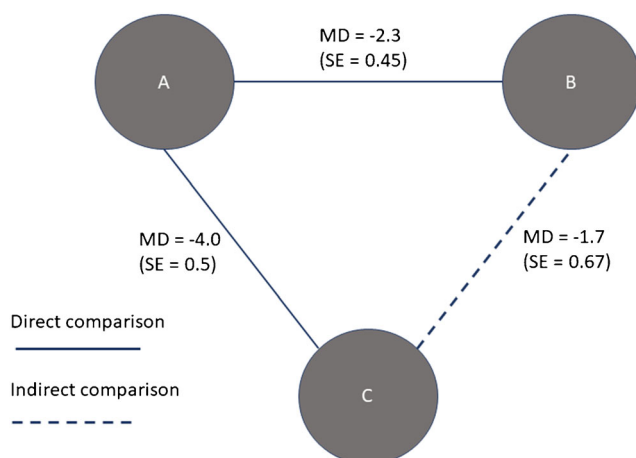


meta-analysis' and 'mixed treatment comparison', which are often used interchangeably (Salanti 2012).

As explained by Tonin et al. (2017), the approach evolved from the initial work of Bucher et al. (1997) on 'adjusted indirect treatment comparison', which was called 'network meta-analysis' after the improvements made by Lumley (2002), and later evolved to 'mixed treatment comparison' by Lu and Ades (2004). A distinctive characteristic of network meta-analysis is the case when both direct and indirect comparisons are available for a given pair of treatments. In this case, evidence from both direct and indirect comparisons is used to do a mixed treatment comparison (Fig. 5), hence the alternative name (Dias and Caldwell 2019). For more details about related terminology on mixed treatment comparisons, we refer to Salanti (2012) and Coleman et al. (2012).

Network meta-analysis can be implemented with two different types of models: (1) contrast-based models, also known as conditional models, in which the treatment effects per trial are estimated as a contrast relative to a baseline treatment to subsequently analyse all the contrasts across studies, and (2) arm-based models, also known as unconditional models, in which the treatment summaries per trial are analysed in a two-way linear mixed model (Piepho et al. 2012b; Madden et al. 2016). Arm-based models are commonly applied for the analysis of multi-environment crop variety trials (Section 4) (Piepho 1997; Albert and Makowski 2018; Damesa et al. 2017). As explained in Section 4, it is possible to use single-stage or two-stage analysis in linear mixed models, although for network meta-analysis, a single-stage analysis might be constrained by the availability of data from individual primary studies, rather than usual summary results such as the estimated effect sizes (Madden et al. 2016).

Both frequentist and Bayesian approaches can be applied to network meta-analysis (Tonin et al. 2017), although the Bayesian approach seems to be more popular (Piepho et al.



**Fig. 5** Example of a network of treatments (varieties) allowing direct and indirect comparisons. Adapted from Dias and Caldwell (2019). MD, mean difference

2012b). Network meta-analysis usually includes the use of network diagrams, where the nodes represent the compared elements (e.g. treatments or varieties), and the lines (edges) connecting the nodes represent the direct comparison of elements, to evaluate network connectivity. This is relevant in network meta-analysis, especially because poorly connected networks might provide less reliable results compared to a strongly connected network (Tonin et al. 2017). It is also possible the computation of ranking probabilities for each treatment to be assigned a particular position in a ranking from best to worst (Tonin et al. 2017).

Based on yield data obtained from 28 published papers selected through a systematic literature review, Laurent et al. (2015) applied both direct and indirect comparisons in a meta-analysis for ranking crop species based on yield. Direct comparisons compare crops which were grown at the same site and in the same year, whereas indirect comparisons compare crops grown at different sites or in different years, using a third crop grown at all sites as a reference (Laurent et al. 2015). In this case, only results from experimental sites were considered (no farmers' fields), resulting in a database containing 856 records of yield for 36 crop species (Laurent et al. 2015). Mean yield was estimated using a linear mixed effect model, with a log transformation to normalize the yield data (Laurent et al. 2015). For the direct comparison, four crops species (*Miscanthus × giganteus*, *Panicum virgatum*, *Triticosecale*, *Salix*) were selected to be used as reference crops, as they were included in the higher number of comparisons with other crops for the same site-years (Laurent et al. 2015). A model was fitted for each reference crop using restricted maximum likelihood. Then, yield ratios of the mean yield of each evaluated crop (except reference crops) to the mean yield of a reference crop grown in the same site and year were calculated (Laurent et al. 2015).

Since direct comparison allows to compare only a limited number of species, indirect comparison was used to compare the yields of a crop of interest, *Miscanthus × giganteus*, to yields of crops that were not grown in the same site-years as the crop of interest. Three reference crops were selected for the indirect comparison: *Panicum virgatum*, *Triticosecale* and *Salix*. Therefore, *Miscanthus × giganteus* was compared to crops not grown in the same site-years, by indirect comparison using the reference crops, allowing to include more crop species than using direct comparison only.

Albert and Makowski (2018) recently published a paper describing the use of Bayesian mixed treatment comparison models for ranking crop species. According to Albert and Makowski (2018), the dataset used is the same as that analysed in Laurent et al. (2015), although they also indicate that 639 yield observations were analysed, which are less than the 856 yield data observations analysed in Laurent et al. (2015). Mixed treatment comparison combines direct and indirect evidence (Dias and Caldwell 2019). Five different models were fitted (Table 4), of which four were contrast-

based models and one was an arm-based model. According to Albert and Makowski (2018), the Bayesian contrast-based models (1 to 4) are variants of the model presented by Dias et al. (2010), while the arm-based is a Bayesian two-way model. The model estimation was done using Markov chain Monte Carlo simulations, while model assessment was made using the deviance information criterion (DIC), in which the models with the lowest DIC are preferred (Albert and Makowski 2018). Compared to the rankings obtained by Laurent et al. (2015) using direct and indirect comparison, the results are very similar for the two species with higher yields (*Pennisetum purpureum* and *Arundo donax*) when compared against *Miscanthus × giganteus*.

### 5.3 Assessment of available data synthesis methods

The methods reviewed above address the challenge of combining crop variety trial data from multiple and independent sources. In Section 3, we presented a set of challenges identified by Simko and Pechenick (2010) that arise when aiming to combine data from different trials. Here, we assess both rank aggregation and network meta-analysis as solutions to those problems. Additionally, we provide an overview of the relative strengths and weaknesses of data synthesis methods.

