DOI: 10.1111/epp.13037

PP 1/333 (1) Adoption of digital technology for data generation for the efficacy evaluation of plant protection products

Specific scope: This Standard describes the validation, verification, and calibration of digital technologies that may be used to assess the efficacy of plant protection products (PPP). Currently, efficacy data are collected through human observation or other documented methods of assessment in the Good Experimental Practice (GEP) system as described in PP 1/181 *Conduct and reporting of efficacy evaluation trials, including good experimental practice*.[1]

Hardware or sensors which directly produce a measurement (e.g. scales for weighing, thermometers) are out of the scope of this Standard, as their verification and calibration are already covered in the GEP system. New technologies with parameters that are currently not covered by EPPO PP1 Standards are also out of this scope but could be covered in future by specific EPPO Standards.

Specific approval and amendment: First approved in 2024–09.

1 | INTRODUCTION

The development and integration of digital technology is growing across a wide range of industries including agriculture. In crop protection and plant phenotyping, digital technologies are already well established in research and at grower level. Usage of digital technology for the assessment of the efficacy of plant protection products is also rapidly increasing.

This Standard focuses on how digital technologies used in efficacy trials can be accepted within GEP systems and by regulators. The Standard also includes processes to validate, verify, and calibrate digital technologies, relevant for the GEP system.

Digital technologies are electronic devices used to collect, process, analyse, transmit, receive and store data. They include:

- Hardware: equipment (e.g. drones, tractors, handheld devices) and sensors (e.g. Red Green Blue (RGB) sensors and multi/hyperspectral cameras),
- Software: any algorithm approach (e.g. object recognition and classification, regression, segmentation, classical approaches, machine or deep learning as well as simple mathematical indexes such as the Normalized Difference Vegetation Index).

When used in efficacy evaluation trials, digital technologies should produce an outcome which is comparable to the data being currently collected by human observation or by other methods currently accepted in the GEP system. The data obtained by human observation which are used as comparisons to validate or verify data obtained by digital technologies can be referred to as reference values (sometimes referred to as 'ground truth'). Reference value data enables validation of algorithms in the development stage and verification of digital technology during use.

As described in PP 1/152 *Design and analysis of evaluation of trials*, the qualities to be considered in observations for efficacy evaluations are: accuracy, reliability, precision, sensitivity, repeatability and reproducibility. These qualities should also be considered in the evaluation of digital technologies used in the GEP system.

2 | VALIDATION

Validation is a crucial process which serves to develop and assess the accuracy and reliability of the digital technology and to define the specification of any hardware required to generate the data for the algorithm, model or software. Validation is the first step of the digital technology conducted under development conditions to show that it works for the intended use.

The validation process should use a known or predefined dataset or samples and compare the results from the digital technology with the assessment results from a human observation or from other methods currently accepted in the GEP system (the reference value). This should be carried out by experts in that type of assessment and should be conducted in controlled situations. The development of the algorithm, model or software may or may not include an element of machine learning or artificial intelligence.

[1]The terms validation and verification are used in other EPPO Standards e.g. those in the series PM 7 Diagnostics. The definitions used in other EPPO Standards might not be the same as those in this Standard as the context in which these Standards are used is different.

The validation procedure, the dataset and samples used for the development of the model/algorithm/software and the validation criteria will be different for each technology and will vary depending on the stage of development and are defined by the developer.

No further guidance on validation is given in this Standard, since validation is the first step of the digital technology conducted under development conditions, and it is not within scope of GEP.

3 | VERIFICATION

Verification is the process to show that the digital technology is robust enough to be used for efficacy assessment in the GEP system. The data obtained by the digital technology will be compared with the data obtained by human observation or by other methods currently accepted in the GEP system (i.e. the reference value). The data forming the basis for the verification should be generated under GEP conditions.

The digital technology needs to be tested by different users independent of the technology developer and comply with the relevant instructions provided by the digital technology developer.

Based on the verification, the digital technology provider defines the intended use and the conditions under which accuracy of the digital technology is guaranteed and relevant limitations identified.

Compiling a verification report, considering the requirements to prove accuracy and identifying the different responsibilities in the context of GEP are important elements of the verification process.

3.1 | Verification report

The intended use, use conditions, accuracy, and limitations should be described by the provider in a verification report which should contain the following information, if applicable:

a. A full description of the intended use, namely:

- intended type of assessment provided by the digital technology (e.g. incidence, severity, specific counting, density) and unit;
- intended target or objective of the assessment (e.g. crop, pest, growth stage, life stage);
- accordance with a specific EPPO PP1 Standard on Efficacy Evaluation (if applicable).

b. A description of the digital technology, namely:

- hardware specification (e.g. sensor and equipment model, stated accuracy, error margin);
- software name and version.

c. The instructions of use: the description of the standard operating procedure to acquire data, e.g. the distance to the target, the angle, drone speed and image resolution.

d. The operating conditions under which the digital technology should be used to ensure its accuracy, namely:

- The range of relevant environmental and agronomic conditions (e.g. sowing density, pest pressure, temperature, light intensity or cloud cover, humidity, wind speed).

e. The limitations to the use of the technology, namely:

- A description of the limitations (e.g. the conditions not tested during the verification process or conditions that failed to produce accurate data).

f. The accuracy, including information on how accuracy was determined, namely:

- The description of the dataset (e.g. trial location, crop, number of assessors)
- A discussion of the results
- The conclusions.

