Boosting agricultural scientific research and innovation through challenges: the ROSE Challenge example

Guillaume Avrin [a,*], Agnes Delaborde [a], Olivier Galibert [a], Daniel Boffety [b]

[a] Laboratoire national de métrologie et d'essais (LNE), Trappes, France

[b] Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (IRSTEA), Clermont-Ferrand, France

* Corresponding author. Email: guillaume.avrin@lne.fr

Abstract

The pending ban on glyphosate raises the general issue of long-term usability of chemical herbicides and pesticides. Torn between legitimate economic and ecological concerns, EU lawmakers and agri-food stakeholders need to rely on alternatives to chemicals, which entails a strong support to the development of new solutions. Autonomous weed killing solutions are among these concrete solutions, considering that navigation, vision and decision-making technologies have reached a level of mastery that allows considering realistically the deployment of such devices. However, the technical feasibility and overall usability of autonomous agricultural machines need yet to be confirmed.

In this regard, the French ministries of agriculture, ecology and research and the national research agency started in 2018 the four-year "ROSE Challenge". The Challenge addresses the weed killing of intra row of crops by autonomous robots. Four competing teams tackle the problem by developing autonomous solutions, which will be formally assessed throughout the Challenge.

This paper offers an overview of robotics challenges and presents the methodology of the real-life evaluation performed in the context of the ROSE Challenge. The overall object of this contribution aims at raising awareness in the agricultural community about the effectiveness of challenges to support innovation in the agricultural field.

Keywords: Testing, Digital agriculture, Robotics and automation, Performance evaluation, Phytosanitary reduction

1. Introduction

Herbicides account for 40% of crop protection products used in agriculture and are the main pesticides responsible for the contamination of waterways. Some weed control alternatives to herbicides already exist and are operational and used: redesign of crop systems and management methods or introduction of preventive practices. Alternative solutions also involve mechanical weeding solutions integrated with towed tool blocks or autonomous robots. They are constantly being improved through the joint efforts of technical institutes and agricultural professionals and are increasingly used. But these solutions usually concern the inter row of crops (between two rows), while in the intra row of crops (space between plants of the same row) few solutions allow dispensing completely with chemical products.

In 2018, the French ministries in charge of agriculture and ecological transition via the French Biodiversity Agency (AFB), in partnership with the Ministry of Research and the National Research Agency (*Agence Nationale de la Recherche*, ANR), started the "ROSE Challenge"[1]. Four competing consortia have been selected, and their four-year objective consists in developing innovative technological solutions that will help reducing the use of phytosanitary products.

The teams focus on weeding the intra row in field vegetable crops and field crops with wide spacing. The projects have to push back the limits of the state of the art in several fields: sensors, modelling, robotics and their combination.

In the course of the challenge, the progress of the teams will be assessed on three components of weeding: observation, interpretation/decision and weeding action. The effectiveness of weeding, while respecting existing crops, will be the subject of annual evaluation campaigns conducted by the French national laboratory for metrology and testing (LNE) and the national research institute of science and technology for environment and agriculture (Irstea), with the participation of the VetAgro Sup engineering school for agronomy. Each year, the teams will face real field events, on experimental plots provided by the Irstea laboratory in Montoldre (France) to enable an agronomy evaluation.

The study will first present a literature review of the challenges carried out internationally in the robotic field and underline the specificities and the interest of the ROSE Challenge. Set in the context of the Challenge, the study will then provide the methods used to rigorously and reproducibly evaluate the performance of the different technological blocks of agricultural robotic systems, including infrared or hyperspectral cameras and their use in multimodal detection systems, dynamic mapping tools and automated platforms combined with precision processing strategies.

The ROSE Challenge constitutes a significant first step in the efforts to finance and support research on the targeted objectives of the national Ecophyto II plan carried out by the French Ministries in charge of Research, Agriculture and Ecology for the reduction of the use of phytosanitary products. The Challenge paves the way for other initiatives that will

[1] Project website: http://challenge-rose.fr/en/home/

learn from it. The present contribution aims at raising awareness in the agricultural community about the effectiveness of challenges to support innovation in the agricultural field, by fostering collaborations between actors that are not necessarily accustomed to working together: researchers in agronomy and ecology, researchers in digital sciences and robotics, farm enginery manufacturers, farmers, or professional agricultural organizations.