#### 5.3.1 Partial overlap in evaluated accessions between trials

The problem of partial overlap in the varieties evaluated across trials can be solved by exploiting the capacity of rank aggregation methods to handle partially ranked lists, although the specific approach depends on the particular rank aggregation method. For example, some models are based on pairwise comparison such as the Bradley-Terry model, while others allow multiple comparisons, such as models based on the Plackett-Luce model. In the case of network meta-analysis, the problem of partial overlap is solved by indirect comparison. For example, in Fig. 5, items B and C are indirectly compared with A. Examples are Laurent et al. (2015) and Albert and Makowski (2018) who used reference crop species to allow the comparison of crop species not tested in the same trial.

#### 5.3.2 Measurements based on different rating scales or different methods

Rank aggregation methods solve the problems of measurements in different rating scales or traits evaluated with different methodologies, replacing the original raw data from each trial by relative rankings (Simko and Piepho 2011; Simko et al. 2012). In the work of Laurent et al. (2015) and Albert and Makowski (2018) this was no issue, however, because the data from the different yield studies were in the same units, tons of dry matter per ha per year (Laurent et al. 2015). Network meta-analysis and meta-analysis, in general, can deal with measurements in different units by estimating either the standardized mean difference or the response ratio (Borenstein et al. 2009; Makowski et al. 2019; Murad et al. 2019).

#### 5.3.3 Relative strengths and weaknesses of data synthesis methods

There are a few studies applying either rank aggregation or network meta-analysis to crop variety evaluation. Future studies will need to consider the relative merits of each.

Network meta-analysis can provide absolute values (yield differences in tons per hectare), which is difficult to obtain with rank-based models. Even so, the item ‘worth’ estimated by the rank-based methods is linearly correlated with the underlying latent variable (for example, yield) (Coe 2002; Fadda and van Etten 2019). Also, in theory, it should be possible to combine ranking data and continuous variables in the same model (Böckenholt 2004), but this is still challenging in practice, as such models have not been implemented in general use software.

A useful output that can be obtained from both rank aggregation and network meta-analysis is ranking probabilities, the probability of each variety to be ranked first. Ranking probabilities are related to the concept of *reliability* in plant breeding, the probability of outperforming a check variety (a reference; for example a previously released variety, commonly used variety or market leader). The concept of reliability was proposed by Eskridge (1990) in the context of crop improvement as a ‘safety-first’ approach, with subsequent applications by Eskridge and Mumm (1992) and Eskridge (1997).

**Table 4** Description of models used by Albert and Makowski (2018)

Model number	Model type	Effect	Variance	DIC
1	Contrast-based	Fixed	Common residual	912
2	Contrast-based	Random	Common residual	348
3	Contrast-based	Random	Species-specific residual	287
4	Contrast-based	Random	Study-specific residual	214
5	Arm-based	Two-way model		348

Data synthesis approaches should consider the ease of use and interpretation by decision-makers in crop variety evaluation. In that sense, the complexity of network meta-analysis can lead to confusion on model implementation and interpretation (Madden et al. 2016). This complexity might be a barrier to its wider adoption as a tool for data synthesis in crop variety evaluation, just like the low level of expertise of users, is limiting the uptake of more sophisticated  $G \times E$  analysis methods (Lecomte et al. 2010). Rank aggregation methods might be easier to implement but have implicit trade-offs such as information loss and less power to detect existing differences if compared to parametric methods (Simko and Linacre 2010; Whitley and Ball 2002; Sabaghnia 2016).

## 6 Conclusions and recommendations

We structure this section around three main statements based on our review, which derive conclusions from our main findings and translate these into recommendations.

### 6.1 Elements for a data synthesis approach are available and aligned around ranks and reliability

Based on our review, we assert that the main elements are available for data synthesis as an overarching approach that integrates different components, such as data, models and knowledge from experts (farmers and breeders), to efficiently extract useful information to support decision-making. We remarked in Section 5 data synthesis methods that have been tested, exist and can integrate well with existing trial analysis approaches. In particular, rank-based approaches fit within a conceptual framework to analyse variety superiority based on reliability (probability of outperforming a check). A rank-based framework would be able to make versatile use of data from different sources, without complex transformations or doubtful assumptions, and would facilitate the integration of objective measurements and preference data.

### 6.2 Data synthesis should progress from general to specific and from simple to complex

Explicit crop growth modelling has been proposed more than once as a way forward to integrate different types of data into a single conceptual framework for the evaluation of variety performance. However, model building starting from a detailed crop model is not parsimonious and does not build up complexity in a gradual way. For many crops, growth models are not available, hence requiring a large upfront investment in basic (eco)physiological research to enable model building. Also, as shown in Section 4, it seems that progress in this field is mainly theoretical, and that practical advances are limited. Even for the attempts that result in generalizable results, the

focus is solely on yield and excludes user perspectives. For crop variety evaluation, it seems more logical to start with the ‘big picture’ and work down to the details based on better information indicating where the largest gain in accuracy can be obtained (Section 4). This may involve some type of explicit, physiological modelling, but perhaps of a limited number of aspects, not requiring a fully fledged crop growth simulation model. Therefore, we think that a further investment in simpler methods is warranted. This may be less stimulating from a basic research point of view but may give rise to new questions and priorities and give a better sense of the societal relevance and external validity of data synthesis efforts.

### 6.3 Use cases can spur further data sharing and model development

Our review shows that engaging the research community in data sharing is a major challenge (Section 3). Most efforts, however, have focused on the supply side: by encouraging/obligating researchers to share data and by providing the infrastructure to do so. While these efforts are certainly important, in crop science, few concomitant efforts have looked at the prospects of successful use of shared data that would drive citations of data papers, shape collaborations around data analysis, and increase researchers’ motivation for further sharing. Success may at least partially be the result of a siphon effect: some early use cases can perhaps inspire other researchers to engage in sharing and start a virtuous cycle, as described in Section 3. Therefore, investment in a few use cases that use relatively simple methods to show the potential benefits of data synthesis for crop variety evaluation is needed. Our review has shown that those methods are available in principle (Section 5). Even so, they need a modest investment to be adapted and demonstrated for this field of application. Next steps would involve stepwise refinements to address components of variety performance that substantially improve the accuracy of predictions. Close collaboration with the decision-makers interested in such evaluations could also spur further interest in this area of research and demonstrate the relevance of further investment.

**Acknowledgements** We thank Olga Spellman for the editorial support.

**Authors’ contributions** Conceptualization: D.B., I.V.B., S.d.B. and J.V.E.; Writing of the original draft: D.B., I.V.B., S.d.B. and J.V.E.; Writing, review and editing: D.B., I.V.B., S.d.B., L.M. and J.V.E.