The minimum requirements to be followed to show accuracy are described in point 3.2 and an example of a template of a verification report is provided in Appendix 1.

3.2 | Requirements to show accuracy

The minimum requirements to demonstrate accuracy of the digital technology (3.1.f), are:

- Supply data from a minimum of 60 data points from at least 3 trials. The number of data points and trials necessary to verify the digital technology should be adequate to the tool and conditions of the intended use. to represent the variability of the intended use conditions as described in 3.1.a and 3.1.d.
- Rely on data generated with no deviations by the digital technology specifications as described in 3.1.b and by its use as described in 3.1.c.
- Observations should be gathered according to the relevant specific EPPO PP1 Standard(s) by experienced GEP personnel and also by means of the digital technology to be verified.
- Assessment data from the reference value and digital technology should be conducted preferably on the same day under comparable conditions.

In the assessment of the effect of products in an efficacy evaluation trial, ‘variables’ are assessed by four

'modes of observation': measurement, visual estimation, ranking, and scoring (see PP 1/152). If the assessment provides a continuous variable, the verification of reproducibility should be tested using regression analysis techniques with the calculation of the coefficient of determination (R^2) between visual observation and digital technology observation.

If the data are binary, they should be tested by the accuracy metric. If the data are nominal or ordinal variable, they should be tested by the Cohen's kappa. The coefficient of determination (R^2) should not be less than 0.85, the accuracy should not be lower than 0.85 and kappa not lower than 0.7 (see Appendix 2).

3.3 | Responsibilities in the context of GEP

In principle, a GEP unit can only use a digital technology in the scope that it is verified and proven to be accurate. The verification report and the underlying data should be available on request of the GEP auditor. The provider of the digital technology is responsible for verification and making the verification report available to the user. The GEP unit needs to have access to the most recent version of the verification report at any time and make the verification report available for subsequent inspection by relevant authorities if needed.

4 | CALIBRATION

Within a GEP system, evidence is required that equipment used in a GEP accredited trial has been subject to an appropriate calibration procedure on a regular basis, with relevant details recorded as evidence that the equipment was operating correctly and within defined parameters at the time used in the trial.

Calibration is also a required procedure with digital technologies to ensure that the hardware used works accurately within the required specification to feed data of the required quality into the model, algorithm or software.

Calibration is the process of evaluating and adjusting the precision and accuracy of measurement of the hardware, to ensure it is within the acceptable range. The proper calibration of the hardware ensures that valid data is produced for GEP efficacy trials. The relevant manufacturers' instruction on calibration of the hardware should be followed in conjunction with any additional specific requirements from the provider of the algorithm/model/software.

The calibration procedure should be reflected in a relevant standard operational procedure of the trial facility.

A record should be made of the conduct of the calibration, noting relevant information including the identification of the hardware equipment being calibrated and the version of the algorithm/model/software being used. Such information should also be recorded as part of the assessment data when digital technology are used for an assessment in a GEP trial. In case of results outside the range prescribed by the provider of the hardware or the algorithm/model/software the provider(s) should be contacted and the tool verified again.

In addition to equipment calibration, checking the accuracy of the digital technology also needs to be performed regularly, to confirm that it continues to match as minimum the accepted way of assessment in the GEP system e.g., accuracy of human observation, and there is no degradation in performance.

5 | SOFTWARE VERSIONING

The technology provider is responsible for ensuring that the software is functioning properly and up to date and to make the most recent software version available to the user. The user is responsible for using the most appropriate and compatible version. Software versioning should be documented in the verification report.

In case of significant deviation of the digital technology measurement from the expected value, the provider should be contacted and the technology verified again.

APPENDIX 1 - EXAMPLE OF TEMPLATE OF A VERIFICATION REPORT

This appendix provides an example of a template that may be used to prepare a verification report.

Verification report

Title:
Name and contact details of the provider:
Name of the author:
Signature and date:
Intended Use:

Crop (and BBCH growth stage range)	Pest (and life stage if needed)	Assessment	Assessment unit	EPPO PP1 Standard (if applicable)

Digital technology:

Hardware specification (including model, stated accuracy, error margin and other relevant information):
Software specification (including the version number and the scope of application):

Instructions of use:

Standard way of use (data acquisition):
[Description of how to acquire data in the field trial. To include here, for example, the distance to the target, the angle, device speed, image resolution, process to get data managed by the software, time of response].

