



# Using a qualitative and quantitative validation methodology to evaluate a drone detection system

Daniela Doroftei<sup>1</sup>, Geert De Cubber<sup>1</sup>

<sup>1</sup> Royal Military Academy, Avenue de la Renaissance 30, 1000 Brussels, Belgium

## ABSTRACT

Now that the use of drones is becoming more common, the need to regulate the access to airspace for these systems is becoming more pressing. A necessary tool in order to do this is a means of detecting drones. Numerous parties have started the development of such drone detection systems. A big problem with these systems is that the evaluation of the performance of drone detection systems is a difficult operation that requires the careful consideration of all technical and non-technical aspects of the system under test. Indeed, weather conditions and small variations in the appearance of the targets can have a huge difference on the performance of the systems. In order to provide a fair evaluation, it is therefore paramount that a validation procedure that finds a compromise between the requirements of end users (who want tests to be performed in operational conditions) and platform developers (who want statistically relevant tests) is followed. Therefore, we propose in this article a qualitative and quantitative validation methodology for drone detection systems. The proposed validation methodology seeks to find this compromise between operationally relevant benchmarking (by providing qualitative benchmarking under varying environmental conditions) and statistically relevant evaluation (by providing quantitative score sheets under strictly described conditions).

**Section:** RESEARCH PAPER

**Keywords:** drone detection; validation methodology

**Citation:** Daniela Doroftei, Geert De Cubber, Using a qualitative and quantitative validation methodology to evaluate a drone detection system, Acta IMEKO, vol. 8, no. 4, article 5, December 2019, identifier: IMEKO-ACTA-08 (2019)-04-05

**Editor:** Yvan Baudoin, International CBRNE Institute, Belgium

**Received** November 23, 2018; **In final form** February 12, 2019; **Published** December 2019

**Copyright:** This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement N700643 (SafeShore).

**Corresponding author:** Daniela Doroftei, e-mail: [liliana.doroftei@rma.ac.be](mailto:liliana.doroftei@rma.ac.be)

## 1. INTRODUCTION

### 1.1. Problem statement

Drones are gradually becoming commodities. This is a positive evolution, as these tools have many positive uses, and the affordability of the current systems means that new business opportunities arise. However, we cannot also be blind to the negative aspects that these novel tools may induce into our society. Indeed, next to the many airspace infringements, where uneducated hobbyists inadvertently enter potentially dangerous airspace (e.g. near airports, manned aviation), we also see an increasing use of drone technology by criminals [1], [2]. Most countries have now established rules for access to airspace by unmanned aerial vehicles/drones/Remotely Piloted Aircraft Systems (RPASs). The challenge is now to enforce these rules, as law enforcers lack the means to automatically detect airspace infringements. Indeed, something like a car traffic speed camera

for the air does not really exist yet, but it seems to be an essential tool if one wants to ensure safe airspace operations for everyone.

### 1.2. Previous work on drone detection and the scope of the SafeShore project

Numerous commercial and non-commercial parties have noted this gap in the market and have started the development of drone detection systems [1]. There are two main difficulties related to the detection of drones. First, the cross-section/detection baseline for these systems is generally very limited, whatever sensing technology is used. Indeed, drones have a small RADAR cross-section, a small acoustic signature (from a relevant distance), a small visual/infrared signature, they use common radio signal frequencies, etc. Of course, it would be possible to make the detection methodologies extremely sensitive, but this then leads to a second difficulty: how can false positives be avoided? Indeed, the signature of many drones is quite close to that of birds, so it is very difficult to filter out these false positives [3].

Sensing modalities that can be used to solve drone detection problem are typically RADAR [4], acoustics [5], visual [6], IR [7] (thermal and short-wave), sensing of the radio spectrum [8], LIDAR [9], etc. However, as the problem is so difficult to solve in realistic operating conditions, most of the existing solutions rely on a mix of different sensing methodologies in order to solve the drone detection problem [1] and use a mix of traditional detection and tracking methodologies [10], [11] originating from computer vision to achieve multi-sensor tracking.

As no satisfying solution to this problem currently exists, the European Commission decided to fund the EU-H2020-SafeShore project [12], which serves as a case study for this paper.

The main objective of the SafeShore project is to cover existing gaps in coastal border surveillance, increasing internal security by preventing cross-border crime, such trafficking in human beings and the smuggling of drugs. It is designed to be integrated with existing systems and create a continuous detection line along the border.