**Funding information** This research was undertaken as part of, and funded by, the CGIAR Research Program on Roots, Tubers and Bananas (RTB) and supported by the CGIAR Trust Fund contributors (<https://www.cgiar.org/funders/>). This research was also part of the PhD research project of David Brown, who received funding from the Bill & Melinda Gates Foundation through the Project ‘Improvement of Banana for Smallholder Farmers in the Great Lakes Region of Africa’.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abeyasekera S, Ritchie JM, Lawson-McDowall J (2002) Combining ranks and scores to determine farmers' preferences for bean varieties in southern Malawi. *Exp Agric* 38(1):97–109. <https://doi.org/10.1017/S0014479702000182>
- Acquaah G (2012) Performance evaluation for crop cultivar release. In: Acquaah G (ed) *Principles of plant genetics and breeding*, 2nd edn. John Wiley & Sons, Ltd, Oxford, pp 491–510. <https://doi.org/10.1002/9781118313718.ch26>
- Albert I, Makowski D (2018) Ranking crop species using mixed treatment comparisons. *Res Synth Methods* 0 (0). doi:<https://doi.org/10.1002/jrsm.1328>
- Anonymous (1975) Minutes of division business meeting. Institute of Food Technologists Sensory Evaluation Division, IFT, Chicago, IL
- Arief VN, DeLacy IH, Crossa J, Payne T, Singh R, Braun H-J, Tian T, Basford KE, Dieters MJ (2015) Evaluating testing strategies for plant breeding field trials: redesigning a CIMMYT international wheat nursery. *Crop Sci* 55(1):164–177. <https://doi.org/10.2135/cropsci2014.06.0415>
- Atlin GN, Cooper M, Bjørnstad Å (2001) A comparison of formal and participatory breeding approaches using selection theory. *Euphytica* 122(3):463–475. <https://doi.org/10.1023/a:1017557307800>
- Billiau K, Sprenger H, Schudoma C, Walther D, Köhl KI (2012) Data management pipeline for plant phenotyping in a multisite project. *Funct Plant Biol* 39(11):948–957. <https://doi.org/10.1071/FP12009>
- Böckenholt U (2004) Comparative judgments as an alternative to ratings: identifying the scale origin. *Psychol Methods* 9(4):453–465. <https://doi.org/10.1037/1082-989X.9.4.453>
- Bolger AM, Poorter H, Dumschott K, Bolger ME, Arend D, Osorio S, Gundlach H, Mayer KFX, Lange M, Scholz U, Usadel B (2019) Computational aspects underlying genome to phenome analysis in plants. *Plant J* 97(1):182–198. <https://doi.org/10.1111/tbj.14179>
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) When does it make sense to perform a meta-analysis? In: *Introduction to meta-analysis*. John Wiley & Sons, Chichester, pp 357–364
- Bradley RA, Terry ME (1952) Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* 39(3/4):324
- Brancourt-Hulmel M, Biarnès-Dumoulin V, Denis JB (1997) Points de repère dans l'analyse de la stabilité et de l'interaction génotype-milieu en amélioration des plantes. *Agronomie* 17(4):219–246
- Breiman L (2001) Statistical modeling: the two cultures. *Stat Sci* 16(3): 199–215
- Bucher HC, Guyatt GH, Griffith LE, Walter SD (1997) The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 50(6):683–691. [https://doi.org/10.1016/S0895-4356\(97\)00049-8](https://doi.org/10.1016/S0895-4356(97)00049-8)
- Bustos-Korts D, Romagosa I, Borràs-Gelonch G, Casas AM, Slafer GA, van Eeuwijk F (2019) Genotype by environment interaction and adaptation. In: Savin R, Slafer GA (eds) *Crop science*. Springer New York, New York, pp 29–71. [https://doi.org/10.1007/978-1-4939-8621-7\\_199](https://doi.org/10.1007/978-1-4939-8621-7_199)
- Cadotte MW, Mehrkens LR, Menge DNL (2012) Gauging the impact of meta-analysis on ecology. *Evol Ecol* 26(5):1153–1167. <https://doi.org/10.1007/s10682-012-9585-z>
- Candela L, Castelli D, Manghi P, Tani A (2015) Data journals: a survey. *J Assoc Inf Sci Technol* 66(9):1747–1762. <https://doi.org/10.1002/asi.23358>
- Carpenter SR, Armbrust EV, Arzberger PW, Chapin FS III, Elser JJ, Hackett EJ, Ives AR, Kareiva PM, Leibold MA, Lundberg P, Mangel M, Merchant N, Murdoch WW, Palmer MA, Peters DPC, Pickett STA, Smith KK, Wall DH, Zimmerman AS (2009) Accelerate synthesis in ecology and environmental sciences. *BioScience* 59(8):699–701. <https://doi.org/10.1525/bio.2009.59.8.11>
- Ceccarelli S (1989) Wide adaptation: how wide? *Euphytica* 40(3):197–205. <https://doi.org/10.1007/bf00024512>
- Ceccarelli S (2012) *Plant breeding with farmers – a technical manual*. ICARDA, Aleppo
- Ceccarelli S, Grando S (2007) Decentralized-participatory plant breeding: an example of demand driven research. *Euphytica* 155(3):349–360. <https://doi.org/10.1007/s10681-006-9336-8>
- Ceccarelli S, Guimarães EP, Weltzien E (2009) *Plant breeding and farmer participation*. Food and Agriculture Organization of the United Nations, Rome
- Chawade A, van Ham J, Blomquist H, Bagge O, Alexandersson E, Ortiz R (2019) High-throughput field-phenotyping tools for plant breeding and precision agriculture. *Agronomy* 9(5):258
- Chenu K (2015) Chapter 13 - Characterizing the crop environment – nature, significance and applications. In: Sadras VO, Calderini DF (eds) *Crop physiology*, Second edn. Academic, San Diego, pp 321–348. <https://doi.org/10.1016/B978-0-12-417104-6.00013-3>
- Christinck A, Weltzien E, Rattunde F, Ashby J (2017) Gender differentiation of farmer preferences for varietal traits in crop improvement: evidence and issues. CGIAR System Management Office and International Center for Tropical Agriculture (CIAT), Cali
- Coe R (2007) Analysing data from participatory on-farm trials. *Afr Stat J* 4
- Coe RD (2002) Analyzing ranking and rating data from participatory on-farm trials. In: Bellon MR, Reeves J (eds) *Quantitative analysis of data from participatory methods in plant breeding*. CIMMYT, Mexico, pp 44–65
- Cohn LD, Becker BJ (2003) How meta-analysis increases statistical power. *Psychol Methods* 8(3):243–253. <https://doi.org/10.1037/1082-989X.8.3.243>
- Coleman CI, Phung OJ, Cappelleri JC, Baker WL, Kluger J, White CM, Sobieraj DM (2012) Use of mixed treatment comparisons in systematic reviews. University of Connecticut/Hartford Hospital Evidence-based Practice Center, Agency for Healthcare Research and Quality (US), Rockville
- Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos T R Soc B Biol Sci* 363(1491):557–572. <https://doi.org/10.1098/rstb.2007.2170>
- Coumou D, Rahmstorf S (2012) A decade of weather extremes. *Nat Clim Chang* 2:491–496. <https://doi.org/10.1038/nclimate1452>
- Ćwiek-Kupczyńska H, Altmann T, Arend D, Arnaud E, Chen D, Cornut G, Fiorani F, Frohberg W, Junker A, Klukas C, Lange M, Mazurek C, Nafissi A, Neveu P, van Oeveren J, Pommier C,