2. Robotic challenges

2.1. Challenges: definition and objectives

Challenges aim at comparing different systems performing a same task. The comparison may concern the whole system or a module of it, and allows the estimation of its performance, its quality or its safety.

Challenges take different forms, according to several distinctions:

1. The initiation of the challenge. There can be a principal who asks for several systems to be tested, for example in a pre-purchase process. In this case, the principal commissions an expert for the organization of the evaluation. In another setting, the objectives of the challenge can also be set by the organizers themselves, for example for research purposes, so as to develop knowledge on specific topics. In Natural Language Processing domains for example, this type of evaluation campaigns is highly common; one can for example name the Interspeech challenges (Schuller et al., 2018), or the IWSLT evaluation campaigns (Jan et al., 2018). In this context, the organizers set the objectives of the evaluation, the comparison metrics, and provide the reference and test data. In other words, they are both the principals and the evaluation organizer.
2. The selection of candidates. Either the principal selects several systems of interest that need to be tested, or the participation to the evaluation is open to any candidates. In the latter context, the evaluation plan is not designed in line with the technologies developed by the participants: since the specificities of the systems may not be anticipated, the systems must strictly match the specificities of the plan. The evaluation mainly concerns the overall behavior of the system, unless a common framework allows a more detailed comparison between the solutions. When one wants to boost creative solutions, one can also resort to closed challenges, where only a few selected participants tackle the proposed issue. This type of challenge encourages the development of creative solutions while still preserving the quality of the evaluation, since the reduced number of competitors allows an adaptation of the evaluation plan to encompass the variations in the technical choices made by the competitors.
3. The incentive to take part to the challenge. It may take the form of a final prize to the winner(s): participants may vie for a monetary prize, equipment, or contracts. Participants may also be funded throughout the challenge for the development of their solutions. There are also sometimes no specific incentive, except scientific or commercial recognition and development of knowledge in a community.
4. The program of the challenge. The evaluation performed during the challenge can take place in a single stage session: participants prepare to the challenge beforehand, and produce a solution that is evaluated at the moment of the competition. Participants can also be accompanied throughout the development of their solution, through regular evaluations that allows the estimation of their progression.

One may note that the organization of the process of comparing systems may be referred to as “challenges”, “competitions” or “evaluation campaign”. These terms are often used alternately in the community, depending on the finality of the comparison, the public addressed and the nature of the organizers.

2.2. Challenges in robotics

For more than twenty years, challenges have been booming in the robotic community. Fostered by the development of ICT technologies, the evolutions of artificial intelligence and the democratization of robotics components, challenges gather each year millions of participants, spectators and sponsors. The topics of the competitions are closely related to current scientific, societal and technical issues. As an example, the focus of the 2019 edition of RobotNation SeaPerch competition (https://www.seaperch.org/challenge) is directly inspired by Thailand events in 2018, when students got trapped in a submerged cave. The participants to SeaPerch 2019 are thus encouraged to develop solutions for underwater rescue and recovery.

Challenges are mainly organized by networks of expert roboticists (industrials, academics) who set the objectives of the contest and coordinate the definition of the rules and the evaluation process.

Several challenges are explicitly dedicated to the public dissemination of robotics, and rely on the organization of world-wide events, while others are meant for a more specialized audience of researchers and industrials in robotics.

The RoboCup (http://www.robocup.org), initiated in 1997, first focused on robotic soccer competitions (Kitano et al., 1997). The competition now addresses a wide range of robotic applications, including soccer, rescue, service, simulation and industry, where both students and senior researchers can compete.

The Defense Advanced Research Projects Agency (DARPA) regularly launches challenges in robotics and artificial intelligence. Among them, the DARPA Robotics Challenge (Pratt and Manzo, 2013) executed between 2012 and 2015 was addressed to teams of expert roboticists. The challenge aimed the development of semi-autonomous ground robots for disaster response, and participants were funded by DARPA.

The evaluation of both the RoboCup and the DARPA Challenges is partly organized by the National Institute of Standards and Technology (NIST), who relies notably on its expertise in the development of testing methods for robotics in hostile environment (Jacoff et al., 2012).

