ANYWHERE

Project Number 700099
Call: H2020-DRS-01-2015

Project Title:

ANYWHERE

EnhANcing emergencY management and response to extreme WeatHER and climate Events

Subject:

Deliverable 1.4: Results of the co-evaluation of the ANYWHERE tools, products and services on the pilot sites

Dissemination Level:
PU Public

Delivery date: **24th December 2019**
Month: **Month 43**
Organisation name of lead contractor for this deliverable: **UFZ**

European Commission

This project has received funding from the European Union's H2020 Programme under the topic of potential of current and new measures and technologies to respond to extreme weather and climate events under grant agreement no. 700099.

Document Information

Title	Results of the co-evaluation of the ANYWHERE tools, products and services at the pilot sites
Lead Author	Gebhardt, Oliver
Contributors	Kuhlicke, Christian Wolf, Lukas Claudia Vitolo Enrico Duo van Lanen, Henny A.J. Rohrer, Mario Sutanto, Samuel J. Stoffel, Markus
Distribution	PUBLIC
Document Reference	D1.4

Document History

Date	Revision	Prepared by	Organisation	Approved by
27.11.2019	Rev_1	Oliver Gebhardt	UFZ	Christian Kuhlicke
13.12.2019	Rev_2	Oliver Gebhardt	UFZ	Christian Kuhlicke
23.12.2019	Rev_3	Christian Kuhlicke	UFZ	Oliver Gebhardt
24.12.2019	Rev_4	Oliver Gebhardt	UFZ	Daniel Sempere-Torres (CRAHI-UPC)

Related Documents:

This report and others are available from the **ANYWHERE** Project Website at: http://www.anywhere-h2020.eu/

Summary

This document introduces readers to the results of the co-evaluation activities of the ANYWHERE distributed platform for Emergency Management Operation Services at the pilot sites. It reflects upon the application of the co-evaluation methodology outlined in the ANYWHERE project Deliverable 1.2 *Report on needs and requirements from the users.* The main focus is on the comparative assessment of the emergency management systems including the ANYWHERE platform at the pilot sites and the locally implemented Legacy systems. The relative advantages and disadvantages of the respective systems are analysed and documented. The analysis revealed that the new ANYWHERE systems perform better with respect to 14 out of 21 criteria, it performs similar with respect to four criteria and there is room for improvement with respect to three criteria. However, it is to be expected that this improvement is to be achieved within a more intense use of the ANYWHERE system as the criteria related above all to the trust in the new system, which will grow through use.

Table of Contents

Document Information i
Document History i
Related Documents: i
Summary ii
Table of Contents iii
List of Figures v
List of Tables vi
1 Introduction 7
1.1 Purpose of the document 7
1.2 Structure of the document 8
2 Methodology 9
2.1 Overview 9
2.2 Criteria and indicators 10
2.3 Evaluation method 13
2.4 Data collection 14
3 Results 15
3.1 A4ALPS 15
3.1.1 General impressions 15
3.1.2 Usefulness of relevant features 16
3.1.3 Social and technical aspects 17
3.1.4 Results of the Multi-criteria analysis 21
3.2 A4CAT 24
3.2.1 General impressions 24
3.2.2 Usefulness of relevant features 25
3.2.3 Social and technical aspects 26
3.2.4 Results of the Multi-criteria analysis 30
3.3 A4LIG 33
3.3.1 General impressions 33
3.3.2 Usefulness of relevant features 33
3.3.3 Social and technical aspects 35
3.3.4 Results of the Multi-criteria analysis 39
3.4 A4FINN 42
3.4.1 General impressions 42
3.4.2 Usefulness of relevant features 43
3.4.3 Social and technical aspects 44
3.4.4 Results of the Multi-criteria analysis 48
3.5 A4NOR 50
3.5.1 General impressions 50
3.5.2 Usefulness of relevant features 50
3.5.3 Social and technical aspects 51
3.5.4 Results of the Multi-criteria analysis 55
3.6 A4COR 57
3.6.1 General impressions 57
3.6.2 Usefulness of relevant features 58
3.6.3 Social and technical aspects 59
3.6.4 Results of the Multi-criteria analysis 63
3.7 A4CENEM 66
3.7.1 General impressions 66

3.7.2 Usefulness of relevant features ... 66
3.8 Comparison of results across pilot sites ... 68
4 Discussion and conclusion ... 72
4.1 Challenges ... 72
4.1 Further benefits of the co-production of the ANYWHERE platform ... 73
4.2 Conclusion ... 73
5 Scientific advancements of ANYWHERE ... 74
6 References ... 78
7 ANNEX 1 ... 79
7.1 Overview of criteria used to co-evaluate ANYWHERE platform at the pilot sites ... 80
7.2 A4CAT – Overview of criteria categories ... 86
7.3 A4LIG – Overview of criteria categories ... 87
7.4 A4FINN – Overview of criteria categories ... 88
7.5 A4NOR – Overview of criteria categories ... 89
7.6 A4COR – Overview of criteria categories ... 90
7.7 A4ALPS - Relative importance of all co-evaluation criteria ... 91
7.8 A4ALPS - Relative importance of criteria for comparison of operator's and coordinator's perspectives ... 92
7.9 A4ALPS - Performance comparison with Legacy system (operator) ... 93
7.10 A4ALPS - Performance comparison with Legacy system (coordinator) ... 94
7.11 A4CAT - Relative importance of all co-evaluation criteria ... 95
7.12 A4CAT - Relative importance of criteria comparing operator's and coordinator's perspectives ... 96
7.13 A4CAT - Performance comparison with Legacy system (operator) ... 97
7.14 A4CAT - Performance comparison with Legacy system (coordinator) ... 98
7.15 A4LIG - Relative importance of all co-evaluation criteria ... 99
7.16 A4FINN - Relative importance of all co-evaluation criteria ... 100
7.17 A4FINN - Relative importance of criteria comparing operator's and coordinator's perspectives ... 101
7.18 A4FINN - Performance comparison with Legacy system (operator) ... 102
7.19 A4FINN - Performance comparison with Legacy system (coordinator) ... 103
7.20 A4NOR - Relative importance of all co-evaluation criteria ... 104
7.21 A4COR - Relative importance of all co-evaluation criteria ... 105
7.22 A4COR - Relative importance of criteria comparing operator's and coordinator's perspectives ... 106
7.23 A4COR - Performance comparison with Legacy system (operator) ... 107
7.24 A4COR - Performance comparison with Legacy system (coordinator) ... 108

List of Figures

Figure 1: Criteria by impact areas: Criteria selected 11
Figure 2: Criteria by impact areas: Comprehensive list 11
Figure 3: A4ALPS - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom) 13
Figure 4: A4ALPS - Usefulness of relevant features (mean values) 17
Figure 5: A4ALPS - Working routine-related criteria (mean values) 19
Figure 6: A4ALPS - Knowledge production-related criteria (mean values) 20
Figure 7: A4ALPS - Technical criteria (mean values) 21
Figure 8: Performance comparison A4ALPS and Legacy system 23
Figure 9: A4CAT - Usefulness of relevant features (mean values) 26
Figure 10: A4CAT - Working routine-related criteria (mean values) 28
Figure 11: A4CAT - Knowledge production-related criteria (mean values) 28
Figure 12: A4CAT - Technical criteria (mean values) 30
Figure 13: Performance comparison A4CAT and Legacy system 32
Figure 14: A4LIG - Usefulness of relevant features (mean values) 35
Figure 15: A4LIG - Working routine-related criteria (mean values) 37
Figure 16: A4LIG - Knowledge production-related criteria (mean values) 37
Figure 17: A4LIG - Technical criteria (mean values) 38
Figure 18: Performance comparison A4LIG and Legacy system 41
Figure 19: A4FINN - Usefulness of relevant features (mean values) 44
Figure 20: A4FINN - Working routine-related criteria (mean values) 46
Figure 21: A4FINN - Knowledge production-related criteria (mean values) 46
Figure 22: A4FINN - Technical criteria (mean values) 47
Figure 23: Performance comparison A4FINN and Legacy system 49
Figure 24: A4NOR - Usefulness of relevant features (mean values) 51
Figure 25: A4NOR - Working routine-related criteria (mean values) 53
Figure 26: A4NOR - Knowledge production-related criteria (mean values) 53
Figure 27: A4NOR - Technical criteria (mean values 54
Figure 28: Performance comparison A4NOR and Legacy system 56
Figure 29: A4COR - Usefulness of relevant features (mean values) 59
Figure 30: A4COR - Working routine-related criteria (mean values) 61
Figure 31: A4COR - Knowledge production-related criteria (mean values) 62
Figure 32: A4COR - Technical criteria (mean values) 63
Figure 33: Performance comparison A4COR and Legacy system 65
Figure 34: A4CENEM - Usefulness of relevant features (mean values) 68
Figure 35: Performance comparison of ANYWHERE systems and Legacy systems: All criteria 70
Figure 36: Performance comparison of ANYWHERE systems and Legacy systems: Mean value differences of criteria 71
Figure 37: Assignment of ANYWHERE scientific papers to journal categories 75
Figure 38: Impact Factor (JIF) of the peer-reviewed scientific journals in which ANYWHERE papers are published 76
Figure 39: Collaboration among partners 76
Figure 40: A4CAT - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom) 86
Figure 41: A4LIG - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom) 87
Figure 42: A4FINN - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom) 88

Figure 43: A4NOR - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom) 89
Figure 44: A4COR - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom) 90
Figure 45: A4ALPS - Relative importance of all co-evaluation criteria 91
Figure 46: A4ALPS - Relative importance of criteria for comparison of operator's and coordinator's perspectives 92
Figure 47: A4ALPS - Performance comparison with Legacy system (operator) 93
Figure 48: A4ALPS - Performance comparison with Legacy system (coordinator) 94
Figure 49: A4CAT - Relative importance of all co-evaluation criteria 95
Figure 50: A4CAT - Relative importance of criteria comparing operator's and coordinator's perspectives 96
Figure 51: Performance comparison A4CAT and Legacy system (operator) 97
Figure 52: A4CAT - Performance comparison with Legacy system (coordinator) 98
Figure 53: A4LIG - Relative importance of all co-evaluation criteria 99
Figure 54: A4FINN - Relative importance of all co-evaluation criteria 100
Figure 55: A4FINN - Relative importance of criteria comparing operator's and coordinator's perspectives 101
Figure 56: A4FINN - Performance comparison with Legacy system (operator) 102
Figure 57: A4FINN - Performance comparison with Legacy system (coordinator) 103
Figure 58: A4NOR - Relative importance of all co-evaluation criteria 104
Figure 59: A4COR - Relative importance of all co-evaluation criteria 105
Figure 60: A4COR - Relative importance of criteria for comparison of operator's and coordinator's perspectives 106
Figure 61: A4COR - Performance comparison with Legacy system (operator) 107
Figure 62: A4COR - Performance comparison with Legacy system (coordinator) 108

List of Tables

Table 1: Distribution of scientific papers over the ANYWHERE project lifetime 74
Table 2: Overview of criteria used to co-evaluate ANYWHERE platform at the pilot sites 80
Table 3: List of papers in peer-reviewed scientific journals by ANYWHERE partners 109

1 Introduction

1.1 Purpose of the document

This document presents the results of the co-evaluation of the ANYWHERE platforms at the seven pilot sites.[1] The main purposes of this co-evaluation are[2]:

Support of the development process

As future end-users are actively involved as co-producers in the development of the ANYWHERE platform co-evaluation activities are one of the options to collect and share highly valuable feedback from and among all parties involved.

Support of market outreach of ANYWHERE

The co-evaluation is supposed to support ANYWHERE's market outreach activities by offering potential customers empirical insights from a diverse sample of settings on which kind of changes be expected when implementing the ANYWHERE platform. For this a joint reflection of the results presented in this report with the comprehensive descriptions of pilot site-specific context conditions documented in the ANYWHERE project Deliverable 1.2 *Report on needs and requirements from the users* is advisable.

Support of the sustainable implementation process

Co-evaluation cannot only support the co-production and, hence, enhance the development of new technologies but also and improve their adoption and diffusion. Empirical evidence shows that the variance in the rate of adoption of an innovation and its diffusion can be explained by the inner-organisational perception of its (1) *relative advantage*[3] compared to Legacy system, (2) *compatibility*[4], (3) *complexity*[5], (4) *trialability*[6] and (5) *observability*[7] (Rogers 2003). There are various ways, in which the perception of these attributes can be influenced in the development and implementation phase. In the context of ANYWHERE the collaborative development process was designed and implemented for enabling compatibility and trialability as well as cater for the inevitable complexity. On the one hand side observability was addressed through presentations, workshops and training involving colleagues of the

[1] The methodological framework of the co-evaluation and its foundations are presented in detail in the project Deliverable 1.3 (D1.3) *Report describing the methodology for the co-evaluation of the ANYWHERE distributed platform for Emergency Management Operation Services (A4DEMOS) at the pilot sites*. As this report presents the results of the application of the co-evaluation framework it only includes a brief methodology section, which is introducing information not available, yet, when D1.3 was compiled.
[2] For more details see section 1 of D1.3.
[3] Degree to which an innovation is perceived as better than the idea it supersedes.
[4] Degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of potential adopters
[5] Degree to which an innovation is perceived as relatively difficult to understand and to use.
[6] Degree to which an innovation may be experimented with on a limited basis.
[7] Degree to which the results of an innovation are visible to others.

users directly involved in the development of the ANYWHERE system but also externals. On the other hand side observability was also addressed through the co-evaluation of the site-specific version of the platform as its main focus is on the analysis of whether, in which regard and to what extent the new system is outperforming the existing one. This enhances the visibility and, thereby, the perception of relative advantage of the ANYWHERE system by the users directly or indirectly involved as well as potential new customers.

This report differs from the ones presenting the validation and assessment activities in work packages 4 and 6 by providing a wider reflection of added value of the use of the platform at the sites based on the experiences made during the demonstration period. It has a strong focus on changes in working routines, knowledge production, trust, and fitting of technical properties to organisational routines, whereas the Deliverables in work packages 4 and 6 complement this perspective by focussing on:

- D4.5: Presentation of the functionalities of the different versions of the platform from the developers' point of view
- D6.4: Presentation of the feedback from the end-users based on their experiences during the demonstration period
- D6.5: Synthesis of D4.5 and D6.4 supplemented with additional performance statistics

1.2 Structure of the document

In chapter 2, this report will briefly summarize the selection process of evaluation criteria and indicators (step 3 of the ANYWHERE co-evaluation methodology; see D1.3), recap on the evaluation method applied (step 4) and describe the data collection process (step 5). The main focus of this report will be on providing an overview of results of the co-evaluation of the ANYWHERE platform at the pilot sites (step 6). These are presented in chapter 3. In a first step, we presented the general impression of the platform by referring to some insights deriving from qualitative interviews with conducted with users of the platform. In a second step, we proceed by focusing on the perceived usefulness of the platform before we unravel more systematically on how the users perceived the social and technical aspects of the platform. The section concludes with an overview the results produced by means of the multi-criteria analysis. In section 4, we discuss the results and provide some conclusion. In chapter 5 we present key scientific advancements we were able to make in the ANYWHERE project.

2 Methodology

2.1 Overview

The basic structure of the ANYWHERE co-evaluation methodology[8] can be described by the following six steps:[9]

- Step 1: Context analysis
- Step 2: Description of baseline and ex post scenario
- Step 3: Selection of indicators
- Step 4: Selection of evaluation method
- Step 5: Data collection
- Step 6: Data aggregation and analysis

The foundation for the *context analysis* was formed by the comprehensive investigation of the working and decision-making context at the pilot sites during the first partner consultations and field visits. The expectations of project partners as well as their process-related, but also some output-related needs were investigated and synthesized in section 3 of ANYWHERE report Deliverable 1.1: Report with the recommendations and feedback obtained from the Workshop 1. Product-related needs were then focussed in more details in ANYWHERE report Deliverable 1.2: Report on needs and requirements from the users (see especially section 6). The report also describes in detail the roles and responsibilities as well as the data sources used, interactions and information flows during the management of hydro-meteorological hazards. It presents the temporal structure of the warning phases and decision-making processes. These two reports form basis for the *context-analysis* as well the *description of the general aspects of the baseline scenarios*, i.e. the emergency management process at the pilot sites using the current system.[10]

The description of the ex-post scenarios, i.e. the specific set-up of the ANYWHERE systems at the pilot sites is presented in the ANYWHERE report Deliverable 4.5: A4EU prototypes (A4Lig, A4Cat, A4CENEM, A4Nor, A4Finn, A4Alps, and A4Cor): final versions and performance assessment with focus on the comparison between the different A4EU systems' capabilities and functionalities".

Hence, this report will focus on the steps 3 to 6. It covers the selection of suitable and feasible criteria for measuring the performance of the Legacy system, i.e. the baseline scenario, against the respective ANYWHERE system, i.e. the ex post scenario (step 3). Criteria selection is based on the expected benefits identified at step 1. Furthermore, the evaluation method chosen in the light of the given context conditions of the assessments is presented (step 4) and the data collection process is reflected

[8] For details see sections 3 and 4 of the ANYWHERE project Deliverable 1.3.
[9] The methodology was applied for six of the seven ANYWHERE pilot sites. Due to the current level of maturity of A4CENEM, the first national level ANYHWERE platform prototype, the co-evaluation was primarily focussed on the regional pilot sites.
[10] Due to the evolution of the role of different pilot sites during the course of the project information on the Norwegian pilot site is less comprehensive and the Corsican as well as the Spanish pilot site are not covered in the report.

upon (step 5). The main focus of the report is on the presentation of the results of the co-evaluation. This includes an aggregated view as well as more differentiated results addressing performance differences for specific criteria (groups) and user types (step 6).

2.2 Criteria and indicators

The performance of the ANYWHERE platform and the respective Legacy systems was planned to be assessed along the lines of the four impact areas “economic”, “social”, “technological” and “environmental”. For these impact areas corresponding “effects”, “criteria”, “sub-criteria” (if applicable) and indicators were specified.

The comprehensive list of co-evaluation criteria[11] was based on inputs from initial enquiry at the beginning of the ANYWHERE project and pilot site visits (bottom-up) as well as the so called SEQUOIA methodology (Monacciani et al. 2011, Cucco et al. 2016)[12] and ISO/IEC norms 25010, 25022, 25023 for evaluating software quality (top-down).

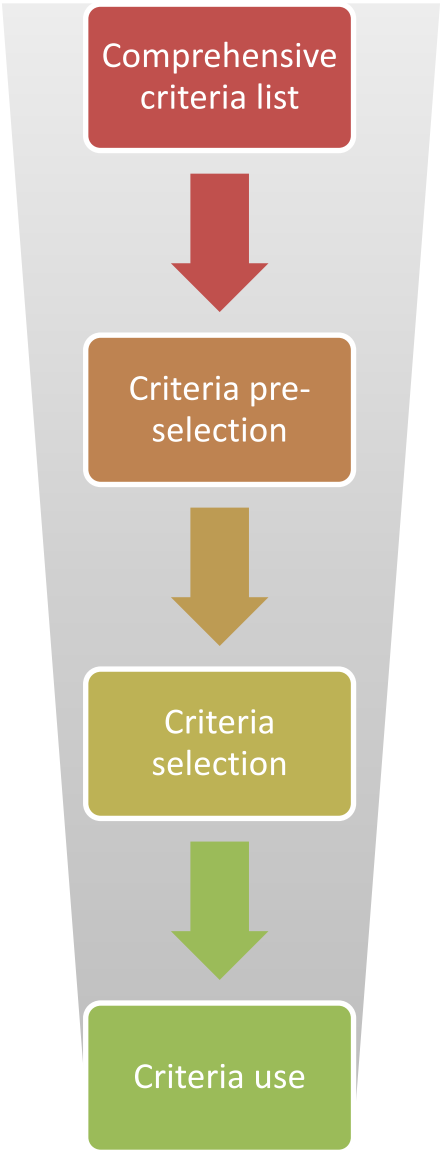
Pilot site partners rated the criteria listed with regard to their suitability and feasibility. Suitability represents the ability of criteria to capture relevant performance aspects of the emergency management systems.[13] Feasibility primarily addresses the data collection process, i.e. requires reflection on data availability and opportunities for data collection in the context of ANYWHERE demonstration activities.[14]

The determination of the final set of co-evaluation criteria was performed in an iterative way. Firstly, a pre-selection of criteria was conducted and presented at the ANYWHERE meeting in Bastia in June 2018. Secondly, based on bilateral consultations with the UFZ team a site-specific set of criteria was determined which formed the basis for the co-evaluation questionnaire. It included criteria which were relevant or highly relevant and for which data colelction would be feasible with high or moderated effort. The results were presented at the ANYWHERE meeting in Barcelona in November 2018 and used to prepare the midterm co-evaluations at the pre-test sites

[11] The criteria list presented in Table 8 of Deliverable 1.3 was supplemented and modified based on ISO/IEC norms 25010, 25022, 25023 and feedback from the pilot site partners.
[12] This approach was originally developed for project evaluation in the context of the EU FP7 research project “Socio-Economic Impact Assessment for Research Projects (SEQUOIA). Since then it was adapted and applied several times in very different fields, e.g. for a Pan-European disaster inventory (SecInCoRe), for Digital Cultural Heritage projects (Maxiculture), for Digital Social Innovation projects (iA4Si), for e-Infrastructure projects (ERINA+) and for automated service-oriented architecture testing infrastructures (MIDAS).
[13] For each criterion it was assessed whether it addresses an (1) irrelevant performance aspect, (2) moderately relevant performance aspect, (3) relevant performance aspect, (4) highly relevant performance aspect.
[14] For each criterion it was assessed whether it is (1) very unlikely that information can be collected, (2) information can be collected with high effort, (3) information can be collected with moderate effort through consultation of actors outside the ANYWHERE consortium or (4) information can be collected with moderate effort through consultation of ANYWHERE partners.

in spring 2019.[15] Thirdly, the actual (and not the expected) ability to collect data for h criterion and corresponding indicator determined the final set of criteria to be used at each pilot site.

As a consequence of this iterative process the number of criteria and share of impact areas covered by the criteria varies across sites. The following figure gives an overview for all sites to what extent the four impact areas mentioned above were covered (1) by the comprehensive criteria list, (2) after the final criteria selection by pilot site partners and (3) by the co-evaluation itself. The comprehensive criteria list is included in the annex as Table 2. Criteria used by each site are indicated.



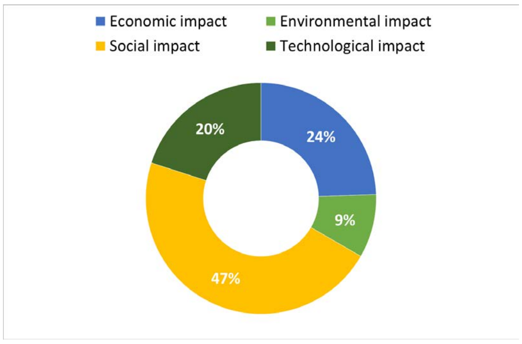


Figure 2: Criteria by impact areas: Comprehensive list

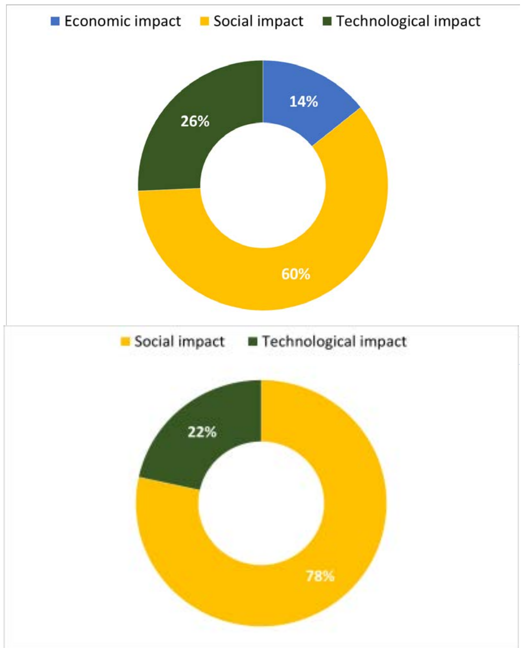


Figure 1: Criteria by impact areas: Criteria used

[15] The full-fledged co-evaluation questionnaire was pre-tested with personal interviews with operators and coordinators in April and May 2019 at two pilot sites, i.e. Finland and Catalonia, which represented the varying degrees of maturity of the prototypes at the different ANYWHERE sites. At the other sites the midterm evaluation was carried out using a condensed version of that questionnaire.

The most important change from the first (comprehensive criteria list) to the second (criteria pre-selection) stage was the exclusion of the environmental impact-related criteria. All pilot site partners considered them to be irrelevant for assessing the performance of an emergency management system. There was a refinement of the pre-selected criteria sets on the basis of an exchange between end-users and evaluators after the pre-selection aimed at clarifying any doubts to ensure a common understanding of the terms used. The result of this process were site-specific sets of evaluation criteria and corresponding indicators. Due to overly optimistic expectations regarding data availability or simply because no changes could be reported, e.g. for different types of costs, for most pilot sites this set differs from the set of criteria which were eventually used for the co-evaluation.

The criteria excluded at this stage primarily concerned economic criteria. There were several reasons for that, e.g.

- As web-based service the ANYWHERE system did not cause any additional investment costs, e.g. for hardware, or operational costs, e.g. for server capacity or licencing fees (during demonstration period).
- Training on ANYWHERE system could be integrated in regular training activities and, hence, did not cause any additional costs.
- Economic effects could not be detected, yet, as demonstration period was too short for compiling reliable statistics. Such statistics require more time and real events which have to be dealt with.

Co-evaluation criteria can be grouped into different categories. Figure 3 gives an overview of different types of categorization of the criteria used at the Swiss pilot site. Due to the before mentioned preferences and restrictions the vast majority of criteria used for the comparative assessment, i.e. at least 3/4 of all criteria considered, address social impacts while the rest focuses on technological impacts.

Within the category of social impacts the shares of working routine-related and knowledge production and sharing-related criteria are quite even. This holds for most of the sites with Norway, Finland and to a lesser extent Catalonia being exceptions in this regard. At these three pilot sites a higher share of knowledge production and sharing-related criteria was used. For an overall assessment of the relevance of different criteria categories and, thereby, specific perspectives on the final result not only the numbers of criteria have to be considered but also their relative weights.

The top row of Figure 3 presents the shares of the various criteria categories by numbers and the bottom row illustrates the relative weight of these criteria categories for A4ALPS. It can be stated that criteria category shares are as follows:

- 4/5 social impact-related, 1/5 technological impact-related
- 2/5 effects on working routines, 2/5 effects on knowledge production and sharing, 1/5 technical effects
- 2/5 knowledge production-related, 1/5 usefulness-related, 1/7 usability-related, 1/7 decision-making-related, 1/10 other categories

The corresponding figures for the other pilot sites are included in the annex (see section 7.2 - 7.6).

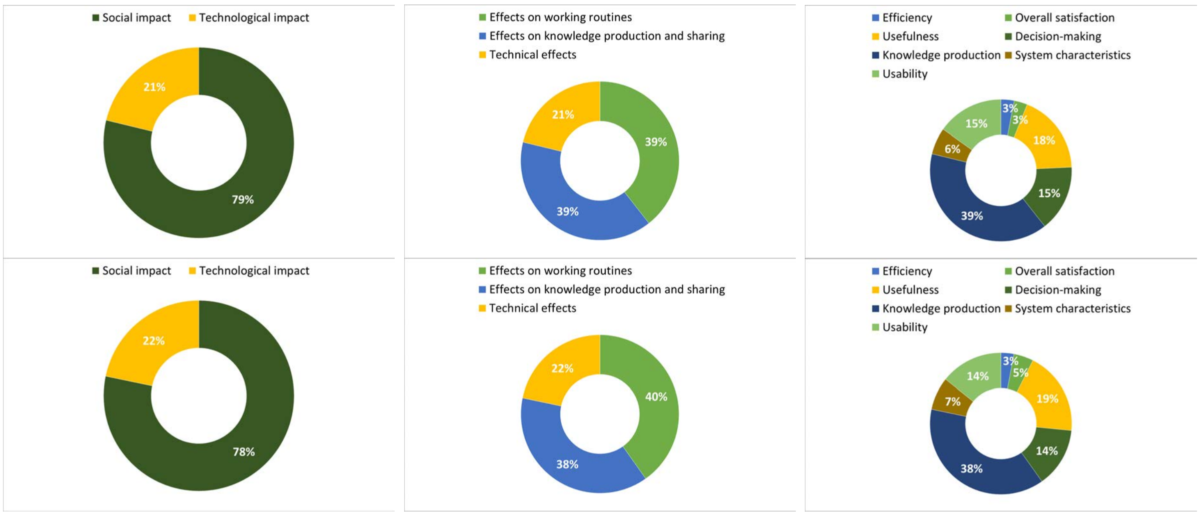


Figure 3: A4ALPS - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom)

2.3 Evaluation method

The different evaluation methods at hand were described in detail in the ANYWHERE project Deliverable 1.3 “Report describing the methodology for the co-evaluation of the ANYWHERE distributed platform for Emergency Management Operation Services (A4DEMOS) at the pilot sites. In section 4.5.4 of D1.3 a comparative appreciation of the different methods was presented and let to the conclusion that weighting the pros and cons of the approaches, acknowledging the properties of the performance data to be dealt with (scales, metrics, level of uncertainty) and reflecting the pre-conditions of the application of the various Multi-criteria approaches at hand the use of a probabilistic Multi-criteria outranking approach seemed to be most advisable.

We decided to use an interactive software for Probabilistic Multi-Attribute Evaluation PRIMATE for the comparative assessment of the two systems at each pilot site as data aggregation and analysis could be substantially facilitated by the use of this tools. The Multi-criteria module of PRIMATE is based on the outranking concept PROMETHEE (Preference Ranking Organisation Method for Enrichment Evaluations), which performs a pairwise comparison of the alternative systems respectively scenarios across all evaluation criteria. A detailed description of the tool and the method is provided in section 4.7.1 of D1.3. Some characteristics of the tool and its application of the probabilistic PROMETHEE II approach shall be explained to facilitate the interpretation of the co-evaluation results presented in the subsection *Results of the Multi-criteria analysis* for each pilot site.

PRIMATE allows for the simultaneous and explicit consideration of varying preferences of different user types by using separate weighting sets for the analysis. Furthermore, it considers uncertain information about the systems‘ performances regarding single aspects, i.e. criteria, probabilistically.[16] For this purpose Monte-Carlo simulations are used. This means that several PROMETHEE analyses are performed for a random sample of criteria values within a pre-defined score range. Based on the performance

[16] Uncertain information is a consequence of diverging scores provided by different users for the same indicator and, thereby, for the same criterion in the data collection process.

data collected this score range is specified as uniform distribution defined by a minimum and a maximum value. PRIMATE randomly selects values out of this range and runs up to 10.000 Multi-criteria analysis (MCA) with this values. Results of all evaluations are then statistically analysed (arithmetic mean, standard deviation, ranking order) and documented. On this basis it can be stated that with a certain probability either the ANYWHERE system or the Legacy system outperforms the respective alternative when considering all performance aspects and their relative importance as specified by the weighting set used. The results of the MCA are presented in PRIMATE in various ways. This includes not only the overall performance of the alternative systems but also their strengths or weaknesses with regard to each performance aspect, i.e. criterion, considered. This provides more insights in the evaluation results than the bare interpretation of ranking probabilities. As a sensitivity analysis separate MCAs can be conducted for the data and weights sourced from each user type which then can be contrasted with the overall results.[17] As there are certain challenges to this approach due to performance data availability for all user types and criteria we conducted separate MCAs for each user type, i.e. coordinators and operators, using the same performance data sets but only the respective distinct user weighting set and compared the results. For all MCAs at all sites the variation from the overall results, which are presented in the *Results of the Multi-criteria analysis* subsections for each pilot site, was absolutely negligible.

2.4 Data collection

Once the Legacy systems and ANYWHERE systems were specified, the performance criteria and indicators were selected and evaluation methods chosen the data collection process was initiated. A specific data collection strategy was determined for each pilot site. For every indicator capturing single a performance aspect of the ANYWHERE system and the Legacy system data was collected. Due to the lack of statistical performance data substantial shares of the information used for the co-evaluation was sourced through questionnaire-based expert interviews with operators and coordinators who were experienced in using the ANYWHERE platform.

We pre-tested the questionnaire in the context of the midterm co-evaluation of A4FINN and A4CAT through personal on site interviews. Based on the results of these pre-tests the questionnaire was revised. Data was collected through personal interviews at the end of the demonstration period, i.e. in September and October 2019.

The team of the Helmholtz Centre for Environmental Research – UFZ conducted the interviews for A4ALPS, A4CAT, A4FINN in person on-site and for A4NOR via a 4 hours

[17] A drawback for this kind of sensitivity analysis is the fact that for several reasons not all operators and coordinators who provided data for the analysis did so for all criteria. Hence, only those criteria could be considered for the separate analysis by user type for which it was possible to source data from the operator and coordinator. Still, the results of this analysis provide some interesting insights in the perspectives of operators and coordinators. Hence, they are included in the annex. No diagrams are available for A4CENEM (no such data was collected), A4NOR (data was only sourced from operators) and A4LIG (many missing criteria scores for coordinator's perspective would result in too few common criteria to make a meaningful comparison).

Skype conference. The UFZ team trained ANYWHERE partners of the European Centre for Medium-Range Weather Forecasts - ECMWF and Consorzio Futuro in Ricerca – CFR to conduct the interviews for the co-evaluation of A4COR and A4LIG. These partners also took part in the interviews for the assessment of A4CAT to ensure that the same interview routine is to be followed. The UFZ team assisted the interview in Liguria via Skype conference. Spanish partners filled the co-evaluation questionnaire to provide data for the very lean version of the co-evaluation of A4CENEM. This was prepared through an intense personal exchange beforehand.

The relative importance of the criteria was determined through an interactive exercise based on the revised SIMOS weighting technique.[18] Criteria cards were ordered from the most to the least important. On the one hand side blank cards could be used to increase the distance between criteria and on the other hand side criteria cards could be put next to each other, if criteria were considered to be equally important. Then importance scores were specified for each criteria row. In general, the score 100 was elicited to the row with the most important criteria. Then interviewees indicated the relative importance of the least important criterion compared to the most important one. If it was considered to be half as important, the score 50 was elicited. Then the scores for all row in between the most and the least important criterion were specified.

All filled questionnaires, documentations of the criteria weighting and other interview data was digitalised for further use. A database was compiled, which formed the basis for the statistical analyses as well as Multi-criteria analyses conducted.

3 Results

3.1 A4ALPS

3.1.1 General impressions

ANYWHERE partners at the Swiss pilot site used A4ALPS on a regular basis, since summer 2018. The frequency of use depended on the particular demand, i.e. varied from daily to weekly. Whenever potentially critical situations were expected A4ALPS was consulted. From the very beginning A4ALPS was supposed to complement instead of substituting the Legacy system. Subsequently, it was considered to be an independent alternative information source for supporting decision-making. The opportunity to mutually validate information provided by A4ALPS and the Legacy system was supposed to decrease the level of uncertainty in decision-making processes.

Even though end-users stressed that the exchange with the partners in charge of system integration aimed at specifying particular needs, e.g. setting site specific thresholds, in the course of the development process was challenging, in the end A4ALPS offered what was expected. Especially the three products developed by Swiss developers met their requirements and worked well. As A4ALPS was developed and

[18] For more details see Zardari et al. (2015).

implemented as a complimentary tool no fundamental changes of established working routines took place.

User highlight the following main advantages of having A4ALPS implemented at their site:

- Availability of more meaningful information on individual hazard types for situation assessments
- Improved confidence in the data available for decision-making due to additional information source
- Exceptionally good performance and decision-support during winter season
- Downscaling of precipitation forecasts, Warning system of heavy rainfall and Precipitation type-nowcast most relevant and most frequently used capabilities

The main drawbacks mentioned include:

- Some reliability, stability and scale issues hampered the use of the platform during the first half of the demonstration period, which were later solved.
- Most of the European models included in A4EU are due to their coarse spatial resolution of very limited use.

3.1.2 Usefulness of relevant features

In the context of the needs analysis the improvement respectively new development of the following features for particular phases of the emergency management process were requested:[19]

- Forecasting capabilities (disposition warning phase)
- Precipitation forecasts for local level (disposition warning phase)
- Disposition warning capabilities for the active permafrost layer (disposition warning phase)
- Landslide warning capabilities (disposition warning phase)
- Nowcast information for smaller catchments (alert phase)
- Impact assessment capabilities (alert phase)

At the end of the demonstration period a coordinator as well as an operator who used A4ALPS on a regular basis rated the usefulness of the emergency system's capabilities with regard to the relevant features. They did so for two management scenarios, one including A4ALPS and another one excluding A4ALPS. For simplicity reasons, the former scenario is termed "A4ALPS" and the latter scenario is called "Legacy system". Experts were explicitly asked to make their judgements on the basis of their hands-on experiences and not based on their assumptions about the unfolding of A4ALPS' potential in the future. Figure 4 illustrates the results of this assessment

[19] For improving the readability of the figures documenting the multi-criteria analysis the following abbreviation were used for the different phases of the emergency management process: disposition warning phase – DWP, hazard warning phase – HWP, alert phase – AP. For details about the phases see ANYWHERE project Deliverable 1.2: Report on needs and requirements from the users.

based on the mean values of criteria scores provided by the operator and the coordinator. The distance between the blue line ("Legacy system") and the red line ("A4ALPS") represents the improvement of the perceived usefulness of the respective features for the management process. So far, due to a lack of critical incidences this assessments primarily related to the management of natural hazards under ordinary conditions.

The most remarkable improvements of the Legacy system's comparatively low performance levels were stated for the *Disposition warning capabilities for the active permafrost layer* and the *Landslide warning capabilities*. The *Usefulness of the nowcast information available for smaller catchments* and *of the impact assessment capabilities* improved moderately. The enhancements of the *Usefulness of forecasting capabilities* as well as the p*recipitation forecasts for local level* were considered to be modest (see Figure 4).