The SafeShore solution for detecting small targets that are flying in low altitude is shown in Figure 1 and consists of a 3D LIDAR that scans the sky and creates a virtual dome shield above the protected area. In order to improve the detection, SafeShore integrated the 3D LIDAR with passive acoustic sensors, passive radio detection, and video analytics. All those technologies can be considered as low-cost green technologies (compared to the traditional RADAR systems). It is expected that a combination of orthogonal technologies, such as LIDAR, passive radio, and acoustic and video analytics, will become mandatory for future border control systems in environmentally sensitive areas.

The SafeShore objective is to demonstrate the detection capabilities in the missing detection gaps of other existing systems, such as coastal radars, thereby demonstrating the capability to detect mini-RPAS along the shore and the sea or departing from civilian boats.

Important SafeShore goals are to ensure the intelligent combination and fusion of information sources; to increase the situational awareness and better implementation of the European Maritime Security Strategy based on the information exchange frameworks; and to ensure privacy of the data and conformity to internationally recognized ethical issues concerning the safety of the information and the equipment subject of the project.

There are two problems with the evaluation of drone detection systems:

- 1) Drone detection systems most often rely on complex data fusion and sensor data processing, which means that it is

necessary to carefully control the test conditions in order to single out the limits of the system under test.

- 2) Drone detection systems need to be operational 24/7 and under all weather conditions i.e. it is necessary to assess their performance within a wide range of conditions.

Clearly, both of these constraints are somewhat contradictory, and it is not easy to find a compromise between these two types of requirements. The objective is therefore to find a validation methodology that satisfies both the requests of the end users for qualitative operational validation of the system and the platform developers for a quantitative, statistically relevant validation.

### 1.3. Previous work and a discussion on quantitative and qualitative operational validation

Historically, representatives of different (scientific) communities used quantitative and qualitative evaluation methodologies. Quantitative approaches have therefore been, in this sense, the favourite among members of the hard sciences community, as such methodologies permit a generalisation to be made about a large population on the basis of a relatively small set of (representative) samples. Given a set of initial or boundary conditions, they can help assess the influence of certain variables on a system under test. In principle, they also allow other researchers to validate the original findings by independently replicating the analysis. Of course, this assumption holds valid only if one has sufficient control over the boundary conditions, which limits the usability of such approaches for near-realistic operations. By collecting and analysing data in numerical form, quantitative researchers argue that they are upholding research standards that are simultaneously empirically rigorous, impartial, and objective [13]. An example of a quantitative evaluation methodology in the field of robotics is the set of standardised test methods for explosive ordnance disposal and search and rescue robots, issued by the US National Institute of Standards and Technology (NIST) [14]. Following these standard tests, test subjects need to perform a well-defined standardised basic action (e.g. drive a figure-eight pattern) a statistically relevant number of times in a closed and well-controlled environment.

In the field of social sciences, on the other hand, qualitative evaluation methods [15] have much stronger support due to the fact that many of the variables studied in these research domains cannot meaningfully be represented by singular metrics. Carefully designed qualitative evaluation methods [16] are able to assess the success or performance of a system under test in a more holistic manner than is possible by means of sheer quantitative methods. This also means that they are able to produce results that are much closer to the performance assessment of the system under test by the actual human system user. A disadvantage is that the absence of a controlled working environment and hard metrics implies that repeatability is no longer ensured. In general, these qualitative assessment methodologies are now based on a 'story' or 'scenario'-based approach [16], as this method has proven to be the most useful to incorporate the views of the human stakeholders in the evaluation process. An example of a quantitative evaluation methodology in the field of robotics is the euRathlon challenge [17], [18]. Following this evaluation model, test subjects need to perform a high-level task (e.g. rescue a victim) in an uncontrolled outdoor environment.

It is clear that both these methodologies have their advantages and disadvantages. This has led in the recent years to the development of so-called 'mixed methods' [19], whereby researchers collect and analyse both quantitative and qualitative



Figure 1. SafeShore concept sketch.