- Poorter H, Rocca-Serra P, Sansone S-A, Scholz U, van Schriek M, Seren Ü, Usadel B, Weise S, Kersey P, Krajewski P (2016) Measures for interoperability of phenotypic data: minimum information requirements and formatting. *Plant Methods* 12(1):44. <https://doi.org/10.1186/s13007-016-0144-4>
- Dalrymple DG (2008) International agricultural research as a global public good: concepts, the CGIAR experience and policy issues. *J Int Dev* 20(3):347–379. <https://doi.org/10.1002/jid.1420>
- Damesa TM, Möhring J, Worku M, Piepho H-P (2017) One step at a time: stage-wise analysis of a series of experiments. *Agron J* 109(3): 845–857. <https://doi.org/10.2134/agronj2016.07.0395>
- Dawson JC, Healy G (2018) Flavour evaluation for plant breeders. In: Goldman I (ed) *Plant breeding reviews*. <https://doi.org/10.1002/9781119414735.ch5>
- de Leon N, Jannink J-L, Edwards JW, Kaeppler SM (2016) Introduction to a special issue on genotype by environment interaction. *Crop Sci* 56(5):2081–2089. <https://doi.org/10.2135/cropsci2016.07.0002in>
- Dembo RS, King AJ (1992) Tracking models and the optimal regret distribution in asset allocation. *Applied Stochastic Models and Data Analysis* 8(3):151–157. <https://doi.org/10.1002/asm.3150080305>
- Desclaux D, Nolot JM, Chiffolleau Y, Gozé E, Leclerc C (2008) Changes in the concept of genotype  $\times$  environment interactions to fit agriculture diversification and decentralized participatory plant breeding: pluridisciplinary point of view. *Euphytica* 163(3):533–546. <https://doi.org/10.1007/s10681-008-9717-2>
- Deterding S, Dixon D, Khaled R, Nacke L (2011) From game design elements to gamefulness: defining “gamification”. In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, Tampere, Finland. Association for Computing Machinery, pp. 9–15. doi:<https://doi.org/10.1145/2181037.2181040>
- Dias S, Caldwell DM (2019) Network meta-analysis explained. *Arch Dis Child Fetal Neonatal Ed* 104(1):F8–F12. <https://doi.org/10.1136/archdischild-2018-315224>
- Dias S, Welton NJ, Caldwell DM, Ades AE (2010) Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 29(7–8): 932–944. <https://doi.org/10.1002/sim.3767>
- Diekmann F (2012) Data practices of agricultural scientists: results from an exploratory study. *J Agr Food Inform* 13(1):14–34. <https://doi.org/10.1080/10496505.2012.636005>
- Dittrich R, Hatzinger R, Katzenbeisser W (1998) Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *J R Stat Soc Ser C Appl Stat* 47(4): 511–525. <https://doi.org/10.1111/1467-9876.00125>
- Elias AA, Robbins KR, Doerge RW, Tuinstra MR (2016) Half a century of studying genotype  $\times$  environment interactions in plant breeding experiments. *Crop Sci* 56(5):2090–2105. <https://doi.org/10.2135/cropsci2015.01.0061>
- Ersoz ES, Martin NF, Stapleton AE (2020) On to the next chapter for crop breeding: convergence with data science. *Crop Sci* 60(2):639–655. <https://doi.org/10.1002/csc2.20054>
- Eskridge KM (1990) Selection of stable cultivars using a safety-first rule. *Crop Sci* 30(2):369–374. <https://doi.org/10.2135/cropsci1990.0011183X003000020025x>
- Eskridge KM (1997) Evaluation of corn hybrids using the probability of outperforming a check based on strip-test data. *J Agric Biol Environ Stat* 2(3):245–254. <https://doi.org/10.2307/1400444>
- Eskridge KM, Mumm RF (1992) Choosing plant cultivars based on the probability of outperforming a check. *Theor Appl Genet* 84(3):494–500. <https://doi.org/10.1007/bf00229512>
- Esposito S, Carputo D, Cardi T, Tripodi P (2020) Applications and trends of machine learning in genomics and phenomics for next-generation breeding. *Plants* 9(1):34
- Fadda C, van Etten J (2019) Generating farm-validated variety recommendations for climate adaptation. In: Rosenstock TS, Nowak A, Girvetz E (eds) *The climate-smart agriculture papers: investigating the business of a productive, resilient and low emission future*. Springer International, Cham, pp 127–138. [https://doi.org/10.1007/978-3-319-92798-5\\_11](https://doi.org/10.1007/978-3-319-92798-5_11)
- Feng P, Wang B, Liu DL, Waters C, Yu Q (2019) Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in South-Eastern Australia. *Agric For Meteorol* 275:100–113. <https://doi.org/10.1016/j.agrformet.2019.05.018>
- Finlay K, Wilkinson G (1963) The analysis of adaptation in a plant-breeding programme. *Aust J Agr Res* 14(6):742–754. <https://doi.org/10.1071/AR9630742>
- Folta KM, Klee HJ (2016) Sensory sacrifices when we mass-produce mass produce. *Hortic Res* 3:16032. <https://doi.org/10.1038/hortres.2016.32>
- Franzel S, Coe R (2002) Participatory on-farm technology testing: the suitability of different types of trials for different objectives. In: Franzel S, Coe R (eds) *Quantitative analysis of data from participatory methods in plant breeding*. CIMMYT, Mexico, pp 1–8
- Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A, Michaelsen J (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data* 2:150066. <https://doi.org/10.1038/sdata.2015.66>
- Gauch HG Jr (1992) *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier, New York
- Germeier CU, Unger S (2019) Modeling crop genetic resources phenotyping information systems. *Front Plant Sci* 10(728). <https://doi.org/10.3389/fpls.2019.00728>
- Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, Duke CS, Porter JH (2013) Big data and the future of ecology. *Front Ecol Environ* 11(3):156–162. <https://doi.org/10.1890/120103>
- Hansen P, Ombler F (2008) A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *Journal of Multi-Criteria Decision Analysis* 15(3–4):87–107. <https://doi.org/10.1002/mcda.428>
- Hartung K, Piepho HP (2005) A threshold model for multiyear genebank data based on different rating scales. *Crop Sci* 45(3):1045–1051
- Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B (2017) SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* 12(2):e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Huehn M (1990) Nonparametric measures of phenotypic stability. Part 1: theory. *Euphytica* 47(3):189–194. <https://doi.org/10.1007/bf00024241>
- Hyman G, Espinosa H, Camargo P, Abreu D, Devare M, Arnaud E, Porter C, Mwanzia L, Sonder K, Traore S (2017) Improving agricultural knowledge management: the AgTrials experience [version 2; peer review: 2 approved]. *F1000Research* 6 (317). doi:<https://doi.org/10.12688/f1000research.11179.2>
- Hyman G, Hodson D, Jones P (2013) Spatial analysis to support geographic targeting of genotypes to environments. *Front Physiol* 4(40). <https://doi.org/10.3389/fphys.2013.00040>
- Jarvis A, Reuter HI, Nelson A, Guevara E (2008) Hole-filled seamless SRTM data V4. International Centre for Tropical Agriculture (CIAT). <http://srtm.csi.cgiar.org/>
- Jeuffroy M-H, Casadebaig P, Debaeke P, Loyce C, Meynard J-M (2014) Agronomic model uses to predict cultivar performance in various environments and cropping systems. A review. *Agron Sustain Dev* 34(1):121–137. <https://doi.org/10.1007/s13593-013-0170-9>