The Robotex International competitions (http://robotex.international), set during the Robotex festival, brings together beginners and experts who are invited to design algorithms and robots to perform activities such as sumo wrestling, line following, drone races, or basketball.

The competitions coordinated by RoboNation, Inc. (https://www.robonation.org/competitions) concern air, land and marine robots. They are addressed to undergraduate and graduate students and relate to topics such as manoeuvrability, speed, obstacle detection, or navigation. While some of RoboNation competitions impose that the teams use the affordable robot architecture provided by the organizers, the robot might be built from scratch by the students or commercially bought.

The FIRA RoboWorld Cup (http://www.firaworldcup.org) has brought together since 1996 skilled researchers and students for the development of autonomous soccer-robots on different robotic and simulation platforms. The competitions are now expanded to various tasks, such as drone emergency service or robot welding. One of the competitions focuses on the design of a robot using a limited set of materials, among which daily items such as rubber bands and scissors.

The SICK Robot Day competition (https://www.sick.com) has been organized periodically since 2007. The 2018 edition, for example, concerned the delivery of objects by autonomous robots, which implies object detection, grasping, navigation, path planning, collision avoidance and self-localization. The competition is open to any participants, and the teams may use a SICK LiDAR sensor to perform the tasks.

Introduced in 2008 as part of the ICRA conference, the ICRA Robot Challenges (Tee & Van Der Kooij, 2017) aim at addressing a large robotics community and drawing the attention of a general public. The topics cover service robotics, manipulation, mobile robots, microrobotics and soft robotics.

The European Robotics League (ERL), run by the European Union's Horizon 2020 project SciRoc, coordinates competitions on service, emergency, industrial and smart cities robotics applications, during national and European tournaments (https://www.eu-robotics.net/robotics_league). Participants are selected according to the number of available slots in the tournament and the quality of the research team.

This list, far from exhaustive, shows the wide range of topics covered by challenges in the robotic community. In comparison to challenges for industry, rescue or service, robotics competitions dedicated to agriculture are relatively scarce and recent.

The annual agBOT Challenge (http://www.agbot.ag), started in 2016 in the United States, focuses on topics such as unmaned crop seeder, pest and weed identification and eradication, and robotics harvest method. The competition is open to any participant upon the payment of a small participation fee, and winners are awarded a monetary prize. The judge team is composed of American scientists in multidisciplinary robotics fields and farm owners. The 2019 edition will address the identification of crops and weeds in corn culture, within and between the rows.

Another example is the Field Robot Event (http://www.fieldrobot.com/event) that started in 2003. The 2018 edition of this competition took place in Germany during the DLG Feldtage exhibition for farming professionals. Mostly dedicated to students, the competition addresses the sensing abilities of the autonomous robots, their navigation between rows of corn on naturally rough terrain and their weed control abilities. Robots are expected to detect colored markers meant to represent weed patches.

2.3. Benefits and limitations of challenges

Challenges represent remarkable opportunities for the robotics community:

- They stimulate the development of knowledge and innovative solutions for complex scientific and technical issues;
- They encourage the development and improvement of methodologies for the evaluation of new technologies;
- They help raising awareness among the public audience, the sponsors and public authorities on the present abilities and limitations of technologies;
- They foster an emulation among researchers, industrials and users;
- In terms of education, they allow students to gain practical experience on realistic industrial issues.

From the great variations of challenges offered by the community, one can notice several limitations. Firstly, most challenges are self-funded by the participants, which represents a hindrance for the development of the solutions. Although robotics is expected to become reasonably affordable along the years, scientific research and development need substantial funding. Besides, as stated in (Anderson et al., 2011), competitions need to be more than "one-off demonstrations" and adopt the form of repeatable and reproducible scientific experiments so as to assess their real impact on robotics development. In the recent past years, a major effort has been put on the dissemination of the benchmarking methods. So as to constitute concrete and measurable scientific advances, a challenge should guarantee the independence of the organizers, the coordinated selection of the competing teams and the evaluation plan needs to gradually accompany their progress.