Table 1. Excerpt of the SafeShore traceability matrix

Identifier	Requirement	Validation Test	Remarks
URG01	Health and safety procedures conforming to the local / national guidelines / legislation should be followed	BE01, BE02, BE03, BE04, BE05, RO01, RO02, RO03, RO04, RO05, RO06, IL01, IL02, IL03	A document will be prepared for each validation test, explaining the followed health and safety procedures
URG02	Ethics approval is required for all tests involving humans	BE01, BE02, BE03, BE04, BE05, RO01, RO02, RO03, RO04, RO05, RO06, IL01, IL02, IL03	A document will be prepared for each validation test, showing the ethical approval from the competent national authorities
URG03	Operational cost of overall system should be low	BE01, BE02, BE03, BE04, BE05, RO01, RO02, RO03, RO04, RO05, RO06, IL01, IL02, IL03	Estimated annual operational cost of the SafeShore System as deployed for the trial is recorded
URG04	Acquisition cost of overall system should be low	BE01, BE02, BE03, BE04, BE05, RO01, RO02, RO03, RO04, RO05, RO06, IL01, IL02, IL03	Estimated acquisition cost of the SafeShore System as deployed for the trial is recorded
URG05	Overall system should have no negative impact on the environment	BE01, BE02, BE03, BE04, BE05, RO01, RO02, RO03, RO04, RO05, RO06, IL01, IL02, IL03	Impact of system on environment will be recorded and reported if negative
URG06	Detection system should be able to cope with important tidal effects	BE01, BE02, BE03, BE04, BE05	Tests will be performed at the North Sea in Belgium where important tidal effects are present

data within the same study. The integration of qualitative and quantitative data provides a richer and more comprehensive overview of the performance of the system under test, taking into consideration several viewpoints: those of the end user and those of the platform developer. Obviously, mixed methods also have their disadvantages, as they render the evaluation process and the subsequent analysis process more difficult, so a cost-benefit analysis is required in order to assess whether the investment is merited. In general, it can be stated that this is mostly the case when the human-system interface is a crucial component of the system design, or when the user acceptance of the system is not straightforward (which also explains why these methods are popular within healthcare) [19]. An example of a mixed qualitative-quantitative evaluation methodology in the field of robotics is the framework proposed in [20], where the performance was assessed of a series of search and rescue robots that work in close collaboration with human search and rescue workers. In this study, the human operators were non-experts and were quite reluctant to accept these new tools in the beginning but were only convinced of the use of the system following the outcome of the mixed qualitative-quantitative evaluation [21].

The remainder of the article is organised as follows. In section 2, we discuss the proposed methodology from a conceptual point of view and compare it to state-of-the-art approaches. In section 3, we validate the methodology, taking as the evaluation of the SafeShore drone detection system as a case study. Finally, in the concluding section, the major lessons learned are summarised.

## 2. PROPOSED METHODOLOGY

### 2.1. Methodology for gathering user requirements

The proposed methodology is a derivative of the work performed in [21] but ported to the domain of maritime threat agent detection [22]. A first step in the development of the

validation framework was the requirements analysis, for which we followed a step-wise approach:

- The different end-user communities and stakeholders were identified.
- The different end-user communities were approached by means of market studies and targeted interviews.
- An early draft methodology proposal was compiled.
- This draft document was extensively discussed with both end users (in this specific case, maritime border management agencies) at relevant events and with platform developers in order to come to target performance levels that were both operationally realistic from an end user's point of view as well as from a platform developer's point of view in terms of the required effort, resources, and state-of-the-art and physical constraints.
- As SafeShore focuses on drone detection for the protection of maritime borders, a number of operational validation scenarios were proposed in order to address major issues the maritime border security community is facing today.
- For each of the validation scenarios, the target performance levels were proposed in discussion with end users and platform developers.
- These validation scenarios and target performance levels were updated throughout the lifetime of the project and were continuously adapted for the inevitably continuously changing operational requirements, technological capabilities, and scientific state of the art.

### 2.2. Concept overview

Two crucial aspects of obtaining realistic results from validation scenarios are that the scenarios should be as close as possible to operational reality and that the validation tests should be repeated enough so as to ensure statistical relevance. These two considerations are often in conflict with one another, as operational testing requires uncontrolled environments, whereas statistical relevance of results can only be obtained in controlled settings.

Within SafeShore, we have aimed to strike a balance between both aspects by providing a qualitative and quantitative assessment of the SafeShore system capabilities and by having multiple repeated experiments in realistic environments, following scenarios described by end users based on their needs and today's practical maritime border security problems.

The different components of the SafeShore validation concept are:

- A traceability matrix that clearly indicates what, for each validation scenario, are the relevant user requirements that need to be tested, allowing for the identification of how (by which validation scenario) each system requirement will be validated. This is important in order to keep track of the different user requirements and to make sure that, for each of the requirements, there is a validation scenario in place that ensures the verification of the attainment of the requirement. An example is shown in Table 1.
- A number of detailed scenarios, each related to maritime border security and safety. In total, SafeShore considers 14 validation scenarios: five to be executed in Belgium, three in Israel, and six in Romania. In this paper, we focus on those executed in Belgium. Each of these scenarios contains:
  - A capability score sheet, allowing for a qualitative assessment of the validation of the target performance

Table 2. Excerpt of the SafeShore capability score sheets.