- Kamilaris A, Kartakoullis A, Prenafeta-Boldú FX (2017) A review on the practice of big data analysis in agriculture. *Comput Electron Agric* 143:23–37. <https://doi.org/10.1016/j.compag.2017.09.037>
- Kang MS (1997) Using genotype-by-environment interaction for crop cultivar development. In: Sparks DL (ed) *Advances in agronomy*, vol 62. Academic, pp 199–252. [https://doi.org/10.1016/S0065-2113\(08\)60569-6](https://doi.org/10.1016/S0065-2113(08)60569-6)
- Kang MS, Gorman DP (1989) Genotype  $\times$  environment interaction in maize. *Agron J* 81(4):662–664. <https://doi.org/10.2134/agronj1989.00021962008100040020x>
- Kehel Z, Crossa J, Reynolds M (2016) Identifying climate patterns during the crop-growing cycle from 30 years of CIMMYT Elite Spring Wheat International Yield Trials. *Applied Mathematics and Omics to Assess Crop Genetic Resources for Climate Change Adaptive Traits*: 151–174
- Kipp S, Mistele B, Baresel P, Schmidhalter U (2014) High-throughput phenotyping early plant vigour of winter wheat. *Eur J Agron* 52: 271–278. <https://doi.org/10.1016/j.eja.2013.08.009>
- Krajewski P, Chen D, Ćwiek H, van Dijk ADJ, Fiorani F, Kersey P, Klukas C, Lange M, Markiewicz A, Nap JP, van Oeveren J, Pommier C, Scholz U, van Schriek M, Usadel B, Weise S (2015) Towards recommendations for metadata and data handling in plant phenotyping. *J Exp Bot* 66(18):5417–5427. <https://doi.org/10.1093/jxb/erv271>
- Krupnik TJ, Andersson JA, Rusinamhodzi L, Corbeels M, Shennan C, GÉRard B (2019) Does size matter? A critical review of meta-analysis in agronomy. *Exp Agric*:1–30. doi:<https://doi.org/10.1017/S0014479719000012>, 55
- Lammerts van Bueren ET, Struik PC, van Eekeren N, Nuijten E (2018) Towards resilience through systems-based plant breeding. A review. *Agron Sustain Dev* 38(5):42. <https://doi.org/10.1007/s13593-018-0522-6>
- Laurent A, Pelzer E, Loyce C, Makowski D (2015) Ranking yields of energy crops: a meta-analysis using direct and indirect comparisons. *Renew Sustain Energy Rev* 46:41–50. <https://doi.org/10.1016/j.rser.2015.02.023>
- Lecomte C, Prost L, Cerf M, Meynard J-M (2010) Basis for designing a tool to evaluate new cultivars. *Agron Sustain Dev* 30(3):667–677. <https://doi.org/10.1051/agro/2009042>
- Leonelli S, Davey RP, Arnaud E, Parry G, Bastow R (2017) Data management and best practice for plant science. *Nat Plants* 3:17086. <https://doi.org/10.1038/nplants.2017.86>
- Lesk C, Rowhani P, Ramankutty N (2016) Influence of extreme weather disasters on global crop production. *Nature* 529:84–87. <https://doi.org/10.1038/nature16467>
- Liakos K, Busato P, Moshou D, Pearson S, Bochtis D (2018) Machine learning in agriculture: a review. *Sensors* 18(8):2674
- Lim J (2011) Hedonic scaling: a review of methods and theory. *Food Qual Prefer* 22(8):733–747. <https://doi.org/10.1016/j.foodqual.2011.05.008>
- Lim J, Wood A, Green BG (2009) Derivation and evaluation of a labeled hedonic scale. *Chem Senses* 34(9):739–751. <https://doi.org/10.1093/chemse/bjp054>
- Lin S (2010) Rank aggregation methods. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (5):555–570. doi:<https://doi.org/10.1002/wics.111>
- Lin S, Ding J (2009) Integration of ranked lists via cross entropy Monte Carlo with applications to mRNA and microRNA studies. *Biometrics* 65(1):9–18. <https://doi.org/10.1111/j.1541-0420.2008.01044.x>
- Linacre JM, Wilson M (1992) Objective measurement of rank-ordered objects. In: *Objective measurement: theory into practice*. Ablex, Norwood, pp 195–209
- Lobell DB, Bänziger M, Magorokosho C, Vivek B (2011) Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nat Clim Chang* 1(1):42–45. <https://doi.org/10.1038/nclimate1043>
- Lu G, Ades AE (2004) Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 23(20):3105–3124. <https://doi.org/10.1002/sim.1875>
- Luce RD (1959) *Individual choice behavior: a theoretical analysis*. John Wiley, New York
- Lumley T (2002) Network meta-analysis for indirect treatment comparisons. *Stat Med* 21(16):2313–2324. <https://doi.org/10.1002/sim.1201>
- Madden LV, Piepho HP, Paul PA (2016) Statistical models and methods for network meta-analysis. *Phytopathology* 106(8):792–806. <https://doi.org/10.1094/PHYTO-12-15-0342-RVW>
- Makowski D, Pirau F, Brun F (2019) Basic concepts in meta-analysis. In: *From experimental network to meta-analysis: methods and applications with R for agronomic and environmental sciences*. Springer Netherlands, Dordrecht, pp 105–126. [https://doi.org/10.1007/978-94-024-1696-1\\_6](https://doi.org/10.1007/978-94-024-1696-1_6)
- Malosetti M, Ribaut J-M, van Eeuwijk FA (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Front Physiol* 4:44–44. <https://doi.org/10.3389/fphys.2013.00044>
- Mark H, Briony F, Jon T, Grace B, Ross W, Barend M, Erik S, Luiz OBdSS, Pavel A, Igor O (2018) The state of open data report 2018. Digital science report. Digital Science doi:<https://doi.org/10.6084/m9.figshare.7195058.v2>
- Matteis L, Chibon PY, Espinosa H, Skofic M, Finkers R, Bruskiwicz R, Hyman G, Arnaud E (2013) Crop ontology: vocabulary for crop-related concepts. In: Larmande P, Arnaud E, Mougenot I, Jonquet C, Libourel T (eds) *1st international workshop on semantics for biodiversity (S4BioDiv)*, Montpellier, France. CEUR, pp 29–38
- McLaren CG, Bruskiwicz RM, Portugal AM, Cosico AB (2005) The International Rice Information System. A platform for meta-analysis of rice crop data. *Plant Physiol* 139(2):637–642. <https://doi.org/10.1104/pp.105.063438>
- Möhrling J, Piepho H-P (2009) Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Sci* 49(6):1977–1988. <https://doi.org/10.2135/cropsci2009.02.0083>
- Murad MH, Wang Z, Chu H, Lin L (2019) When continuous outcomes are measured using different scales: guide for meta-analysis and interpretation. *BMJ* 364:k4817–k4817. <https://doi.org/10.1136/bmj.k4817>
- Nalley LL, Barkley A, Watkins B, Hignight J (2009) Enhancing farm profitability through portfolio analysis: the case of spatial rice variety selection. *J Agric Appl Econ* 41(3):641–652. <https://doi.org/10.1017/S1074070800003126>
- Nalley LL, Barkley AP (2010) Using portfolio theory to enhance wheat yield stability in low-income nations: an application in the Yaqui Valley of northwestern Mexico. *Agric Resour Econ Rev* 35(2):334–347
- Orjeda G (2000) Evaluating bananas: a global partnership: results of IMTP phase II. *Biodiversity International*, Montpellier
- Papoutsoglou EA, Faria D, Arend D, Arnaud E, Athanasiadis IN, Chaves I, Coppens F, Cornut G, Costa BV, Ćwiek Kupczyńska H, Drosbecke B, Finkers R, Gruden K, Junker A, King GJ, Krajewski P, Lange M, Laporte M-A, Michotey C, Oppermann M, Ostler R, Poorter H, Ramirez-Gonzalez R, Ramšak Ž, Reif JC, Roccaserra P, Sansone S-A, Scholz U, Tardieu F, Uauy C, Usadel B, Visser RGF, Weise S, Kersey PJ, Miguel CM, Adamblondon A-F, Pommier C (2020) Enabling reusability of plant phenomic datasets with MIAPPE 1.1. *New Phytologist*:nph.16544. doi:<https://doi.org/10.1111/nph.16544>
- Parkes B, Higginbottom TP, Hufkens K, Ceballos F, Kramer B, Foster T (2019) Weather dataset choice introduces uncertainty to estimates of crop yield responses to climate variability and change. *Environ Res Lett* 14(12):124089. <https://doi.org/10.1088/1748-9326/ab5ebb>
- Peryam D, Girardot N (1952) QM pins food “likes” and “dislikes” with advanced taste-test method. *Food Engineering* 24(58)