3. The ROSE Challenge

3.1. Presentation of the ROSE Challenge

Four teams, each composed of at least one research institute and one company, have been selected for the development of robotic solutions aiming intra-row weed control in vegetable field crops and widely spaced large-scale crops. The objective is to offer an alternative to phytosanitary products, through an intelligent interpretation of the culture environment and an autonomous weeding action. The teams develop ground robotic solutions using RGB, hyperspectral and/or multispectral sensing, and mechanical or electrical weeding effectors. The movements of the robot can be tele-operated, so long as this control does not interfere with the detection, decision-making and weeding actions. The research work of teams is funded through the ANR Challenge funding program.

The Challenge is coordinated and organized by LNE and Irstea. LNE provides the ROSE Challenge with its skills in organizing evaluation campaigns as a trusted third-party, defining test scenarios and designing test environments and verified, annotated test databases. LNE also brings its expertise in metrology applied to Artificial Intelligence to measure the performance of the intelligent systems evaluated in a rigorous, repeatable manner using metrics and uncertainty analysis. Irstea provides its expertise in agricultural robotics, digital agriculture and spreading technologies, and contributes to the ROSE Challenge by allowing the use of the infrastructure at its research and experiment site (experimental plots, metrology, information systems, amenities, etc.), where the evaluation campaigns is organized. In addition, Irstea is calling on the agronomic skills of the Higher education and research institute VetAgro Sup, whose involvement includes annotating and interpreting field data.

3.2. Overall organization

Each year of the Challenge, teams are invited to test the aptitudes of their solutions on the experimental fields made available by the organizers on Irstea’s site. The surface devoted to the Challenge is closely controlled by Irstea teams, who provide a detailed history of the past activities performed on the site (previous cultivations and potential soil treatments), and supervise the technical operations performed on the plots in the course of the Challenge (soil and seedbed preparation with thermal treatment, seeding, fertilizing, maintenance in the rows of crops and in the surrounding areas). In addition to the teams’ plots, a “reference plot” is also maintained by the organizers with conventional weeding means, by using chemicals at the pre-emergence and post-emergence phases. This plot is used as a reference in the evaluation process, to help the appreciation of the efficacy of the robotic solutions.

Figure 1. ROSE Challenge, June 2018: first meeting with the teams on Irstea experimental field in Montoldre (France).

For each selected crop and weed combination, teams are assigned plots. The assignment of a plot may be random among teams. The teams are expected to deploy their technological solution with the objective of weeding their own plots.

3.1. An adaptative evaluation plan

The four-year Challenge is divided into two main phases: the first year (spanning from 2018 to 2019) is dedicated to the dry-run phase that allows the tuning of the evaluation protocol and tools, and the three remaining years constitute the effective competition phase.

During the dry-run phase, teams get prepared to the competition, in conditions as close as possible to the real evaluation so as to allow them to get familiar with the evaluation process. On this occasion, the teams may express the requirements and limits of their systems, discuss the evaluation plan, and proceed to several tests of their prototypes in the field and on data samples. Meanwhile, the organizers acquire or develop the tools that will be required for the evaluation process: tools for the preparation and annotation of the reference data, data collection platform to capture images in the field, cloud platform to share data with the teams, meteorological platforms to gather information on the environmental conditions, and tools to allow semi-automated comparisons between hypotheses and references. The organizers also design the most efficient crop patterns for the evaluation and test them in the fields.

The dry-run phase is a fundamental prerequisite for a smooth conduct of the Challenge. Indeed, this phase allows the organizers to prepare an evaluation plan that is both realistic with respect to the capabilities of the systems, and that is fair among the different technologies used by the teams.

Each year of the Challenge, participants will be invited to test their systems in real field conditions. The evaluation plan is adapted throughout the Challenge so as to accompany the evolutions of the teams' technological solutions. At more mature levels, the nature of the evaluations will turn toward the estimation of the acceptability of the solutions by potential users, the impact of the solutions on soil conditions (pollution, soil erosion and compaction, etc.) and technical and economic criteria such as the level of automation, energy usage and the cost of materials and techniques used.