Capability	Performance	User Requirements Reference
Generic Capabilities – See Appendix A		/ 53
Ability to cope with tidal effects	YES / NO	URG06
Ability to detect drones in the submicro category	YES / NO	URS01
Ability to detect drones in the micro category	YES / NO	URS02
Ability to detect drones in the mini category	YES / NO	URS03
Ability to detect drone operators	YES / NO	URS21
SafeShore system is compatible with the Maritime Information Cross point command and control system	YES / NO	URC23

levels. These capability score sheets allow us to make a binary assessment (YES/NO) as to whether one of the user or system requirements has been attained by the system or not. Table 2 shows a small excerpt of these capability score sheets. Three issues must be noted when analysing this figure:

- This figure only represents a very small fraction of the actual capabilities tested. In reality, there were over 60 capabilities that were evaluated for each scenario.
  - It is not possible to incorporate the actual performance of the system (PASS/NOT PASS) in this article. Indeed, as the SafeShore system is intended to be an operational border protection system, such information is restricted. This also explains the lack of actual data points in Table 3 and Table 4.
  - It is important to link the evaluated capabilities with the user requirements, as shown in Table 2, as this creates a link to the origin of the capability requirement.
- Template forms to be filled in during the validation tests, providing standardised information on the threat agents and the detection results. These template forms contain valuable environmental information, such as weather conditions and sea state. They also provide crucial information on the drones used as test agents: their visual/infrared/radiofrequency/acoustic/LIDAR signature, including ground truth timestamped GPS

Table 3. Excerpt of the SafeShore trial metrics forms.

TRIAL IDENTIFIER:			
DATE:		EVENT:	
TIME:		OPERATOR(S):	
LOCATION:		TRIAL NUMBER:	
WIND SPEED (M/S):		ILLUMINANCE (LUX):	
WAVE HEIGHT (M):		TEMPERATURE (°C):	
START TIME (HH:MM):		FINISH TIME (HH:MM):	
ENVIRONMENTAL NOISE (DB(A)):		DETECTOR NOISE <sup>5</sup> (DB(A)):	
GPS Coordinates	Detector 1	Detector 2	Detector 3
Latitude (°)			
Longitude (°)			
Elevation (m)			
Distance to shoreline (m)			

tracks, which allows for a full quantitative evaluation of the precision of the detection results. These evaluation forms also provide a means of evaluating the human-machine interface, as they gather information on the sample sizes for human verification, detection resolution, video framerates, etc. Table 3 shows a small fraction of a template form for one of the scenarios. These forms were generally completed prior to the actual test for recording the boundary conditions. After the test, they were filled with data concerning the test results. This can be performed easily, as the SafeShore command and control system automatically logs and records all system operations.

- A score sheet for the different metrics (Key Performance Indicators [KPIs]), allowing for a quantitative assessment of the validation of the target performance levels. Table 4 shows an example list of a few KPIs that were measured during one of the tests.
- Detailed target performance levels for each of the measured metrics. For each of the KPIs, three different levels of scoring were assessed in collaboration with the end users:
  - Minimum acceptance level: Performance below this level is not acceptable by the end users in operational conditions. Anything above is considered workable.
  - Goal level: This is the performance level anticipated by the end users.
  - Breakthrough level: This is a performance level beyond initial expectations that end users would one day like to have.

In the three right-hand columns of the table in Table 4, the target performance levels were pre-filled by the end users. For security reasons, this data had to be removed from this article. For a typical scenario, between 30 and 50 performance metrics were recorded in this way and compared to the target performance levels.

### 2.3. Conceptual differences between the proposed method and the state of the art

Table 5 gives an overview of a number of qualitative and quantitative (and mixed) evaluation methodologies. However,

Table 4. Excerpt of the target performance levels.

Measured Metrics	Target Performance Levels		
	Min. Acceptance Level	Goal Level	Break-through Level
Capabilities met ( /59)			
System Power Consumption (W)			
Difference between detector noise and environmental noise (DB(A))			
System Mass (kg)			
Estimated Acquisition cost of the SafeShore System as deployed for this trial (k€)			
Estimated Operational cost of the SafeShore System for 1 static VIP protection event as deployed for this trial (k€)			
Number of drones tested during trial			
% of false positives during test duration:			
% of false negatives (missed detections) during trial			

Table 5. Overview and comparison of qualitative and quantitative evaluation methodologies.

