

**Service ID** S00108

**Location** Italy, Remote



## Design of evaluation metrics for physical testing

### Provider service

Politecnico di Milano (POLIMI)

### Link to content

<https://www.agrifoodtef.eu/services/design-evaluation-metrics-physical-testing>

### Type of Sector

Arable farming, Food processing, Greenhouse, Horticulture, Livestock farming, Tree Crops, Viticulture

### Accepted type of products

Data, Design / Documentation

### Type of service

Test design

### Description

Any physical test activity involves three main components: environment (where the tests take place), protocol (defining what tests are executed and how) and evaluation metrics (used to assess the results of the tests). This service concerns the last element; its goal is to design the best metrics to evaluate the performance of a customer solution taking into consideration the use cases specified by the customer and the environment and protocol chosen for the tests (which, if needed, can be designed via services S00106 and S00107). Our team will identify and define with customers the most adequate set of quantitative (i.e., based on instrumental measurements) and/or qualitative (i.e., relying on expert human judgement) metrics to assess the system functionalities of interest. This phase will involve, in particular, agronomists and experts in agricultural machinery. Based on the defined evaluation metrics, a set of requirements for the collection of required data and ground truth annotations will also be defined accordingly. For instance, the service may lay out the specifications for dedicated data collection campaigns (possibly executed via service S00113). This phase will involve engineers and experts in AI and robotics. On request, the output of the service will include analyses on additional environmental factors than those directly tracked through the designed metrics (e.g., seasonal effects, impact of test distribution over time on results).

## How can the service help you

Building a system (e.g., a machine) that solves a problem and designing the mathematical and data-processing operations needed to analyse data collected during experimental activities to evaluate the performance of the machine are two very different activities and involve very different competencies.

Additionally, evaluation metrics are crucial to the identification of issues and ways to improve the performance of the system, so their choice has a strong impact on product development. This service supports customers who developed a solution in designing the evaluation metrics necessary to process experimental data to evaluate system performance and suitability for the task. At the end of the service, customers are provided with a set of metrics tailored to their own systems and necessities, which can be used to collect and process data in order to assess quantitatively the performance of their system.

If required, AgrifoodTEF can support the customer in designing also the environment and testing protocol for tests (via services S00106 and S00107), in the setup of the experimental activities (via services S00110 and S00111), and in the execution of the tests (service S00112) and associated data collection (service S00113). If needed, AgrifoodTEF can also

## How the service will be delivered

This service description is intentionally generic. Every instance of this service is, in fact, customised to adapt it to the needs and requirements of the specific customer.

The following is an example of a service instance.

Example service: The customer is interested in testing the intra-row navigation module for their own agricultural robots. To achieve this goal, we define with them the performance objectives: for instance, maximising success rate in terms of traversed waypoints and minimising the percentage of damaged plants.

Subsequently, suitable quantitative performance metrics are defined to evaluate how well the system attains the chosen objectives. The service can also identify datasets where these metrics have already been annotated. In the absence of any gold standard information about optimal metric values to use as ground truth, an ancillary dedicated data collection campaign can optionally be defined, whether conducted at physical facilities (e.g., via service S00113) or in simulation (e.g., via service S00183).

## Service customisation

The duration of this service is on average 3-6 weeks. The first phase involves one or more interviews, in person or remote, where the customer provides information about the features of the system(s) to be tested, the performance elements of interest and the type of data to be processed for performance evaluation.

Subsequently we design the evaluation metrics and check their compliance with the requirements by executing preliminary processing tests on data fragments provided by the customer (under NDA if needed). During this phase we may provide the customer with feedback about data quality and suitability for the purpose. At the end of the service, the customer receives a report with the design of the performance metrics and the outcomes of the preliminary processing tests.