3.2. Crops and weeds distribution

As mentioned above, the plots are expected to present combinations of crops and weeds. The Challenge addresses both large spacing crops and vegetable field crops. Different weeds are selected, so as to represent the "natural" weed environment of the crops, and several weeds are selected as "models" for the evaluation needs. The weed selection also represents the growth pattern of weeds: upright or spreading. Planting density varies uniformly along the stripes. The Table 1 presents the different plants that have been considered for the Challenge. Research and experimentations are carried out in order to establish relevant combinations of weeds and crops for the evaluation. During the dry-run, the teams are asked to focus on corn, beans, lamb's quarter and matricaria.

Table 1. Categories of crops and weeds envisioned for the ROSE Challenge.
The selection is experimented during the Challenge preparation.

Class	**Type**	**Growth**	**Plant name**	**Alternatives**
Crops	Large spacing		Corn	
			Field beans	
	Vegetable field crops		Beans	
			Peas	
Weeds	Natural	Upright	Lamb's quarter	Foxtail, nightshade, crabgrass
		Spreading	Matricaria	Knotweed
	Model	Upright	Ryegrass	Oats, fescue, dactylis
		Spreading	Wild mustard	Vetch, clover

The plan of the plots of land is established so as to present combinations of crops and weeds equally distributed into patches that are assigned to the competing teams. Figure 2 pictures a stripe of the experimental field.

Figure 2. Each stripe is dedicated to a combination of specific crop and weed; the stripe is divided into four patches so as to allow each team to test their solution.

3.3. Evaluation process

The Challenge focuses on the performance of autonomous weed killing robots for the intra row of crops. The evaluation addresses the three main tasks of the sensorimotor loop performed by a robotic platform:

5. The detection of weeds and crops.
6. The robot's decision to proceed to weeding.
7. The weeding action in itself.

3.3.1. Detection task

The detection task is divided into two sub-tasks. The objective of the first one is to estimate the ability of the technological solution to use its detection system in order to perform the weeding action. This sub-task is evaluated on a test image dataset. The second sub-task aims at estimating the ability of the solution to use its detection system to facilitate decision-making. This is carried out in the field.

<u>First sub-task: detection on test dataset</u>

The evaluation on a dataset requires the collection of test images in the visible, multi-spectral and hyperspectral ranges. This collection is performed by the organizers. During the first phase of the dry-run, only images in the visible range were used. On each image in the dataset, the teams' detection system is expected to:

- detect the presence of plants (weeds and crops) and locate them on the image. Teams are expected to provide masks for each plant detected.
- determine their class (weed or crop).

The output data provided by the teams are called "hypothesis". The evaluation consists of a data mapping between the masks in the "hypothesis", and the plants identified in the "reference". Figures Figure 3, Figure 4, Figure 5 and Figure 6 illustrate a sample of the test dataset, the annotation performed to produce the "reference", a fictional "hypothesis" that may be produced by a team's detection system, and a visual rendering of the comparison performed.

Figure 3. The "source" image shot in the fields.

Figure 4. The "reference" image. Dots and lines represent the cutting performed by an expert labeller so as to mark out the plants. Tags indicate the class that has been selected by the labeller ("PI": crop, "A": weed).

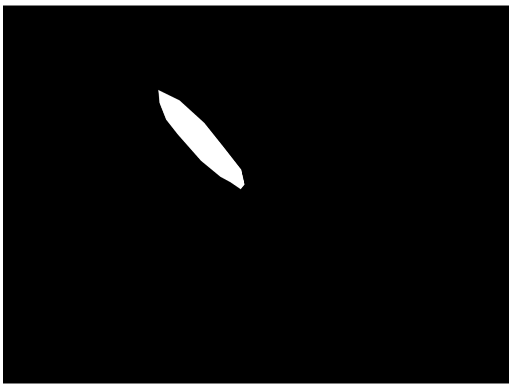

Figure 5. A visual rendering of a "hypothesis" mask (fictional). One can notice that the full surface of the plant has not been detected.

Figure 6. A visual rendering of the comparison process. The striped area represents the part of the plant that has been ignored by the detection system.

The performance of the detection system is scored through the EGER metric (Estimated Global Error Rate). Each plant in the reference is supposed to correspond to a plant of the same class in the hypothesis. The EGER metric takes into account if the system mistakes a plant type for another, forgets a plant, or on the contrary detects a plant where there is none. Each error type is weighted: for example, the confusion between plant types is considered as more important than ignoring a plant, since it may induce the destruction of crops.