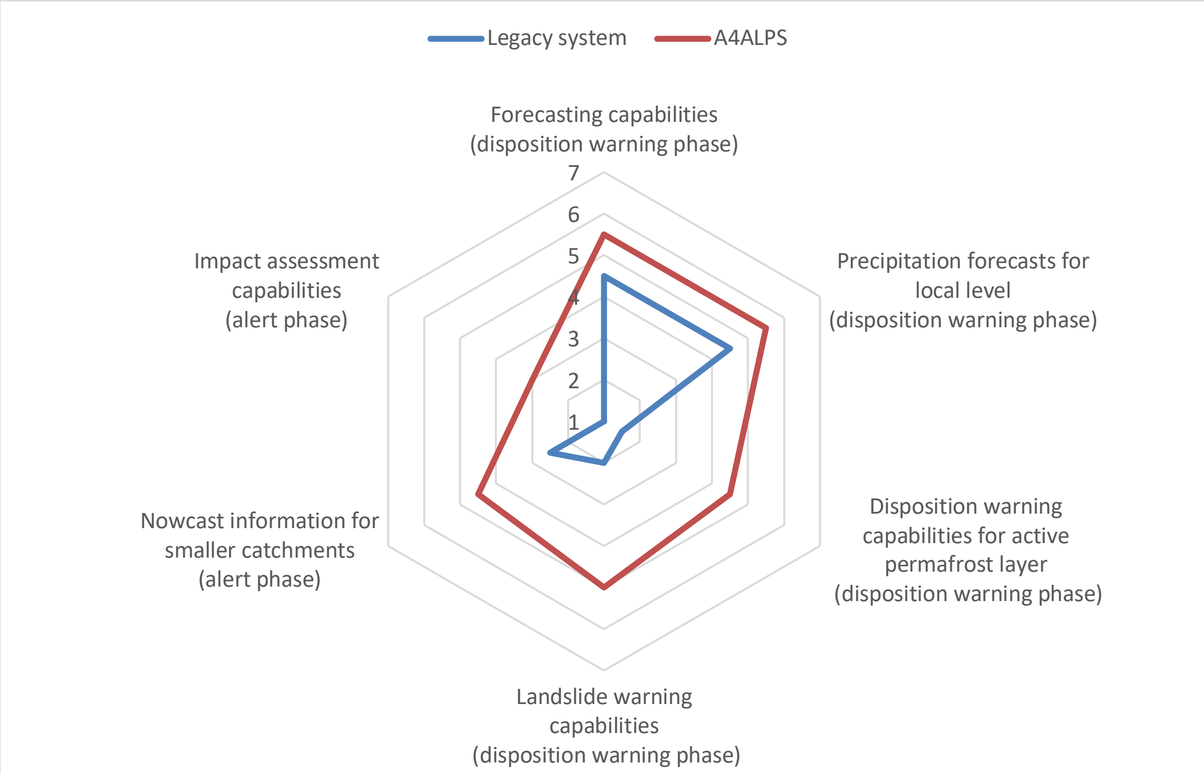


Figure 4: A4ALPS - Usefulness of relevant features (mean values)

3.1.3 Social and technical aspects

Social aspects

The following figures visualise the comparison of A4ALPS and the Legacy system for the two main dimensions of social impacts, i.e. working routines and knowledge production.[20] The comparison is based on an analysis of the mean values of the

[20] Those criteria, which do not feature the 7-point Likert scale but different metrics, e.g. such as e.g. *Time efficiency* were transformed in the following way to match the scale.

respective criteria scores provided by operator and coordinator for a bundle of sub-criteria.
Working routine-related criteria include the following criteria categories *Time efficiency*[21], *Usefulness of features*[22], *Support of decision-making*[23] and *Transparency of decision-making*[24]. Figure 5 shows that the largest improvement was made with regard to the usefulness of the system capabilities which were identified as weak points of the Legacy system and for which improvements were requested for A4ALPS. The performance of the two systems with regard to *Transparency of decision-making* and *Support decision-making* remained unchanged. Figure 5 presents the results.[25]
The inferior performance of A4ALPS for *Time efficiency* is closely related to the fact that A4ALPS is in the short term meant to complement rather than substitute the Legacy system. Therefore, the work flow including A4ALPS is less time efficient that the alternative scenario without operating the ANYWHERE platform. The extent of the decrease in performance is at least partly linked to the normalization method described in the footnote below.

Results for the four sub-categories of knowledge production-related criteria, i.e. *Data inputs*, *Data processing*, *System outputs* and *Trust,* are presented in Figure 6.[26] There are minimal advantages of A4ALPS with regard to *Data inputs* for decision-making[27] as well as notable disadvantages regarding the *System outputs*[28], i.e. information for particular risks, which can be assessed in different ways by the two systems.[29] The

---

Highest/lowest criteria score for each alternative - no matter whether stated by the operator or the coordinator - was set as maximum/minimum value. Average scores of each alternative were then transformed to fit the scale. Therefore, for those criteria absolute values cannot be interpreted in the same way as for the criteria using the 7-point Likert scale. Nevertheless, the visual representation is considered to be suitable for interpreting the differences between the systems.

[21] *Time efficiency* covers the time needed to get an overview of all relevant data sources.

[22] *Usefulness of features* is based on the usefulness scores of the 6 features requested for A4ALPS which were discussed above.

[23] *Support of decision-making* includes the three criteria covering the assessment of the decision support provided by the systems for the distinct emergency management phases.

[24] *Transparency of decision-making* is based on criteria *Transparency of decision-making for insider* and *Transparency of decision-making for outsiders.*

[25] Differences between the mean values of the scores for A4ALPS and Legacy system: *Time efficiency* -3.0, *Transparency of decision-making* 0, *Support of decision-making* +0.08, *Usefulness* +2.0. Positive/negative scores indicate a superior/inferior performance of A4ALPS.

[26] Differences between the mean values of the scores for A4ALPS and Legacy system: *System outputs* -0.88, *Trust* -0.15, *Data processing* 0, *Data inputs* +0.08. Positive/negative scores indicate a superior/inferior performance of A4ALPS.

[27] *Data input* includes the quantity and quality of data inputs for decision-making for the three emergency management phases.

[28] System outputs covers all risk-related criteria, i.e. *Information avalanche, snow gliding risk, Information rockfall, rockslide, debris avalanche, icefall risk, Information landslide risk, Information forest fire risk.*

[29] The lower performance of A4LPS is primarily a consequence of the operator's more critical assessments for the systems capability of providing information for the risk of forest fires and avalanches/snow gliding.

slightly lower mean value for the criterion *Trust* is little surprising as building trust in such a vital management tool takes time. No performance difference exists with regard to *Data processing*.[30]

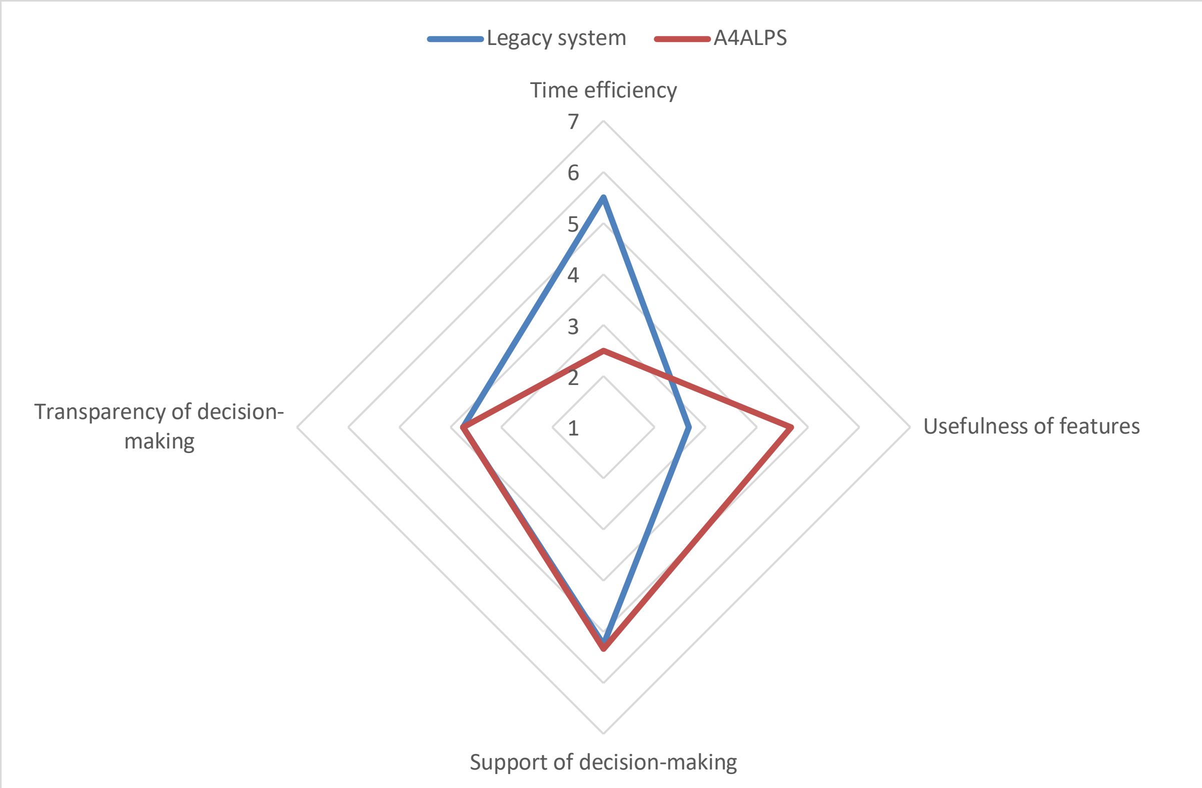


Figure 5: A4ALPS - Working routine-related criteria (mean values)

[30] *Data processing* is based on the criterion *Ease of deciding about issuing disposition warning.*

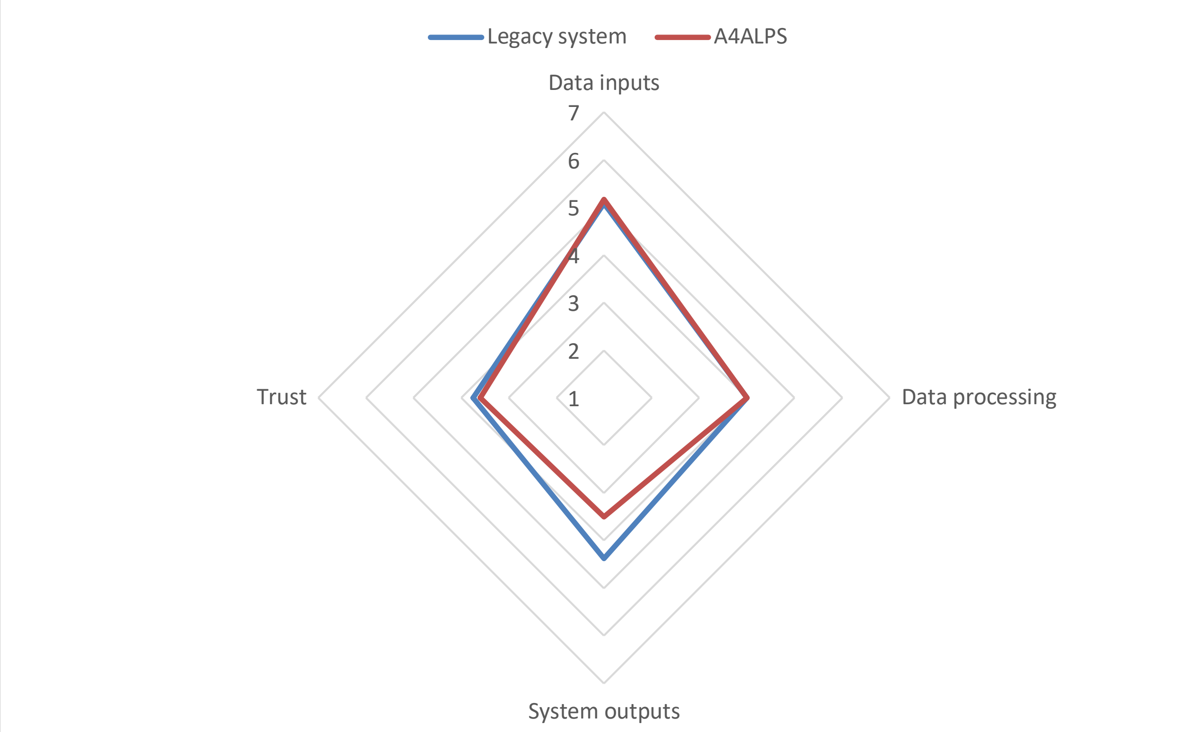


Figure 6: A4ALPS - Knowledge production-related criteria (mean values)

Technical aspects

The analysis of the mean value scores of the technical criteria unveiled that for the sub-categories *Flexibility*[31], *Learnability*[32] and *User interface aesthetics*[33] the Legacy systems slightly outperforms A4ALPS. For the sub-category *Functional appropriateness*[34] both systems perform equally well. Figure 7 depicts the results.[35]

31 Degree to which users with different levels of proficiency can operate the system.
32 Degree to which the user interface is self-explanatory and the user guidance is complete.
33 Degree to which user interfaces and the overall design are aesthetically pleasing in appearance.
34 Degree to which number of functions provided by the system is adequate for users to achieve their objectives appropriately.
35 Differences between the mean values of the scores for A4ALPS and Legacy system: *Flexibility* -0.50, *Learnability* -0.50, *User interface aesthetics* -0.50, *Functional appropriateness* 0. Positive/negative scores indicate a superior/inferior performance of A4ALPS.

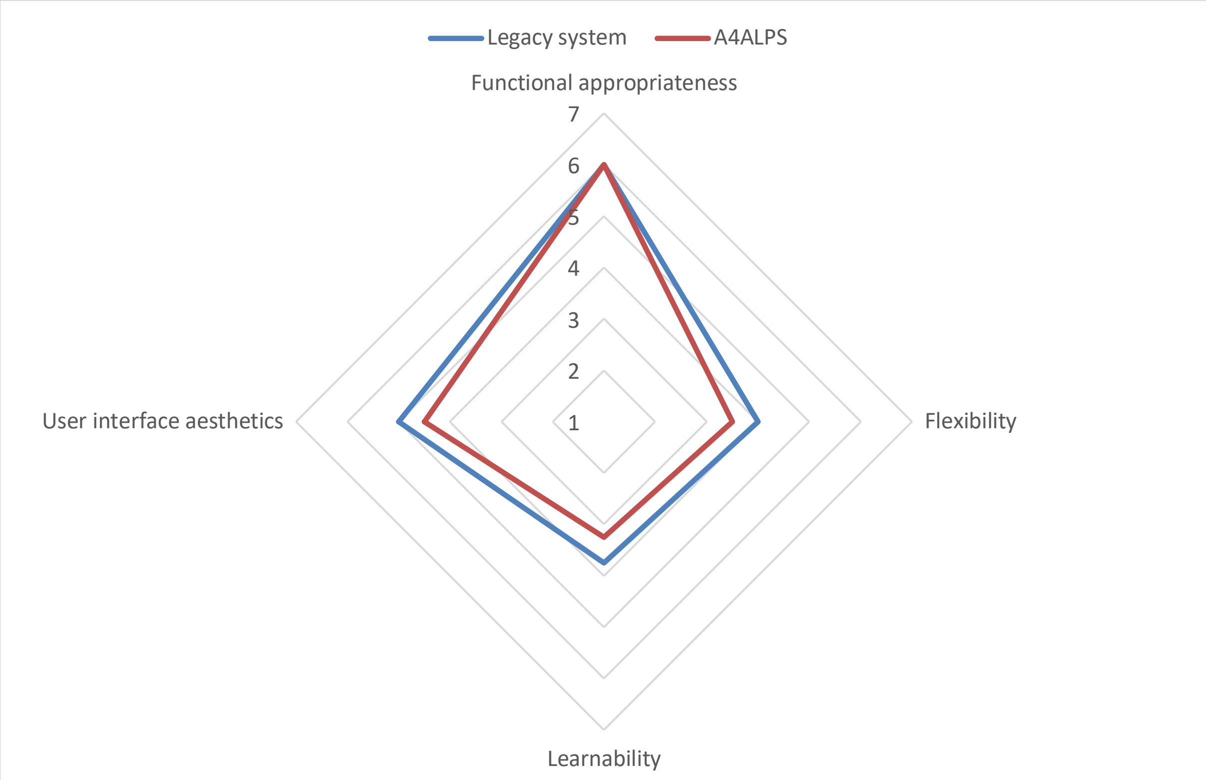


Figure 7: A4ALPS - Technical criteria (mean values)

3.1.4 Results of the Multi-criteria analysis

In contrast to the analysis of the mean values of specific criteria sub-categories the Multi-criteria analysis aims to compare the performance of A4ALPS and the Legacy system in a comprehensive manner. Hence, the data for all criteria selected by the end-users as being relevant for assessing the performance of an emergency management system and for which data could be collected was considered as described in detail in section 2.3.

As not all performance aspects to be considered for the analysis are equally important the operator and the coordinator elicited weights to all criteria. Overall the most important aspects were the *Trust-related criteria* and the criterion *Service availability.* The criteria considered to be least relevant were *Usefulness of impact estimations (alert phase)*, *Ease of deciding about issuing disposition warning* and *Transparency of decision-making for outsiders*.

There were some differences between the perspectives of the operator and coordinator. The importance of *information related to risks of different types of natural hazards*, *Functional appropriateness* and *Quantity* as well as *Quality of data inputs for decision-making in the disposition warning phase* were rated substantially higher by the coordinator. The operator gave higher weight to *Decision-support in the disposition warning and alert phase*, *Quality of data inputs in the hazard warning and alert phase*, *Flexibility* and *Learnability*. For an overview of all criteria weights see Figure 45 in Annex 7.7

The overall result of the probabilistic Multi-criteria analysis substantiated that A4ALPS has a slight advantage over the Legacy system by outperforming it with a probability of 55.22% to 44.78%. This means that when considering both weighting sets and the performance dataset out of the 10.000 single Multi-criteria analyses conducted A4ALPS outperformed the Legacy system 5,522 times.

A breakdown of the co-evaluation results by criteria reveals that the superior performance of A4ALPS is primarily driven by the usefulness of the improved or newly developed features (for details see Figure 8). Hence, A4ALPS has the most profound advantage over the Legacy system with regard to the criteria *Usefulness of landslide warning capabilities (disposition warning phase)*, *Usefulness of disposition warning capabilities for the active permafrost layer (disposition warning phase)* and *Usefulness of nowcast information for smaller catchments (alert phase).*

For now the Legacy systems still outperforms the current A4ALPS prototype in several other aspects, the most relevant being *Supported language adequacy*, *Time efficiency - Overview of all relevant data sources (disposition warning phase)* and *Overall satisfaction with management system.* The lagging behind of A4ALPS in these regards can easily be explained. From the beginning the end-users in Switzerland expressed comparatively high levels of satisfaction with the Legacy system and asked for add-ons to complement their management system. It takes time to adjust prototypes, build trust and, eventually, reach the levels of satisfaction, which are comparable to the one of the Legacy system. The fact that for now A4ALPS does not substitute parts of the Legacy system but complements the management system as an alternative source of information leads to its lower time-efficiency. The Legacy system outperforms A4ALPS with regard to the *Language support adequacy* for the simple reason that the current version of the A4ALPS prototype is only available in English and not in any other of the official Swiss languages. This limitation was agreed upon with the end-users for the prototype phase at the beginning of the platform development process.

No substantial performance differences exist with regard to the criteria categories Decision-support, Data quality and quantity, Risk and impact information.

Figure 8 presents the net preference flows of the two systems for each criterion. The higher/lower the respective bar above/below the zero line the better/worse the performance of the respective system compared to the alternative in this aspect.

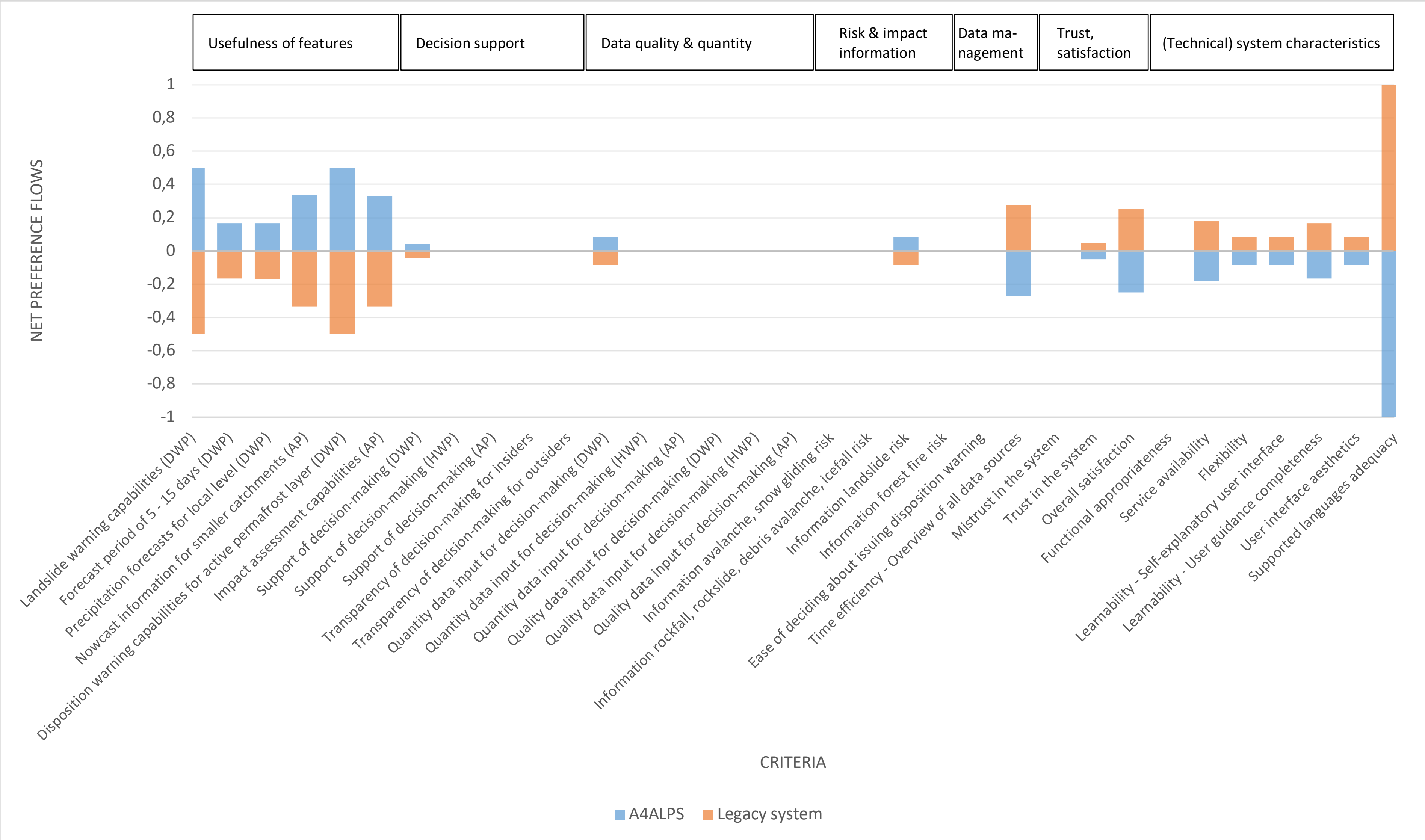


Figure 8: Performance comparison A4ALPS and Legacy system

3.2 A4CAT

3.2.1 General impressions

A4CAT was used through the demonstration phase all the time. It ran in parallel to the Legacy system, and it is fully operational at the moment and used on a daily basis. The operators consulted A4CAT particularly during ongoing rain events

The first impression of A4CAT, was surprisingly positive, and this particularly with respect to improved forecasts with respect to rain and floods. According to the users, it was what we envisioned to be achieved through A4CAT and it was a solution for many of the Legacy system's drawbacks.

Also the first experiences were rather positive, while it became quickly apparent, that a lot of things still needed to be developed. The flood-related capabilities were already really good from the beginning; also applied in practise. The tool was perceived as being easy to use, to be powerful but simple. However, there were also aspects that needed to be improved. For flash floods and accumulation in rivers the new system was working really well from the beginning, but there were some problems with respect to the accumulation in metropolitan zones.

The implementation of the ANYWHERE system had a strong impact on the working routines. Before the instalment of A4CAT, staff was waiting for the calls reporting the impact and then started to monitor the situation and checked for information. This means, alerts can be made public before the actual impact occurs. Now authorities can send information about impacts in advance, for example they can warn the municipalities before heavy rain occurs in the field, because A4CAT shows them the areas, where problems could occur. It is reported that behaviour changed from reactive to proactive.

User highlight the following main advantages of having A4CAT implemented at their site:

- Data integration, i.e. that all relevant is available through one platform.
- High resolution information for small rivers.
- Lead time improved substantially. Impacts are projected reliably 1-2 hours before they occur.
- ANYWHERE system is of practical use.
- Foreseen integration of the information of the Catalonian water agency into A4CAT in form of a map, instead of getting informed by email.

The main drawbacks mentioned include:

- System does not perform equally well for all hazard types. It is less powerful for winds and storms.
- Level of integration with other systems is currently still too low. Feedback from all levels of public administration should be integrated, e.g. from municipal civil protection authorities.

3.2.2 Usefulness of relevant features

In the context of the needs analysis the improvement respectively new development of the following features for particular phases of the emergency management process were requested:[36]

- Rainfall predictions (monitoring phase)
- Rainfall predictions (pre-warning phase)
- Rainfall-related run-off prediction (monitoring phase)
- Run-off predictions (pre-warning phase)
- Flood forecasts (pre-warning phase)
- Real time information about flood-related impacts (warning phase)
- Data integration capabilities (warning phase)
- Creation of risk maps (pre-warning phase)

At the end of the demonstration period a coordinator as well as an operator who used A4CAT on a regular basis rated the usefulness of the emergency system's capabilities with regard to the relevant features. They did so for two management scenarios, one including A4CAT and another one excluding A4CAT. For simplicity reasons, the former scenario is termed "A4CAT" and the latter scenario is called "Legacy system". Experts were explicitly asked to make their judgements on the basis of their hands-on experiences and not based on their assumptions about the unfolding of A4CAT's potential in the future. Figure 9 illustrates the results of this assessment based on the mean values of criteria scores provided by the operator and the coordinator. The distance between the blue line ("Legacy system") and the red line ("A4CAT") represents the improvement of the perceived usefulness of the respective features for the management process.

Figure 9 gives an overview of how the operator and coordinator comparatively assessed the usefulness of particular features for the Legacy system as well as for A4CAT. It is evident that A4CAT outperforms the Legacy system with regard to all features except the *Usefulness of creation of risk maps*, which continues to be identical in both scenarios. Improvements are substantial across the board being most remarkable for *Usefulness of data integration capabilities*, *Usefulness of real time information about flood-related impacts (warning phase)* and *Usefulness of flood forecasts (pre-warning phase)*.[37]

---

[36] For improving the readability of the figures documenting the Multi-criteria analysis the following abbreviation were used for the different phases of the emergency management process: monitoring phase – MP, pre-warning phase – PWP, warning phase – WP. For details about the phases see ANYWHERE project Deliverable 1.2: Report on needs and requirements from the users.

[37] Differences between the mean values of the scores for A4CAT and Legacy system: *Visualisation of risk levels* +1.0, *Access to reports of past events* +1.0, *Decision support by on-call engineer* +1.0, *Flood forecasts* +1.5, *Data integration capabilities* +1.5, *Rainfall predictions* +2.0, *Run-off predictions* +2.0, *Overview of current & expected risks* +2.0, *Transmission of mobilization levels & recommendations to municipalities* +2.0, *Real time information about flood-related impacts* +2.50, *Field feedback capabilities* +3.0, *Overview of warning states of all*

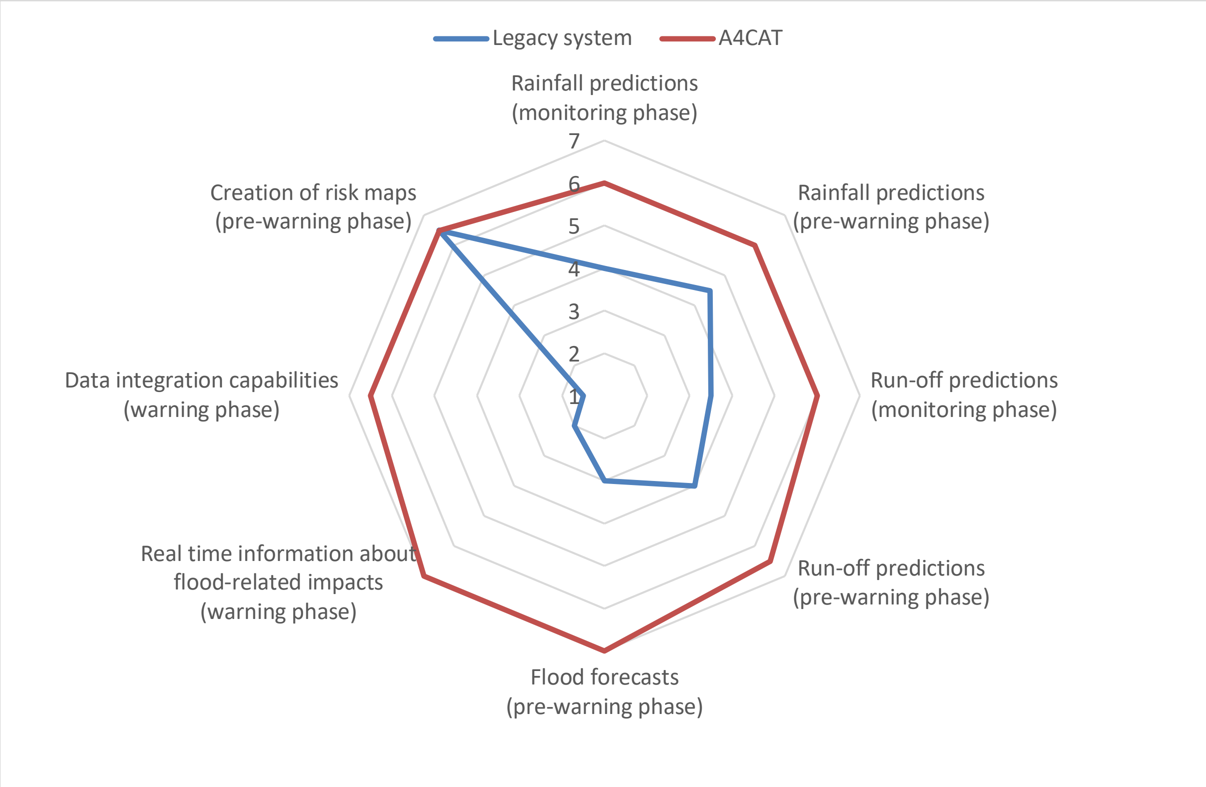


Figure 9: A4CAT - Usefulness of relevant features (mean values)

3.2.3 Social and technical aspects

Social aspects

The following figures visualise the comparison of A4CAT and the Legacy system for the two main dimensions of social impacts, i.e. working routines and knowledge production.[38] The comparison is based on an analysis of the mean values of the respective criteria scores provided by operator and coordinator for a bundle of sub-criteria.

---

municipalities +4.0, *Access to local safeguard plans and vulnerability information* +4.5. Positive/negative scores indicate a superior/inferior performance of A4CAT.

[38] Those criteria, which do not feature the 7-point Likert scale but different metrics, e.g. such as e.g. *Time efficiency* were transformed in the following way to match the scale. Highest/lowest criteria score for each alternative - no matter whether stated by the operator or the coordinator - was set as maximum/minimum value. Average scores of each alternative were then transformed to fit the scale. Therefore, for those criteria absolute values cannot be interpreted in the same way as for the criteria using the 7-point Likert scale. Nevertheless, the visual representation is considered to be suitable for interpreting the differences between the systems.

Working routine-related criteria include the following categories *Time efficiency*[39], *Stress*[40], *Usefulness of features*[41], *Support of decision-making*[42] and *Transparency of decision-making*[43]. The performance comparison of A4CAT and Legacy system for these criteria provides evidence for the immense improvements witnessed after the implementation of A4CAT (see Figure 10).[44] Enhancements are not only linked to the *Usefulness* of the relevant features but also relate to *Time efficiency*, *Stress* and *Support of decision-making*. The inferior performance with regard to Transparency of the decision-making is linked to the fact that insiders, i.e. personnel at the regional civil protection authority CECAT, as well as outsiders, i.e. people outside of CECAT such as citizens or politicians, are not familiar with the system, yet. Hence, the traceability and understanding of decisions taken with the support of the Legacy system are still higher.

Results for the four sub-categories of knowledge production-related criteria, i.e. *Data inputs*, *Data processing*, *System outputs* and *Trust,* are presented in Figure 11.[45] Trust remains on the same level. There are notable advantages of A4CAT with regard to *Data inputs* for decision-making[46] as well as for *System outputs*, i.e. information for particular risks, which can be assessed in different ways by the two systems.[47] Massive improvements can be depicted for *Data processing.*[48]

[39] *Time efficiency* covers the time needed to get an overview of all relevant data sources as well as the time to create risk maps in the pre-warning phase.
[40] Level of stress caused by the operation of the system.
[41] *Usefulness of features* is based on the usefulness scores of the 8 features requested for A4CAT which were discussed above.
[42] *Support of decision-making* includes the three criteria covering the assessment of the decision support provided by the systems for the distinct emergency management phases.
[43] *Transparency of decision-making* is based on criteria *Transparency of decision-making for insider* and *Transparency of decision-making for outsiders.*
[44] Differences between the mean values of the scores for A4CAT and Legacy system: *Transparency of decision-making* -1.5, *Support of decision-making* +2.25, *Usefulness* +2.81, *Time efficiency* +3.15, *Stress* +3.5. Positive/negative scores indicate a superior/inferior performance of A4CAT.
[45] Differences between the mean values of the scores for A4CAT and Legacy system: *Trust* -0.05*, System outputs* +1.19, *Data inputs* +1.75, *Data processing* +3.13. Positive/negative scores indicate a superior/inferior performance of A4CAT.
[46] *Data input* includes the quantity and quality of data inputs for decision-making for the three emergency management phases.
[47] *System outputs* covers all risk and impact-related criteria, i.e. *Information about flood risks and impacts, Information about wind risks and impacts, Information about forest fire risks and impacts, Information about snowstorm risks and impacts.*
[48] *Data processing* is based on the criteria *Ease of managing data* sources in the three phases of the emergency management process, *Ease of creating risk maps (pre-warning phase)* and *Ease of information exchange within and with partners outside of CECAT.*

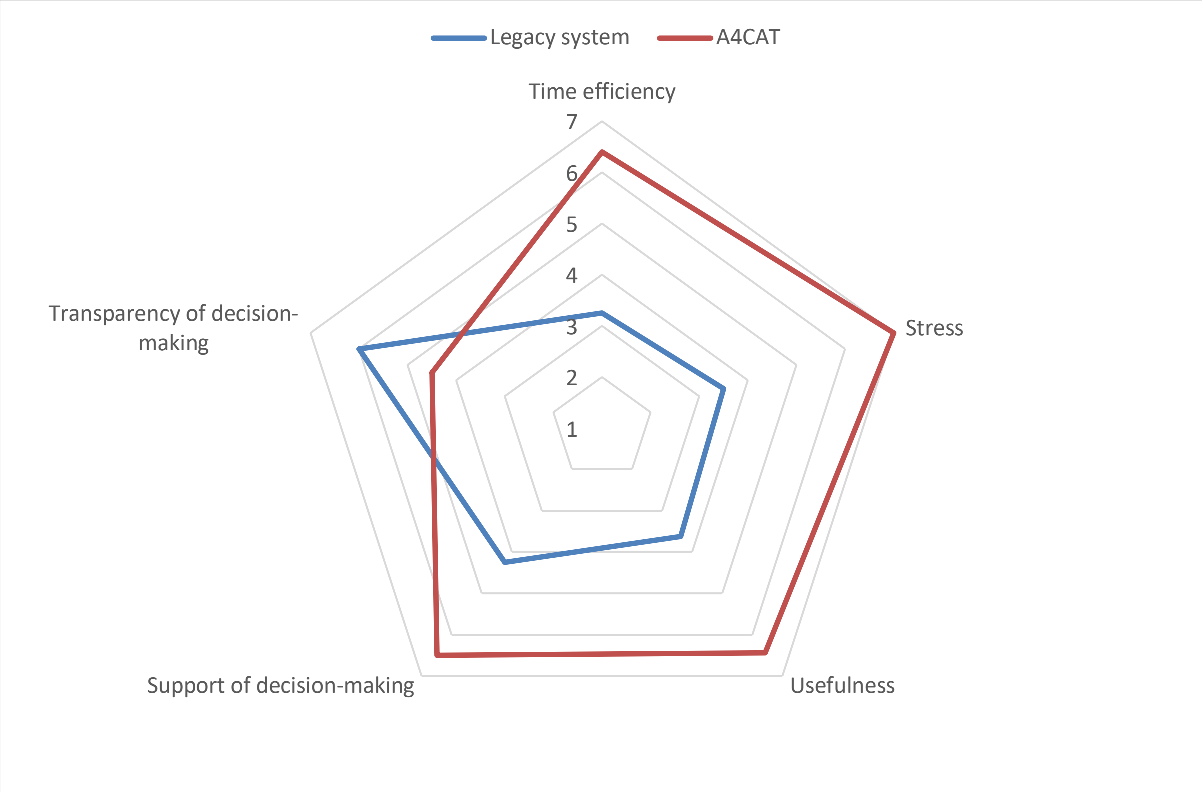


Figure 10: A4CAT - Working routine-related criteria (mean values)

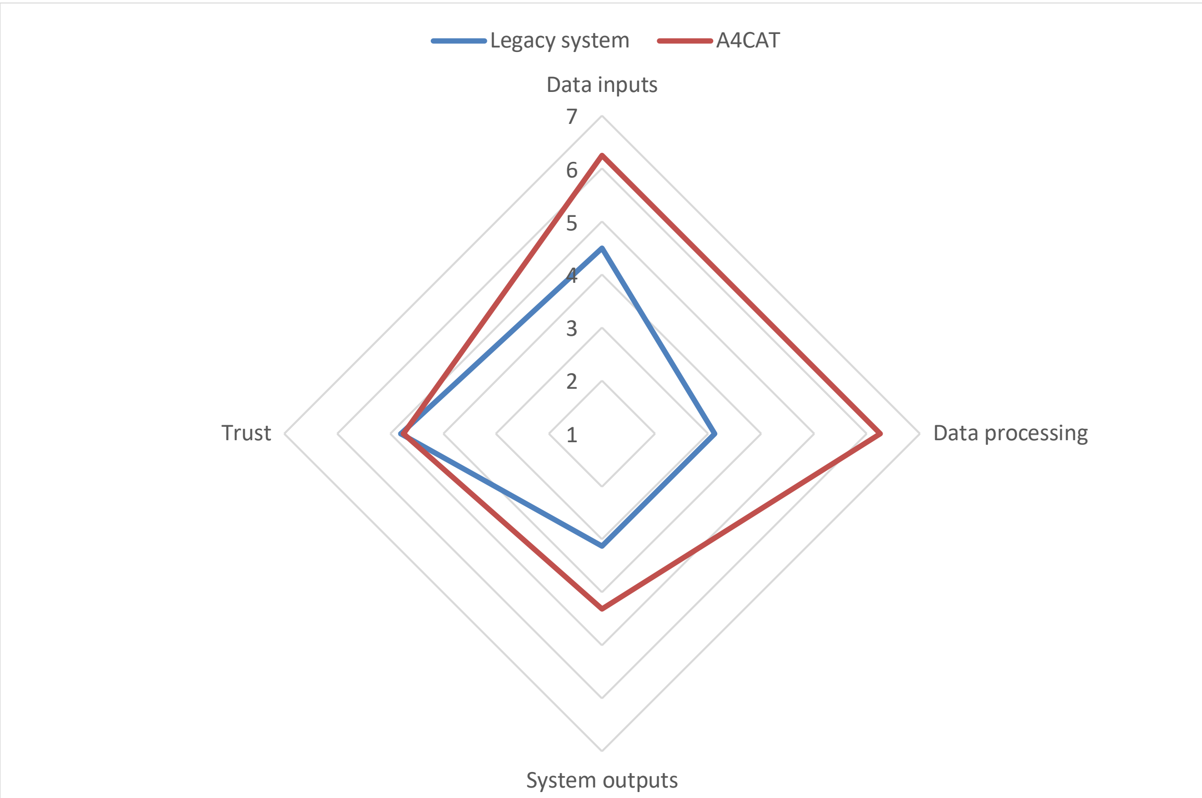


Figure 11: A4CAT - Knowledge production-related criteria (mean values)

Technical aspects

The technical criteria are clustered in the following categories *Hazard coverage*, *Functional appropriateness*[49], *Flexibility*[50], *Learnability*[51], *User interface aesthetics*[52], and *Accessibility*[53]. The analysis of the mean value scores of the technical criteria unveiled that for the sub-categories *Flexibility*, *Learnability*, *Accessibility, User interface aesthetics* and *Functional appropriateness* A4CAT performs notably better than the Legacy system. Due to the fact that the Legacy system enables the management of more types of natural hazards than A4CAT it outperforms the ANYWHERE platform in this regard, i.e. *Hazard coverage*. Figure 12 depicts the results.[54]

[49] Degree to which number of functions provided by the system is adequate for users to achieve their objectives appropriately.
[50] Degree to which users with different levels of proficiency can operate the system.
[51] Degree to which the user interface is self-explanatory and the user guidance is complete.
[52] Degree to which user interfaces and the overall design are aesthetically pleasing in appearance.
[53] *Accessibility* includes *Accessibility for users with visual impairment*, i.e. degree to which users with visual impairment can successfully operate the system, and *Supported language adequacy,* i.e. ratio of languages supported by the system to languages, in which system should be available for being operated by any given (potential) user at the civil protection authority.
[54] Differences between the mean values of the scores for A4CAT and Legacy system: *Hazard coverage* -1.0*, Functional appropriateness +*0.50*, User interface aesthetics +*0.50*, Flexibility +*1.0, *Learnability +*1.0, *Accessibility +*1.0. Positive/negative scores indicate a superior/inferior performance of A4CAT.