Method	WinField et al. [17]	Jacoff et al. [14]	Rao et al. [23]	Goldman et al. [24]	Doroftei et al. [20]	Doroftei et al. Proposed method
Type of method	Qualitative	Quantitative	Mixed	Mixed	Mixed	Mixed
Repeatability	Low	High	Medium	Medium	Medium	Medium
Statistical significance	Low	High	Medium	Medium	Medium	Medium
Potential for standardisation	Low	High	Low	Low	Low	Low
Realistic	High	Low	High	High	High	High
Story-based	Yes	No	Yes	Yes	Yes	Yes
Comprehensiveness	Medium	Medium	High	High	High	High
Cost	Very high	Low	Very high	Very high	Very high	High
User involvement	Low	Low	Very high	Very high	High	Medium
Methodological flexibility	Low	Low	High	High	High	High
Interdisciplinary constraints	Low	Low	High	High	High	High
Various viewpoints	No	No	Yes	Yes	Yes	Yes
Complexity	High	Low	Very high	High	Very high	High
Application field	Search and rescue robotics	Robotics	Social investment	Healthcare	Search and rescue robotics	Threat detection

the purpose of this table and this section is not – as is usual in a research article – to argue why one or other method is the best one in all cases. Indeed, each of these evaluation methodologies has its merits and value, and each could be the best choice for a given application. The purpose of the table is to better indicate for the reader the differentiating factors between the different evaluation methodologies such that an educated choice can be made between each of these options in a given case. Therefore, we will first explain the terminology of the different criteria used in Table 5:

- **Repeatability:** Considers how well the performance evaluation results can be replicated by other researchers. Quantitative methods are obviously much better in this regard than qualitative methods, with mixed methodologies somewhere in between.
- **Statistical significance:** Considers from a statistical point of view whether the produced results are significant or not. Quantitative methods are obviously much better in this than qualitative methods, with mixed methodologies somewhere in between.
- **Potential for standardisation:** Considers whether the approach can lead to a widely adopted test method standard. This requires repeatability and statistical significance, so it is related to the previous two criteria.
- **Realistic:** Considers whether the methodology takes into account realistic scenarios. Here, quantitative methods often fail to reproduce realistic operational conditions, as they need to stick to well-constrained boundary conditions.
- **Story-based:** Considers whether a story-based approach is followed in which end users' perspectives are incorporated, which is mostly the case for qualitative and mixed approaches.
- **Comprehensiveness:** Considers whether the evaluation results effectively showcase the performance of the system under test in a wide parameter domain, which is the strong point of the mixed-method approach.
- **Cost:** Considers the cost of doing the test. Here, quantitative methods excel, as they can generally keep the costs down, whereas qualitative and mixed approaches can be very expensive.
- **User involvement:** Considers the amount by which end users of the product under test are involved in the evaluation process. This is typically very high in mixed-method approaches (and in many qualitative approaches, but less in [17]). It is related to the level of realism of the test as well as the cost.
- **Methodological flexibility:** Considers the adaptability to different study designs. Mixed-method approaches are typically capable of elucidating more information than can be obtained in only quantitative research and therefore excel in this domain.
- **Interdisciplinary constraints:** Considers the requirement that a multi-disciplinary team is required to perform the evaluation. This is typically the case for the mixed-method approach. It has an impact on the comprehensiveness of the data as well as the cost.
- **Various viewpoints:** Considers the capability to give a voice to study participants and to ensure that the study findings are grounded in user experiences.
- **Complexity:** Considers the complexity of setting up the evaluation. This is obviously related to the cost as well, which is typically very high for the qualitative and mixed-method approaches, as a realistic setting needs to be reproduced, and many stakeholders need to be involved.

Comparing the proposed methodology to the other mixed methods [23], [24] and the earlier evaluation method proposed by the authors [20] (from which the approach proposed in this study is derived), it becomes obvious how this proposed evaluation methodology has been geared towards the application at hand. Indeed, it is important to note that all other mixed methods feature a higher level of user involvement than the proposed method. This is exactly due to the application domain: Rao considers social investment as an application domain in [23], so stakeholder interaction is crucial. Furthermore, in the healthcare domain [24], the primary actor is the patient, so the evaluation focused on this aspect. In [20], there was a great reluctance between the search and rescue workers to use the unmanned tools, so too here was user involvement (to foster user acceptance) crucial. For the SafeShore case, the user involvement was lower on the priority list, as the tool to be developed was essentially an autonomous sensor system that raises alarms for border security agents from time to time. Furthermore, there is

no specific reluctance to use such technology, among these border security agents (even quite to the contrary). Therefore, the user involvement aspect within the proposed evaluation methodology was lowered with respect to [20]. The advantage of this approach was that costs can be saved, as the complexity of the evaluation is somewhat reduced.