- Philibert A, Loyce C, Makowski D (2012) Assessment of the quality of meta-analysis in agronomy. *Agric Ecosyst Environ* 148:72–82. <https://doi.org/10.1016/j.agee.2011.12.003>
- Pickett STA, Kolasa J, Jones CG (2007) Integration and synthesis. In: Pickett STA, Kolasa J, Jones CG (eds) *Ecological understanding*, Second edn. Academic, San Diego, pp 146–167. <https://doi.org/10.1016/B978-012554522-8.50009-1>
- Piepho H-P, Denis J-B, van Eeuwijk FA (1998) Predicting cultivar differences using covariates. *J Agric Biol Environ Stat* 3(2):151–162. <https://doi.org/10.2307/1400648>
- Piepho HP (1997) Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics* 53(2):761–766. <https://doi.org/10.2307/2533976>
- Piepho HP (2000) Exact confidence limits for covariate-dependent risk in cultivar trials. *J Agric Biol Environ Stat* 5(2):202–213. <https://doi.org/10.2307/1400531>
- Piepho HP, Möhring J, Schulz-Streck T, Ogutu JO (2012a) A stage-wise approach for the analysis of multi-environment trials. *Biom J* 54(6): 844–860. <https://doi.org/10.1002/bimj.201100219>
- Piepho HP, Williams ER, Madden LV (2012b) The use of two-way linear mixed models in multitreatment meta-analysis. *Biometrics* 68(4): 1269–1277. <https://doi.org/10.1111/j.1541-0420.2012.01786.x>
- Pillemer D, Light R (1980) Synthesizing outcomes: how to use research evidence from many studies. *Harvard Educ Rev* 50(2):176–195. <https://doi.org/10.17763/haer.50.2.v755316522jkup33>
- Plackett RL (1975) The analysis of permutations. *J R Stat Soc Ser C Appl Stat* 24(2):193–202. <https://doi.org/10.2307/2346567>
- Priyadarshan PM (2019) Maintenance breeding and variety release. In: *PLANT BREEDING: classical to modern*. Springer Singapore, Singapore, pp 561–570. [https://doi.org/10.1007/978-981-13-7095-3\\_25](https://doi.org/10.1007/978-981-13-7095-3_25)
- Rangarajan A (2002) Vigor or rigor? The Competing Goals of Variety Trials 12(4):562. <https://doi.org/10.21273/horttech.12.4.562>
- Rasch G (1960) Probabilistic models for some intelligence and educational tests. The Danish Institute for Education Research, Copenhagen
- Reynolds M, Chapman S, Crespo-Herrera L, Molero G, Mondal S, Pequeno DNL, Pinto F, Pinera-Chavez FJ, Poland J, Rivera-Amado C, Saint Pierre C, Sukumaran S (2020) Breeder friendly phenotyping. *Plant Sci*:110396. doi:<https://doi.org/10.1016/j.plantsci.2019.110396>
- Rijgersberg H, Top JL (2000) Exchanging crop trials information: standardization by means of data model templates. *Comput Electron Agric* 25(3):221–231. [https://doi.org/10.1016/S0168-1699\(99\)00070-8](https://doi.org/10.1016/S0168-1699(99)00070-8)
- Ritchie JT (1995) International consortium for agricultural systems applications (ICASA): establishment and purpose. *Agr Syst* 49(4):329–335. [https://doi.org/10.1016/0308-521X\(95\)00028-4](https://doi.org/10.1016/0308-521X(95)00028-4)
- Sabaghnia N (2016) Nonparametric statistical methods for analysis of genotype × environment interactions in plant pathology. *Australas Plant Pathol* 45(6):571–580. <https://doi.org/10.1007/s13313-016-0453-0>
- Salanti G (2012) Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods* 3(2):80–97. <https://doi.org/10.1002/jrsm.1037>
- Schutz HG, Cardello AV (2001) A labeled affective magnitude (LAM) scale for assessing food liking/disliking. *J Sens Stud* 16(2):117–159. <https://doi.org/10.1111/j.1745-459X.2001.tb00293.x>
- Shorter R, Lawn RJ, Hammer GL (1991) Improving genotypic adaptation in crops – a role for breeders, physiologists and Modellers. *Exp Agric* 27(2):155–175. <https://doi.org/10.1017/S0014479700018810>
- Shrestha R, Sanchez H, Ayala C, Wenzl P, Arnaud E (2010) Ontology-driven International Maize Information System (IMIS) for phenotypic and genotypic data exchange. *Nature Precedings*. <https://doi.org/10.1038/npre.2010.5029.1>