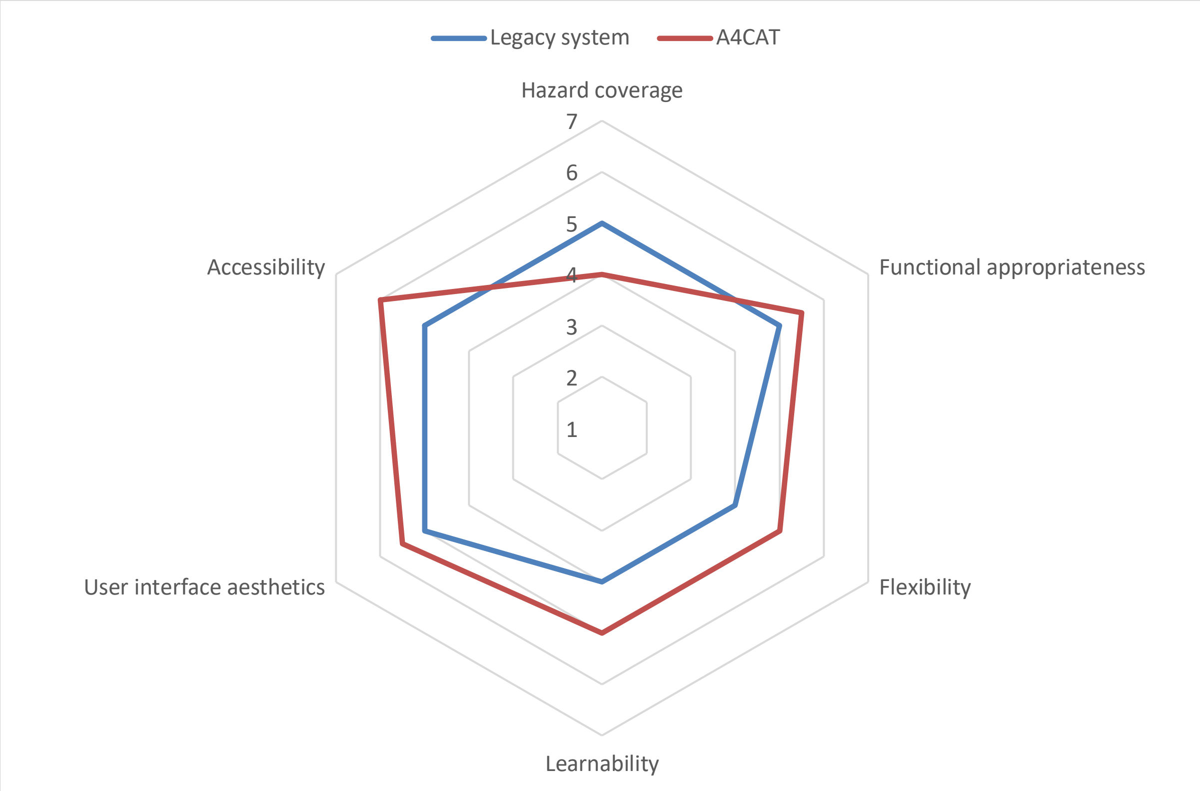


Figure 12: A4CAT - Technical criteria (mean values)

3.2.4 Results of the Multi-criteria analysis

In contrast to the analysis of the mean values of specific criteria sub-categories the Multi-criteria analysis aims to compare the performance of A4CAT and the Legacy system in a comprehensive manner. Hence, the data for all criteria selected by the end-users as being relevant for assessing the performance of an emergency management system and for which data could be collected was considered as described in detail in section 2.3.

As not all performance aspects to be considered for the analysis are equally important the operator and the coordinator elicited weights to all criteria. Overall the most important aspects were *Service availability*, *Trust*-related criteria, *Knowledge about different types of impacts (floods, winds, forest fires, snowstorms).* The criteria considered to be least relevant were *Supported language adequacy, Accessibility for users with visual impairment, Ease of managing data sources (monitoring phase), Transparency of the decision-making process for insiders.*

There were some differences between the perspectives of the operator and coordinator. The importance of *Usefulness creation risk maps (pre-warning phase)*, *Quality data input for decision-making* and *Support of decision-making (warning phase)* were rated substantially higher by the coordinator. The operator gave higher weight to *Quantity data input for decision making, User interface aesthetics, Learnability - Self-explanatory user interface* and *Usefulness rainfall predictions*. For an overview of all criteria weights see Figure 49 in Annex 7.11

The overall result of the probabilistic Multi-criteria analysis is that A4CAT outperforms the Legacy system with a probability of 100%. This means that when considering both weighting sets and the performance dataset out of the 10.000 single Multi-criteria analyses conducted A4CAT outperformed the Legacy system every time.

A breakdown of the co-evaluation results by criteria reveals that the superior performance of ACAT is primarily driven by *Data management & processing* and *Usefulness of the features* of the emergency management system (see Figure 13). A4CAT has the most profound advantage over the Legacy system with regard to the criteria *Time efficiency - Overview all data sources (warning phase)*, *Usefulness integration data sources (warning phase)*, *Usefulness provision of flood-related impacts in real time (warning phase)* and *Ease of managing data sources (warning phase)*.

For the reasons mentioned in section 3.2.3 the Legacy system has a slight advantage over A4CAT regarding the *Transparency of the decision-making processes for insiders and outsiders*, *Information about wind impacts* and *Hazard coverage*. No substantial performance differences exist with regard to the criteria *Quantity data input for decision-making (pre-warning phase)*, *Ease of creating risk maps (pre-warning phase)*, *Service availability*, *Supported languages adequacy*, *Overall satisfaction* and *Trust*.

For details see Figure 13, which presents the net preference flows of the two systems for each criterion. The higher/lower the respective bar above/below the zero line the better/worse the performance of the respective system compared to the alternative in this aspect.

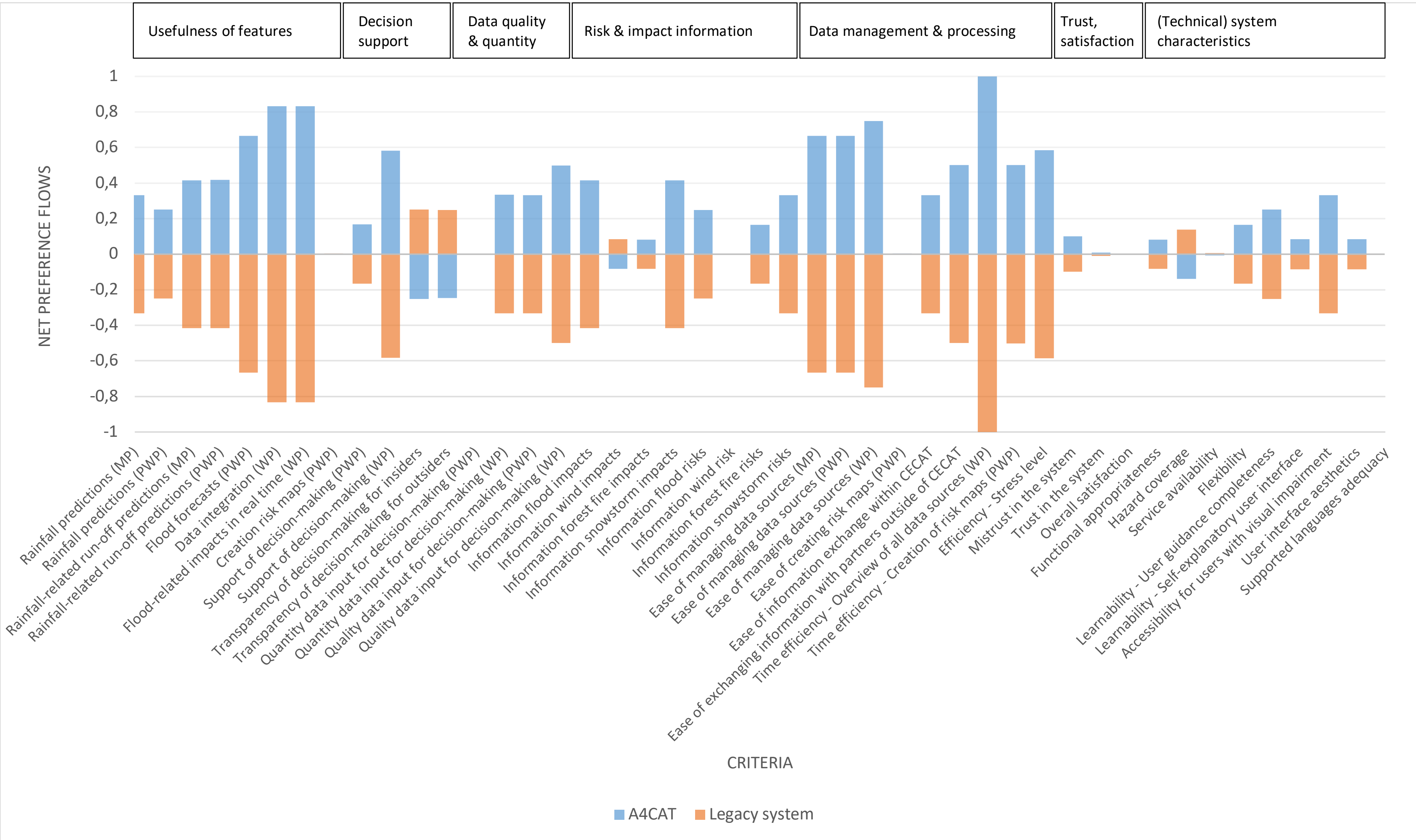


Figure 13: Performance comparison A4CAT and Legacy system

3.3 A4LIG

3.3.1 General impressions

A4LIG was used through the demonstration phase particularly with respect to nowcasting of heavy rainfall events. Furthermore, operator tested A4LIG alongside the Legacy system during two fire events (in operation). For other types of hazards they tested the system using historical events data (for training).

The first impression of A4LIG was positive especially because of the monitoring tools and nowcasting tools included.

Also the first experiences were rather positive. It is a powerful tool for monitoring (heavy) rainfall events. However, the nowcasting tool was perceived initially as less useful for fast responding catchments in Genova (because the delay time is very short).

User highlighted the following main advantages of having A4LIG implemented at their site:

- Provision of very quick overview of relevant data
- Good decision-making support, e.g. for closing roads
- Higher potential for further improvement e.g. by offering access to information for historical events

The main drawbacks mentioned included:

- Low capability of offering support for fast-responding catchments (floods)
- Fire simulation capabilities could be further improved, e.g. improvement of consideration of wind directions

3.3.2 Usefulness of relevant features

In the context of the needs analysis the improvement respectively new development of the following features for particular phases of the emergency management process were requested:[55]

- Impact assessment capabilities (monitoring phase)
- Flood risk information (alert phase)
- Flood impact information (alert phase
- Forest fire risk information (alert phase)
- Forest fire impact information (alert phase)
- Data collection capabilities (alert phase)

[55] For improving the readability of the figures documenting the multi-criteria analysis the following abbreviation were used for the different phases of the emergency management process: monitoring phase – MP, pre-alert phase – PAP, alert phase – AP. For details about the phases see ANYWHERE project Deliverable 1.2: Report on needs and requirements from the users.

- Data integration capabilities (alert phase)

At the end of the demonstration period a coordinator as well as an operator who occasionally used A4LIG rated the usefulness of the emergency system's capabilities with regard to the relevant features. They did so for two management scenarios, one including A4LIG and another one excluding A4LIG For simplicity reasons, the former scenario is termed "A4LIG" and the latter scenario is called "Legacy system". Experts were explicitly asked to make their judgements on the basis of their hands-on experiences and not based on their assumptions about the unfolding of A4LIG's potential in the future. Figure 14 illustrates the results of this assessment based on the mean values of the criteria scores provided by the operator and the coordinator. The distance between the blue line ("Legacy system") and the red line ("A4LIG") represents the improvement of the perceived usefulness of the respective features for the management process. So far, due to a lack of critical incidences this assessments primarily related to the management of natural hazards under ordinary conditions.

Figure 14 gives an overview of how the operator and coordinator comparatively assessed the usefulness of particular features for the Legacy system as well as for A4LIG. The most remarkable improvements of the Legacy system's comparatively low performance levels were stated for the usefulness of A4LIG's *Forest fire risk and impact information (alert phase)*. A4LIG's *Data integration and data collection capabilities* are considered to be moderately more useful than the ones of the Legacy system. There are no differences between the usefulness of systems' *Impact assessment capabilities*. The *Flood risk and impact information (alert phase)* provided by the Legacy system is slightly more useful than the information offered by A4LIG.[56]

[56] Differences between the mean values of the scores for A4LIG and Legacy system*: Flood impact information (alert phase)* -1.0, *Flood risk information (alert phase)* -0.50, *Impact assessment capabilities (monitoring phase)* 0, *Data collection capabilities (alert phase)* +0.50, *Data integration capabilities (alert phase)* +1.0, *Forest fire impact information (alert phase)* +3.0, *Forest fire risk information (alert phase)* +4.0. Positive/negative scores indicate a superior/inferior performance of A4LIG.

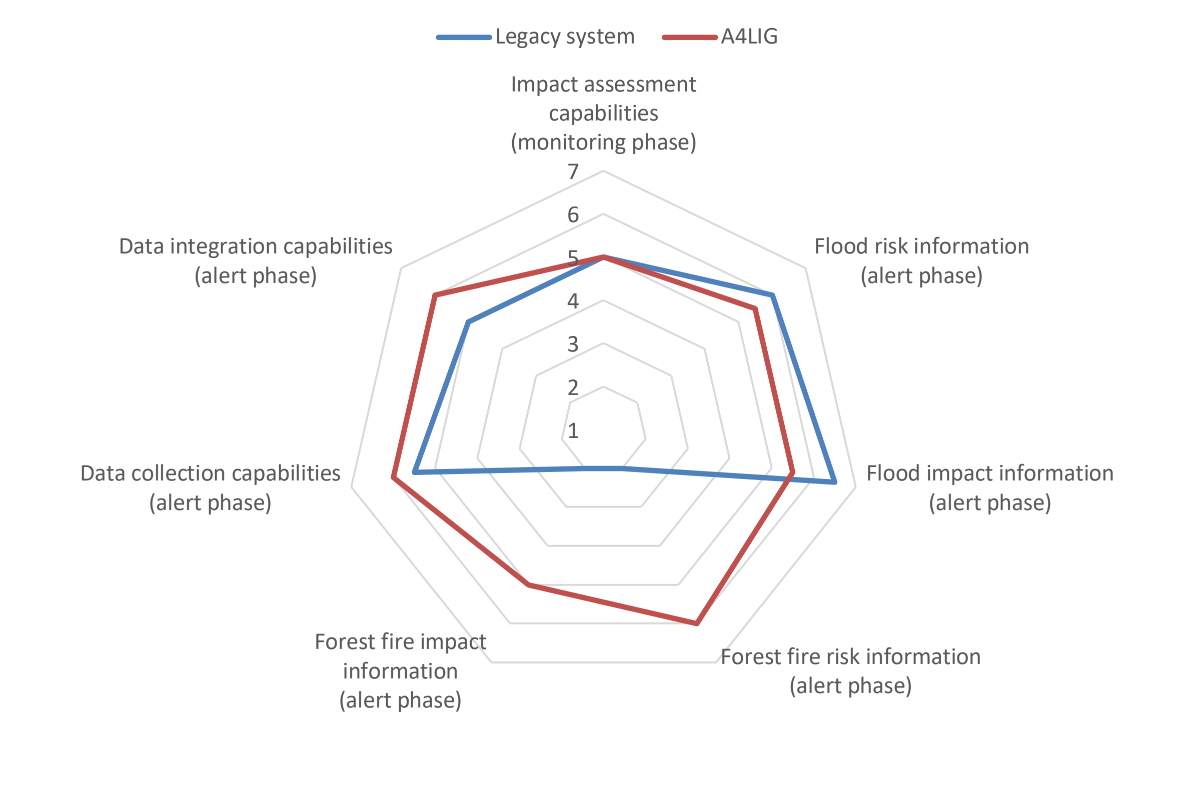


Figure 14: A4LIG - Usefulness of relevant features (mean values)

3.3.3 Social and technical aspects

Social aspects

The following figures visualise the comparison of A4LIG and the Legacy system for the two main dimensions of social impacts, i.e. working routines and knowledge production.[57] The comparison is based on an analysis of the mean values of the respective criteria scores provided by operator and coordinator for a bundle of sub-criteria.

[57] Those criteria, which do not feature the 7-point Likert scale but different metrics, e.g. such as e.g. *Time efficiency* were transformed in the following way to match the scale. Highest/lowest criteria score for each alternative - no matter whether stated by the operator or the coordinator - was set as maximum/minimum value. Average scores of each alternative were then transformed to fit the scale. Therefore, for those criteria absolute values cannot be interpreted in the same way as for the criteria using the 7-point Likert scale. Nevertheless, the visual representation is considered to be suitable for interpreting the differences between the systems.

Working routine-related criteria include the following categories *Time efficiency*[58], *Stress*[59], *Usefulness of features*[60], *Support of decision-making*[61] and *Transparency of decision-making*[62]. Moderate improvements of the criteria *Usefulness* of the relevant features, *Time efficiency* and *Stress* are reported due to the implementation of A4LIG (see Figure 15).[63] Whereas the Legacy system performs much better than A4LIG regarding the criteria *Transparency of decision-making* and *Support of decision-making*. As in the case of A4CAT the inferior performance regarding *Transparency of decision-making* is linked to the fact that at the pilot site neither people working in the field of civil protection nor outsiders such as citizens or politicians are familiar with the system, yet. Hence, the traceability and understanding of decisions taken with the support of the Legacy system are still higher.

Results for the four sub-categories of knowledge production-related criteria, i.e. *Data inputs* for decision-making[64], *Data processing*[65], *System outputs*, i.e. information for particular risks and impacts, which can be assessed in different ways by the two systems[66] and *Trust,* show that A4LIG has only a slight advantage over the Legacy system when it comes to the *System outputs*. Regarding all other categories the Legacy system outperforms A4LIG (see Figure 16).[67] The notably lower mean value for the criterion *Trust* is not surprising. The fact that building trust in such a system takes time is confirmed across all pilot sites.

---

[58] *Time efficiency* covers the time needed to get an overview of all relevant data sources.

[59] Level of stress caused by the operation of the system.

[60] *Usefulness of features* is based on the usefulness scores of the 6 features requested for A4ALIG which were discussed above.

[61] *Support of decision-making* includes the three criteria covering the assessment of the decision support provided by the systems for the distinct emergency management phases.

[62] *Transparency of decision-making* is based on criteria *Transparency of decision-making for insider* and *Transparency of decision-making for outsiders.*

[63] Differences between the mean values of the scores for A4LIG and Legacy system: *Transparency of decision-making* -3.5, *Support of decision-making* -2.0, *Time efficiency* +0.60, *Stress* +1.0, *Usefulness* +1.0. Positive/negative scores indicate a superior/inferior performance of A4LIG.

[64] *Data input* includes the quantity and quality of data inputs for decision-making for the three emergency management phases.

[65] *Data processing* is based on the criteria *Ease of managing data* sources in the three phases of the emergency management process and *Ease of determining adequate operational phase (alert phase).*

[66] *System outputs* covers all risk and impact-related criteria, i.e. *Information about flood risks and impacts, Information about wind risks and impacts, Information about forest fire risks and impacts, Information about snowstorm risks and impacts.*

[67] Differences between the mean values of the scores for A4LIG and Legacy system: *Trust* -1.29, *Data inputs* -1.25, *Data processing* -0.63, *System outputs* +0.80. Positive/negative scores indicate a superior/inferior performance of A4LIG.

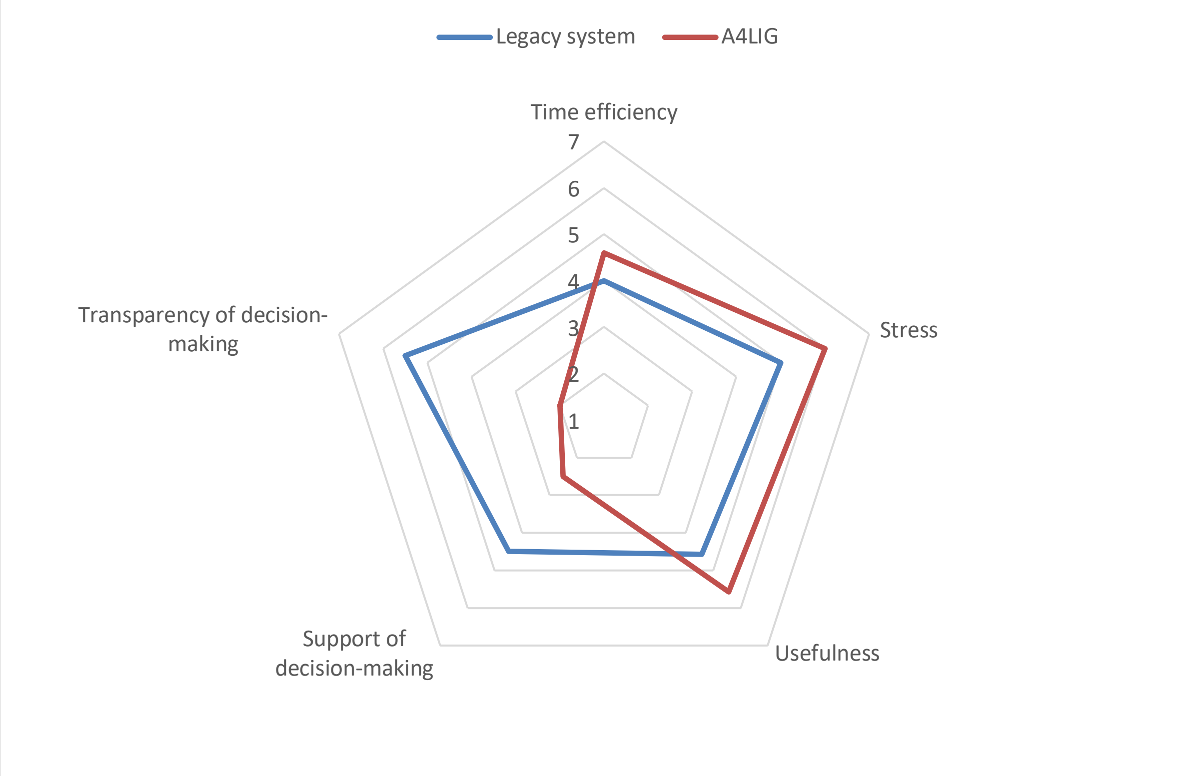


Figure 15: A4LIG - Working routine-related criteria (mean values)

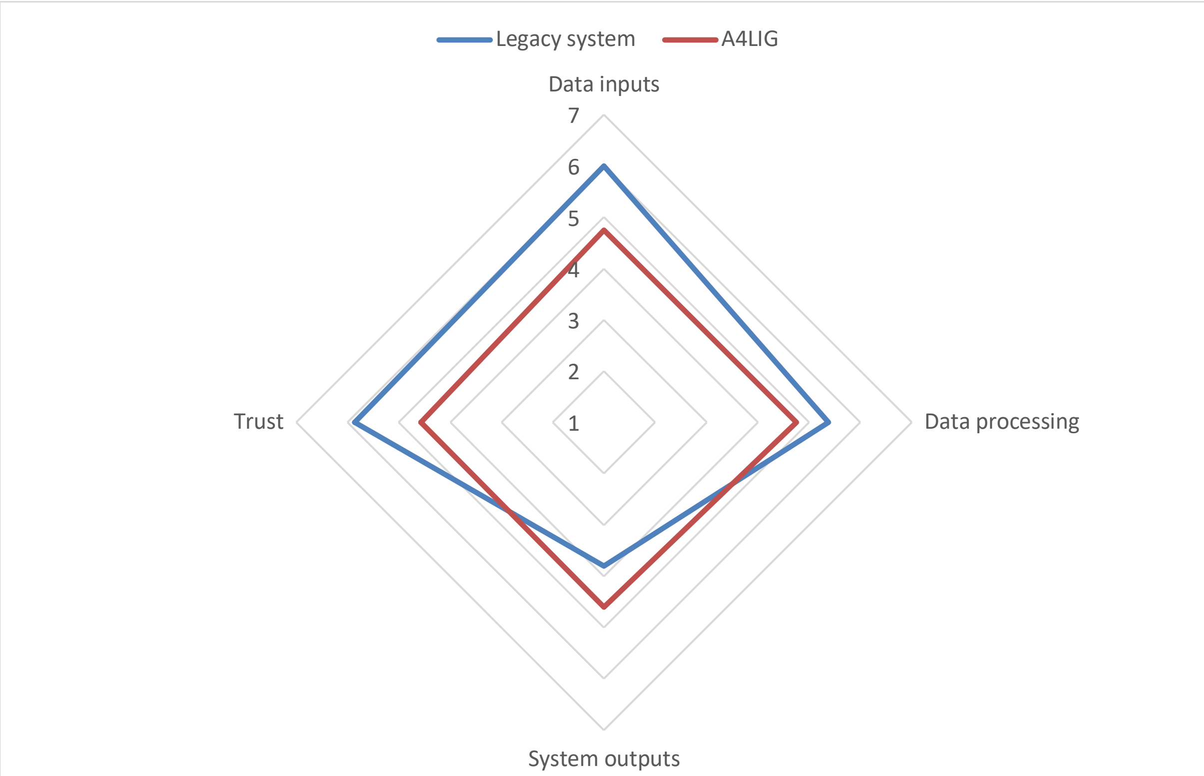


Figure 16: A4LIG - Knowledge production-related criteria (mean values)

Technical aspects

The technical criteria are clustered in the following categories *Hazard coverage*, *Functional appropriateness*[68], *Flexibility*[69], *Learnability*[70], *User interface aesthetics*[71], and *Accessibility*[72]. The analysis of the mean value scores of the technical criteria unveiled that for the sub-categories *User interface aesthetics* and *Accessibility* A4LIG performs notably better and for *Learnability* and *Hazard coverage* still slightly better than the Legacy system. There is no difference regarding the *Flexibility* of the systems and the A4LIG has a notable disadvantage when it comes to the *Functional appropriateness* of the two systems. Figure 17 depicts the results.[73]

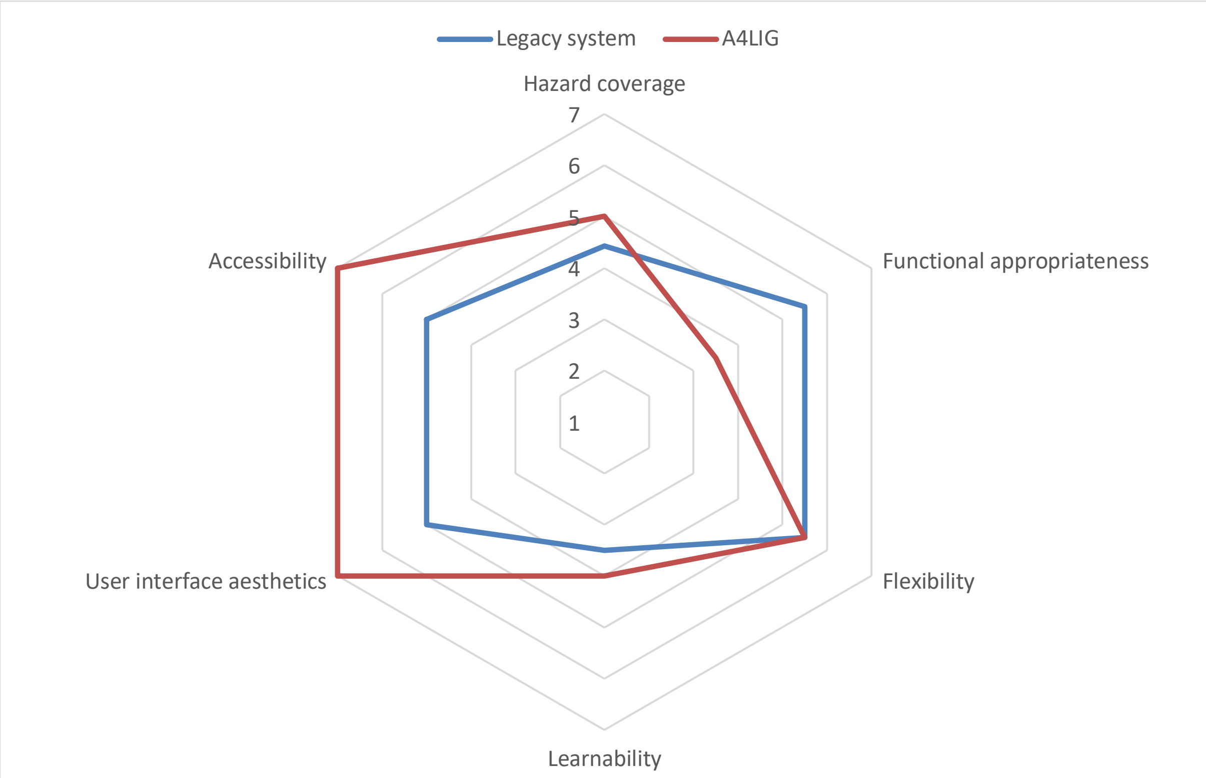


Figure 17: A4LIG - Technical criteria (mean values)

[68] Degree to which number of functions provided by the system is adequate for users to achieve their objectives appropriately.
[69] Degree to which users with different levels of proficiency can operate the system.
[70] Degree to which the user interface is self-explanatory and the user guidance is complete.
[71] Degree to which user interfaces and the overall design are aesthetically pleasing in appearance.
[72] *Accessibility* includes *Accessibility for users with visual impairment*, i.e. degree to which users with visual impairment can successfully operate the system, and *Supported language adequacy*, i.e. ratio of languages supported by the system to languages, in which system should be available for being operated by any given (potential) user at the civil protection authority.
[73] Differences between the mean values of the scores for A4LIG and Legacy system: *Functional appropriateness* -2.0, *Flexibility 0, Learnability* +0.50, *Hazard coverage* +0.58, *User interface aesthetics* +2.0, *Accessibility* +2.0. Positive/negative scores indicate a superior/inferior performance of A4LIG.

3.3.4 Results of the Multi-criteria analysis

In contrast to the analysis of the mean values of specific criteria sub-categories the Multi-criteria analysis aims to compare the performance of A4LIG and the Legacy system in a comprehensive manner. Hence, the data for all criteria selected by the end-users as being relevant for assessing the performance of an emergency management system and for which data could be collected was considered as described in detail in section 2.3.

As not all performance aspects to be considered for the analysis are equally important the operator and the coordinator elicited weights to all criteria. Overall the most important aspects are *Quantity & Quality data input for decision-making (alert phase)*, *Ease of managing data sources (alert phase)* and *Ease of determining adequate operational phase (alert phase).* The criteria considered to be least relevant are *Ease of managing data sources (monitoring & pre-alert phase)* and *Useful impact assessments (monitoring phase).*

There were some differences between the perspectives of the operator and coordinator. The importance of *Usefulness data collection and data integration capabilities (alert phase), Usefulness flood impact & risk information, Usefulness forest fire risk information, Stress and Service availability* were rated substantially higher by the coordinator. The operator gave higher weight to *Accessibility for users with visual impairments, Supported language adequacy* and *Ease of determining adequate operational phase (alert phase)*. For an overview of all criteria weights see Figure 53 in Annex 7.15.

The overall result of the probabilistic Multi-criteria analysis is that A4LIG outperforms the Legacy system with a probability of 79.83% to 20.17%. This means that when considering both weighting sets and the performance dataset out of the 10.000 single Multi-criteria analyses conducted A4LIG outperformed the Legacy system 7,983 times.

A breakdown of the co-evaluation results by criteria reveals that the superior performance of ALIG is primarily driven by *Forest fire*-related criteria (see Figure 18). A4LIG has the most profound advantage over the Legacy system with regard to the criteria *Usefulness forest fire risk information*, *Information about forest fire risks and impacts*, *Accessibility for users with visual impairments* and *User interface aesthetics*.

The Legacy system has the most pronounced advantage over A4LIG regarding *Transparency for decision-making for insiders*, *Support of decision-making (alert phase)* and *Functional appropriateness.*

No performance differences exist with regard to the criteria *Impact assessment capabilities (monitoring phase), Information flood risk, Ease of determining adequate operational phase (alert phase), Overall satisfaction, Flexibility* and *Supported languages adequacy*

For details see Figure 18, which presents the net preference flows of the two systems for each criterion. The higher/lower the respective bar above/below the zero line the better/worse the performance of the respective system compared to the alternative in this aspect.