### 3. VALIDATION OF THE METHODOLOGY

#### 3.1. Trial concept and execution

As discussed above, five different trial scenarios related to maritime border security and safety were validated during the SafeShore trial in Belgium, which was the first in a series of three trial events of the project from which this validation methodology was applied.

For this operational field test, 11 different drone platforms (rotary wing, fixed wing, systems made of different materials, very fast drones, slow drones, etc) were deployed during a two-week measurement campaign in order to grasp different kinds of system capabilities as well as meteorological and operational conditions. For these series of tests, the Belgian Local Police West Coast acted as an end user, using the SafeShore detector as an integrated part of their surveillance capability. This allowed us to validate the system performance in a near-operational context.

Figure 2 shows the SafeShore prototype as it was installed on the beach in Belgium for a period of two weeks.

#### 3.2. Trial results

As this was the first of a series of three successive test campaigns, it was to be expected that the system would have some quirks and teething problems. The performance validation methodology was therefore essential in order to identify these issues and to indicate the causes for these problems.

Thanks to the proposed validation, at the end of each validation day, it was possible to provide an overview of the performance of the system, both from a qualitative as well as from a quantitative point of view. This analysis is shown in Figure 3. As stated above, the SafeShore validation trial in Belgium lasted for two weeks, consisting of ten working days. Figure 3 indicates for each working day the percentage of qualitative capabilities and quantitative metrics that were evaluated using the proposed methodology. Based on this daily real-time update, daily debriefings between SafeShore developers and SafeShore end users could be held in order to discuss the possibilities of and deficiencies with the system. As such, an action plan could be set up on a daily basis in order to improve the performance of the system. Due to this iterative review of the system, the performance of the SafeShore system improved on a daily basis.

At the end of the trial, the proposed validation methodology enabled us to provide a full overview of the performance of the system for all five scenarios, both from a qualitative as well as from a quantitative point of view. However, as this was the very first trial, it was not possible to sort out all problems with the system by the end of the trial period (which is why the 100 % mark was not reached, as shown in Figure 3). Based on the results of the validation method, a new action plan was therefore elaborated between end users and developers in order to improve