- Simko I, Hayes RJ, Kramer M (2012) Computing integrated ratings from heterogeneous phenotypic assessments: a case study of lettuce post-harvest quality and downy mildew resistance. *Crop Sci* 52(5):2131–2142. <https://doi.org/10.2135/cropsci2012.02.0111>
- Simko I, Linacre JM (2010) Combining partially ranked data in plant breeding and biology: II. Analysis with Rasch model. *Communications in Biometry & Crop Science* 5 (1)
- Simko I, Pechenick DA (2010) Combining partially ranked data in plant breeding and biology: I. Rank aggregating methods. *International Journal of the Faculty of Agriculture and Biology, Warsaw University of Life Sciences, Poland* 5(1):41–55
- Simko I, Piepho H-P (2011) Combining phenotypic data from ordinal rating scales in multiple plant experiments. *Trends Plant Sci* 16(5): 235–237. <https://doi.org/10.1016/j.tplants.2011.02.001>
- Singh YP, Nayak AK, Sharma DK, Gautam RK, Singh RK, Singh R, Mishra VK, Paris T, Ismail AM (2014) Farmers’ participatory varietal selection: a sustainable crop improvement approach for the 21st century. *Agroecol Sustain Food Syst* 38(4):427–444. <https://doi.org/10.1080/21683565.2013.870101>
- Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *J Agric Sci* 143(6):449–462. <https://doi.org/10.1017/S0021859605005587>
- Smith KF, Fennessy PF (2011) The use of conjoint analysis to determine the relative importance of specific traits as selection criteria for the improvement of perennial pasture species in Australia. *Crop Pasture Sci* 62(4):355–365. <https://doi.org/10.1071/CP10320>
- Snapp S (2002) Quantifying farmer evaluation of technologies: the mother and baby trial design. *Quantitative analysis of data from participatory methods in plant breeding*
- Spyns P, Meersman R, Jarrar M (2002) Data modelling versus ontology engineering. *SIGMOD Rec* 31(4):12–17. <https://doi.org/10.1145/637411.637413>
- Steinke J, van Etten J (2017) Gamification of farmer-participatory priority setting in plant breeding: design and validation of “AgroDuos”. *J Crop Improv* 31(3):356–378. <https://doi.org/10.1080/15427528.2017.1303801>
- Stevens SS (1946) On the theory of scales of measurement. *Science* 103(2684):677–680. <https://doi.org/10.1126/science.103.2684.677>
- Stone H, Bleibaum RN, Thomas HA (2012) Chapter 1 - Introduction to sensory evaluation. In: Stone H, Bleibaum RN, Thomas HA (eds) *Sensory evaluation practices*, Fourth edn. Academic, San Diego, pp 1–21. <https://doi.org/10.1016/B978-0-12-382086-0.00001-7>
- Streck EA, de Magalhaes Jr AM, Aguiar GA, Henrique Facchinello PK, Reis Fagundes PR, Franco DF, Nardino M, de Oliveira AC (2018) Genetic progress in 45 years of irrigated rice breeding in southern Brazil. *Crop Sci* 58(3):1094–1105. <https://doi.org/10.2135/cropsci2017.06.0383>
- Strobl C, Wickelmaier F, Zeileis A (2011) Accounting for individual differences in Bradley-Terry models by means of recursive partitioning. *J Educ Behav Stat* 36(2):135–153
- Sukcharoen K, Leatham D (2016) Mean-variance versus mean-expected shortfall models: an application to wheat variety selection. *J Agric Appl Econ* 48(2):148–172. <https://doi.org/10.1017/aae.2016.8>
- Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M (2017) Plant phenomics, from sensors to knowledge. *Curr Biol* 27(15):R770–R783. <https://doi.org/10.1016/j.cub.2017.05.055>
- Tenkouano A, Ortiz R, Nokoe S (2012) Repeatability and optimum trial configuration for field-testing of banana and plantain. *Sci Hortic* 140:39–44. <https://doi.org/10.1016/j.scienta.2012.03.023>
- Tesfaye K, Sonder K, Cairns J, Magorokosho C, Tarekegn A, Kassie GT, Getaneh F, Abdoulaye T, Abate T, Erenstein O (2016) Targeting drought-tolerant maize varieties in Southern Africa: a geospatial crop modeling approach using big data. *International Food and Agribusiness Management Review* 19(A):75–92