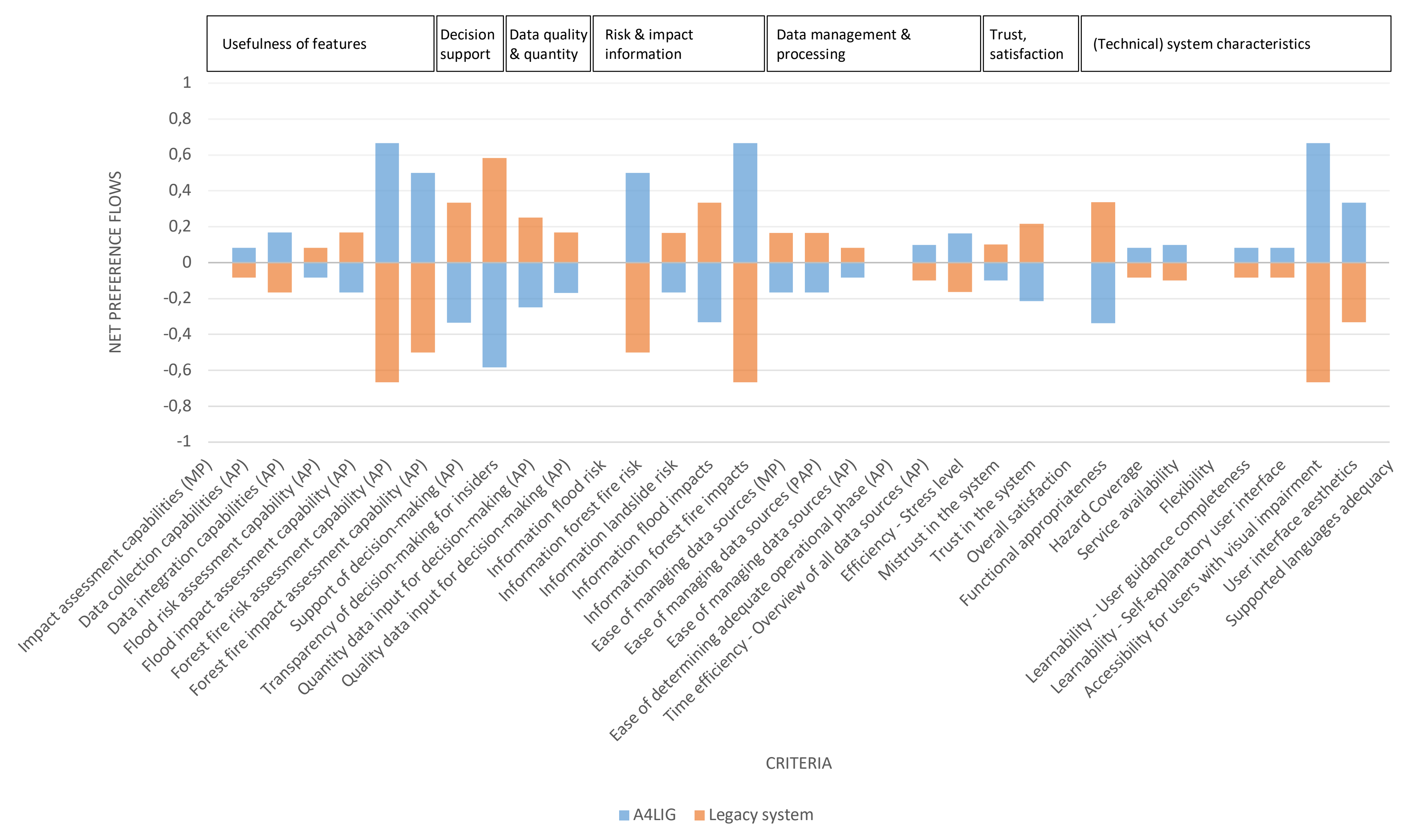


Figure 18: Performance comparison A4LIG and Legacy system

3.4 A4FINN

3.4.1 General impressions

During the demonstration period A4FINN was used on a regular basis whenever the operator or the coordinator were on ISTIKE shift. Irregularly A4FINN was also used when some event such as a storm was expected.

First impressions of the system were mixed as it did not become evident on first sight how it was different from the powerful Legacy system. After getting to know the system a bit better the potential added value of operating A4FINN became clear but the during the first time the systems was not as reliable as expected.

First hands-on experiences were positive. Users were impressed by the fire propagation simulation and the satellite maps, but also had some doubts whether PROPAGATOR would be able to produce reliable results in the Finnish setting. At the beginning of the demonstration period reaction speed of the system was considered to be somewhat low. The new capabilities created new wishes by the users, e.g. notification functionality for the officers in the field including recommendations for action.

The implementation of A4FINN changed the working routines even though not as radically as elsewhere. The main difference is that users can assess different types of risks by looking at one tool and screen, which makes relevant information more accessible.

User highlight the following main advantages of having A4FINN implemented at their site:

- Overview of risks in one map
- Possibility to see risks evolving over time
- Intuitive presentation of data through adequate colour codes and the timeline functionality
- Accessibility from everywhere as it is a web based service

The main drawbacks mentioned include:

- Limited options for customizing the system, e.g. wind speed currently cannot be changed from km/h to m/s (the commonly used at ISTIKE)
- Operational readiness level is not available for sub-regions, but only for the entire region
- Fire propagation simulation of limited use as different types of forest fires occur, e.g. burning peat
- As not all information provided by the Legacy system can be accessed through the A4FINN user interface both systems have to be used in a complementary way simultaneously

- Spatial resolution is still too low, neighbourhood or municipality level would be ideal

3.4.2 Usefulness of relevant features

In the context of the needs analysis the improvement respectively new development of the following features for particular phases of the emergency management process were requested:

- Crowdsourcing data module
- Impact assessment capabilities
- Determination of operational readiness level

At the end of the demonstration period a coordinator as well as an operator who used A4FINN on a regular basis rated the usefulness of the emergency system's capabilities with regard to the relevant features. They did so for two management scenarios, one including A4FINN and another one excluding A4FINN. For simplicity reasons, the former scenario is termed "A4FINN" and the latter scenario is called "Legacy system". Experts were explicitly asked to make their judgements on the basis of their hands-on experiences and not based on their assumptions about the unfolding of A4FINN's potential in the future. Figure 19 illustrates the results of this assessment based on the mean values of the criteria scores provided by the operator and the coordinator. The distance between the blue line ("Legacy system") and the red line ("A4FINN") represents the improvement of the perceived usefulness of the respective features for the management process. So far, due to a lack of critical incidences this assessments primarily related to the management of natural hazards under ordinary conditions.

Figure 19 gives an overview of how the operator and coordinator comparatively assessed the usefulness of particular features for the Legacy system as well as for A4FINN.[74] It is evident that A4FINN outperforms the Legacy system regarding the *Usefulness of impact assessment capabilities*, but the advantage is somewhat less pronounced for the *Usefulness of the determination of operational readiness level*. The *Usefulness of the crowdsourcing data module* is mentioned here for the sake of completeness as this feature was initially requested. Later in the development process efforts were concentrate on the capabilities of A4FINN and, hence, no performance difference can be detected.[75]

[74] Differences between the mean values of the scores for A4FINN and Legacy system*: Crowdsourcing data module 0, Determination of operational readiness level* +0.50, *Impact assessment capabilities* +3.0. Positive/negative scores indicate a superior/inferior performance of A4FINN.

[75] For this reason this criterion was excluded from the Multi-criteria analysis.

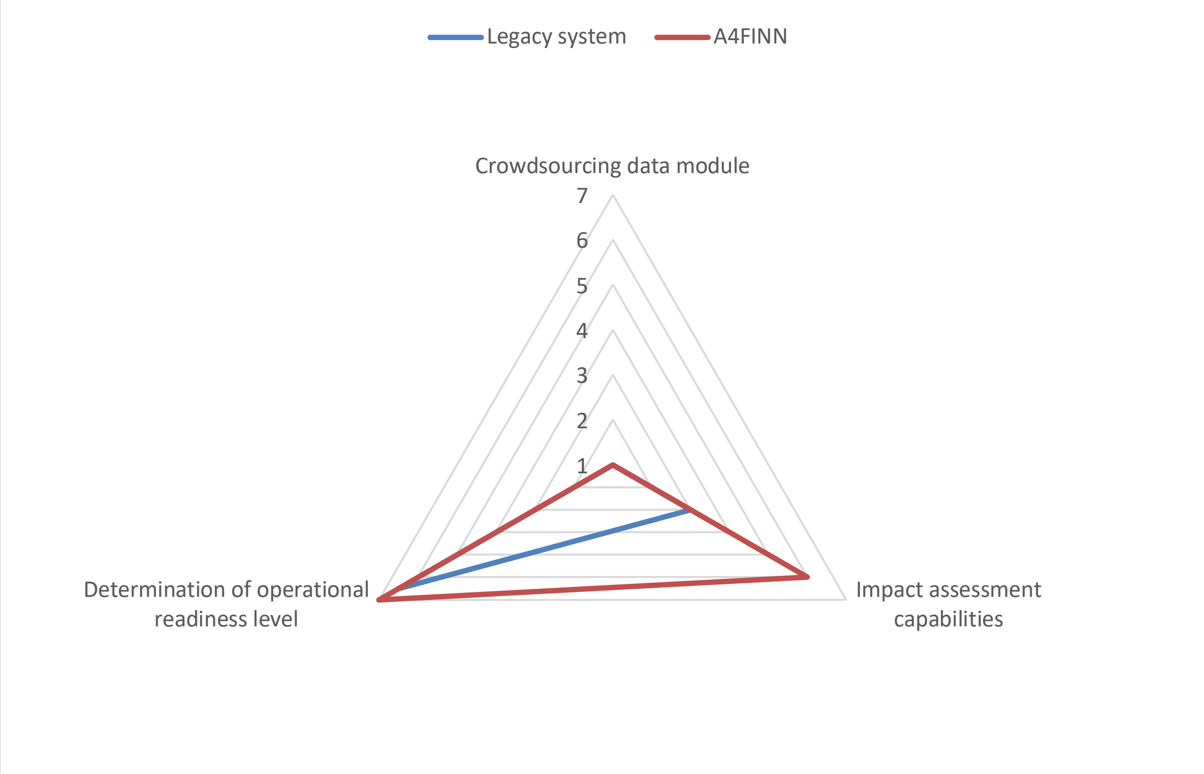


Figure 19: A4FINN - Usefulness of relevant features (mean values)

3.4.3 Social and technical aspects

Social aspects

The following figures visualise the comparison of A4FINN and the Legacy system for the two main dimensions of social impacts, i.e. working routines and knowledge production.[76] The comparison is based on an analysis of the mean values of the respective criteria scores provided by operator and coordinator for a bundle of sub-criteria.

[76] Those criteria, which do not feature the 7-point Likert scale but different metrics, e.g. such as e.g. *Time efficiency* were transformed in the following way to match the scale. Highest/lowest criteria score for each alternative - no matter whether stated by the operator or the coordinator - was set as maximum/minimum value. Average scores of each alternative were then transformed to fit the scale. Therefore, for those criteria absolute values cannot be interpreted in the same way as for the criteria using the 7-point Likert scale. Nevertheless, the visual representation is considered to be suitable for interpreting the differences between the systems.

Working routine-related criteria include the following categories *Time efficiency*[77], *Stress*[78], *Usefulness of features*[79], *Support of decision-making*[80] and *Transparency of decision-making*[81]. The performance comparison of A4FINN and Legacy system for these criteria provides evidence for the substantial improvements witnessed after the implementation of A4FINN (see Figure 20). Superior performance of A4FINN is stated for all categories. The most pronounced advantage is the substantially higher Time efficiency when operating A4FINN compared to the Legacy system. A4FINN is also outperforming the Legacy system notably with regard to the criteria categories *Stress*, *Transparency of decision-making* and *Support of decision-making*.[82]

Results for the four sub-categories of knowledge production-related criteria, i.e. *Data inputs*, *Data processing*, *System outputs* and *Trust,* are presented in Figure 21.[83] Trust remains on the same level. A4FINN outperforms the Legacy system slightly for System outputs[84] and more substantially for *Data inputs* for decision-making[85] as well as for *Data processing.*[86]

[77] *Time efficiency* relates to time determining the operational readiness level.
[78] Level of stress caused by the operation of the system.
[79] *Usefulness of features* is based on the usefulness scores of the 2 features requested for A4FINN which were discussed above.
[80] *Support of decision-making* includes the three criteria covering the assessment of the decision support provided by the systems for the distinct emergency management phases.
[81] *Transparency of decision-making* is based on criteria *Transparency of decision-making for insider* and *Transparency of decision-making for outsiders.*
[82] Differences between the mean values of the scores for A4FINN and Legacy system: Support of decision-making +1.0, Transparency of decision-making +1.38, Stress +1.5, Usefulness +1.75, Time efficiency +3.67. Positive/negative scores indicate a superior/inferior performance of A4FINN.
[83] Differences between the mean values of the scores for A4FINN and Legacy system: *Trust 0, System outputs* +0.54, *Data inputs* +1.5, *Data processing* +2.50. Positive/negative scores indicate a superior/inferior performance of A4FINN.
[84] *System outputs* covers all risk and impact-related criteria, i.e. *Information about risks and impacts of convective storms, thunderstorms, snowstorms, heavy rain, floods* and *forest fires.*
[85] *Data input* includes the quantity and quality of data inputs for decision-making for the three emergency management phases.
[86] *Data processing* is based on the criterion *Ease of determination of operational readiness level.*

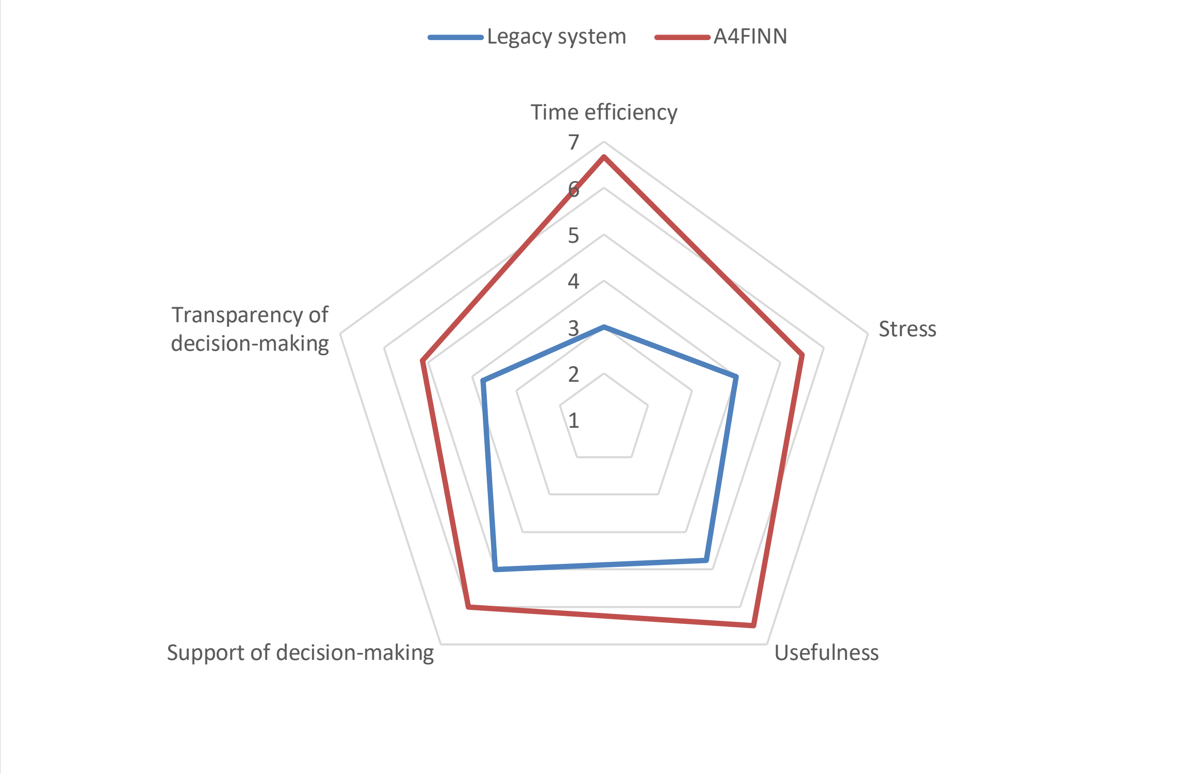


Figure 20: A4FINN - Working routine-related criteria (mean values)

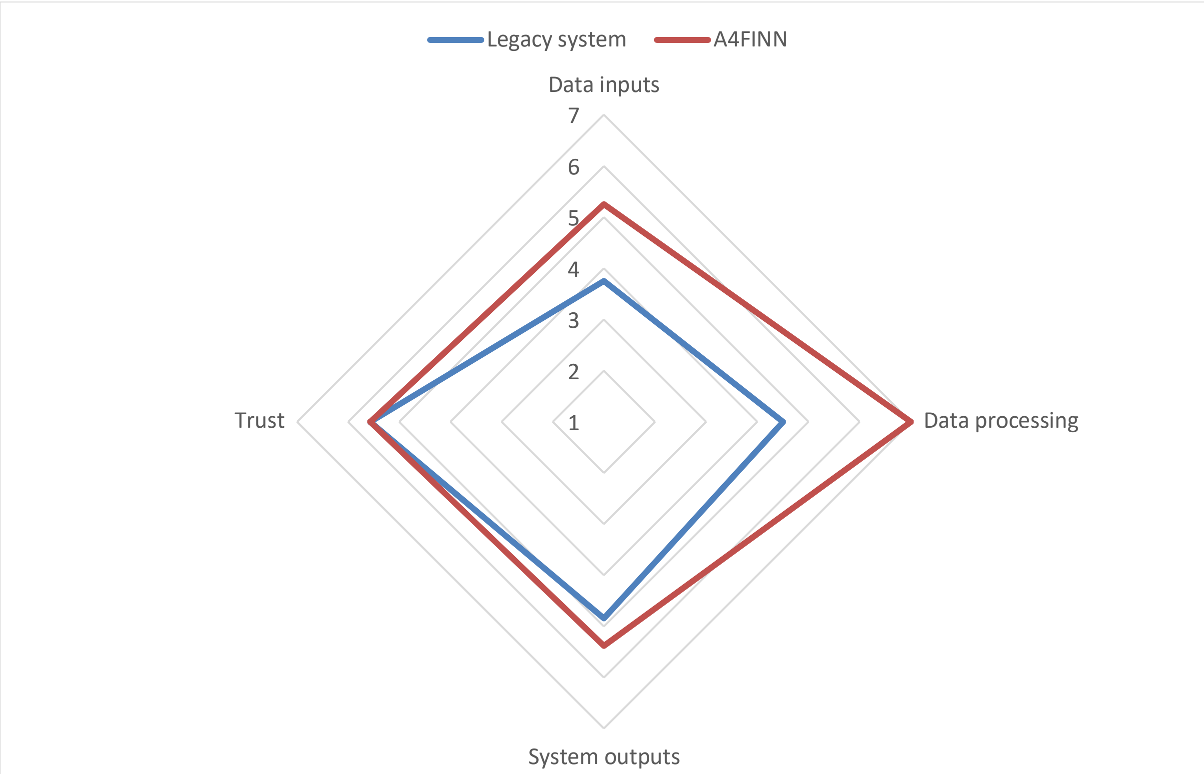


Figure 21: A4FINN - Knowledge production-related criteria (mean values)

Technical effects

The technical criteria are clustered in the following categories *Hazard coverage*, *Functional appropriateness*[87], *Flexibility*[88], *Learnability*[89], *User interface aesthetics*[90], and *Accessibility*[91]. The analysis of the mean value scores of the technical criteria unveiled that for all sub-categories except *Accessibility* A4FINN performs better than the Legacy system. The difference is substantial with regard to the criteria *User interface aesthetics* and *Learnability* and somewhat less pronounced for *Functional appropriateness*, *Flexibility* and *Hazard coverage*. A4FINN has a notable disadvantage when it comes to the *Accessibility* of the two systems. Figure 22 depicts the results.[92]

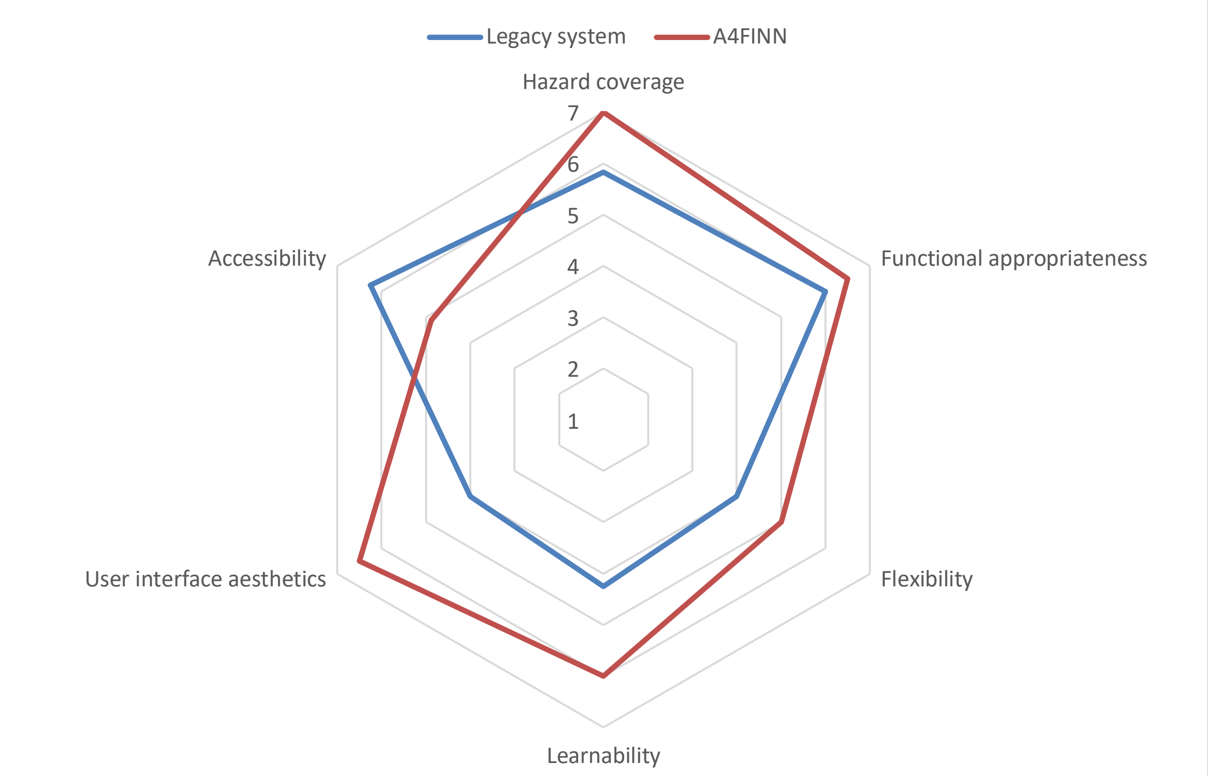


Figure 22: A4FINN - Technical criteria (mean values)

[87] Degree to which number of functions provided by the system is adequate for users to achieve their objectives appropriately.
[88] Degree to which users with different levels of proficiency can operate the system.
[89] Degree to which the user interface is self-explanatory and the user guidance is complete.
[90] Degree to which user interfaces and the overall design are aesthetically pleasing in appearance.
[91] *Accessibility* includes *Accessibility for users with visual impairment*, i.e. degree to which users with visual impairment can successfully operate the system, and *Supported language adequacy*, i.e. ratio of languages supported by the system to languages, in which system should be available for being operated by any given (potential) user at the civil protection authority.
[92] Differences between the mean values of the scores for A4FINN and Legacy system: *Accessibility* -1.38, *Functional appropriateness* +0.50, *Flexibility* +1.0, *Hazard coverage* +1.17, *Learnability* +1.75, *User interface aesthetics* +2.50. Positive/negative scores indicate a superior/inferior performance of A4FINN.

3.4.4 Results of the Multi-criteria analysis

In contrast to the analysis of the mean values of specific criteria sub-categories the Multi-criteria analysis aims to compare the performance of A4FINN and the Legacy system in a comprehensive manner. Hence, the data for all criteria selected by the end-users as being relevant for assessing the performance of an emergency management system and for which data could be collected was considered as described in detail in section 2.3.

As not all performance aspects to be considered for the analysis are equally important the operator and the coordinator elicited weights to all criteria. Overall the most important aspects are *Service availability, Trust-related criteria, Stress* and *Usefulness of determination of operational readiness level.* The criteria considered to be least relevant are *Flexibility, User interface aesthetics and Transparency of the decision-making process for insiders and outsiders.*

There were some differences between the perspectives of the operator and coordinator. The importance of *Transparency of the decision-making process for insiders, User interface aesthetics* and *Flexibility* were rated substantially higher by the coordinator. The operator gave higher weight to *Stress, Supported language adequacy level, Trust-related criteria and Overall satisfaction*. For an overview of all criteria weights see Figure 54 in in Annex 7.16.

The overall result of the probabilistic Multi-criteria analysis is that A4FINN outperforms the Legacy system with a probability of 100%. This means that when considering both weighting sets and the performance dataset out of the 10.000 single Multi-criteria analyses conducted A4FINN outperformed the Legacy system every time.

A breakdown of the co-evaluation results by criteria reveals that A4FINN has the most profound advantage over the Legacy system with regard to the criteria *Hazard coverage and Time efficiency* and *Ease of the determination of the operational readiness level (alert phase)* and *Usefulness of impact assessment.* The only advantage of the Legacy system is its *Supported language adequacy.* This is due to the fact that some of the functionalities of the ANYWHERE platform for now are only available in English.

No performance differences exist with regard to the criteria *Determination of operational readiness level*, *Information convective storm risk* and *impacts*, *Information thunderstorm risk* and *impacts*, *Information heavy rain risk* and *impacts*, *Information flood risk* and *impacts.*

For details see Figure 23, which presents the net preference flows of the two systems for each criterion. The higher/lower the respective bar above/below the zero line the better/worse the performance of the respective system compared to the alternative in this aspect.

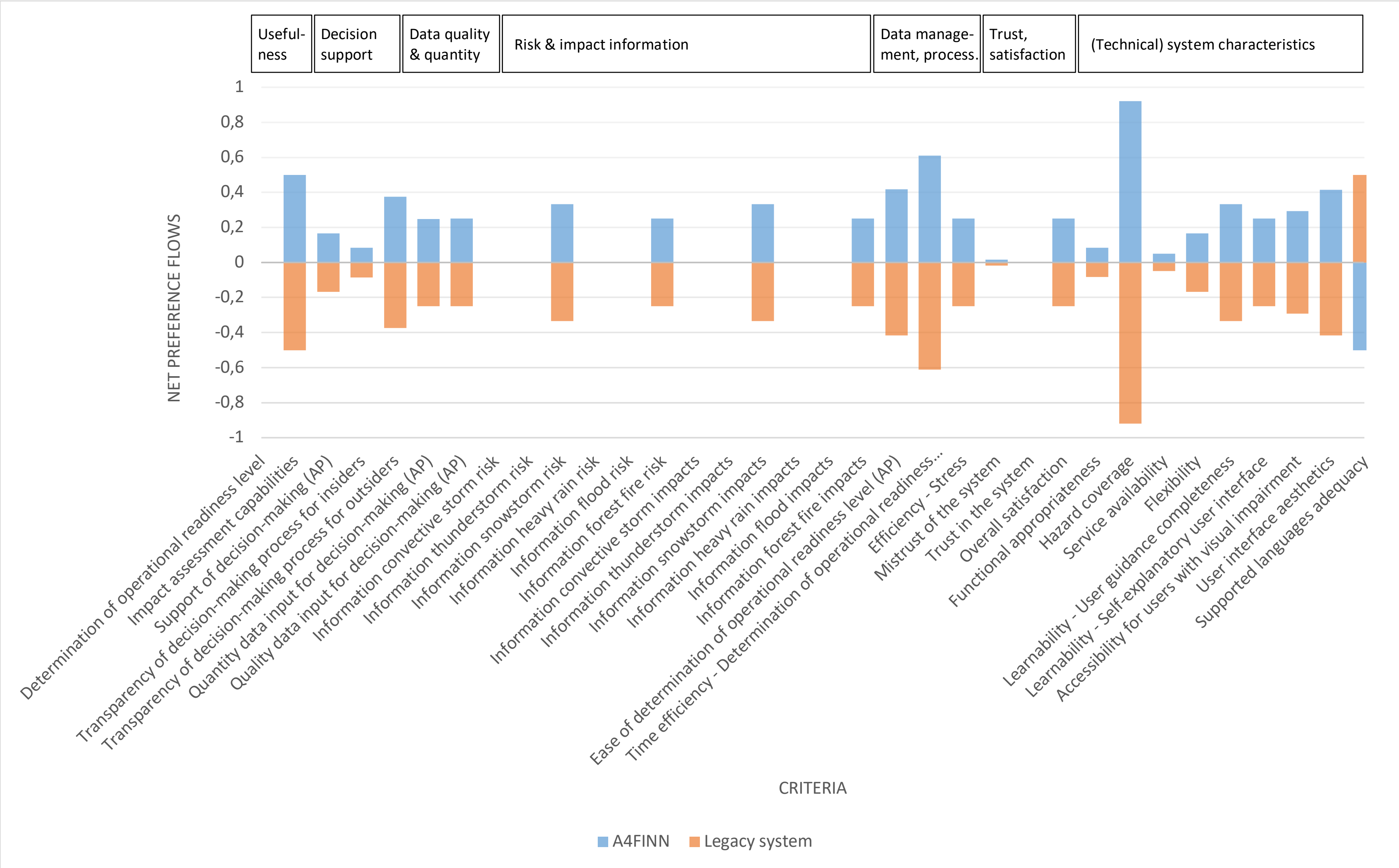


Figure 23: Performance comparison A4FINN and Legacy system

3.5 A4NOR

3.5.1 General impressions

A4NOR was not yet in actual use. *However, first impressions were quite positive. It is perceived as a p*otentially good tool, which however, was lacking functionality the first time it was seen. After some time, there were more features implemented (wave directions, wave height), then it began to make more sense to the users.

User highlighted the following (potential) main advantages of having A4NOR implemented at their site:

- The wave modelling capabilities are a big asset for improving the management of some hospital services, e.g. the operation of boats, helicopters.
- Usefulness could be further improved by implementing an email alert function.

The main drawbacks mentioned included:

- At the beginning of the demonstration the reaction of tools and updating of information was a bit slow, but meanwhile this has been fixed.

3.5.2 Usefulness of relevant features

In the context of the needs analysis the improvement respectively new development of the following features for particular phases of the emergency management process were requested:

- Communication of weather situation (phase A)
- Description of weather event for local level (phase A)
- Impact assessment capabilities (phase B)
- Real time visualisation of risk levels (phase C)
- Data integration capabilities

At the end of the demonstration period two operators who are familiar with A4NOR and had used it during the demonstration period rated the usefulness of the emergency system's capabilities with regard to the relevant features. They did so for two management scenarios, one including A4NOR and another one excluding A4NOR. For simplicity reasons, the former scenario is termed "A4NOR" and the latter scenario is called "Legacy system". Experts were explicitly asked to make their judgements on the basis of their hands-on experiences and not based on their assumptions about the unfolding of A4NOR's potential in the future. Figure 24 illustrates the results of this assessment based on the mean values of the criteria scores provided by the operators. The distance between the blue line ("Legacy system") and the red line ("A4NOR") represents the improvement of the perceived usefulness of the respective features for the management process. So far, due to a lack of critical incidences this assessments primarily related to the management of natural hazards under ordinary conditions.

Figure 24 gives an overview of how the operators comparatively assessed the usefulness of particular features for A4NOR and the Legacy system. It is evident that A4NOR outperforms the Legacy system with regard to *Usefulness of the communication of weather situation (phase A)* and of the *real time visualisation of risk levels (phase C).* The *Description of weather event for local level (phase A)* provided by the A4FINN is slightly more useful than the information offered by Legacy system. There are no differences between the usefulness of the systems' *Impact assessment and Data integration capabilities*.[93]

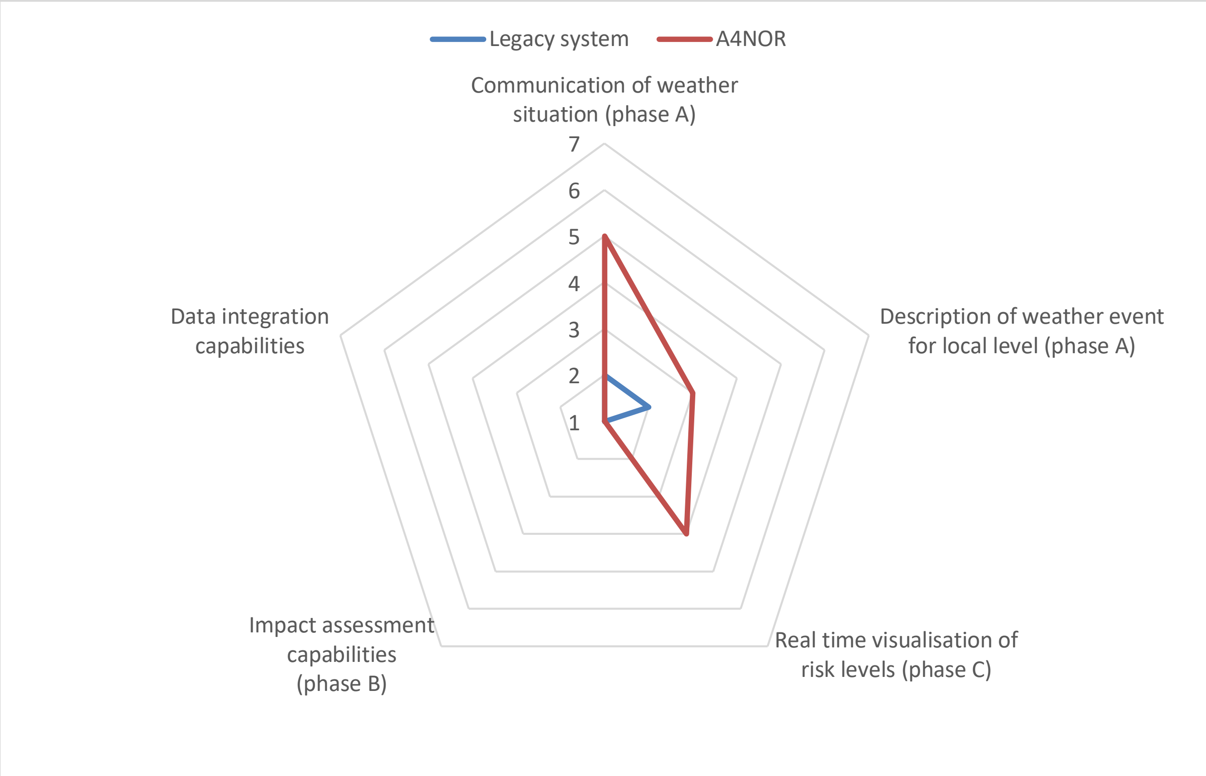


Figure 24: A4NOR - Usefulness of relevant features (mean values)

3.5.3 Social and technical aspects

Social aspects

The following figures visualise the comparison of A4NOR and the Legacy system for the two main dimensions of social impacts, i.e. working routines and knowledge production.[94] The comparison is based on an analysis of the mean values of the

[93] Differences between the mean values of the scores for A4NOR and Legacy system: *Communication of weather situation (phase A)* +3.0 and *Real time visualisation of risk levels (phase C)* +3.0, *Description of weather event for local level (phase A)* +1.0, *Impact assessment capabilities (phase B)* 0, *Data integration capabilities* 0. Positive/negative scores indicate a superior/inferior performance of A4NOR.

[94] Those criteria, which do not feature the 7-point Likert scale but different metrics, e.g. such as e.g. *Time efficiency* were transformed in the following way to match the scale. Highest/lowest criteria score for each alternative - no matter whether stated by the operator or

respective criteria scores provided by operator and coordinator for a bundle of sub-criteria.

Working routine-related criteria include the following categories *Time efficiency*[95], *Usefulness of features*[96] and *Support of decision-making.*[97] The performance comparison of A4NOR and Legacy system for these criteria provides evidence for the substantial improvements witnessed after the implementation of A4NOR (see Figure 20). Superior performance of A4NOR is stated for all categories. The most pronounced advantage is the substantially higher *Time efficiency* when operating A4NOR compared to the Legacy system. A4NOR is also outperforming the Legacy system notably with regard to the criteria *Support of decision-making* and *Usefulness of relevant features*.[98]

Results for the four sub-categories of knowledge production-related criteria, i.e. *Data inputs*, *Data processing*, *System outputs* and *Trust,* are presented in Figure 26.[99]
A4NOR outperforms the Legacy system substantially for System outputs[100], *Data inputs* for decision-making[101] as well as for *Data processing.*[102] In contrast to this for now operators have much more *Trust* in the Legacy system than in A4NOR.

---

the coordinator - was set as maximum/minimum value. Average scores of each alternative were then transformed to fit the scale. Therefore, for those criteria absolute values cannot be interpreted in the same way as for the criteria using the 7-point Likert scale. Nevertheless, the visual representation is considered to be suitable for interpreting the differences between the systems.

[95] *Time efficiency* covers the time needed to get an overview of all relevant data sources.

[96] *Usefulness of features* is based on the usefulness scores of the 5 features requested for A4FINN which were discussed above.

[97] *Support of decision-making* includes the three criteria covering the assessment of the decision support provided by the systems for the distinct emergency management phases.

[98] Differences between the mean values of the scores for A4NOR and Legacy system: *Usefulness* +1.4, *Support of decision-making* +2.0, *Time efficiency* +6.0. Positive/negative scores indicate a superior/inferior performance of A4NOR.

[99] Differences between the mean values of the scores for A4NOR and Legacy system: *Trust* -3.0, *System outputs* +1.94, *Data processing* +2.0, *Data inputs* +2.25. Positive/negative scores indicate a superior/inferior performance of A4NOR.

[100] *System outputs* covers all risk and impact-related criteria, i.e. *Information about risks and impacts of storm surges, heavy rain, floods* and *strong winds.*

[101] *Data input* includes the quantity and quality of data inputs for decision-making for the three emergency management phases.