Figure 2. SafeShore system as installed on the beach in Belgium, by Daniel Orban

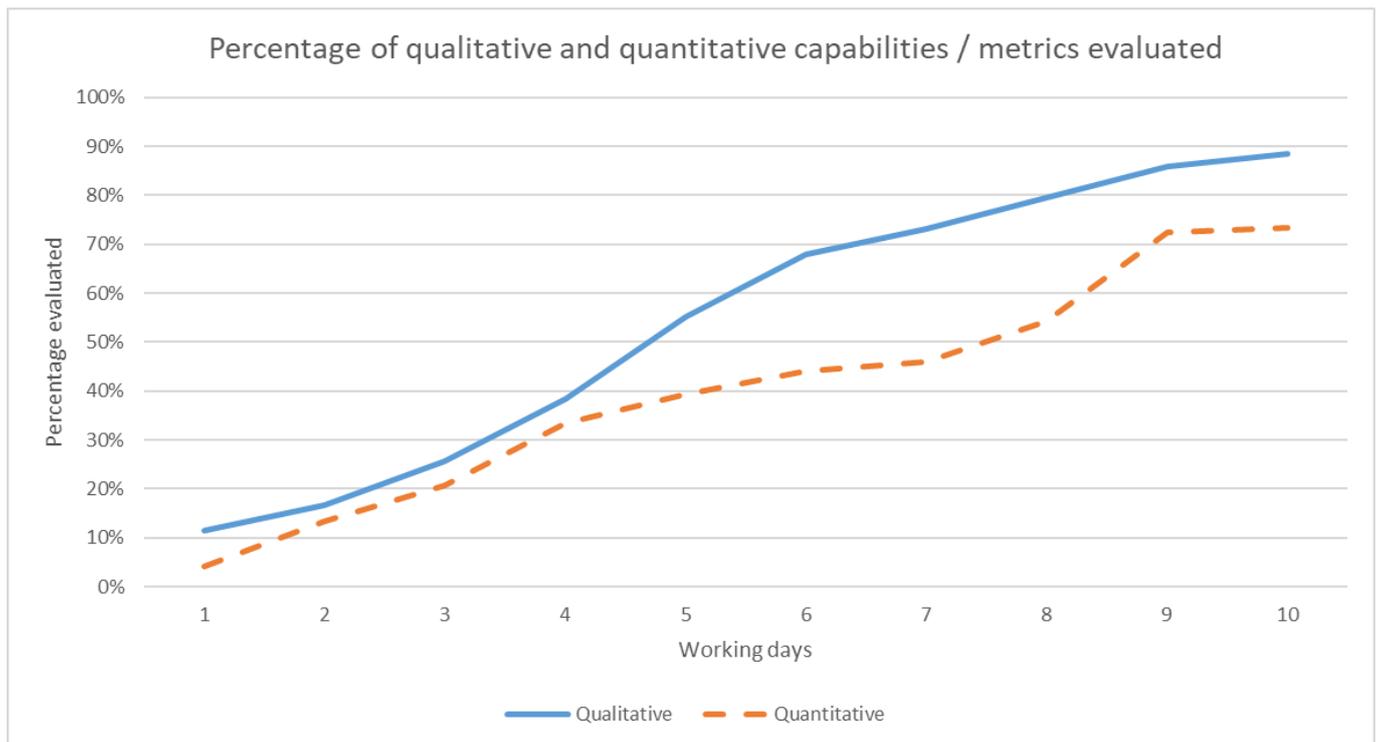


Figure 3. Evolution of the cumulative percentage of qualitative and quantitative capabilities/metrics successfully validated over the different working days of the SafeShore North Sea trial period.

the performance of the system during the next trials in Israel and Romania.

#### 4. CONCLUSIONS

In this paper, a validation methodology for evaluating complex systems was proposed, aiming to strike a balance between the rigorous, scientifically correct, and statistically relevant evaluation methodologies requested by platform developers in the iterative design stage on the one hand and the requirements of the end users on the other hand. The end users require field tests in operational conditions in order to evaluate the real-life performance of the system.

The proposed methodology achieves this objective by incorporating and integrating qualitative and quantitative aspects in the validation process.

The comparison of the proposed approach to state-of-the-art approaches showed that the proposed methodology has its value in reducing the cost related to the execution of qualitative and quantitative evaluation processes, while maintaining a realistic outlook and while staying close to operational conditions.

The proposed methodology was tested on a drone detection system in the context of the EU-H2020 SafeShore project and allowed the project participants (a heterogeneous mix of end users and platform developers) to improve the performance of the system on a daily basis during the operational field tests of the system, thereby proving the value of the proposed methodology. During the SafeShore trial, the validation methodology proved to work as expected: the number of validated capabilities (both quantitative and qualitative) rose on a daily basis, as indicated clearly by Figure 3. It is also clear from Figure 3 that the proposed methodology achieves a balance between the qualitative and quantitative aspects of the validation.

Concerning future work, we will focus on improving the proposed methodology to better incorporate the different priorities among requirements. Indeed, the current version does not take into consideration that some capabilities/metrics are more important than others (stemming from optional, desired, or mandatory requirements). This can sometimes lead to a twisted view on reality and will be improved in a future iteration.

#### ACKNOWLEDGEMENT

This work was supported by the European Union Horizon 2020 research and innovation programme under grant agreement N700643 (SafeShore).

#### REFERENCES

- [1] U. Franke, Drone proliferation: a cause for concern?, ISN Articles, November 2014.
- [2] M. Buric, G. de Cubber, Counter remotely piloted aircraft systems, MTA Review 27(1) (2017).
- [3] A. Coluccia, M. Ghenescu, T. Piatrik, G. De Cubber, A. Schumann, L. Sommer, J. Klatté, T. Schuchert, J. Beyerer, M. Farhadi, R. Amandi, C. Aker, S. Kalkan, M. Saqib, N. Sharma, S. Daud, K. Makkah, M. Blumenstein, Drone-vs-bird detection challenge at IEEE AVSS 2017, Proc. of the 14th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS) 2017, pp. 1-6.
- [4] C. J. Li, H. Ling, An investigation on the radar signatures of small consumer drones, IEEE Antennas and Wireless Propagation Letters (16) (2017) pp. 649-652.
- [5] J. Mezei, A. Molnar, Drone sound detection by correlation, Proc. of the 2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI), May, 2016, pp. 509-518.
- [6] A. Rozantsev, Vision-based detection of aircrafts and UAVs, [Master's thesis] EPFL, Lausanne, 2017.