- Theobald CM, Talbot M, Nabugoomu F (2002) A Bayesian approach to regional and local-area prediction from crop variety trials. *J Agric Biol Environ Stat* 7(3):403–419. <https://doi.org/10.1198/108571102230>
- Thessen AE (2016) Adoption of machine learning techniques in ecology and earth science. *One Ecosyst* 1. <https://doi.org/10.3897/oneeco.1.e8621>
- Tomlins K, Rwiza E, Nyango A, Amour R, Ngendello T, Kapinga R, Rees D, Jolliffe F (2004) The use of sensory evaluation and consumer preference for the selection of sweetpotato cultivars in East Africa. *J Sci Food Agric* 84(8):791–799. <https://doi.org/10.1002/jsfa.1712>
- Tonin FS, Rotta I, Mendes AM, Pontarolo R (2017) Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. *Pharm Pract* 15(1):943–943. <https://doi.org/10.18549/PharmPract.2017.01.943>
- Turner HL, van Etten J, Firth D, Kosmidis I (2020) Modelling rankings in R: the PlackettLuce package. *Comput Stat*. <https://doi.org/10.1007/s00180-020-00959-3>
- van Eeuwijk FA, Bustos-Korts DV, Malosetti M (2016) What should students in plant breeding know about the statistical aspects of genotype  $\times$  environment interactions? *Crop Sci* 56(5):2119–2140. <https://doi.org/10.2135/cropsci2015.06.0375>
- van Eeuwijk FA, Malosetti M, Yin X, Struik PC, Stam P (2005) Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. *Aust J Agr Res* 56(9):883–894. <https://doi.org/10.1071/AR05153>
- Van Etten J, Beza E, Calderer L, Van Duijvendijk K, Fadda C, Fantahun B, Kidane YG, Van De Gevel J, Gupta A, Mengistu DK, Kiambi DAN, Mathur PN, Mercado L, Mittra S, Mollé MJ, Rosas JC, Steinke J, Suchini JG, Zimmerer KS (2016) First experiences with a novel farmer citizen science approach: crowdsourcing participatory variety selection through on-farm triadic comparisons of technologies (tricot). *Exp Agric* 55:1–22. <https://doi.org/10.1017/S0014479716000739>
- van Etten J, de Sousa K, Aguilar A, Barrios M, Coto A, Dell'Acqua M, Fadda C, Gebrehawaryat Y, van de Gevel J, Gupta A, Kiros AY, Madriz B, Mathur P, Mengistu DK, Mercado L, Nurhisn Mohammed J, Paliwal A, Pè ME, Quirós CF, Rosas JC, Sharma N, Singh SS, Solanki IS, Steinke J (2019) Crop variety management for climate adaptation supported by citizen science. *Proc Natl Acad Sci* 116(10):4194–4199. <https://doi.org/10.1073/pnas.1813720116>
- van Etten J, Steinke J, van Wijk M (2017) How can the data revolution contribute to climate action in smallholder agriculture? *Agricult Dev*:44–48
- Vargas M, Crossa J, Reynolds MP, Dhungana P, Eskridge KM (2007) Structural equation modelling for studying genotype  $\times$  environment interactions of physiological traits affecting yield in wheat. *J Agric Sci* 145(2):151–161. <https://doi.org/10.1017/S0021859607006806>
- Vargas M, Crossa J, van Eeuwijk FA, Ramírez ME, Sayre K (1999) Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype  $\times$  environment interaction. *Crop Sci* 39(4):955–967. <https://doi.org/10.2135/cropsci1999.0011183X003900040002x>
- Virk DS, Pandit DB, Sufian MA, Ahmed F, Siddique MAB, Samad MA, Rahman MM, Islam MM, Ortiz-Ferrara G, Joshi KD, Witcombe JR (2009) REML is an effective analysis for mixed modelling of unbalanced on-farm varietal trials. *Exp Agric* 45(1):77–91. <https://doi.org/10.1017/S0014479708007047>
- Waldman KB, Kerr JM, Isaacs KB (2014) Combining participatory crop trials and experimental auctions to estimate farmer preferences for improved common bean in Rwanda. *Food Policy* 46:183–192. <https://doi.org/10.1016/j.foodpol.2014.03.015>
- Wan Z, Hook S, Hulley G (2015) MYD11A1 MODIS/aqua land surface temperature/emissivity daily L3 global 1km SIN grid V006. NASA EOSDIS Land Processes DAAC
- Weltzien E, Christinck A (2017) Participatory breeding: developing improved and relevant crop varieties with farmers. In: *Agricultural systems* (Second edition). Elsevier, Amsterdam, pp 259–301
- White JW, Hunt LA, Boote KJ, Jones JW, Koo J, Kim S, Porter CH, Wilkens PW, Hoogenboom G (2013) Integrated description of agricultural field experiments and production: the ICASA version 2.0 data standards. *Comput Electron Agric* 96:1–12. <https://doi.org/10.1016/j.compag.2013.04.003>
- White JW, van Evert FK (2008) Publishing agronomic data. *Agron J* 100(5):1396–1400
- Whitley E, Ball J (2002) Statistics review 6: nonparametric methods. *Crit Care* 6(6):509–513. <https://doi.org/10.1186/cc1820>
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3:160018. <https://doi.org/10.1038/sdata.2016.18>
- Williams SC (2012) Data practices in the crop sciences: a review of selected faculty publications. *J Agr Food Inform* 13(4):308–325. <https://doi.org/10.1080/10496505.2012.717846>
- Wyborn C, Louder E, Harrison J, Montambault J, Montana J, Ryan M, Bednarek A, Nesshöver C, Pullin A, Reed M, Dellecker E, Kramer J, Boyd J, Dellecker A, Hutton J (2018) Understanding the impacts of research synthesis. *Environ Sci Policy* 86:72–84. <https://doi.org/10.1016/j.envsci.2018.04.013>
- Xu Y (2016) Envirotyping for deciphering environmental impacts on crop plants. *Theor Appl Genet* 129(4):653–673. <https://doi.org/10.1007/s00122-016-2691-5>
- Yan W (2014a) An overview of variety trial data and analyses. In: Yan W (ed) *Crop variety trials: data management and analysis*, pp 23–30. <https://doi.org/10.1002/9781118688571.ch2>
- Yan W (2014b) Theoretical framework for crop variety trials. In: *Crop variety trials: data management and analysis*. Wiley Online Books. <https://doi.org/10.1002/9781118688571.ch1>
- Yan W, Hunt LA, Sheng Q, Szlavnicz Z (2000) Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci* 40(3):597–605. <https://doi.org/10.2135/cropsci2000.403597x>
- Yates F, Cochran WG (1938) The analysis of groups of experiments. *J Agric Sci* 28(4):556–580. <https://doi.org/10.1017/S0021859600050978>
- Yu PLH, Gu J, Xu H (2019) Analysis of ranking data. *WIREs Computational Statistics* 11(6):e1483. <https://doi.org/10.1002/wics.1483>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.