[102] *Data processing* is based on the criterion *Ease of managing data sources.*

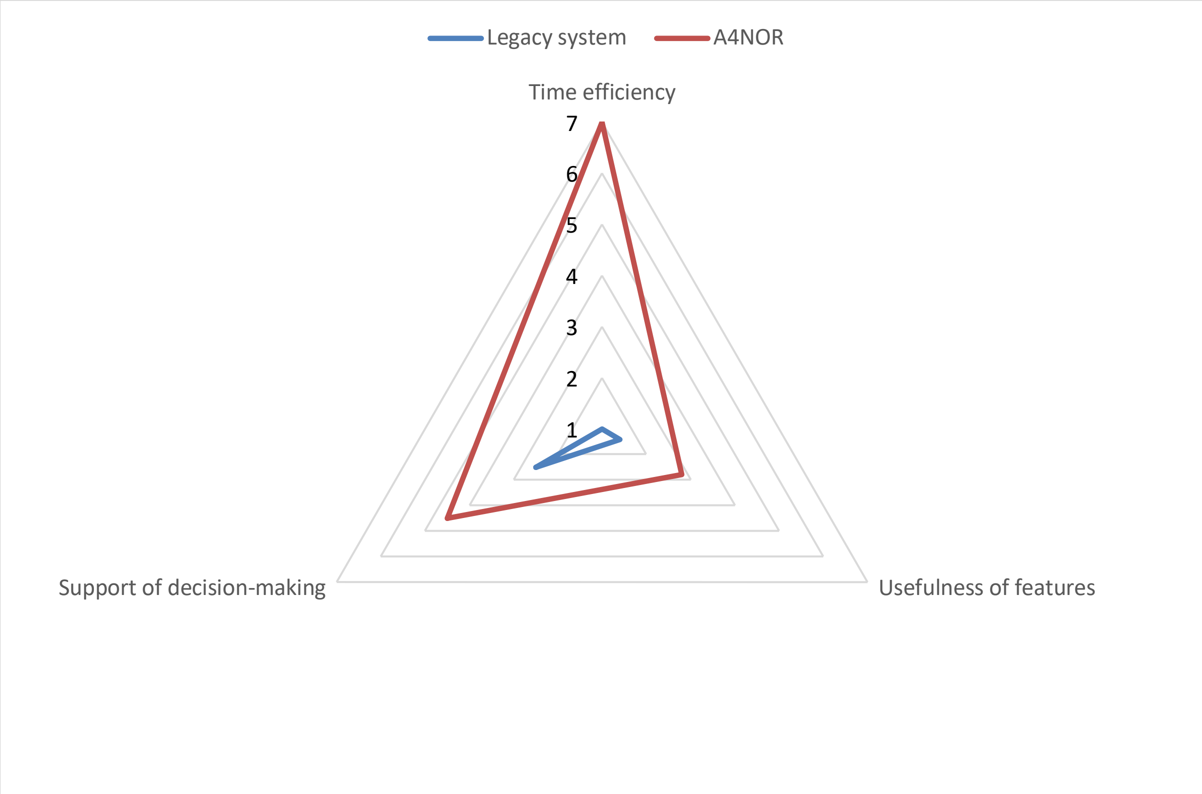


Figure 25: A4NOR - Working routine-related criteria (mean values)

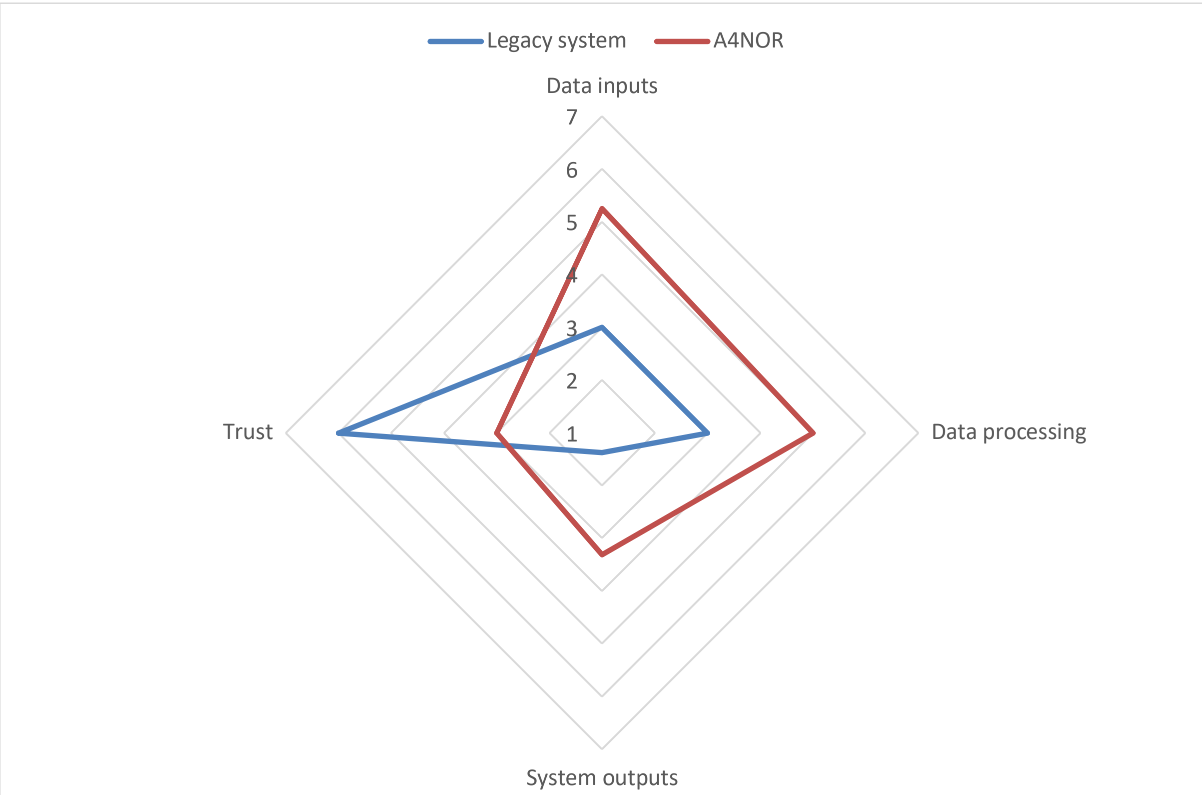


Figure 26: A4NOR - Knowledge production-related criteria (mean values)

Technical effects

The technical criteria are clustered in the following categories *Functional appropriateness*[103], *Flexibility*[104], *Learnability*[105], *User interface aesthetics*[106], and *Accessibility*[107]. The analysis of the mean value scores of the technical criteria unveiled that only for *Functional appropriateness* A4NOR performs better than the Legacy system. The difference is substantial. A4NOR has notable disadvantages when it comes to *Flexibility*, *Accessibility* and *Learnability*. There is no difference regarding the *User interface aesthetics* of the systems. Figure 27 depicts the results.[108]

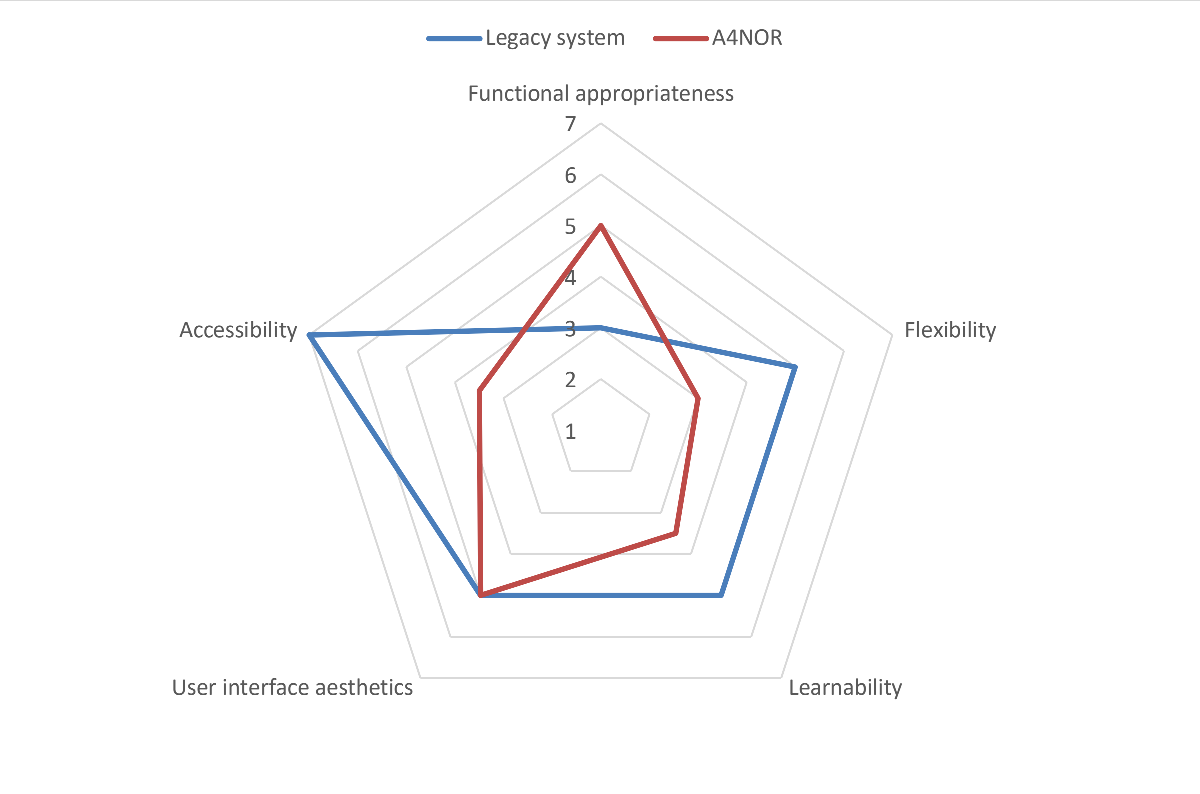


Figure 27: A4NOR - Technical criteria (mean values

[103] Degree to which number of functions provided by the system is adequate for users to achieve their objectives appropriately.
[104] Degree to which users with different levels of proficiency can operate the system.
[105] Degree to which the user interface is self-explanatory and the user guidance is complete.
[106] Degree to which user interfaces and the overall design are aesthetically pleasing in appearance.
[107] *Accessibility* includes *Accessibility for users with visual impairment*, i.e. degree to which users with visual impairment can successfully operate the system, and *Supported language adequacy*, i.e. ratio of languages supported by the system to languages, in which system should be available for being operated by any given (potential) user at the civil protection authority.
[108] Differences between the mean values of the scores for A4NOR and Legacy system: *Accessibility* -3.5*,* *Flexibility* -2.0*,* *Learnability* -1.5*,* *User interface aesthetics* 0*,* *Functional appropriateness* 2.0. Positive/negative scores indicate a superior/inferior performance of A4FINN.

3.5.4 Results of the Multi-criteria analysis

In contrast to the analysis of the mean values of specific criteria sub-categories the Multi-criteria analysis aims to compare the performance of A4NOR and the Legacy system in a comprehensive manner. Hence, the data for all criteria selected by the end-users as being relevant for assessing the performance of an emergency management system and for which data could be collected was considered as described in detail in section 2.3.

The weights, which were elicited by the operators were quite even for all criteria the only exception being the following criteria, which were considered to be about half as important as the others: *Functional appropriateness*, *Flexibility*, *User interface aesthetics* and *Supported languages adequacy*. For an overview of all criteria weights see Figure 58 in Annex 7.20.

The overall result of the probabilistic Multi-criteria analysis is that A4NOR outperforms the Legacy system with a probability of 100%. This means that when considering both weighting sets and the performance dataset out of the 10.000 single Multi-criteria analyses conducted A4NOR outperformed the Legacy system every time.

A breakdown of the co-evaluation results by criteria reveals that the superior performance of ANOR is primarily driven by the social and not by the technical criteria. (see Figure 28). A4NOR has the most substantial advantage over the Legacy system with regard to the criteria *Time efficiency - overview all relevant data sources, Information about storm surge risk and impacts, Usefulness of communication of the weather situation (phase A), Usefulness of real time representation of the risk level (phase C)*.

The Legacy system has the biggest advantage over A4NOR regarding *Supported language adequacy*, *Learnability - user guidance completeness* and *Trust*.

No performance differences exist with regard to the criteria *Usefulness of impact assessment capabilities (phase B)*, *Information flood risk and impacts*, *Information heavy rain impacts*, *Information strong wind impacts*, *User interface aesthetics* and *Overall satisfaction*.

For details see Figure 28, which presents the net preference flows of the two systems for each criterion. The higher/lower the respective bar above/below the zero line the better/worse the performance of the respective system compared to the alternative in this aspect.

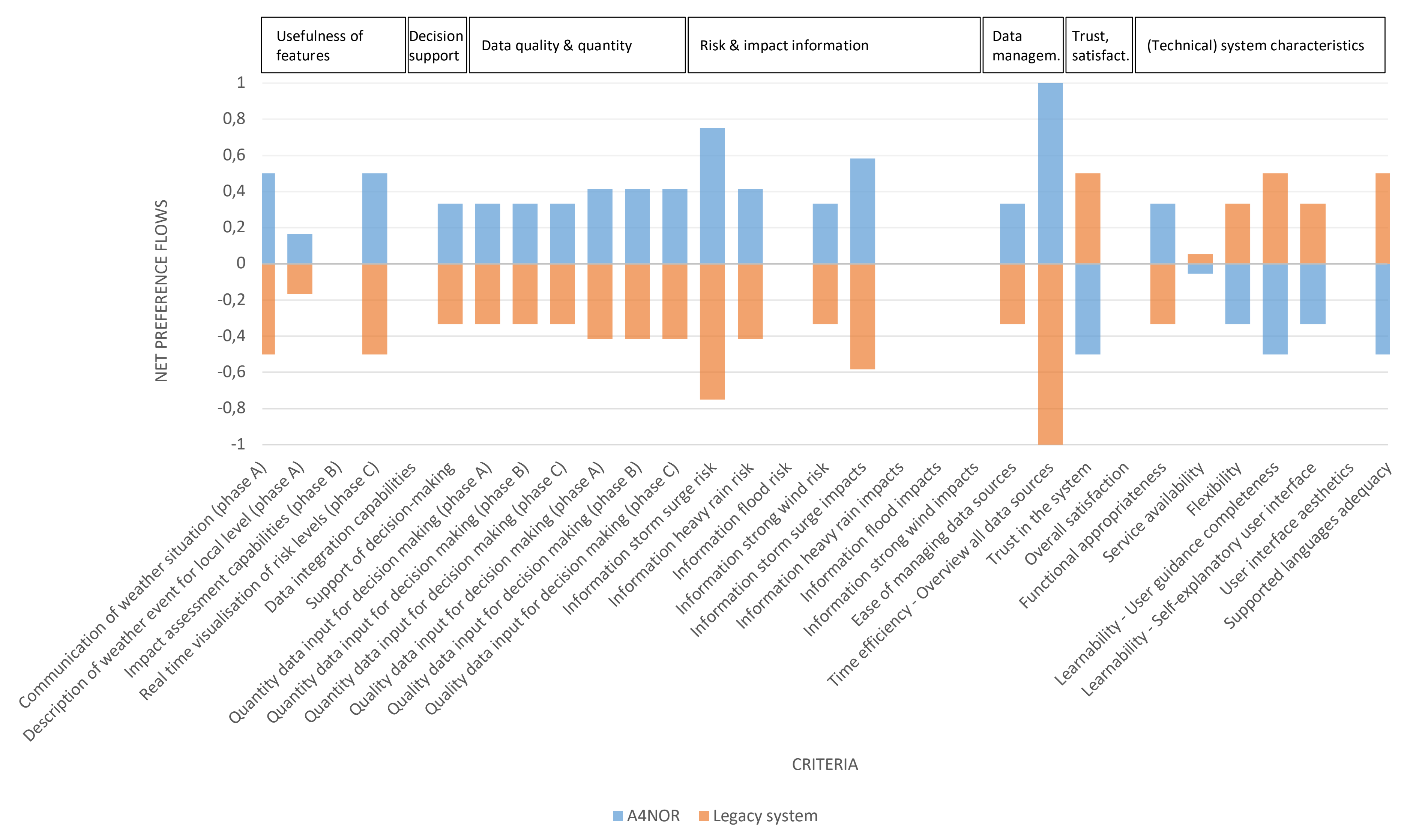


Figure 28: Performance comparison A4NOR and Legacy system

3.6 A4COR

3.6.1 General impressions

During the demonstration period A4COR was used to organize the operational response for floods (impacted areas) and storms, e.g. in October 2018. Even though there haven't been too many fires in the last 2 years the system was PROPAGATOR tested when managing a fire in February 2019.

On the one hand side the first impression of an early prototype was that the system was not as simple and information-rich as hoped for, but on the other hand side the first use of PROPAGATOR was very impressive. The reason was that the tool was capable of re-simulating a – back then – recent fire event. Users were convinced right away that they would be able to use PROPAGATOR to prepare the operational response.

The experiences during the last 20 months were very positive. A4COR is considered to be very useful and is almost constantly used by the coordinator on his smartphone. The system provides more added value to the operational people then to the officers in the call centre.

The implementation of A4COR had a substantial impact on the working routines. The coordinator is using it almost constantly on his smartphone to cross-check information from other sources. The fire propagation capability greatly supports decision-making when managing the resources. In case of floods mayors can be contacted directly. Even if there is no change of official procedures, A4COR has a substantial impact on many of the established routines.

User highlight the following main advantages of having A4COR implemented at their site:

- Easy to use if internet connection is available
- Possibility to select the variety of information to be visualized on a single map
- Integration of links to other databases.
- Currently use of the system free of charge
- Fire propagation tool is a bit asset.

The main drawbacks mentioned include:

- System is not working without an internet connection.
- Becoming familiar with the system is time consuming.
- Difficult to create awareness for the huge potential of the tool, especially among young colleagues
- Possibility to add additional geo-referenced information is required, e.g. position of firemen and cars or other operational resources
- Limited usability on mobile devices such as smartphones

- Would be helpful to have tool for floods which works in a similar way as PROPAGATOR for forest fires.

3.6.2 Usefulness of relevant features

In the context of the needs analysis the improvement respectively new development of the following features for particular phases of the emergency management process were requested:[109]

- Rainfall predictions
- Rainfall-related run-off predictions
- Overview of current & expected risks
- Visualisation of risk levels
- Flood forecasts
- Real time information about flood-related impacts
- Overview of warning states of all municipalities
- Access to local safeguard plans and vulnerability information
- Access to reports of past events
- Data integration capabilities
- Field feedback capabilities
- Decision support by on-call engineer
- Transmission of mobilization levels & recommendations to municipalities

At the end of the demonstration period a coordinator as well as an operator who used A4COR on a regular basis rated the usefulness of the emergency system's capabilities with regard to the relevant features. They did so for two management scenarios, one including A4COR and another one excluding A4COR. For simplicity reasons, the former scenario is termed "A4COR" and the latter scenario is called "Legacy system". Experts were explicitly asked to make their judgements on the basis of their hands-on experiences and not based on their assumptions about the unfolding of A4COR's potential in the future. Figure 29 illustrates the results of this assessment based on the mean values of the criteria scores provided by the operator and the coordinator. The distance between the blue line ("Legacy system") and the red line ("A4COR") represents the improvement of the perceived usefulness of the respective features for the management process.

Figure 29 gives an overview of how the operator and coordinator comparatively assessed the usefulness of particular features for A4COR and for the Legacy system. It is evident that A4COR outperforms the Legacy system with regard to all features. Improvements are substantial across the board being most remarkable for *Usefulness of real time information about flood-related impacts, Usefulness of field feedback*

[109] For improving the readability of the figures documenting the multi-criteria analysis the following abbreviation were used for the different phases of the emergency management process: Awareness phase – AWP, preparation phase – PP, alert phase – ALP. For details about the phases see ANYWHERE project Deliverable 1.2: Report on needs and requirements from the users.

capabilities, *Usefulness of* overview of warning states of all municipalities and *Usefulness of a*ccess to local safeguard plans and vulnerability information.[110] Advantages of A4COR over the Legacy system are also notable for the *Usefulness* of the following feature: *Flood forecasts*, *Data integration capabilities, Rainfall predictions*, *Run-off predictions, Overview of current & expected risks* and *Transmission of mobilization levels & recommendations to municipalities*. Minor advantages of A4COR are mentioned with regard to *Usefulness of visualisation of risk levels*, *Usefulness of access to reports of past events* and *Usefulness of decision support by on-call engineer.*

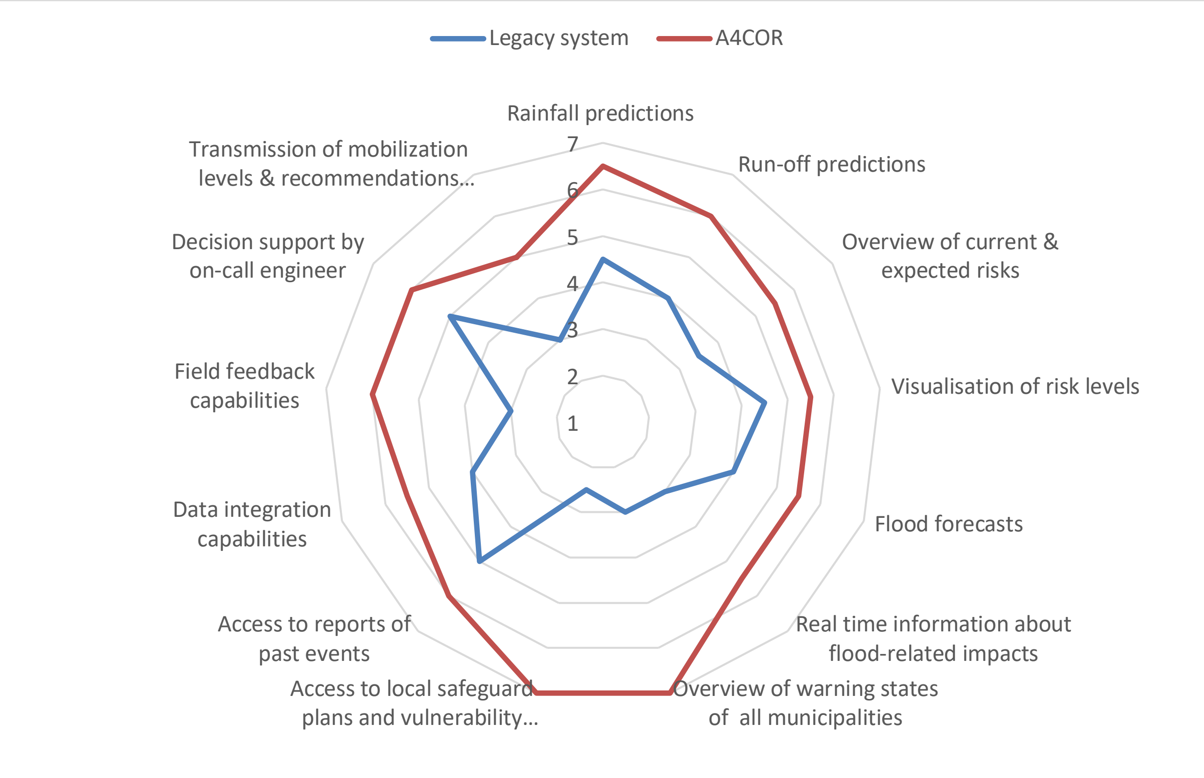


Figure 29: A4COR - Usefulness of relevant features (mean values)

3.6.3 Social and technical aspects

Social aspects

The following figures visualise the comparison of A4COR and the Legacy system for the two main dimensions of social impacts, i.e. working routines and knowledge

110 Differences between the mean values of the scores for A4COR and Legacy system: *Visualisation of risk levels* +1.0, *Access to reports of past events* +1.0, *Decision support by on-call engineer* +1.0, *Flood forecasts* +1.5, *Data integration capabilities* +1.5, *Rainfall predictions* +2.0, *Run-off predictions* +2.0, *Overview of current & expected risks* +2.0, *Transmission of mobilization levels & recommendations to municipalities* +2.0, *Real time information about flood-related impacts* +2.50, *Field feedback capabilities* +3.0, *Overview of warning states of all municipalities* +4.0, *Access to local safeguard plans and vulnerability information* +4.5. Positive/negative scores indicate a superior/inferior performance of A4COR.

production.[111] The comparison is based on an analysis of the mean values of the respective criteria scores provided by operator and coordinator for a bundle of sub-criteria.

Working routine-related criteria include the following categories *Time efficiency*[112], *Stress*[113], *Usefulness of features*[114], *Support of decision-making*[115] and *Transparency of decision-making*[116]. The performance comparison of A4COR and Legacy system for these criteria provides evidence for the substantial improvements witnessed after the implementation of A4COR (see Figure 30). Superior performance of A4COR is stated for all categories. The most pronounced advantage is the substantially higher *Time efficiency* when operating A4COR compared to the Legacy system and the *Usefulness of relevant features*. A4COR is also outperforming the Legacy system notably with regard to the criteria *Stress, Support of decision-making* and *Transparency of decision-making*.[117]

Results for the four sub-categories of knowledge production-related criteria, i.e. *Data inputs* for decision-making[118], *Data processing*[119], *System outputs*, i.e. information for particular risks and impacts, which can be assessed in different ways by the two

[111] Those criteria, which do not feature the 7-point Likert scale but different metrics, e.g. such as e.g. *Time efficiency* were transformed in the following way to match the scale. Highest/lowest criteria score for each alternative - no matter whether stated by the operator or the coordinator - was set as maximum/minimum value. Average scores of each alternative were then transformed to fit the scale. Therefore, for those criteria absolute values cannot be interpreted in the same way as for the criteria using the 7-point Likert scale. Nevertheless, the visual representation is considered to be suitable for interpreting the differences between the systems.

[112] *Time efficiency* covers the *Time efficiency - Overview of all data sources*, *Time efficiency - Overview of warning states of municipalities* and *Time efficiency - Overview of current & expected risks*.

[113] Level of stress caused by the operation of the system.

[114] *Usefulness of features* is based on the usefulness scores of the 13 features requested for A4COR which were discussed above.

[115] *Support of decision-making* includes the three criteria covering the assessment of the decision support provided by the systems for the distinct emergency management phases.

[116] *Transparency of decision-making* is based on criteria *Transparency of decision-making for insider* and *Transparency of decision-making for outsiders.*

[117] Differences between the mean values of the scores for A4COR and Legacy system: *Stress* +1.0, *Support of decision-making* +1.67, *Transparency of decision-making* +1.75, *Time efficiency* +2.0, *Usefulness* +2.13. Positive/negative scores indicate a superior/inferior performance of A4COR.

[118] *Data input* includes the quantity and quality of data inputs for decision-making for the three emergency management phases.

[119] *Data processing* is based on the criteria *Ease of managing data* sources in the three phases of the emergency management process, *Ease of getting overview of municipalities' warning states, Ease of transmitting mobilization levels & recommendations* and *Ease of getting overview of current & expected risks.*

systems [120] and *Trust,* show that A4COR has only a slight advantage over the Legacy system when it comes to *Trust* and *System outputs*, but regarding *Data inputs* and *Data processing* the performance difference is substantial (see Figure 31).[121]

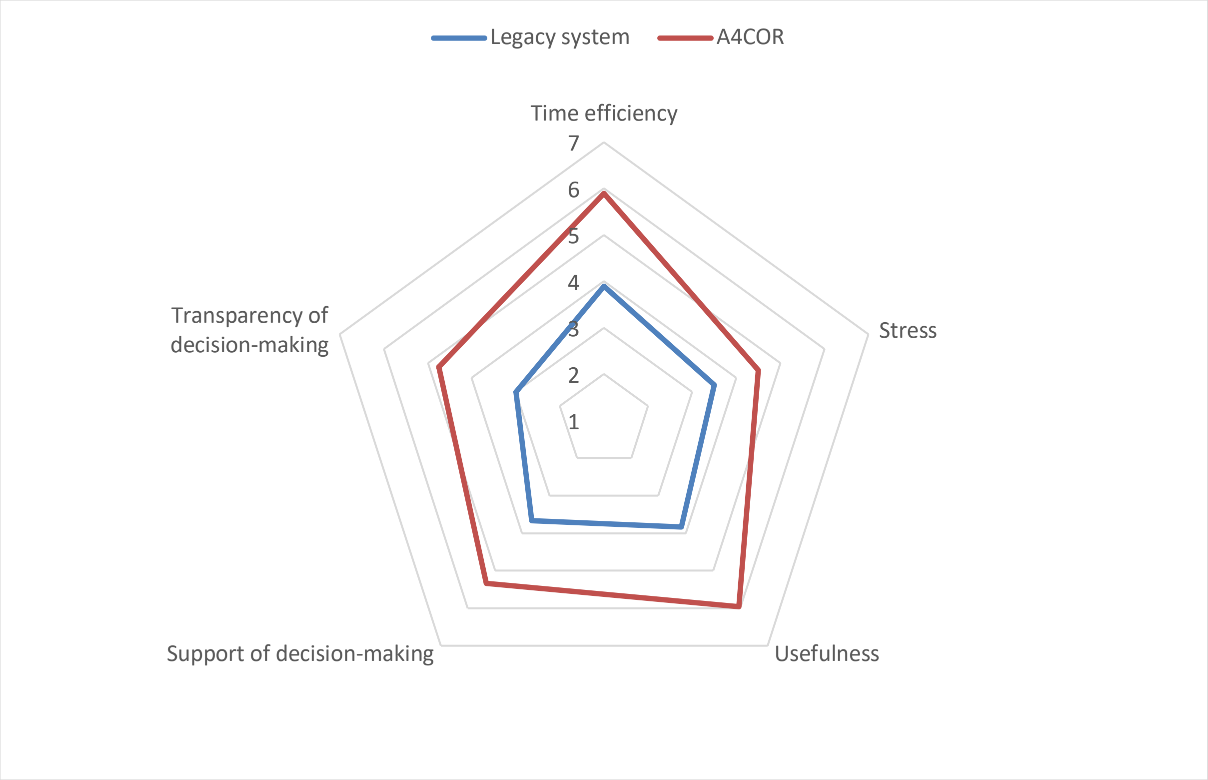


Figure 30: A4COR - Working routine-related criteria (mean values)

[120] *System outputs* covers all risk and impact-related criteria, i.e. *Information about risks and impacts of floods, convective storms, strong winds and forest fires.*

[121] Differences between the mean values of the scores for A4COR and Legacy system: *Trust* +0.5, *System outputs* +0.75, *Data inputs* +2.25, *Data processing* +2.58. Positive/negative scores indicate a superior/inferior performance of A4COR.

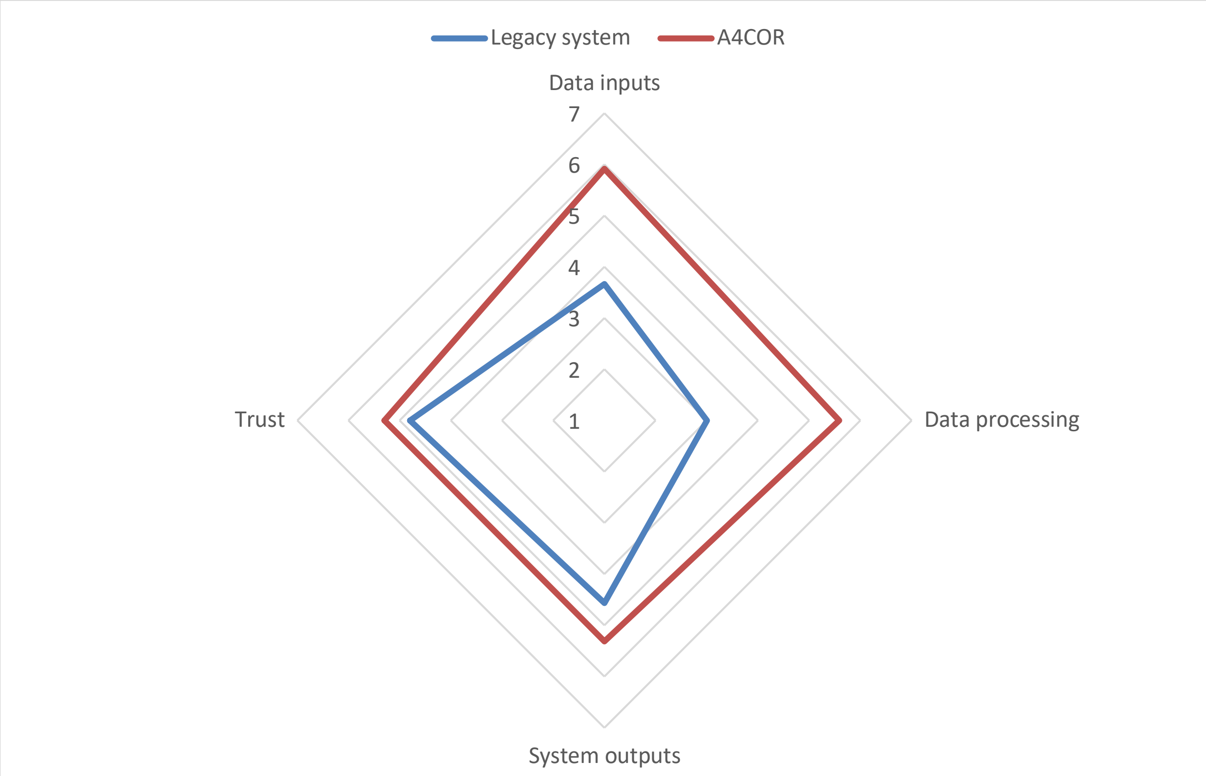


Figure 31: A4COR - Knowledge production-related criteria (mean values)

Technical aspects

The technical criteria are clustered in the following categories *Hazard coverage*, *Functional appropriateness* [122], *Flexibility* [123], *Learnability* [124], *User interface aesthetics*[125] and *Hazard coverage*[126]. The analysis of the mean value scores of the technical criteria shows that for the sub-categories *User interface aesthetics* and *Flexibility* A4COR performs substantially and for *Learnability* and *Hazard coverage* still notably better than the Legacy system. Whereas A4COR has a notable disadvantage when it comes to the *Functional appropriateness* of the two systems. Figure 32 depicts the results.[127]

[122] Degree to which number of functions provided by the system is adequate for users to achieve their objectives appropriately.
[123] Degree to which users with different levels of proficiency can operate the system.
[124] Degree to which the user interface is self-explanatory and the user guidance is complete.
[125] Degree to which user interfaces and the overall design are aesthetically pleasing in appearance.
[126] Ratio of hazards, which can be managed using system to total number of relevant hazards, which need to be managed.
[127] Differences between the mean values of the scores for A4COR and Legacy system: *Functional appropriateness* -1.5, *Learnability* +1.0, *Hazard coverage* +1.16, *Flexibility* +2.5, *User interface aesthetics* +2.5. Positive/negative scores indicate a superior/inferior performance of A4COR.

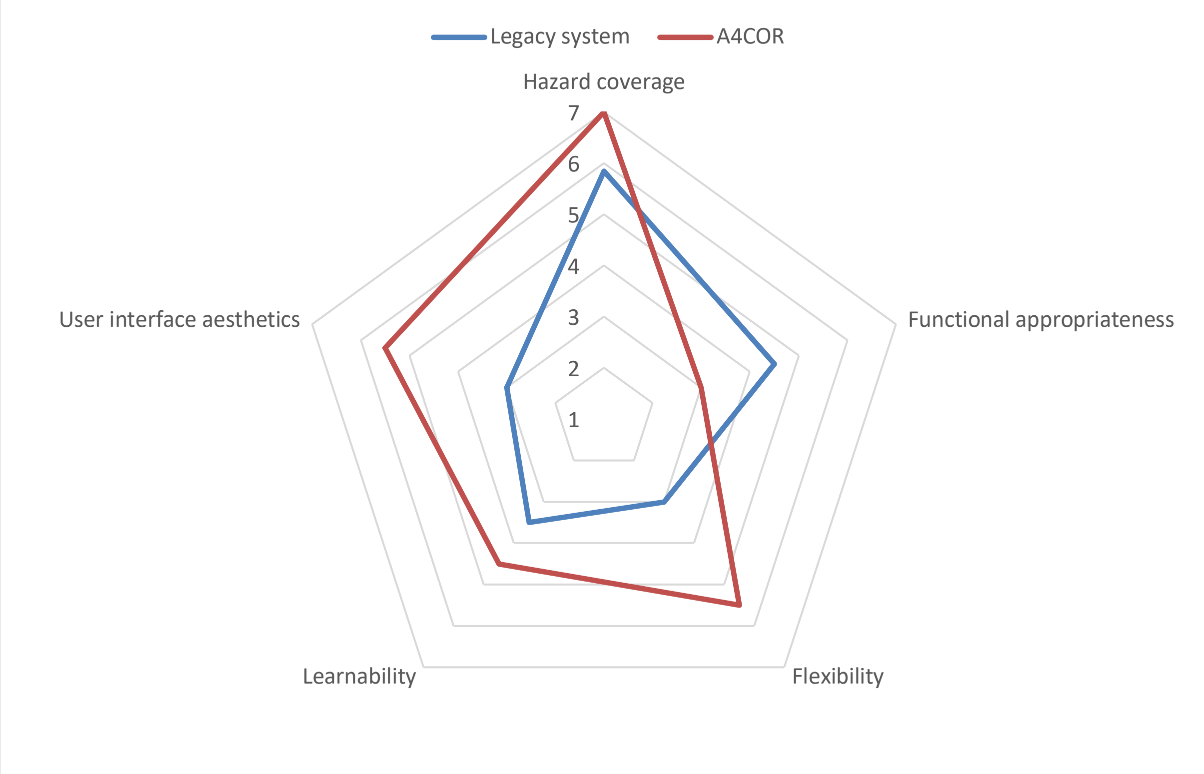


Figure 32: A4COR - Technical criteria (mean values)

3.6.4 Results of the Multi-criteria analysis

In contrast to the analysis of the mean values of specific criteria sub-categories the Multi-criteria analysis aims to compare the performance of A4COR and the Legacy system in a comprehensive manner. Hence, the data for all criteria selected by the end-users as being relevant for assessing the performance of an emergency management system and for which data could be collected was considered as described in detail in section 2.3.