- [7] P. Andrai, T. Raddi, M. Mutra, J. Ivoevi, Nighttime detection of UAVs using thermal infrared camera. *Transportation Research Procedia (INAIR)* 28 (2017) pp. 183-190. [Online] <http://www.sciencedirect.com/science/article/pii/S2352146517311043>
- [8] Y. L. Sit, B. Nuss, S. Basak, M. Orzol, W. Wiesbeck, T. Zwick, Real-time 2d+velocity localization measurement of a simultaneous-transmit OFDM MIMO radar using software defined radios, *Proc. of the 2016 European Radar Conference (EuRAD)*, Oct, 2016, pp. 21-24.
- [9] M. U. de Haag, C. G. Bartone, M. S. Braasch, Flight-test evaluation of small form-factor lidar and radar sensors for SUAS detect-and-avoid applications, *Proc. of the 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, Sept, 2016, pp. 1-11.
- [10] G. de Cubber, S. A. Berrabah, H. Sahli, Color-based visual servoing under varying illumination conditions. *Robotics and Autonomous Systems* 47(4) (2004) pp. 225-249.
- [11] V. Enescu, G. de Cubber, K. Cauwerts, H. Sahli, E. Demeester, D. Vanhooydonck, M. Nuttin, Active stereo vision-based mobile robot navigation for person tracking. *Integrated Computer-Aided Engineering* 13(3) (2006) pp. 203-222.
- [12] G. de Cubber, R. Shalom, A. Coluccia, O. Borcan, R. Chamrad, T. Radulescu, E. Izquierdo, Z. Gagov, The SafeShore system for the detection of threat agents in a maritime border environment, *Proc. of the IARP Workshop on Risky Interventions and Environmental Surveillance*, 2017.
- [13] V. Rao, M. Woolcock, *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*. World Bank, Washington DC, USA, 2003, ISBN 978-0-8213-5491-9, p. 165-190.
- [14] A. Jacoff, Guide for evaluating, purchasing, and training with response robots using DHS-NIST-ASTM international standard test methods, in: *Standard Test Methods for Response Robots*, National Institute of Standards and Technology, Gaithersburg MD, USA, 2014.
- [15] M. Q. Patton, *Qualitative Research and Evaluation Methods*, Sage Publications Ltd., Thousand Oaks, 2002.
- [16] F. Vanclay, Guidance for the design of qualitative case study evaluation, Department of Cultural Geography, University of Groningen, February 2012.
- [17] A. F. T. Winfield, M. P. Franco, B. Brueggemann, A. Castro, G. Ferri, F. Ferreira, A. Viguria, euRathlon and ERL emergency: a multi-domain multi-robot grand challenge for search and rescue robots, *Proc. of ROBOT 2017: Third Iberian Robotics Conference vol. 2*, 2017, pp. 263-271. (*Advances in Intelligent Systems and Computing* 694(1)). Springer. DOI: 10.1007/978-3-319-70836-2\_22
- [18] M. M. Marques, R. Parreira, V. Lobo, A. Martins, A. Matos, N. Cruz, J. M. Almeida, J. C. Alves, E. Silva, J. Będkowski, K. Majek, M. Pelka, P. Musialik, H. Ferreira, A. Dias, B. Ferreira, G. Amaral, A. Figueiredo, R. Almeida, F. Silva, D. Serrano, G. Moreno, G. de Cubber, H. Balta, H. Beglerović, S. Govindaraj, J. M. Sanchez, M. Tosa, Use of multi-domain robots in search and rescue operations – contributions of the ICARUS team to the euRathlon 2015 challenge, *Proc. of the IEEE OCEANS 2016*, Apr., 2016, Shanghai, China.
- [19] A. Shorten, J. Smith, Mixed methods research: expanding the evidence base. *Evidence-based Nursing* 20(3) (2017) pp. 74-75. ISSN 1367-6539
- [20] D. Doroftei, A. Matos, E. Silva, V. Lobo, R. Wagemans, G. de Cubber, Operational validation of robots for risky environments, *Proc. of the 8th IARP Workshop on Robotics for Risky Environments*, 2015.
- [21] G. de Cubber, D. Doroftei, H. Balta, A. Matos, E. Silva, D. Serrano, S. Govindaraj, R. Roda, V. Lobo, M. Marques, R. Wagemans, Operational validation of search and rescue robots, in: *Search and Rescue Robotics: From Theory to Practice*, InTech, 2017.
- [22] D. Doroftei, G. de Cubber, Qualitative and quantitative validation of drone detection systems, *Proc. of the International Symposium on Measurement and Control in Robotics ISMCR2018*, September 2018, Mons, Belgium.
- [23] V. Rao, A. M. Ibáñez, The social impact of social funds in Jamaica: a mixed-methods analysis of participation, targeting and collective action in community-driven development. Policy Research Working Paper 2970. World Bank, Development Research Group, Washington, D.C [Processed].
- [24] R. E. Goldman, D. R. Parker, J. Brown, J. Walker, C. B. Eaton, J. M. Borkan, Recommendations for a mixed methods approach to evaluating the patient-centered medical home. *Ann. Fam. Med.*, 13(2) (2015) pp. 168-75. doi: 10.1370/afm.1765.