Conditions of guaranteed accuracy:

Environmental and agronomic conditions*	Range of guaranteed use	Number of tests done under this condition

* Environmental and agronomic conditions may include sowing density, pest infestation, temperature, light intensity/cloud cover, humidity, wind speed.

Identified limitations:

[Description of the limitation of uses that do not ensure valid results of the digital technology compared to conventional assessment process].

Accuracy verification data

[Include in the table the list of tests carried out to check the accuracy of the assessment. The table can be adapted with any relevant parameter tested in addition to those listed below].

Material and methods

Crop	Pest	Trials ID	Location	Assessment type	Number of observations in this pool of trials

Results

Crop	Pest	Assessment type	Number of observations on this pool of trials	Digital technology value (mean – min/max)	Conventional assessment value (mean – min/max)	R^2	Difference acceptance[a]

[a] This column could be used to indicate if the difference between the assessment results of the digital technology and of the conventional assessment is correct.

[A correlation graph can be included].

Discussion
Conclusions

APPENDIX 2 - METRICS COMPARISON

The following table defines which metrics may be used during verification depending on the variable nature of Reference Values (i.e. the categories of variables according to EPPO Standard PP 1/152 *Design and analysis of evaluation of trials*). It is to be noted that the examples are provided to illustrate the methodology that can be followed and they do not reflect an entire set of Digital Technology and Reference Value observations.

Variable	Metric
Binary	Accuracy
Nominal	Cohen's Kappa
Ordinal	Cohen's Kappa
Quantitative	R^2

Hereafter, the metrics formulas are reported with one example for each.
Examples data reports Digital Technology observations and Reference Values.

Accuracy

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

where TP, True Positive; TN, True Negative; FP, False Positive; FN, False Negative.

Example: Plant infested out of 4 randomly selected within a plot.

Observation number	Digital technology	Reference values
1	Infested	Infested
2	Infested	Infested
3	Not infested	Not infested
4	Not infested	Infested

$$\text{Accuracy} = \frac{2+1}{4} = 0.75$$

Cohen's Kappa

$$K = \frac{\text{GP} / N_{\text{obs}} - \text{Marginal} / N_{\text{obs}}^2}{1 - \text{Marginal} / N_{\text{obs}}^2}$$

$$\text{Marginal} = \sum row1 \times \sum col1 + \ldots + \sum rowN \times \sum colN$$

GP, good predictions; N_{obs}, number of observations.

Example: Leaf discoloration in potato (PP 1/135 Phytotoxicity assessment).

Observation number	Digital technology	Reference values
1	Chlorosis	Chlorosis
2	Yellow veins	Yellow veins
3	Yellow spots	Yellow spots
4	Yellow spots	Yellow spots
5	Whitening	Yellow spots

		Reference Values				
		Chlorosis	**Yellow veins**	**Yellow spots**	**Whitening**	
Digital Technology	**Chlorosis**	1	0	0	0	Sum row 1=1
	Yellow veins	0	1	0	0	Sum row 2=1
	Yellow spots	0	0	2	1	Sum row 3=3
	Whitening	0	0	0	0	Sum col 4=0
		Sum col 1=1	Sum col 2=1	Sum col 3=2	Sum col 4=1	

GP in green and not correct predictions in red.

$$\text{Marginal} = (1 \times 1) + (1 \times 1) + (3 \times 2) + (0 \times 1) = 8$$

$$\text{GP} = 4;$$

$$N_{\text{obs}} = 5$$

$$K = \frac{4/5 - 8/5^2}{1 - 8/5^2} = 0.71$$

R^2 (coefficient of determination)

$$R^2 = 1 - \frac{\sum_{i=1}^{N_{\text{obs}}} \left(\text{RV}_i - \text{DT}_i\right)^2}{\sum_{i=1}^{N_{\text{obs}}} \left(\text{RV}_i - \overline{\text{RV}}\right)^2}$$

$$\overline{\text{RV}} = \frac{1}{N_{\text{obs}}} \sum_{i=1}^{N_{\text{obs}}} \text{RV}_i = \text{ReferenceValuesaverage}$$

$$\overline{\text{DT}} = \frac{1}{N_{\text{obs}}} \sum_{i=1}^{N_{\text{obs}}} \text{DT}_i = \text{DigitalTechnologyobservationsaverage}$$

Example: number of larvae

Observation number	Digital technology	Reference values
1	1	0
2	2	2
3	4	4

$$R^2 = 1 - \frac{(0-1)^2 + (2-2)^2 + (4-4)^2}{(0-2)^2 + (2-2)^2 + (4-2)^2} = 0.88$$