As not all performance aspects to be considered for the analysis are equally important the operator and the coordinator elicited weights to all criteria. When taking the perspective of the coordinator as well as the operator into consideration the following criteria are judged as being the most important ones to assess the performance of an emergency management system: *Usefulness visualisation of risk levels*, *Trust* and *Service availability*. In contrast, the following are the least relevant ones*: User interface aesthetics*, *Transparency of the decision-making processes for outsiders* and *Flexibility*.

There are also some differences evident. The coordinator values the following criteria substantially higher than the operator: *Flexibility*, *Support of decision-making (awareness phase)*, *Information about various risks (convective storms, strong winds, forest fires)* and *impacts (floods, convective storms, strong winds).* In turn, the operator considers the following criteria to be substantially more important for assessing the performance of an emergency management system than the coordinator: *User* *interface aesthetics*, *Stress*, *Learnability*[128] and *Usefulness of an overview of the warning states for all municipalities* offered by the system. For an overview of all criteria weights see Figure 59 in Annex 7.21.

The overall co-evaluation result is unambiguous. With certainty, i.e. a probability of 100%, A4COR outperforms the Legacy system. This means that when considering both weighting sets and the performance dataset out of the 10.000 single Multi-criteria analyses conducted A4COR outperformed the Legacy system every time.

A breakdown of the co-evaluation results by criteria reveals that A4COR outperforms the Legacy system for all criteria except the *Functional appropriateness*. A4COR has the most profound advantage over the Legacy system with regard to the criteria *Overall satisfaction, Time efficiency - getting an overview about all relevant data sources & warning states of all municipalities & current and expected risks and Usefulness of information about local safeguard plans and vulnerabilities.* No substantial performance differences exist with regard to the criteria *Information forest fire risk and impacts* and *Service availability*.

For details see Figure 33 which presents the net preference flows of the two systems for each criterion. The higher/lower the respective bar above/below the zero line the better/worse the performance of the respective system compared to the alternative in this aspect.

[128] Degree to which the user interface is self-explanatory and the user guidance is complete.

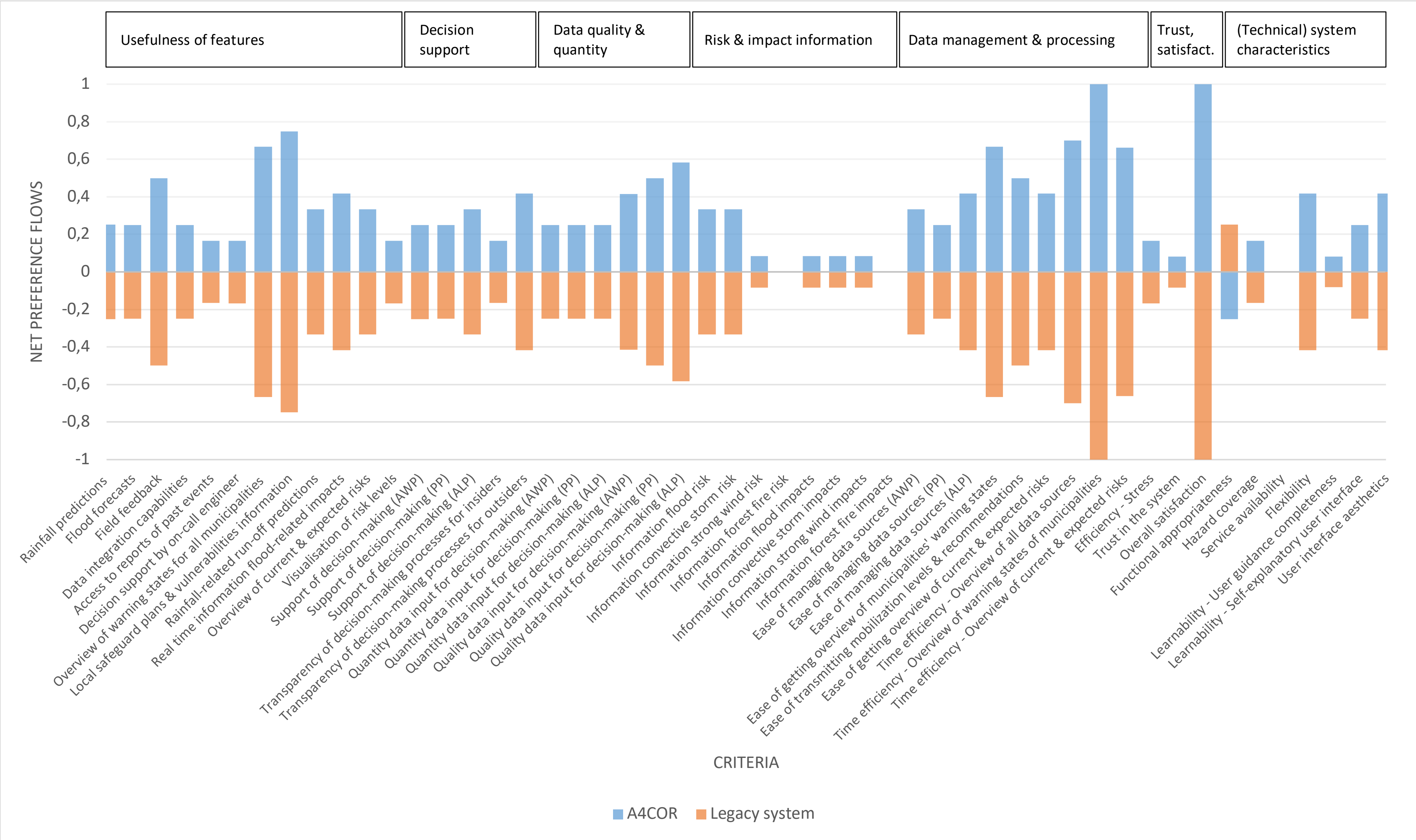


Figure 33: Performance comparison A4COR and Legacy system

3.7 A4CENEM

3.7.1 General impressions

A4CENEM was used through the demonstration phase on a regular basis. More specifically, it was consulted in case it was need to obtain additional data sources and information and receive new insight (unprecedented tools) in order to derive to more robust decisions.

The first impressions, was positive, and this particularly with respect to improved information provision through a single viewer.

A4CENEM was used the first time during floods in Valencia in April 2019, which represented a real time experience.

A4CENEM also had an impact on the working routines, as at the national level it allows staff to anticipate warnings for decision makers (authorities) much earlier. User highlighted the following main advantages of having A4CENEM implemented at their site:

- Data integration and presentation of risks of all natural hazards in one platform
- Facilitation of exchange of emergency-related information among regional operational centres

The main drawbacks mentioned included:

- At present data of past events cannot be stored.
- High usefulness of the system for flood events cannot compare with usefulness for other hazards.
- Simultaneous examination of data of different sensors is not possible.

3.7.2 Usefulness of relevant features

In the context of the needs analysis the improvement respectively new development of the following features for particular phases of the emergency management process were requested:

- Rainfall predictions
- Run-off predictions
- Flood-related information
- Forest fire-related information
- Snowfall-related information
- Heatwave-related information
- Coldwave-related information
- Data integration capabilities
- Visualisation of meteorological warnings
- Communication of meteorological warnings

At the end of the demonstration period a coordinator as well as an operator who are familiar with A4CENEM rated the usefulness of the emergency system's capabilities with regard to the relevant features. They did so for two management scenarios, one including A4CENEM and another one excluding A4CENEM. For simplicity reasons, the former scenario is termed "A4CENEM" and the latter scenario is called "Legacy system". Experts were explicitly asked to make their judgements on the basis of their hands-on experiences and not based on their assumptions about the unfolding of A4CENEM's potential in the future. Figure 34 illustrates the results of this assessment based on the mean values of the criteria scores provided by the operator and the coordinator. The distance between the blue line ("Legacy system") and the red line ("A4CENEM") represents the improvement of the perceived usefulness of the respective features for the management process. So far, due to a lack of critical incidences this assessments primarily related to the management of natural hazards under ordinary conditions.

Figure 34 gives an overview of how the operator and coordinator comparatively assessed the usefulness of particular features for the Legacy system as well as for A4CENEM. It is evident that A4CENEM outperforms the Legacy system with regard to all features. Improvements are substantial across the board being most remarkable for *Usefulness of data integration capabilities*, *Usefulness of run-off predictions, Usefulness of flood-related information and Usefulness of rainfall predictions real time*. Minor advantages of A4COR are mentioned with regard to *Usefulness of snowfall-related, Heatwave-related and Coldwave-related information.*[129]

[129] Differences between the mean values of the scores for A4CENEM and Legacy system: *Snowfall-related information* +0.5, *Heatwave-related information* +0.5, *Coldwave-related information* +0.5, *Forest fire-related information* +1.0, *Communication of meteorological warnings* +1.0, *Visualisation of meteorological warnings* +1.5, *Run-off predictions* +2.5, *Flood-related information* +2.75, *Rainfall predictions* +3.5, *Data integration capabilities* +4.5. Positive/negative scores indicate a superior/inferior performance of A4CENEM.

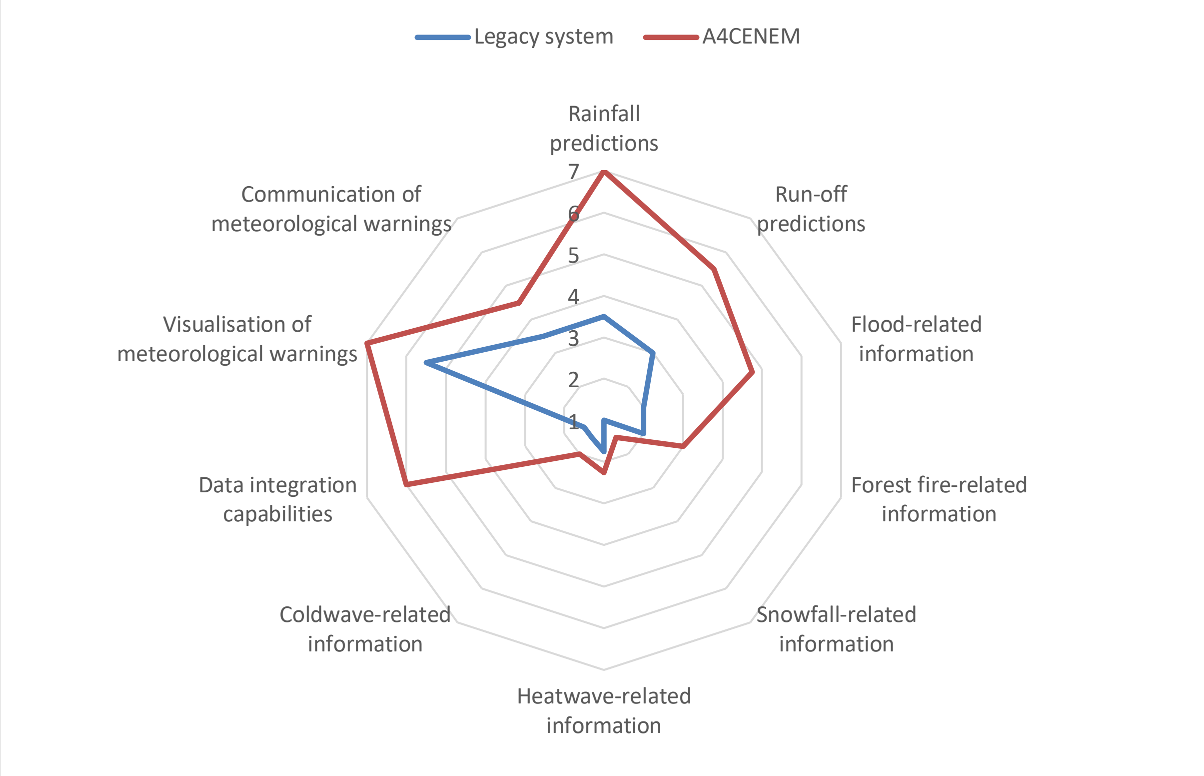


Figure 34: A4CENEM - Usefulness of relevant features (mean values)

3.8 Comparison of results across pilot sites

Performance comparison based on all criteria

In this section, we provide a generic overview of how the newly developed ANYWHERE systems are operating in comparison to the Legacy system in place in the pilot sites. Figure 35 shows how the ANYWHERE systems are performing across all pilot sites and Figure 36 shows relative differences between the new and the existing systems.[130] According to the evaluation of users, the new ANYWHERE systems perform better with respect to 14 out of 21 criteria, it performs similar with respect to four criteria and worse with respect to three criteria. However, a closer look reveals that the ANYWHERE systems will eventually perform better with respect to all criteria assessed, as it is to be expected that users' perception will change to the better with respect to the three criteria which are currently significantly negative evaluated, i.e. *Trust*, *Transparency of decision-making for insider*, *Supported languages adequacy*, by simply more often using the ANYWHERE platform in decision-making processes and, thus, build up trust with respect to the new system. More specifically (see Figure 36).

Strong advancements were made with respect to the following aspects: *Time efficiency* (+2.07 on a scale from 1 to 7), *Usefulness* (+1.85), *Ease of data processing* (+1.60),

[130] Criteria are considered for which performance data is available for at least for 4 of the 6 regional pilot sites.

Q*uality of data input for decision-making* (+1.42), *User interface aesthetics* (+1.17), *Overall satisfaction* (+1.0).

Advancements were made with respect to *Support of decision-making* (+0.83), *Knowledge about relevant impacts* (+0.82), *Transparency of decision-making for outsiders* (+0.81), *Knowledge about relevant risks* (+0.81), *Quantity of data input for decision-making* (+0.78), *Hazard coverage* (+0.48), *Flexibility* (+0.33), *Learnability - self-explanatory user interface* (+0.25).

Both systems perform similar with respect to *Learnability - User guidance completeness* (+0.08), *Stress* (+0.00), *Service availability* (-0.08) and *Functional appropriateness* (-0.08).

The Legacy system is performing better with respect to three criteria: *Trust* (-0.66), *Transparency of decision-making for insiders* (-0.70) and *Supported languages adequacy* (-2.4). It is to be expected, that the perception of all three criteria can change to the positive with a continuous and increasingly intensive use of the ANYWHERE systems and with expected updates for the language support.

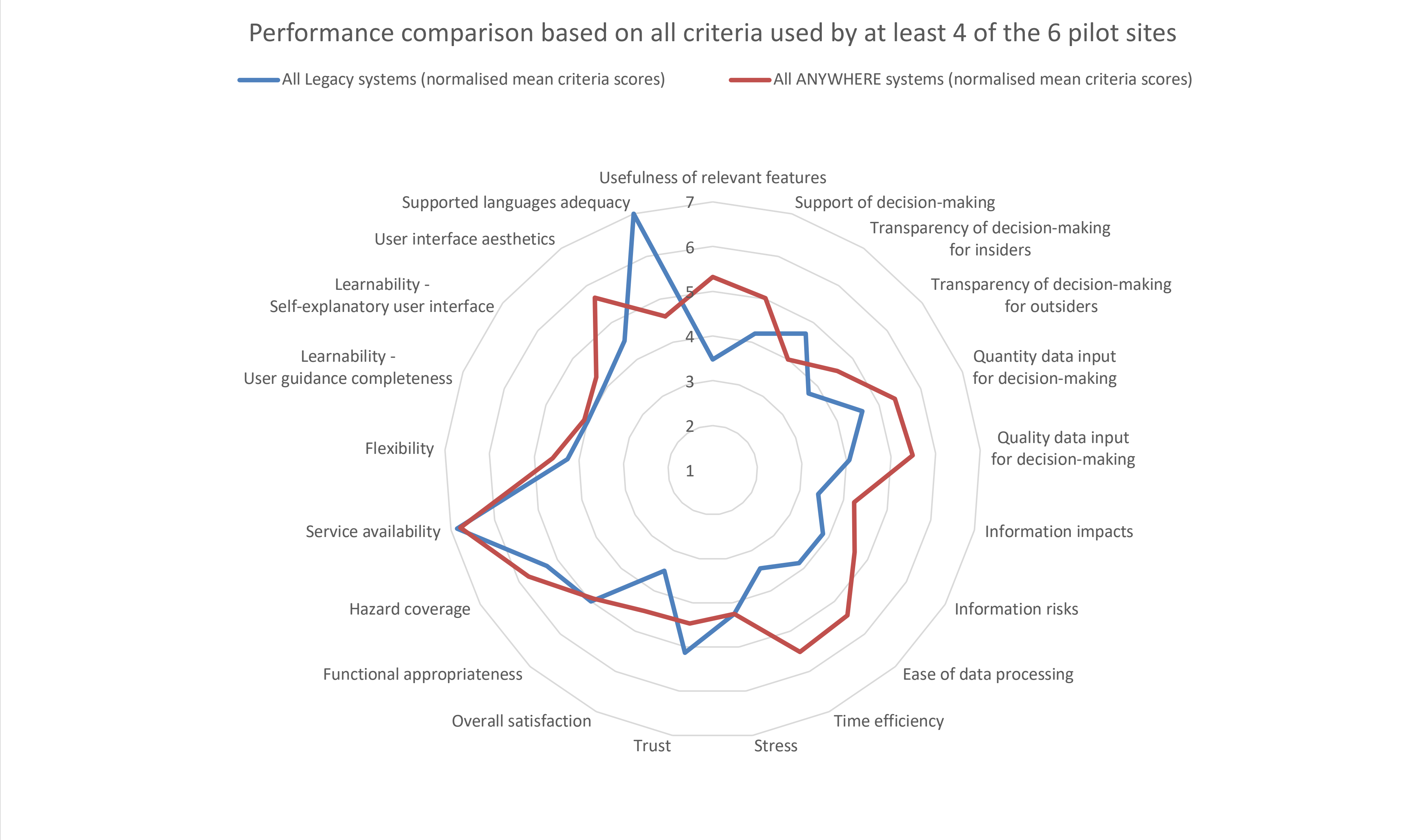


Figure 35: Performance comparison of ANYWHERE systems and Legacy systems: All criteria

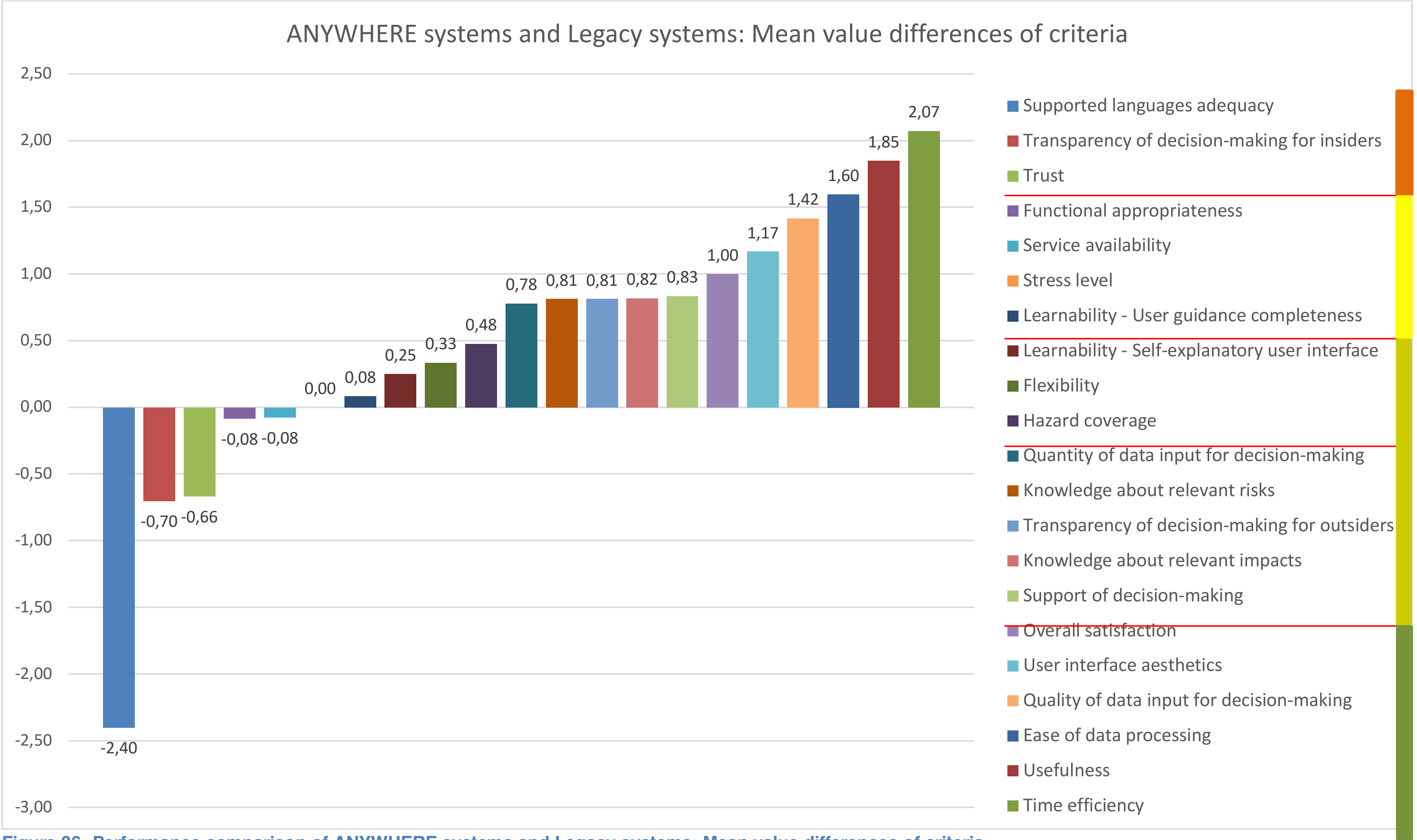


Figure 36: Performance comparison of ANYWHERE systems and Legacy systems: Mean value differences of criteria

4 Discussion and conclusion

4.1 Challenges

Co-evaluation object and context

When conducting such co-evaluations the definition of the objects to be compared forms the first important step of the procedure. In this particular case defining distinct emergency management systems was simply not feasible. Even though there are difference in the degree of substitution for most pilot sites – at least at this stage – the ANYWHERE system rather complements than substitutes the Legacy systems. If substitution occurs it is related to particular functionalities which are offered by both systems. But even in this case information from different, often independent sources is used for cross-checking in order to at least subjectively decrease uncertainty. Hence, we decided to compare scenarios, i.e. the baseline scenario without the ANYWHERE system and the ex-post scenario including the ANYWHERE system.[131] From a theoretical perspective this co-existence made the comparison more fuzzy, but after having clarified at the beginning of the data collection interviews what is to be compared interviewees were very capable of thinking in these two scenarios and answer accordingly.

Another key challenges for the co-evaluation was that each system installed was unique and that each pilot site was uniquely defined by a specific set of context conditions. Therefore, we took great effort to map relevant context conditions and to understand how different stakeholder are interacting and relying on early warning platform. In addition, the ANYWHERE systems were changing and evolving during the demonstration period as the systems were constantly updated and adapted according to the requirement of users. Taking this flux, into account, we also needed to constantly update our data collection strategy as well as the underlying evaluation methodology.

Data collection

Generally, we decided to use qualitative data for the co-evaluation. This decision is grounded both in a lack of statistical, more quantitative data and our ambition to understand the added value of the ANYWHERE system for those in actual use of the system. Therefore, we focused on their perception and valuation of the news systems in comparison to the established emergency management systems. We took great effort to identify the right persons to obtain data from by means of semi-structured interviews containing both closed and a set of open questions. These persons needed to be engaged in the co-development process of the new platform, they needed to be knowledgeable (i.e. use the new platform as well as the already existing platform) and they needed to be able and willing to share experiences. The interviews were

[131] Still we decided that for comprehensibility reasons in this report we speak of systems rather than scenarios.

conducted at the actual location where the platform is used to be able to contextualizes answers-

Co-evaluation method

We decided to apply a multi-criteria analysis as this method has allowed us to use qualitative data sourced through interviews from the users, the consideration of their vision of what performance aspects have to be taken into consideration and what weight they should have in the final aggregation of the preference scores. This method allows us, thus, to systematically capture and compare different aspects and catering for different perspectives that are underlying the results of the co-evaluation methodology. Through this, we were able to make judgement of the experts, who are most familiar with the systems comprehensible and traceable by systematically dismantling the overall assessment.

4.1 Further benefits of the co-production of the ANYWHERE platform

We would like to also highlight an additional asset of the co-production and co-evaluation approach taken by the ANYWHERE project. We observed considerable learning effects as the project was progressing. This included, above all, peer-to-peer learning across sites. Many pilot site partners were actively exchanging with other pilot-site partners both on the usability of the platform, but also on advancements in emergency management in more general terms. This included also the assessments of usefulness of existing extreme weather-related products (not only the ones included in A4*). Relevant for the development process, was also the iterative needs refinement as basis for further developments (tools, systems) after ANYWHERE ends.

4.2 Conclusion

The goal of this task was to develop and apply a generic framework which as on the one hand flexible enough to be adapted to the situation at the different sites defined by different context conditions, specific needs, preference expectations, starting levels etc. and on the other hand the framework should be ‘stable’ enough the rigorously compare evaluation outcomes between the pilot sites. This goal was achieved by applying a Multi-criteria analysis, which also supported the co-production process which was one of the pillars of the ANYWHERE project.

In this report, we shed some light at the end of the demonstration process not only on the overall performance of the ANYWHERE systems compared to the Legacy system but more specifically on the pros and cons, on what was achieved and what aspect there is still room for improvement. The analysis revealed that the new ANYWHERE systems perform better with respect to 14 out of 21 criteria, it performs similar with respect to four criteria and there is room for improvement with respect to three criteria. However, it is to be expected that this improvement is to be achieved within a more intense use of the ANYWHERE system as the criteria related above all to the trust in the new system, which will grow through use.

5 Scientific advancements of ANYWHERE

The ANYWHERE project has been focussing on innovation that developed more effective early warning systems (MH-EWS) and decision-support systems (A4* platforms), and tailored online services developed to support self-preparedness, self-protection, and self-response of citizens across Europe. Although innovation in ANYWHERE was mainly based on existing knowledge, the project also increased scientific knowledge, for instance, by synthesing knowledge, or by progressing knowledge as a response to reported limitations by end-users. Scientific outreach was foreseen in the ANYWHERE project, that is, up to 30 scientific papers with publishing confirmed in first tier journals (no minimum was specified).

We collected information from (reference date: 14 November 2019):

- EU SyGMa: System of Grant Management, Continues Reporting, Publications;
- EVISE: Elsevier Editorial System, Environment International, Virtual Special Issue (EI VSI).

In total 79 ANYWHERE publications are reported in SyGMa. After excluding abstracts, conference proceedings and theses, and adding accepted papers in EI VSI not reported yet in SyGMa, 43 peer-reviewed scientific papers are left (Annex X.1), called the ANYWHERE papers. There is a steady growth of ANYWHERE papers over time (Table X.1). As expected, the number of papers in the last part of the project are higher than in the first part, because it takes some time to develop and to publish knowledge.

Table 1: Distribution of scientific papers over the ANYWHERE project lifetime

Year	Number of papers
2016	1
2017	7
2018	16
2019	19

The number of peer-reviewed scientific papers in 2019 will likely be higher than mentioned in Table 1. There are still some papers under review, among others in the Special Issue of Environment International.

The 43 scientific papers have been published in 26 different peer-reviewed scientific journals (Table 3). Four papers is the maximum number of papers in a specific journal, i.e. in Natural Hazards and Earth System Sciences (NHESS), Hydrology and Earth System Sciences (HESS) and Environment International. We have clustered the journals in a number of categories (Figure 37). Most of the journals, about one third, are climate-weather-related (atmosphere), which are followed by water-related journals (hydrology).

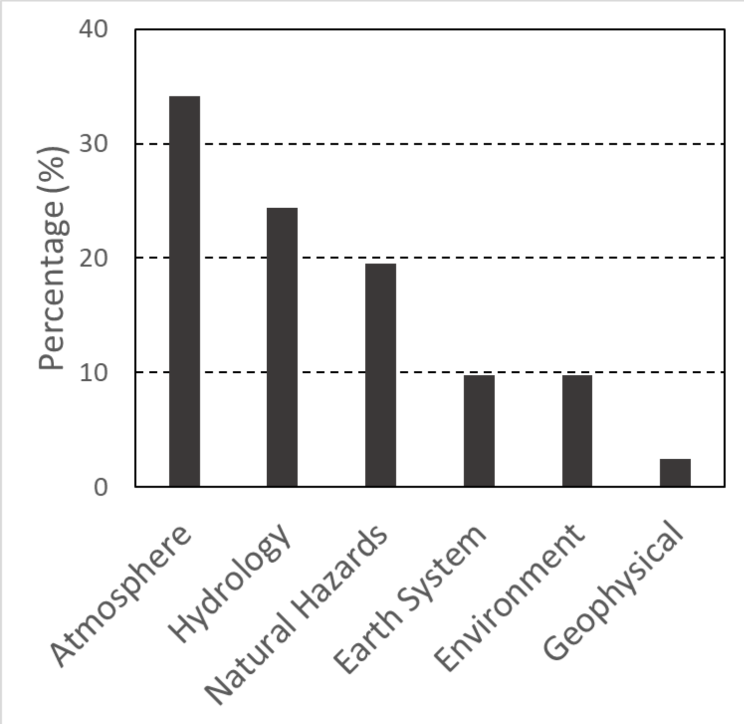


Figure 37: Assignment of ANYWHERE scientific papers to journal categories

We also investigated the importance of the scientific journals in which the ANYWHERE papers are published. There are several metrics to measure the importance or impact of the journals in the academic community. We have selected the frequently-used Journal Impact Factor (JIF). The JIF counts the number of citations received by papers from a journal in a particular year and divides them by the total number of citable papers published in that journal in the past two years. We obtained the JIF for 2018 from the SCImago Journal & Country Rank[132], which is a publicly available portal that includes the journals scientific indicators developed from the information contained in the Scopus® database.

We have plotted the JIF of the journals in which ANYWHERE papers have been published (Fig. X.2, solid line), called ANYWHERE journals. Annex X.1 gives the JIF of each journal. It appears that almost all journals (97.8%) have a JIF ≥ 1. To put this number in context, we used the Journal Citation Reports (JCR)[133] database, which tracked all impact factors for almost 13,000 journals in 2017, called the JIF reference. Figure X.2 (dashed line) shows the percentage of journals that were assigned impact factors ranging from 0 to 10+. Approximately two-thirds of the journals tracked by JCR have a JIF ≥ 1. This implies that the ANYWHERE journals have a higher impact, i.e. about 27% more journals have a JIF ≥ 1 relative to the reference. The higher impact is also supported by that about 60% of the ANYWHERE journals have a more than twice as high JIF than the JCR reference, that is, a JIF of 4.5 versus 2.0. In the high-end the JIFs become similar. The percentage of ANYWHERE journals with a JIF of 8+ is almost equal to the reference, both have about 2.5% journals with a JIF larger than 8+. The

[132] SCImago, (n.d.). SJR — SCImago Journal & Country Rank [Portal]. Retrieved Date you Retrieve, from http://www.scimagojr.com.
[133] Journal Citation Reports (JCR) https://jcr.clarivate.com

highest JIF of a journal in which an ANYWHERE paper has been published, is almost 12 (Annex X.1).

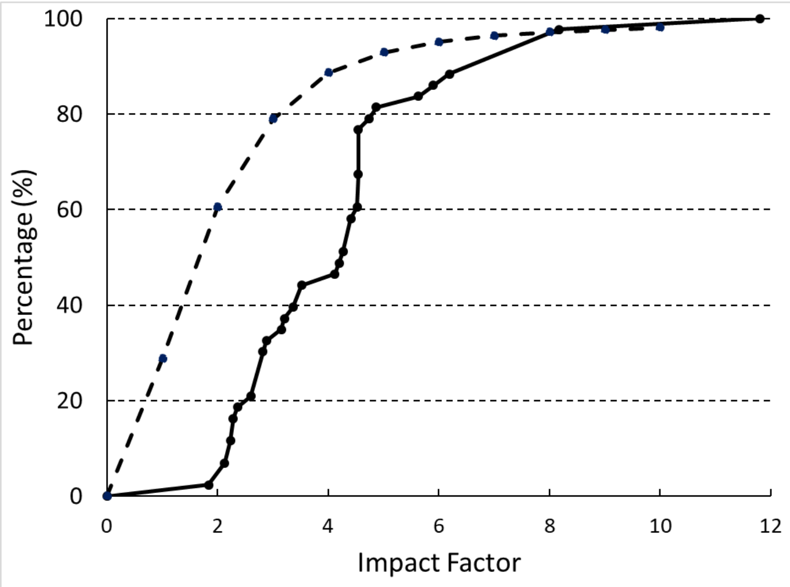


Figure 38: Impact Factor (JIF) of the peer-reviewed scientific journals in which ANYWHERE papers are published

Note: (▬▬), and the percentage of all journals that were assigned Impact Factors ranging from 0 to 10+ (▬▬), JIF reference.

Over 125 authors have contributed to the ANYWHERE papers. Most of the authors are employed by ANYWHERE partners, which, however, implies that also knowledge from beyond the ANYWHERE consortium have been incorporated in the scientific achievements. Co-writing with externals is an important mechanism to extend the knowledge domain. On average each paper has about 3 authors.

The authors list of each of the papers is an indication of collaboration among ANYWHERE partners, which is a key point to synthesize knowledge for development of the integrated MH-EWS and A4* platforms.

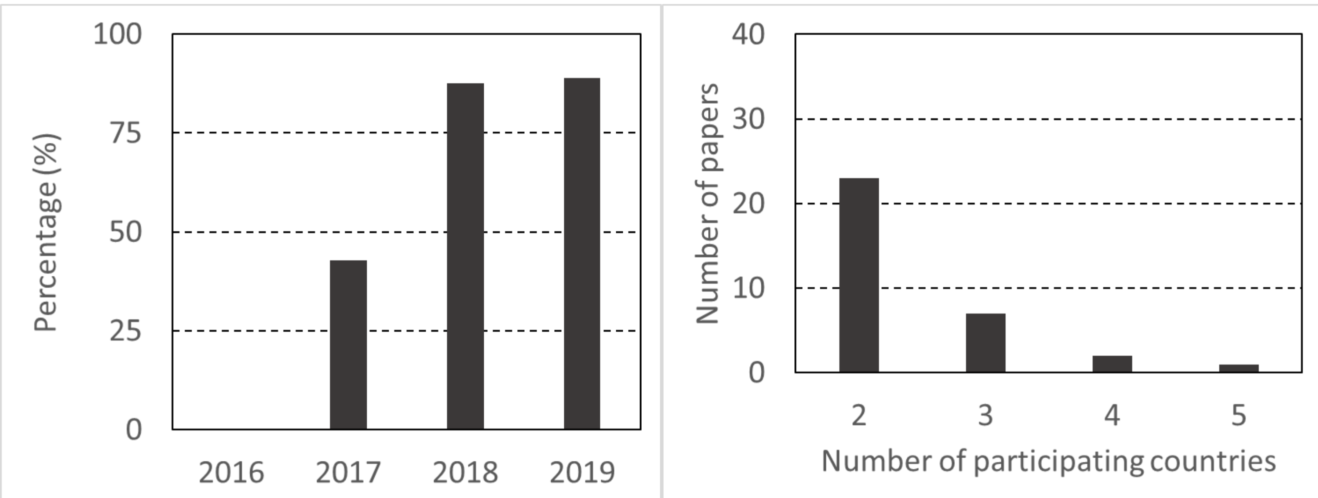


Figure 39: Collaboration among partners

Note: Left: percentage of ANYWHERE papers with authors coming from two or more different countries. Right: number of ANYWHERE papers with authors coming from different countries.

As expected, cooperation of authors from different countries has been growing during the lifetime of ANYWHERE (Fig. X.3, left), because scientists need time to get to know individual competences. In the last part of the project over 80% of the ANYWHERE papers have been written by authors coming from ≥ 2 countries. About three quarter of ANYWHERE papers have been compiled by authors coming from two or more countries. Ten papers (about 25%) by authors coming from ≥ 3 countries.

In conclusion:

ANYWHERE has overperformed by compiling 43 peer-reviewed scientific papers (target was up to 30 articles);

More than 80% of the scientific journals in which the ANYWHERE papers have been published, have a higher impact than the reference;

External knowledge has been incorporated by co-writing with authors from beyond the ANYWHERE consortium;

The cooperation among ANYWHERE partners thorough jointly writing scientific papers has generated added knowledge. The vast majority of the papers in the last part of the project have been written by authors from two or more countries.

6 References

Cucco, I., S. De Rosa, A. Passani, A. Georgiev, J. Pottebaum, C. Schäfer, T. Sauerland, K. Petersen, and I. Danilidis, I., 2016: ***SecInCoRe Deliverable 5.3, Validation strategy and first functional evaluation model of communication system concept.***

Fink, R. D., 2014: **Vertrauen in autonome Technik. Modellierung und Simulation von Mensch-Maschine-Interaktion in experimentell-soziologischer Perspektive.** PhD thesis, TU Dortmund.

Jian, J.-Y., Bisantz, A. M., & Drury, C. G., 2000: **Foundations for an empirically determined scale of trust in automated systems**. International Journal of Cognitive Ergonomics, 4(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Monacciani, F., M. Navarra, A. Passani, and F. Bellini, 2011: ***D3.3a - SEQUOIA Final Self-Assessment Methodology.***

Rogers, E. M., 2003: ***Diffusion of innovations (5th ed.)***, New York: Free Press.