In the course of the Challenge, the dataset is solely constituted for testing purposes; teams are expected to rely on their own training datasets. However, the characteristics of the test dataset are selected collaboratively with the teams (distance, sharpness, resolution, luminosity, etc.), so as to match the capabilities of the systems at best.

The whole ROSE Challenge dataset will be released to the community at the end of the Challenge, along with a full documentation on the collection, preparation and data qualification process. It will include aligned images in several ranges of wavelength, description of plants (cutting-out, class, name and growth stage) and description of the environmental conditions (timestamp, weather, etc.).

Second sub-task: detection in the field

The detection in real conditions constitutes the second part of the evaluation of the detection systems. In this context, the platforms are expected to pass over the culture rows and determine automatically the number of weeds and crop plants detected. So as to estimate the repeatability of the detection, the process is repeated several times.

Before the passage of the teams' platform, the organizers identify manually the number of weeds and crop plants on each patches; the evaluation consists in comparing the numbers provided by the human experts and the systems, for each class of plant.

3.3.2. Weeding action task

So as to evaluate their weeding ability, the platforms are expected to perform weeding actions on plots. A specific focus is given to the ability of the platform to perform the action on weeds while preserving the surrounding crops. The objective is to estimate the regrowth rate of the weeds, and to observe the behaviour of the potential damaged crops.

This task has to be as independent from the detection performance as possible. Indeed, the weeding action needs to be done autonomously, which implies that the system needs to localise the plant that needs to be neutralized. To that end, visual markers have been designed by the organizers and submitted to the approval of the teams, ensuring that all the systems are presumably able to easily spot the base of the plants and perform a weeding action. These markers take the form of plastic disks placed at the bottom of each plant (see Figure 7). These disks are color-coded: yellow represents a weed, and blue indicates a crop.

Figure 7. Markers on the base of three plants: on the left, two yellow markers indicate the presence of weeds. On the right, a blue marker signals the presence of a crop.

The evaluation consists in a comparison of the state of the plot before and after the weeding. So as to assess the effectiveness of the weed destruction and the impact of the surrounding crops, the observation of the plot is performed at several points in time, up to twenty-one days after the weeding. The scoring relies on:

- the count of weeds before and after the action, and the percentage of eliminated weeds,
- the count of undamaged crops before and after the action. The integrity of the crops will be qualitatively assessed.
- the density of weeds before and after the action,
- the density of crops before and after the action.

Since the density of weeds and crops varies across a same plot, the robustness of the weeding action is also estimated. The cost of action will be evaluated through a ratio between the total time of processing for a patch and the number of weeds killed on this patch; another estimation can be performed by computing the ratio between the speed of the platform and the density of weeds.

3.3.3. Global evaluation in the field

The evaluation of the specific decision-making aspect is to be carried out in the later phases of the Challenges. The relevance of performing a weeding action depends on various temporal and spatial considerations. Indeed, checking that the system takes the right decision implies considering several factors, like growth stage or proximity to other plants, which are not expected in the first stages of the Challenge.

In the first stages, the performance of the global detection – decision – action chain is determined through an estimation of the biomass of the plot twenty-one days after the realization of the action. Regularly updated on the state of the cultures, teams are expected to define at what time their platforms may be the most efficient. They are thus invited to have their platform proceed to weeding. In this context, systems are not necessarily expected to take a decision of action autonomously: a human operator may trigger the action. However, there are no markers present on plants to facilitate the detection. The evaluation thus allows assessing the detection system and the weeding effector, but also all the decisions taken during the intervention.

The global criterion for the evaluation is the biomass of the plot, compared to the biomass of the conventionally weeded reference plot. This criterion allows notably the appreciation of:

- the choice in the moment of intervention, which depends on constraints imposed by the culture (for example, the intervention needs to happen when weeds are young) and imposed by the robotic platform (for example, weeds need to be developed enough for detection),
- the choice in weeding or not, depending on the likelihood of deteriorating a crop.

Other criteria may be considered, such as work rate and environmental impact of the technological solution.