Zardari, N. H.; Ahmed, K.; Moniruzzaman Shirazi, S.; Yusop, Z., 2015: **Weighting Methods and their Effects on Multi-Criteria Decision Making Model Outcomes in Water Resources Management**. SpringerBriefs in Water Science and Technology, Springer, https://doi.org/10.1007/978-3-319-12586-2

7 ANNEX 1

7.1 Overview of criteria used to co-evaluate ANYWHERE platform at the pilot sites

Table 2: Overview of criteria used to co-evaluate ANYWHERE platform at the pilot sites

No	Impact area	Effects	Criterion	Sub-criterion	Indicator	Unit	A4CAT	A4LIG	A4ALPS	A4FINN	A4NOR	A4COR
1	Economic impact	Financial effects	False alarm-related costs (organisational level)		Total annual false alarm-related costs at the level of the regional civil protection authority.	EUR						
2	Economic impact	Financial effects	System-related costs	One-time costs	Total onetime investment costs due to the implementation of the emergency management system occurring at the regional civil protection authority.	EUR						
3	Economic impact	Financial effects	System-related costs	Operational costs	Total annual costs due to operation of the emergency management system at the regional civil protection authority.	EUR						
4	Economic impact	Financial effects	System-related costs	Maintenance costs	Total annual costs due to maintenance of the emergency management system at the regional civil protection authority.	EUR						
5	Economic impact	Economic effects	False alarm-related costs (municipal / regional level)		Total annual false alarm-related costs with exception of the costs occurring at the level of the regional civil protection authority.	EUR						
6	Economic impact	Economic effects	Damages occurred	Direct effects	Total annual damages due to natural hazards in the operational area of the regional civil	EUR						

					protection authority (site-specific sets of hazards).							
7	Economic impact	Economic effects	Damages occurred	Indirect effects	Total annual costs due to business interruption.	EUR						
8	Economic impact	Economic effects	People affected	Fatalities	Ratio of fatalities to number of events managed.	%						
9	Economic impact	Economic effects	People affected	Casualties	Ratio of casualties to number of events managed.	%						
10	Economic impact	Economic effects	People affected	Evacuated	Ratio of people evacuated to number of events managed.	%						
11	Economic impact	Economic effects	People affected	Impaired	Ratio of people impaired to number of events managed.	%						
12	Economic impact	Economic effects	Resource usage	Energy consumption	Energy used to operate the system (per hour/ day/annum).	kWh						
13	Economic impact	Economic effects	Resource usage	Consumable supplies	Quantity of consumable supplies used for operating the system per annum.	Divers e						
14	Economic impact	Economic effects	Resource usage	Travel effort	Travel costs occurring per annum for keeping the system operational.	EUR						
15	Economic impact	Economic effects	System outputs	Waste	Quantity of waste produced by operating the system per annum.	Divers e						
16	Social impact	Effects on working routines	Efficiency	Time efficiency	Time spent for the execution of typical management activities.	min						
17	Social impact	Effects on working routines	Efficiency	Stress	Level of stress caused by the operation of the system.	Likert scale score						
18	Social impact	Effects on working routines	Overall satisfaction		Net promoter score (NPS)	NPS						
19	Social impact	Effects on working routines	Usefulness		Level of usefulness of specific features, which are relevant for the emergency	Likert scale score						

					management process (site-specific set of features).							
20	Social impact	Effects on working routines	Support of decision-making		Level of decision-making support in specific phases of the emergency management process.	Likert scale score						
21	Social impact	Effects on working routines	Transparency of decision-making	Transparency of decision-making for insiders	Level of transparency of decision-making for “insiders“, i.e. public protection and disaster relief organizations.	Likert scale score						
22	Social impact	Effects on working routines	Transparency of decision-making	Transparency of decision-making for outsiders	Level of transparency of the decision-making processes for “outsiders”, i.e. citizens, politicians, non-PPDR organizations (e.g. in the aftermath of an event).	Likert scale score						
23	Social impact	Effects on knowledge production and sharing	Knowledge production	Inputs	Quantity of data input for decision making in specific phases of the emergency management process.	Likert scale score						
24	Social impact	Effects on knowledge production and sharing	Knowledge production	Inputs	Quality of data input for decision making in specific phases of the emergency management process.	Likert scale score						
25	Social impact	Effects on knowledge production and sharing	Knowledge production	Processing	Ease of data processing for specific relevant system outputs (site-specific outputs).	Likert scale score						
26	Social impact	Effects on knowledge production and sharing	Knowledge production	Outputs	Level of information about specific relevant hazards (site-specific set of hazards).	Likert scale score						
27	Social impact	Effects on knowledge production and sharing	Knowledge production	Outputs	Level of information about specific relevant risks (site-specific set of risks).	Likert scale score						

28	Social impact	Effects on knowledge production and sharing	Knowledge production	Outputs	Level of information about specific relevant impacts (site-specific set of impacts).	Likert scale score						
29	Social impact	Effects on knowledge production and sharing	Knowledge production	Outputs	Level of information about management options (site-specific set of management options).	Likert scale score						
30	Social impact	Effects on knowledge production and sharing	Knowledge production	Trust, Mistrust	Degree to which the users trust the system using validated trust questionnaire (Jian, Bisantz, Drury 2000; Fink 2014)	Psych o metric scale value						
31	Social impact	Effects on knowledge production and sharing	Knowledge sharing	Knowledge dissemination (one-way communication)	Speed of information dissemination to emergency forces in case of an emergency.	min						
32	Social impact	Effects on knowledge production and sharing	Knowledge sharing	Knowledge dissemination (one-way communication)	Number of recipients of warnings.	no.						
33	Social impact	Effects on knowledge production and sharing	Knowledge sharing	Knowledge dissemination (one-way communication)	Ease of information dissemination to wider public in case of an emergency.	Likert scale score						
34	Social impact	Effects on knowledge production and sharing	Knowledge sharing	Knowledge dissemination (one-way communication)	Ease of information dissemination to emergency forces in case of an emergency.	Likert scale score						
35	Social impact	Effects on knowledge production and sharing	Knowledge sharing	Knowledge exchange (reciprocal communication)	Ease of knowledge exchange with internal partners.	Likert scale score						

36	Social impact	Effects on knowledge production and sharing	Knowledge sharing	Knowledge exchange (reciprocal communication)	Ease of knowledge exchange with external partners.	Likert scale score						
37	Technolog ical impact	Technical effects	System charac-teristics	Hazard coverage	Ratio of hazards, which can be managed using system to total number of relevant hazards, which need to be managed.	%						
38	Technolo-gical impact	Technical effects	System charac-teristics	System Integration	Ratio of interfaces with or links to other tools and platforms, which are relevant for managing disaster risks, to total number of these tools and platforms.	%						
39	Technolo-gical impact	Technical effects	System charac-teristics	Functional appropriate-ness	Degree to which number of functions provided by the system is adequate for users to achieve their objectives appropriately.	Likert scale score						
40	Technolo-gical impact	Technical effects	Flexibility	Proficiency independence	Level of usability of the system by operators with different degrees of proficiency.	Likert scale score						
41	Technolo-gical impact	Technical effects	Learnability	User guidance completeness	Degree to which system functions are explained in sufficient detail to enable user to operate the system (e.g. by pop-up windows).	Likert scale score						
42	Technolo-gical impact	Technical effects	Learnability	Self-explanatory user interface	Degree to which system can be operated by a first-time user without prior study or training or seeking external assistance.	Likert scale score						
43	Technolo-gical impact	Technical effects	User interface aesthetics		Degree to which user interfaces and the overall design are aesthetically pleasing in appearance.	Likert scale score						
44	Technolo-gical impact	Technical effects	Accessibility	Accessibility for users with visual impairment	Degree to which users with visual impairment can successfully operate the system.	Likert scale score						
45	Technolo-gical impact	Technical effects	Accessibility	Supported languages adequacy	Ratio of languages supported by the system to languages, in which system should be available for being operated by any given	%						

					(potential) user at the civil protection authority.							
46*	Technolo-gical impact	Technical effects	Service Availability		Ratio of time in which system was ready to be used to time in which should have been available for use.	%						

*added after official criterion-selection process based on feedback in preparation of data collection interviews

7.2 A4CAT – Overview of criteria categories

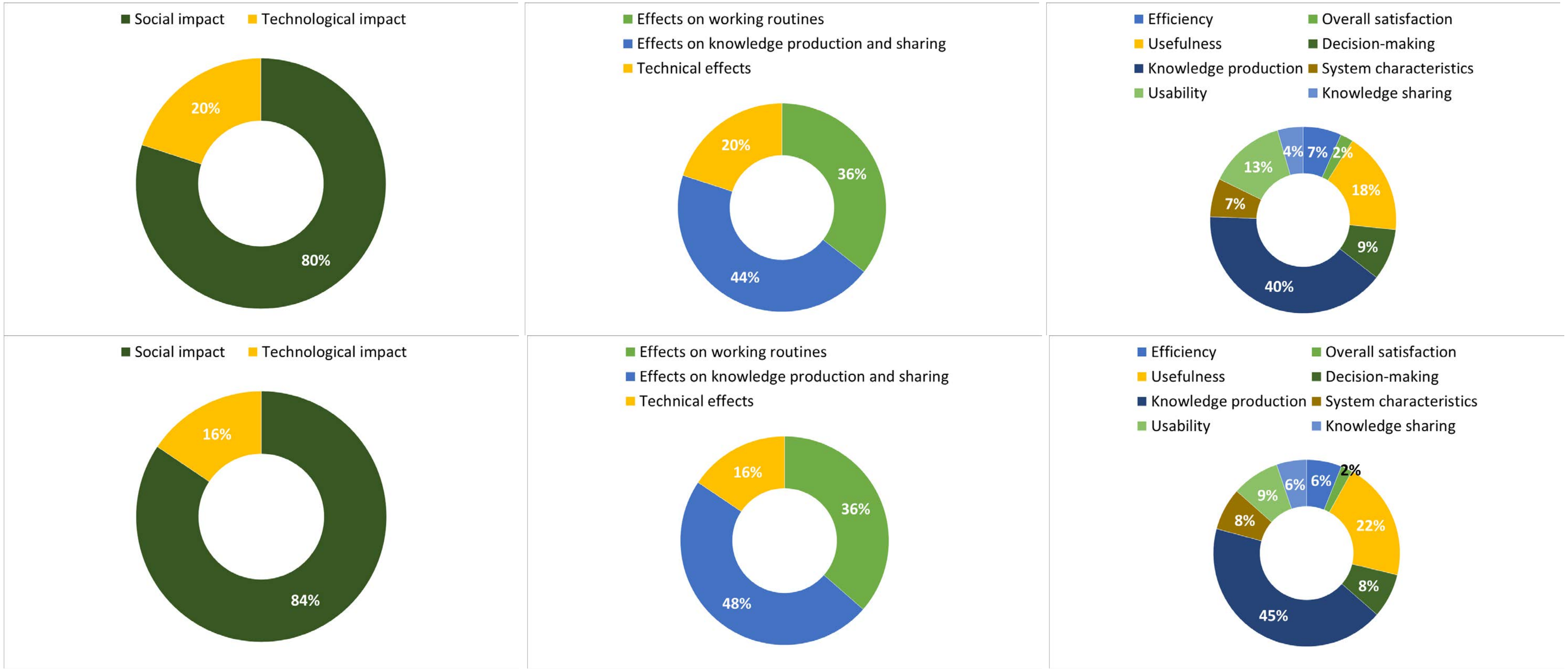


Figure 40: A4CAT - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom)

7.3 A4LIG – Overview of criteria categories

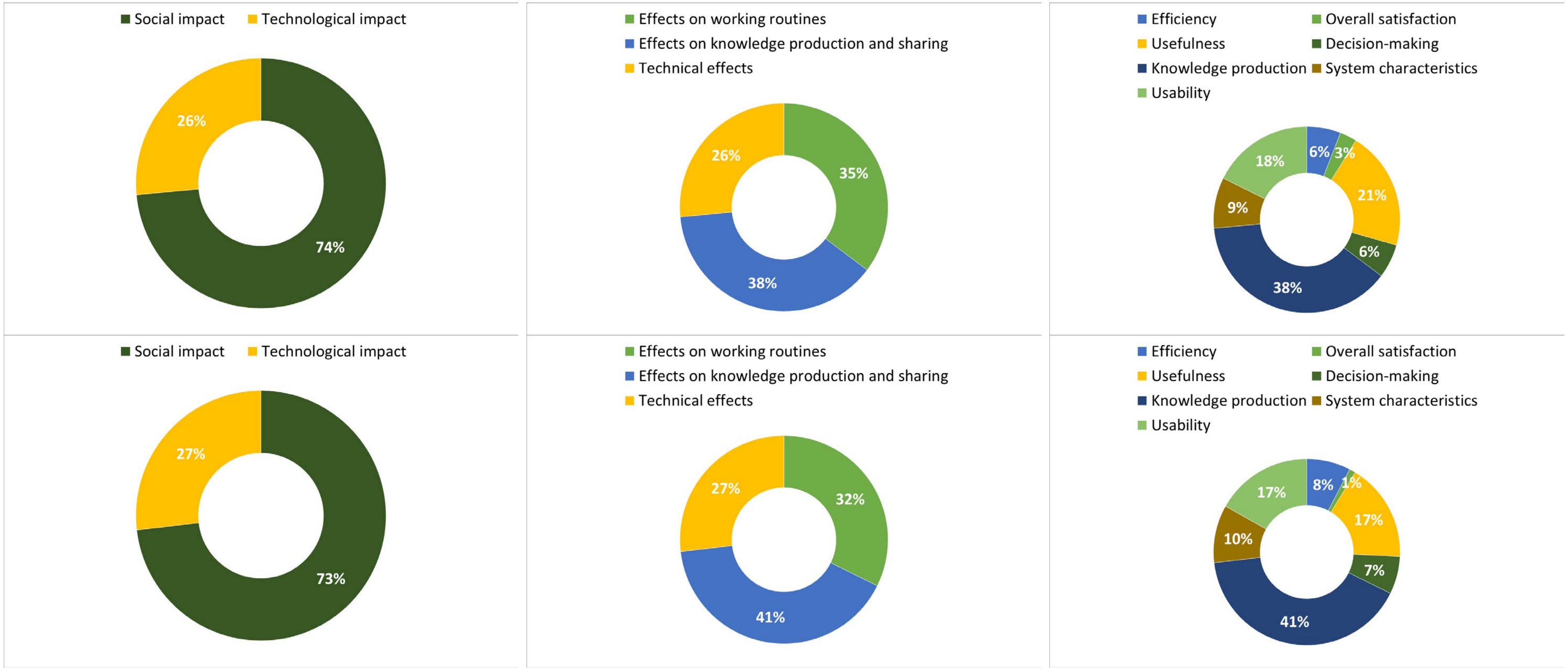


Figure 41: A4LIG - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom)

7.4 A4FINN – Overview of criteria categories

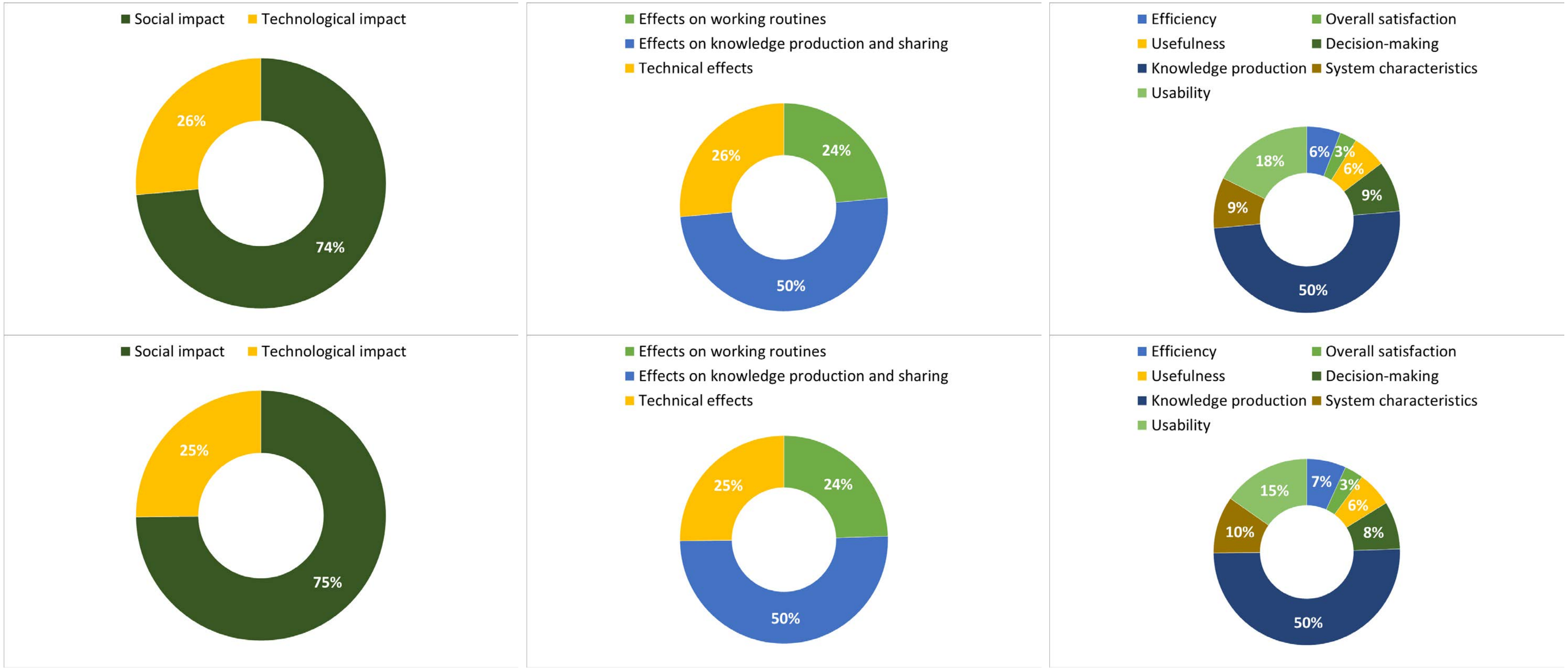


Figure 42: A4FINN - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom)

7.5 A4NOR – Overview of criteria categories

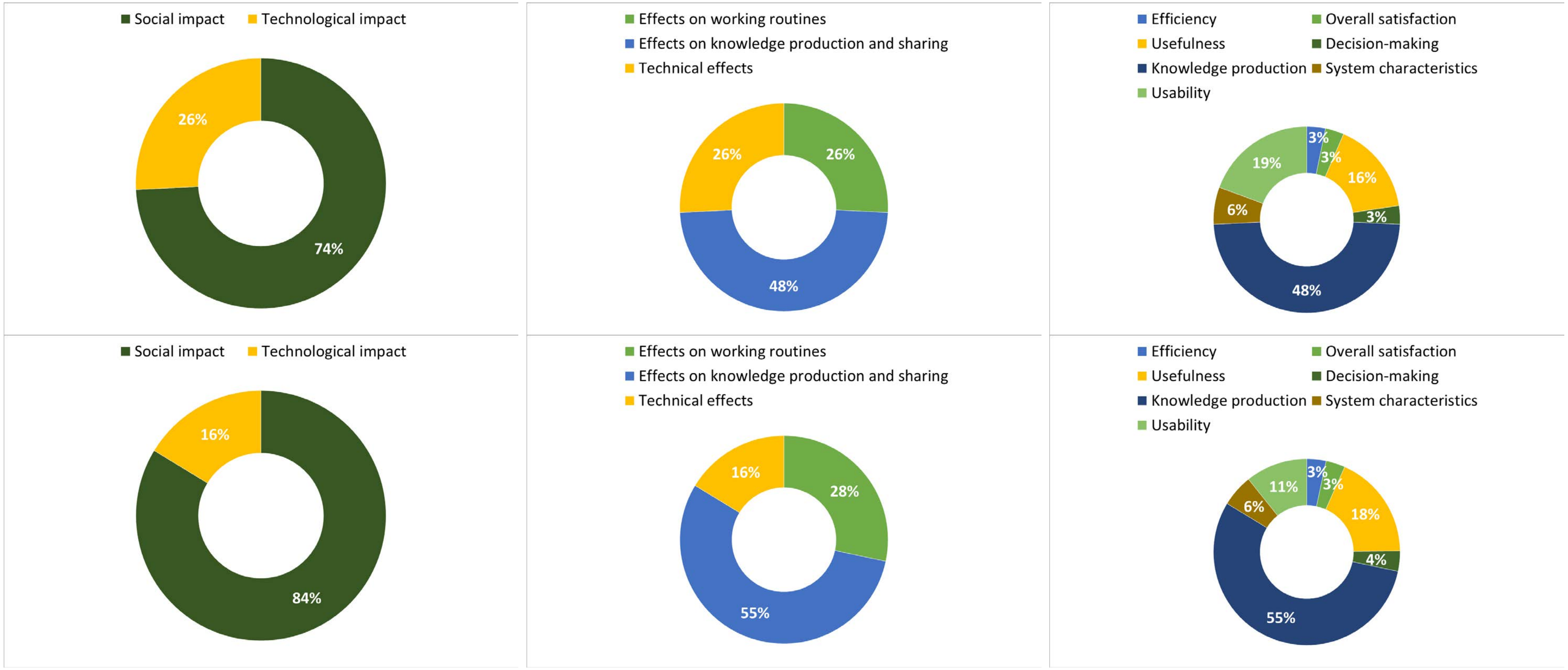


Figure 43: A4NOR - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom)

7.6 A4COR – Overview of criteria categories

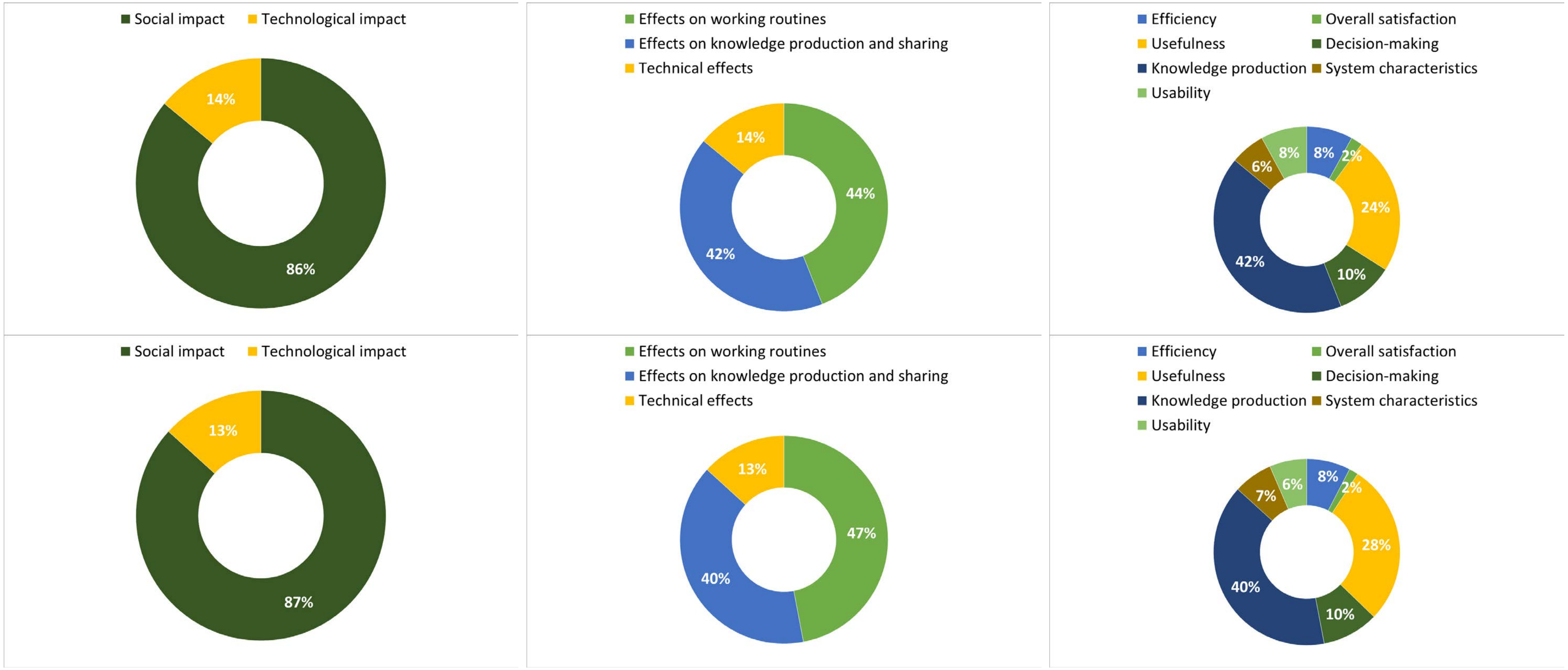


Figure 44: A4COR - Criteria used for the co-evaluation by categories in absolute numbers (top) and by relative importance (bottom)

7.7 A4ALPS - Relative importance of all co-evaluation criteria

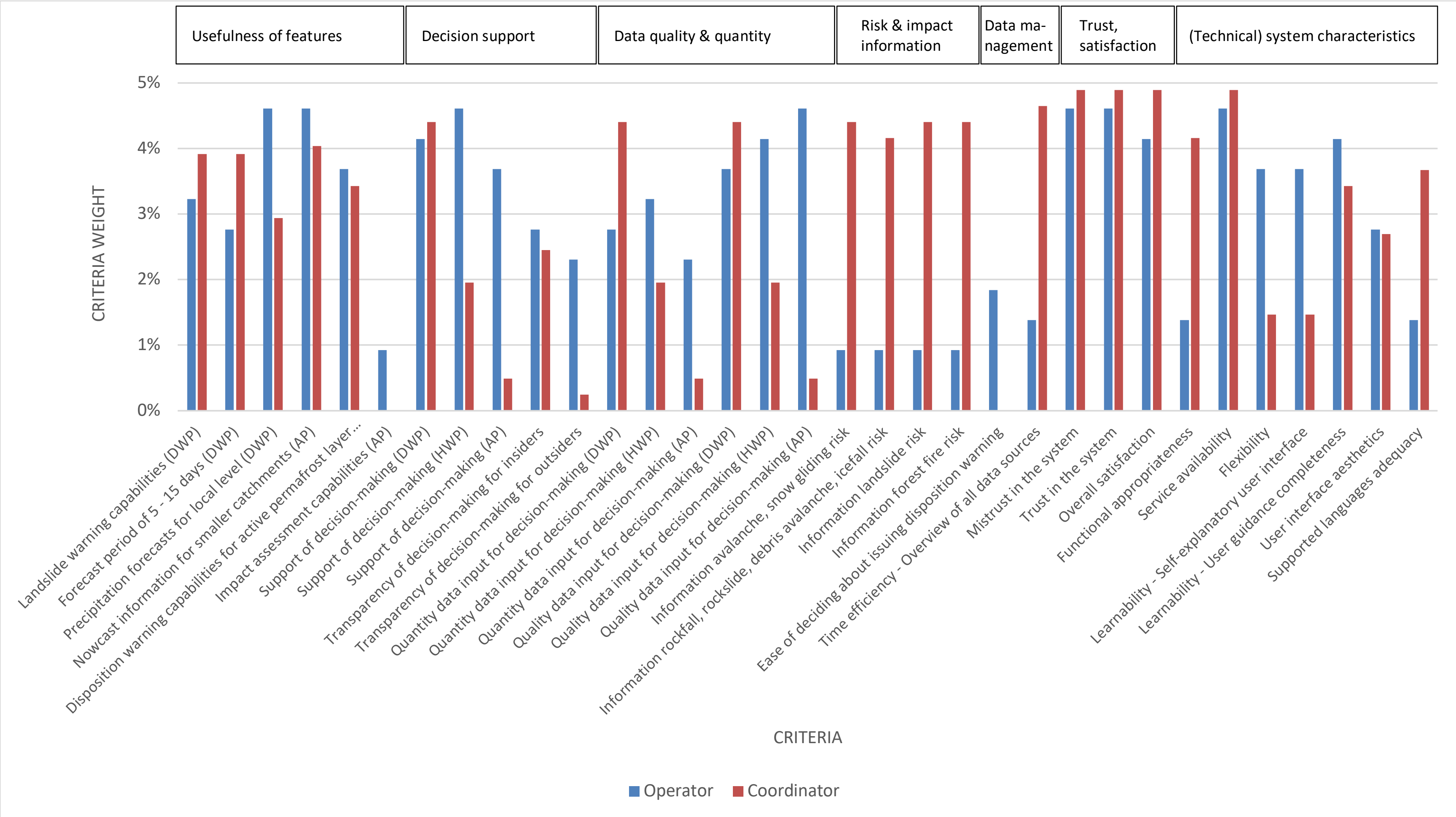


Figure 45: A4ALPS - Relative importance of all co-evaluation criteria

7.8 A4ALPS - Relative importance of criteria for comparison of operator's and coordinator's perspectives

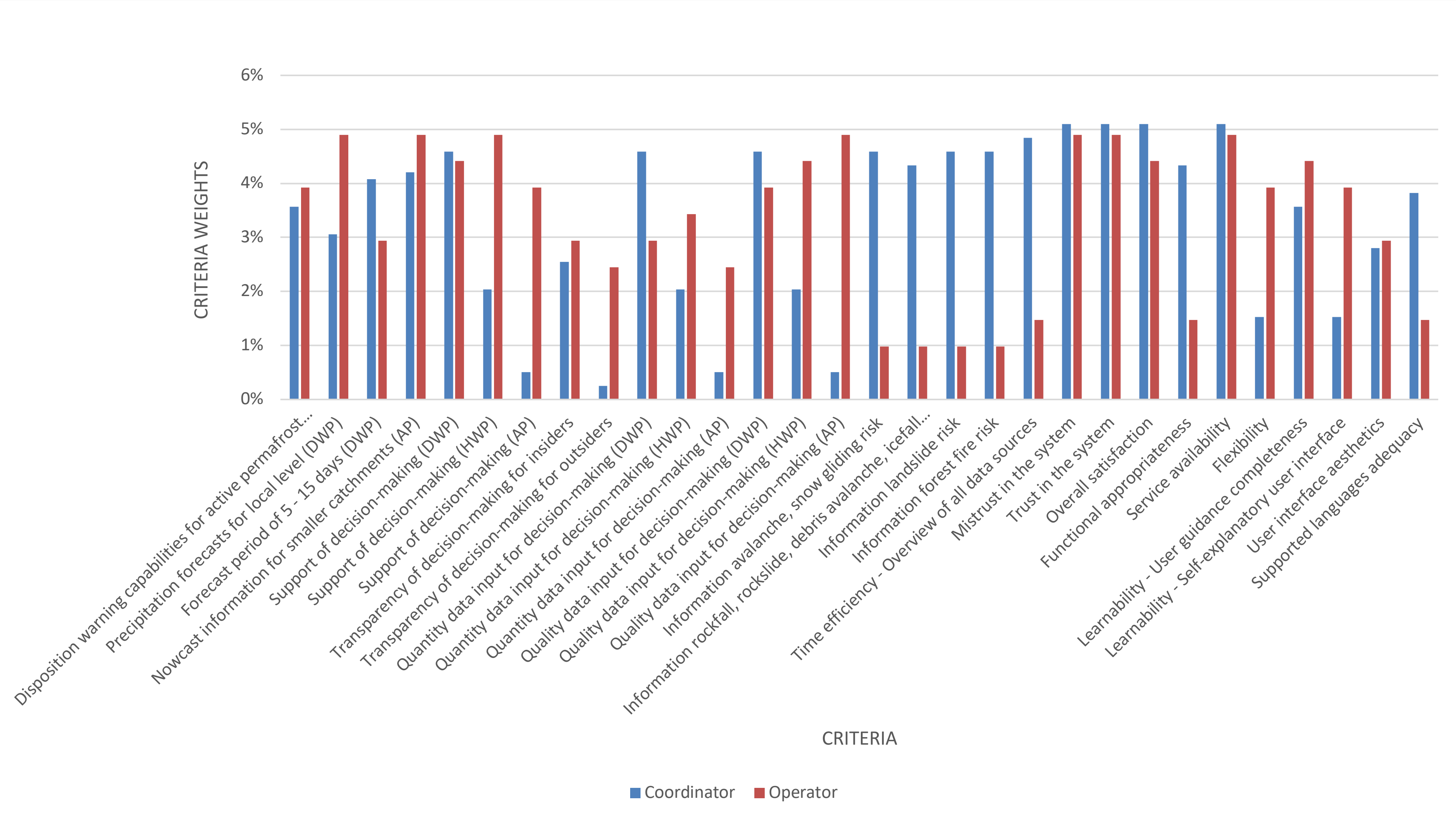


Figure 46: A4ALPS - Relative importance of criteria for comparison of operator's and coordinator's perspectives

7.9 A4ALPS - Performance comparison with Legacy system (operator)

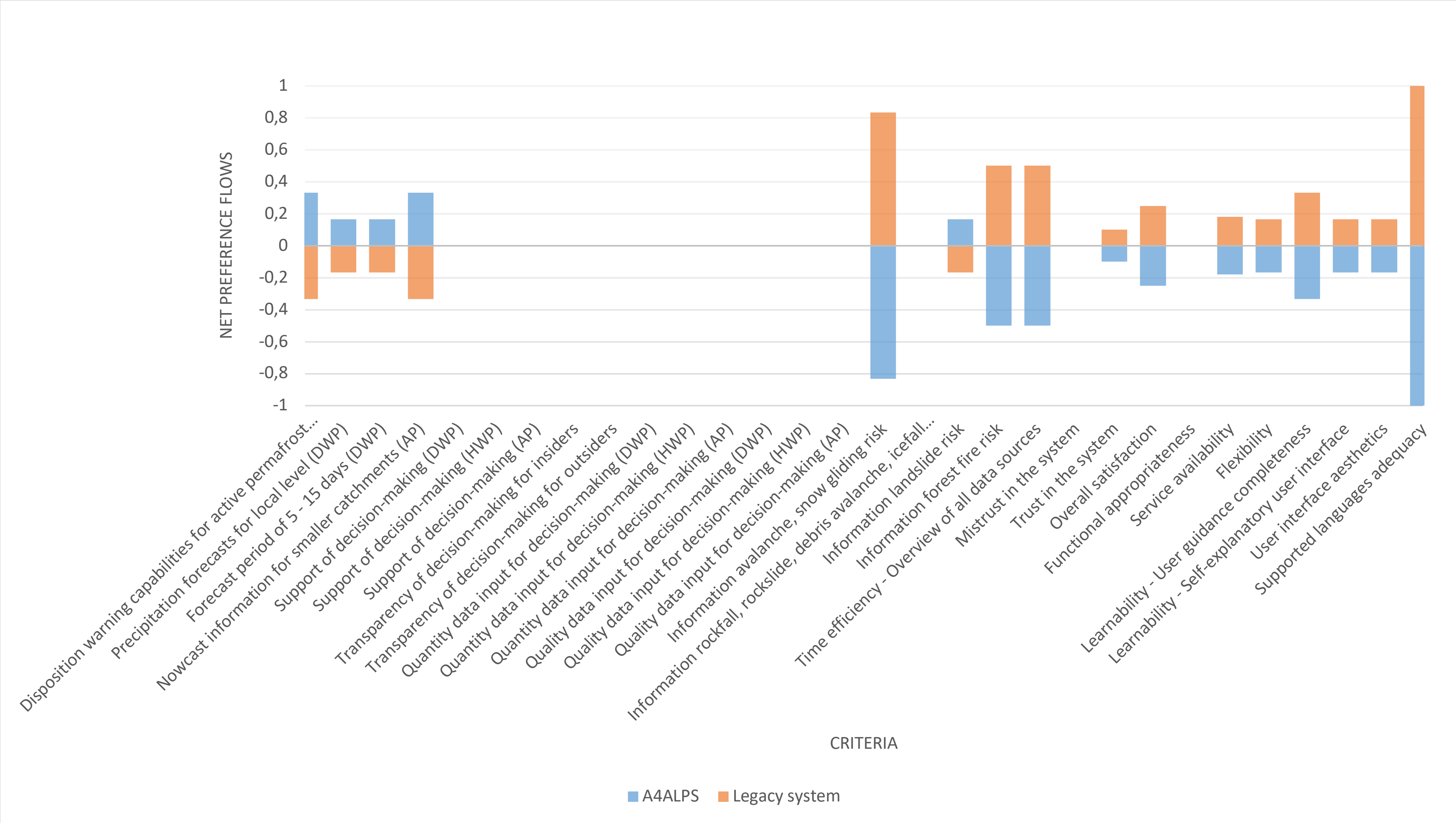


Figure 47: A4ALPS - Performance comparison with Legacy system (operator)