4. Conclusion

4.1. A challenge to accompany research on agricultural innovation

Financially supported by the French National Research Agency ANR, the participants to the ROSE Challenge are expected to develop technological solutions to face the limitation of chemical inputs, by designing automatic weed control robots for intra row of crops. During the four years of the Challenge, their progress will be objectively measured in real field conditions. To this end, the evaluation plan addresses the detection, decision-making and action abilities of the systems. The plan is meant to evolve according to the development of the robots' capabilities, so as to include eventually societal, economic and ecological criteria. This iterative evaluation procedure avoids the tunnel effect which could drive the development of a solution that does not meet in the end the targeted objective and needs.

The Challenge takes the research far from constrained evaluation in artificial conditions: it faces variety in the capabilities and limitations of the solutions progressively developed by the teams, and may face the potential occurrence of natural events affecting the experimental field (meteorological events, pests, diseases, etc.). The organization of the ROSE Challenge relies on the expertise of experts in metrology, evaluation and agronomy, who are able to adapt the evaluation protocol while still ensuring objective measures and evaluation methodology.

The evaluation procedure is not punitive and the whole process is not validated by a final prize; the overall objective is to push back the scientific limitations and establish objectively the abilities of autonomous systems in agriculture. The evaluation allows building cohesion in the comparison of different technical solutions, helps in the definition of measurable objectives, and quantifies the distance towards these objectives.

4.2. References for the evaluation of agricultural robots

The ROSE Challenge will provide the first global level campaign that addresses both an evaluation on image datasets and in real field conditions. The Challenge allows a modular evaluation of the technological solutions, an overall estimation of the weeding effectiveness and an analysis of the economical, sociological and ecological impact.

The tools developed in the context of the Challenge will constitute reference baselines for the characterization of future research in the domain. In particular, test datasets, through the richness of their contents, will present a high potential for reusability in the community. Indeed, the collection of a corpus of images in the visible, multi-spectral and hyperspectral ranges constitutes a novelty which will allows the comparative evaluation of different plant detection technologies. In the context of strong limitation of phytosanitary products, these databases will prove useful for the development of new weed management technologies endowed with the ability to detect automatically weeds.

The evaluation plan designed throughout the Challenge will serve as a baseline for the evaluation of agricultural robots. Dedicated to performance in ROSE, the evaluation procedure may adapt to address the evaluation of other global issues in robotics such as safety and security.

Acknowledgements

This work is carried out with the financial support of the French Ministry of Agriculture and Food, the French Ministry for the Ecological and Solidary Transition, the French Ministry of Higher Education, Research and Innovation, and the French National Research Agency.

References

Anderson, M., Jenkins, O. C., & Osentoski, S. (2011). *Recasting robotics challenges as experiments*. IEEE Robotics and Automation Magazine, 18(2), 10-11.

Jacoff A., Huang H.M., Virts A., Downs A., Sheh R., 2012. *Emergency Response Robot Evaluation Exercise*. Proceedings of the Workshop on Performance Metrics for Intelligent Systems, Pages, 145–154.

Jan, N., Cattoni, R., Sebastian, S., Cettolo, M., Turchi, M., & Federico, M. (2018). *The IWSLT 2018 Evaluation Campaign*. In International Workshop on Spoken Language Translation (pp. 2-6).

Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., & Osawa, E. (1997). *Robocup: The robot world cup initiative*. In Proceedings of the first international conference on Autonomous agents (pp. 340-347). ACM.

Pratt, G., & Manzo, J. (2013). *The DARPA robotics challenge [competitions]*. IEEE Robotics & Automation Magazine, 20(2), 10-12.

Schuller, B. W., Steidl, S., Batliner, A., Marschik, P. B., Baumeister, H., Dong, F., ... & Einspieler, C. (2018). *The Interspeech 2018 computational paralinguistics challenge: Atypical & self-assessed affect, crying & heart beats.* Proceedings of the INTERSPEECH International Conference, Hyderabad, India.

Tee, K. P., & Van Der Kooij, H. (2017). *The ICRA 2017 Robot Challenges [Competitions]*. IEEE robotics & automation magazine, 24(3), 15-21.