7.10 A4ALPS - Performance comparison with Legacy system (coordinator)

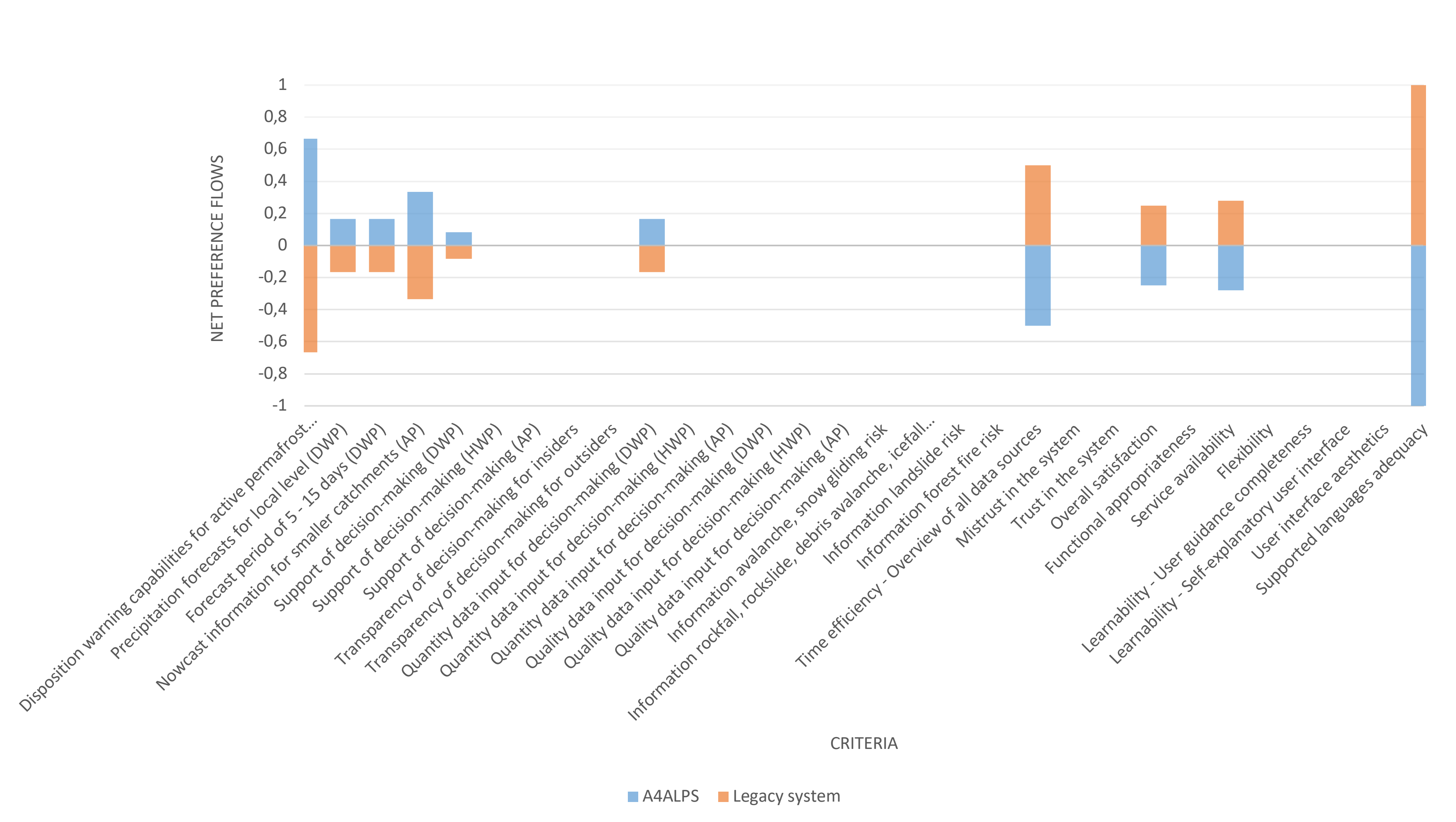


Figure 48: A4ALPS - Performance comparison with Legacy system (coordinator)

7.11 A4CAT - Relative importance of all co-evaluation criteria

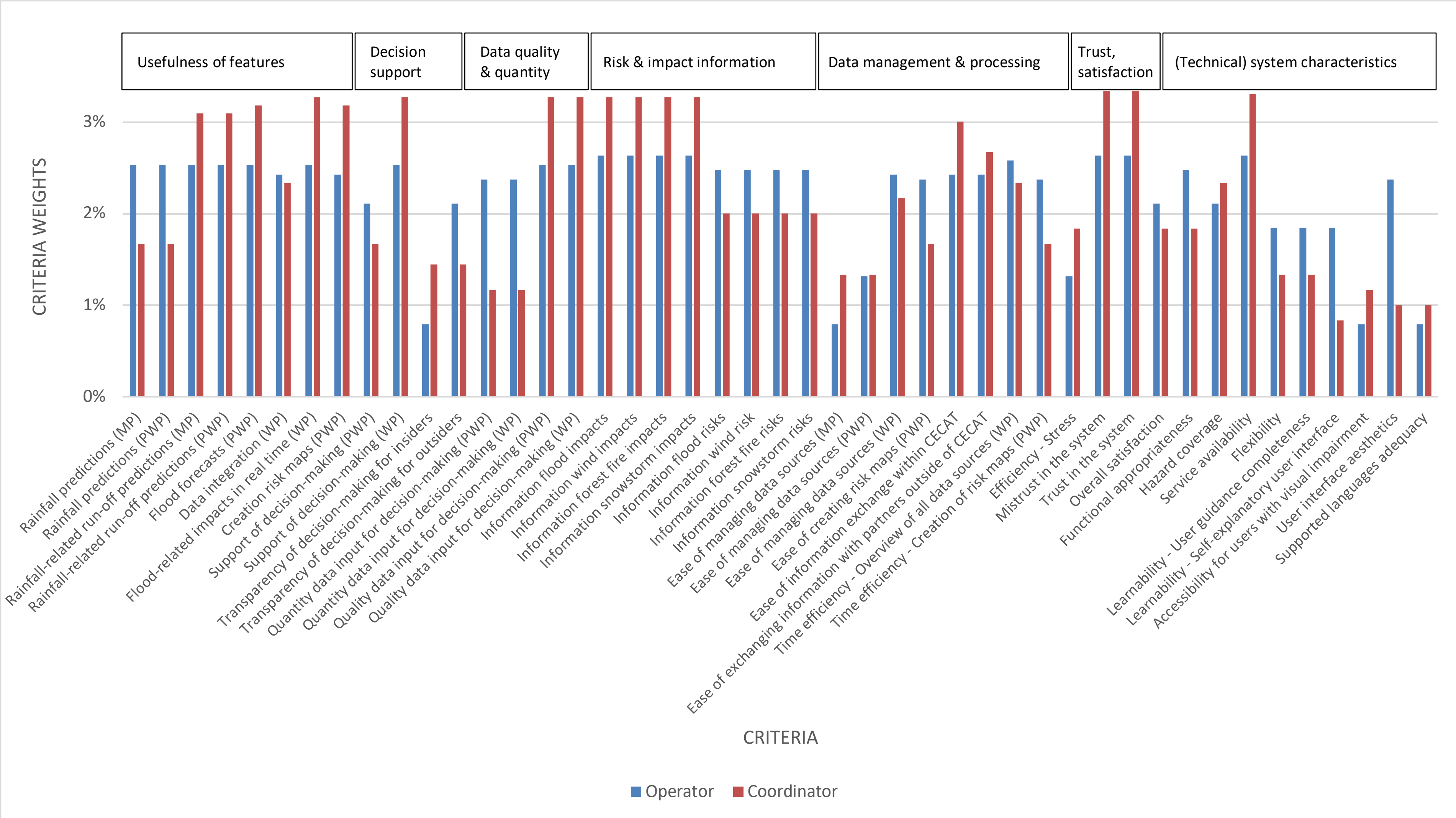


Figure 49: A4CAT - Relative importance of all co-evaluation criteria

7.12 A4CAT - Relative importance of criteria comparing operator's and coordinator's perspectives

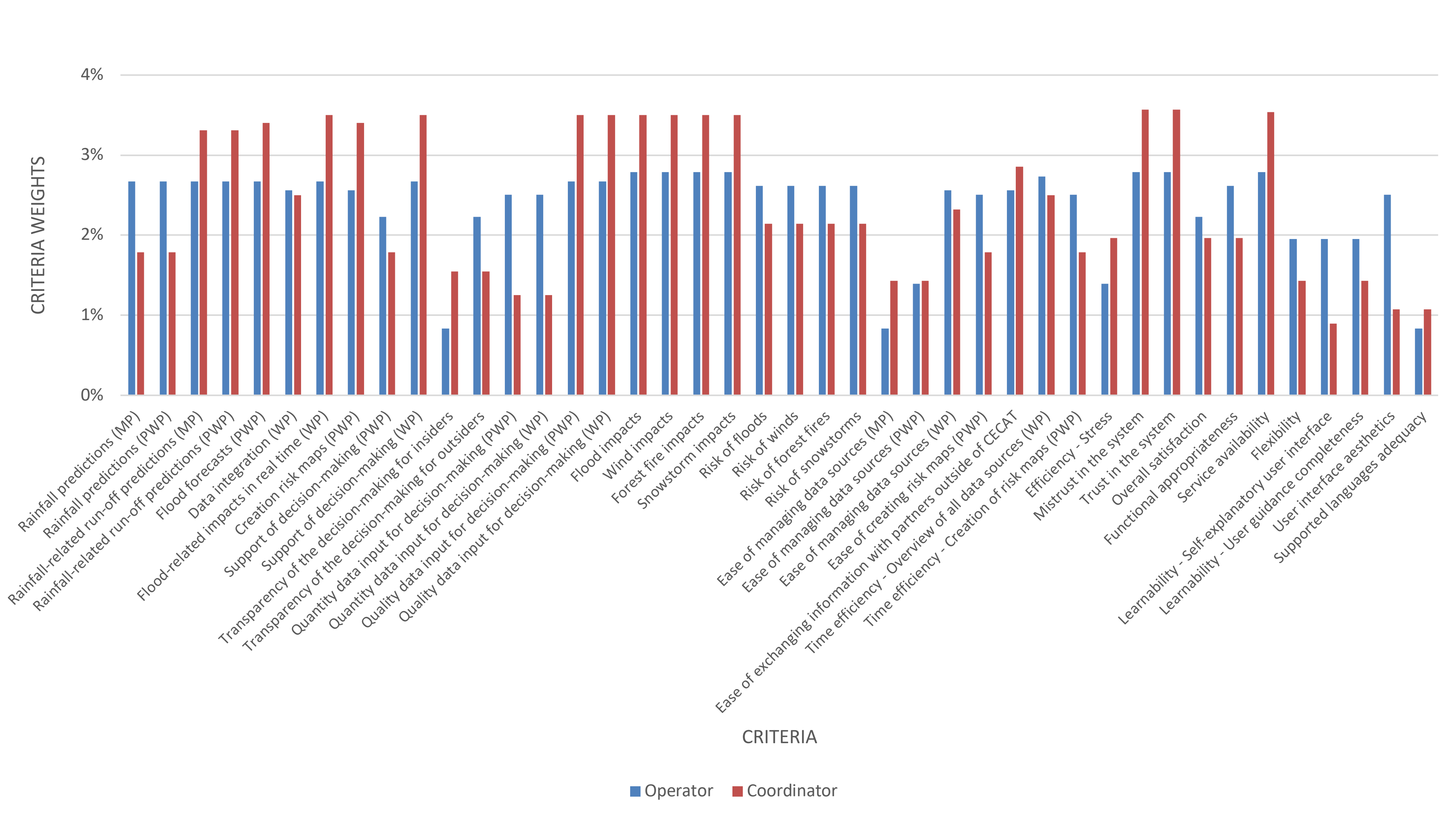


Figure 50: A4CAT - Relative importance of criteria comparing operator's and coordinator's perspectives

7.13 A4CAT - Performance comparison with Legacy system (operator)

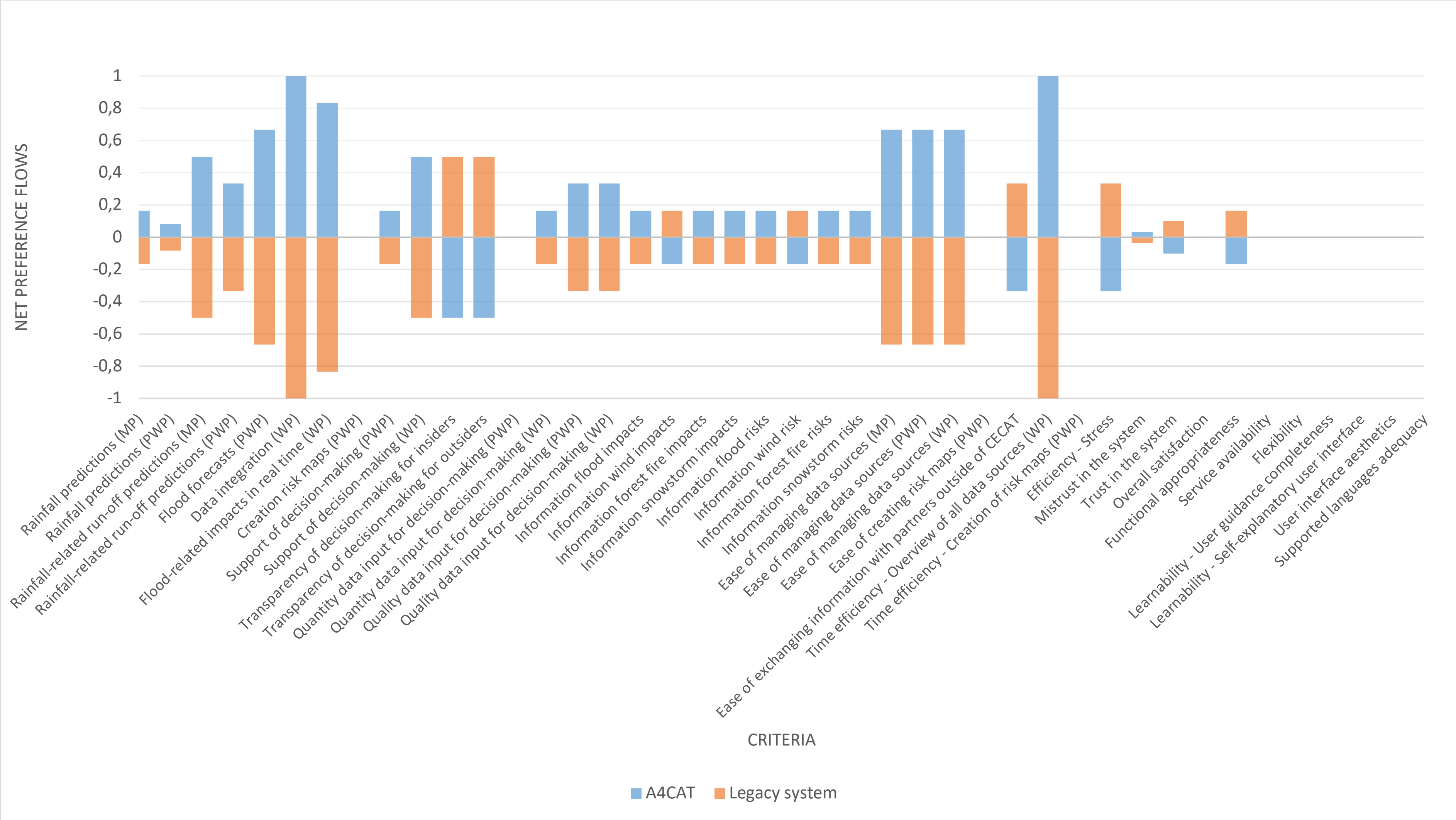


Figure 51: Performance comparison A4CAT and Legacy system (operator)

7.14 A4CAT - Performance comparison with Legacy system (coordinator)

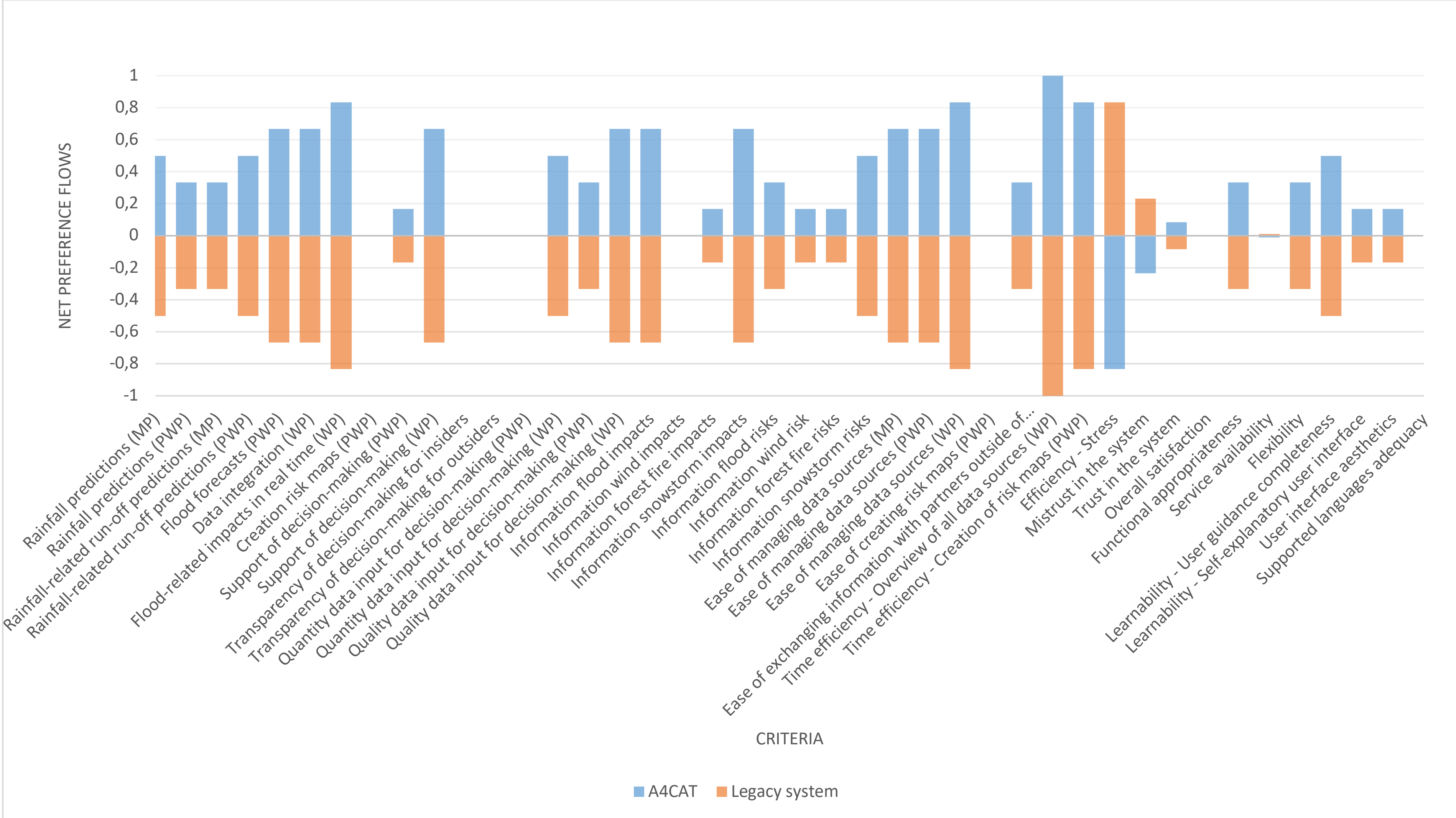


Figure 52: A4CAT - Performance comparison with Legacy system (coordinator)

7.15 A4LIG - Relative importance of all co-evaluation criteria

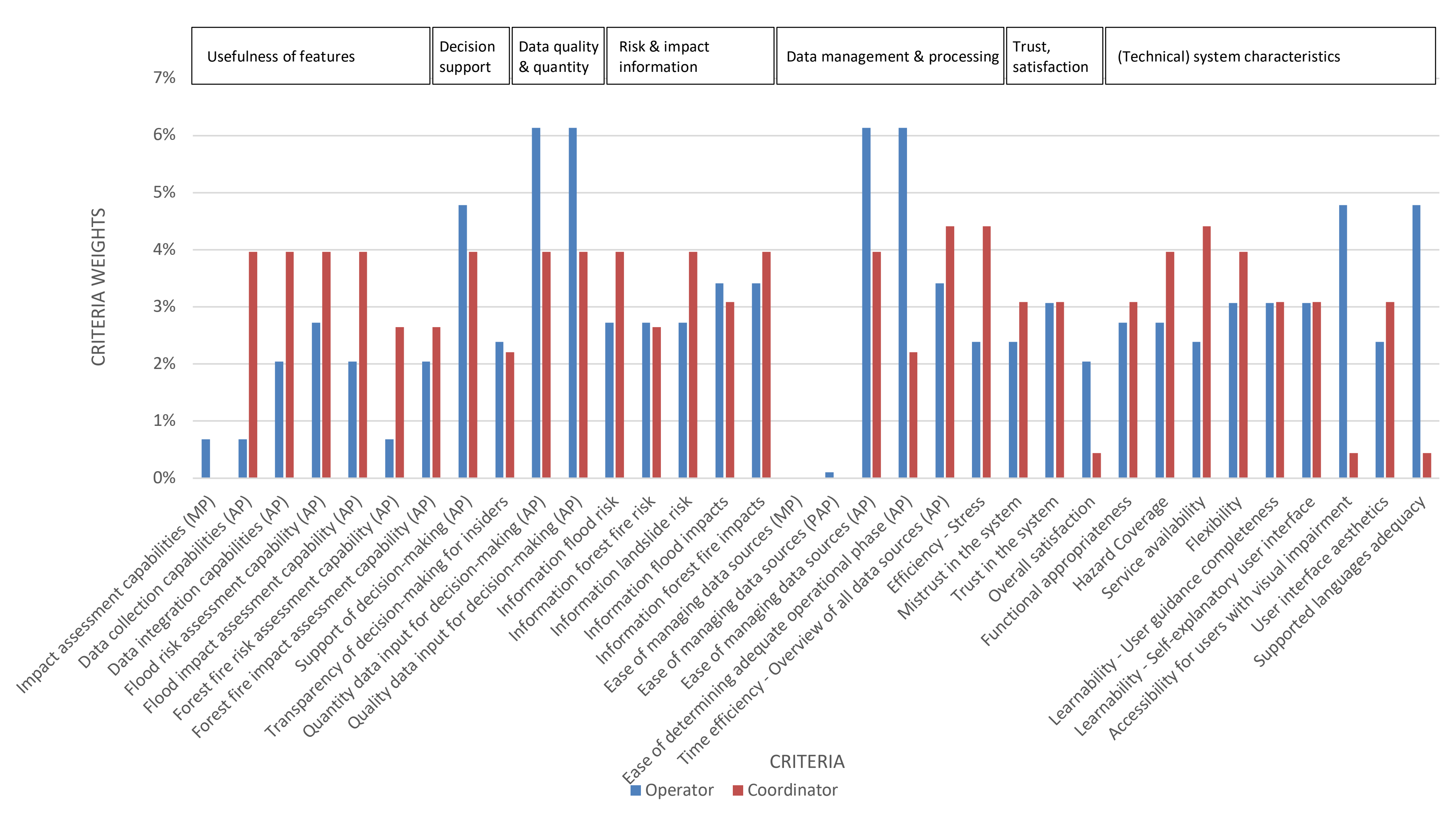


Figure 53: A4LIG - Relative importance of all co-evaluation criteria

7.16 A4FINN - Relative importance of all co-evaluation criteria

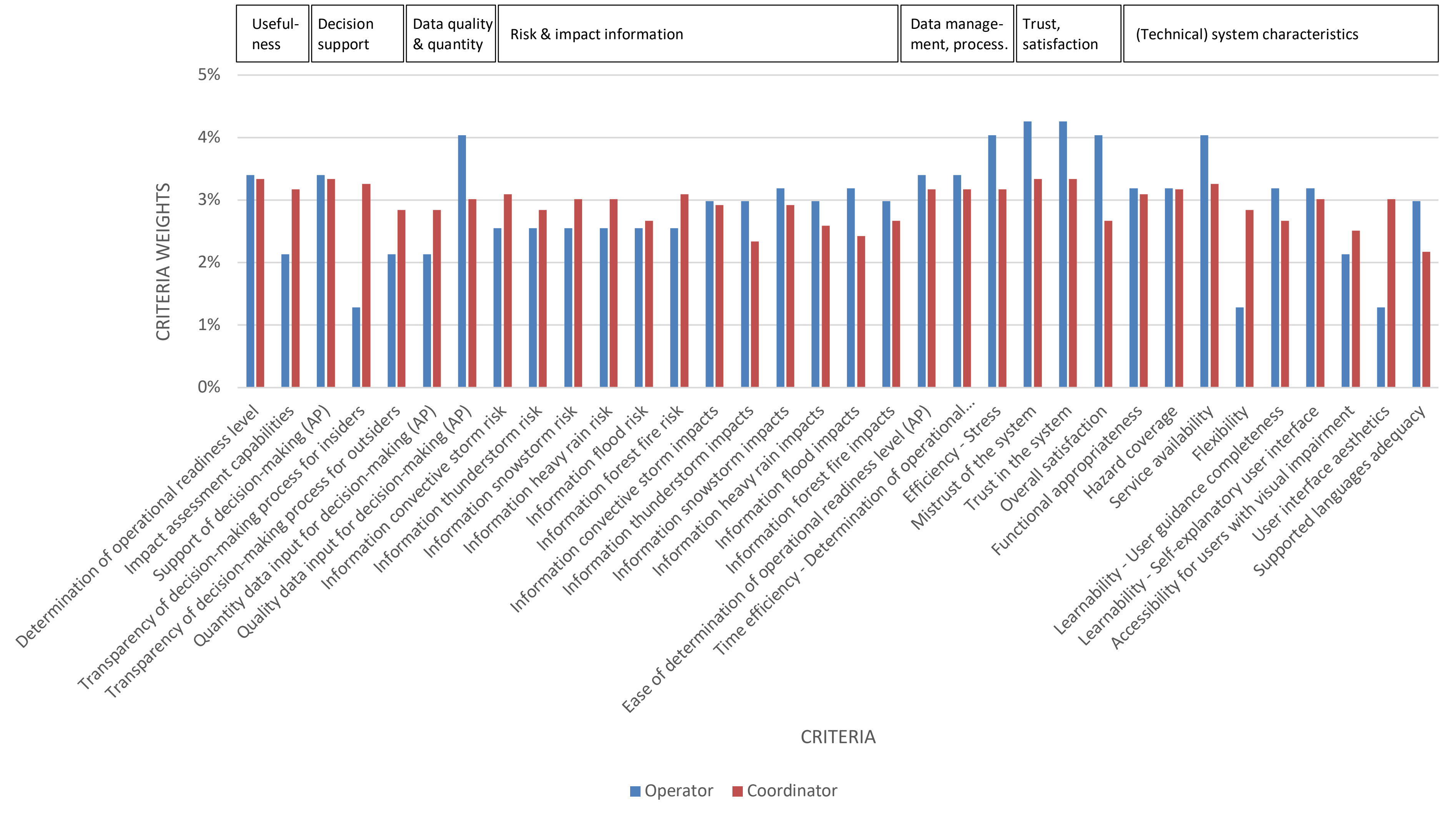


Figure 54: A4FINN - Relative importance of all co-evaluation criteria

7.17 A4FINN - Relative importance of criteria comparing operator's and coordinator's perspectives

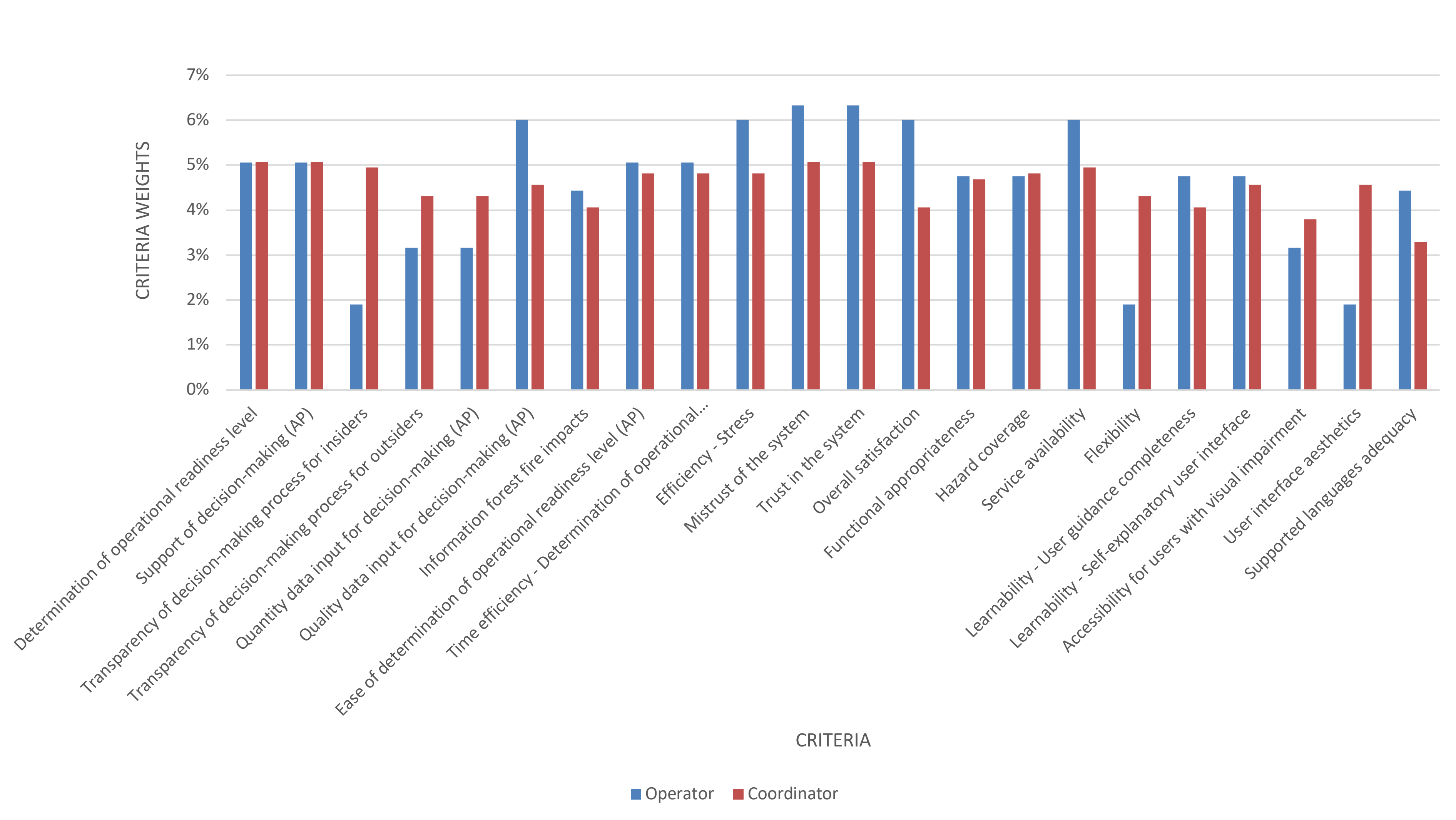


Figure 55: A4FINN - Relative importance of criteria comparing operator's and coordinator's perspectives

7.18 A4FINN - Performance comparison with Legacy system (operator)

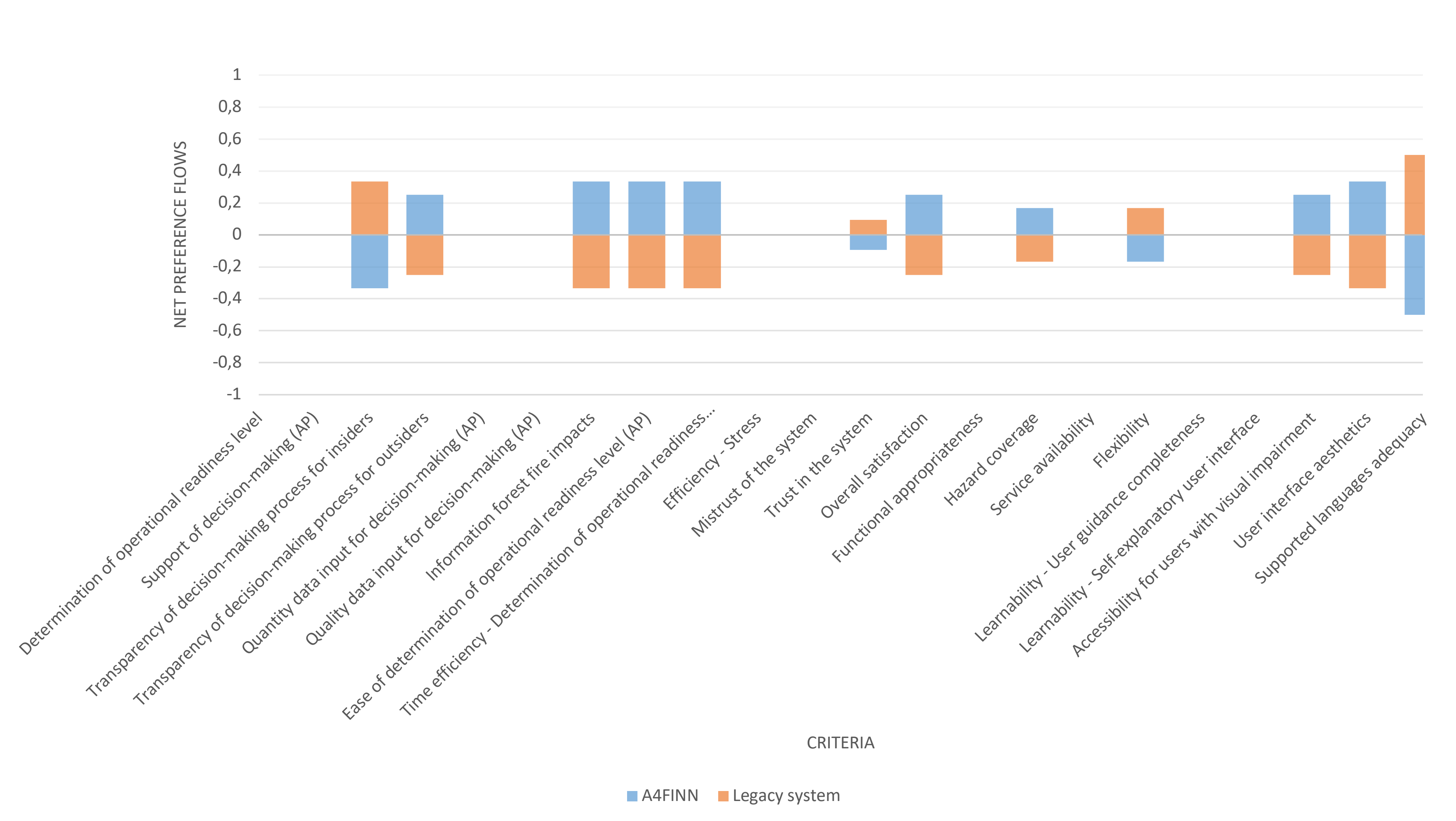


Figure 56: A4FINN - Performance comparison with Legacy system (operator)

7.19 A4FINN - Performance comparison with Legacy system (coordinator)

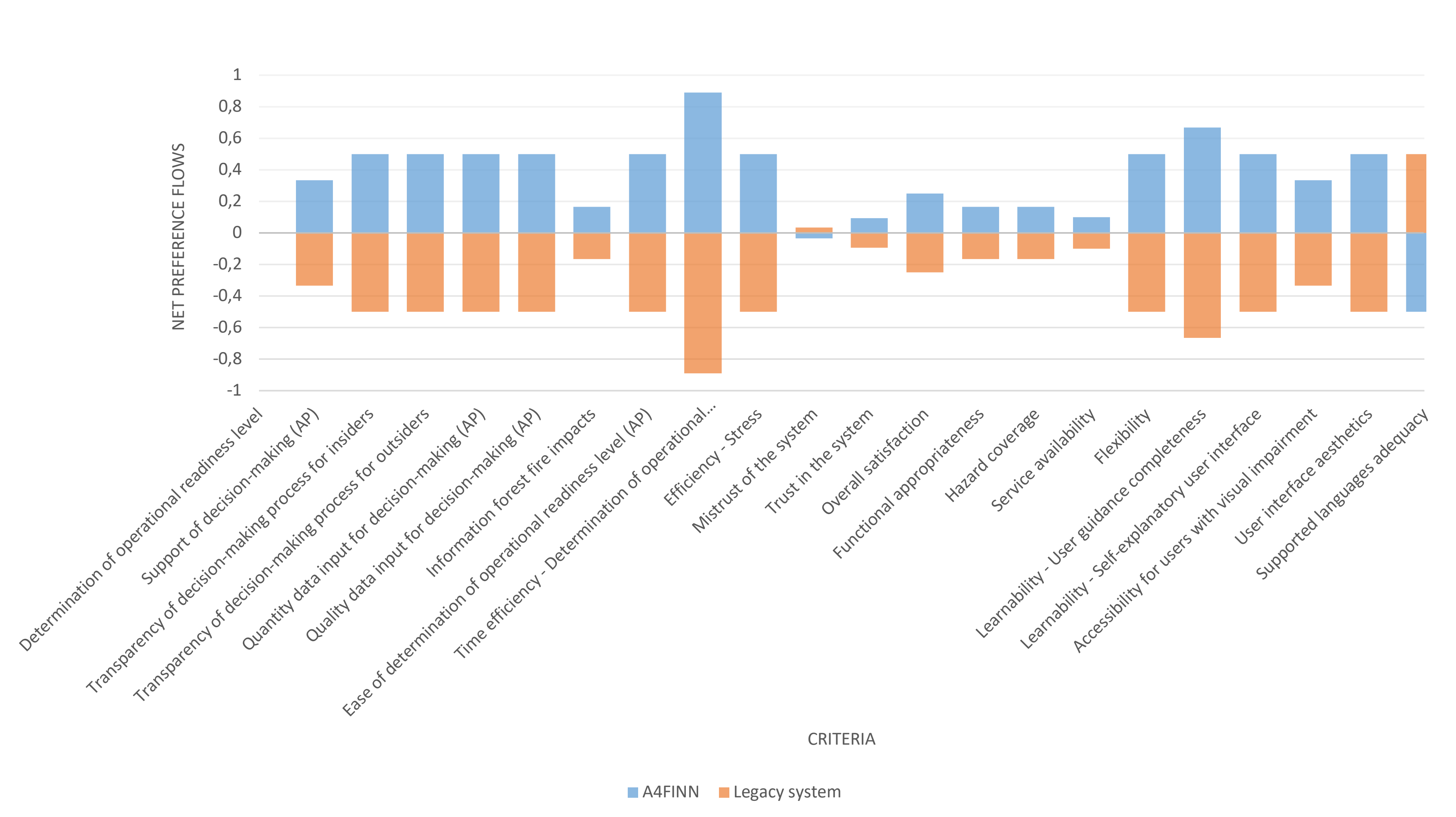


Figure 57: A4FINN - Performance comparison with Legacy system (coordinator)

7.20 A4NOR - Relative importance of all co-evaluation criteria

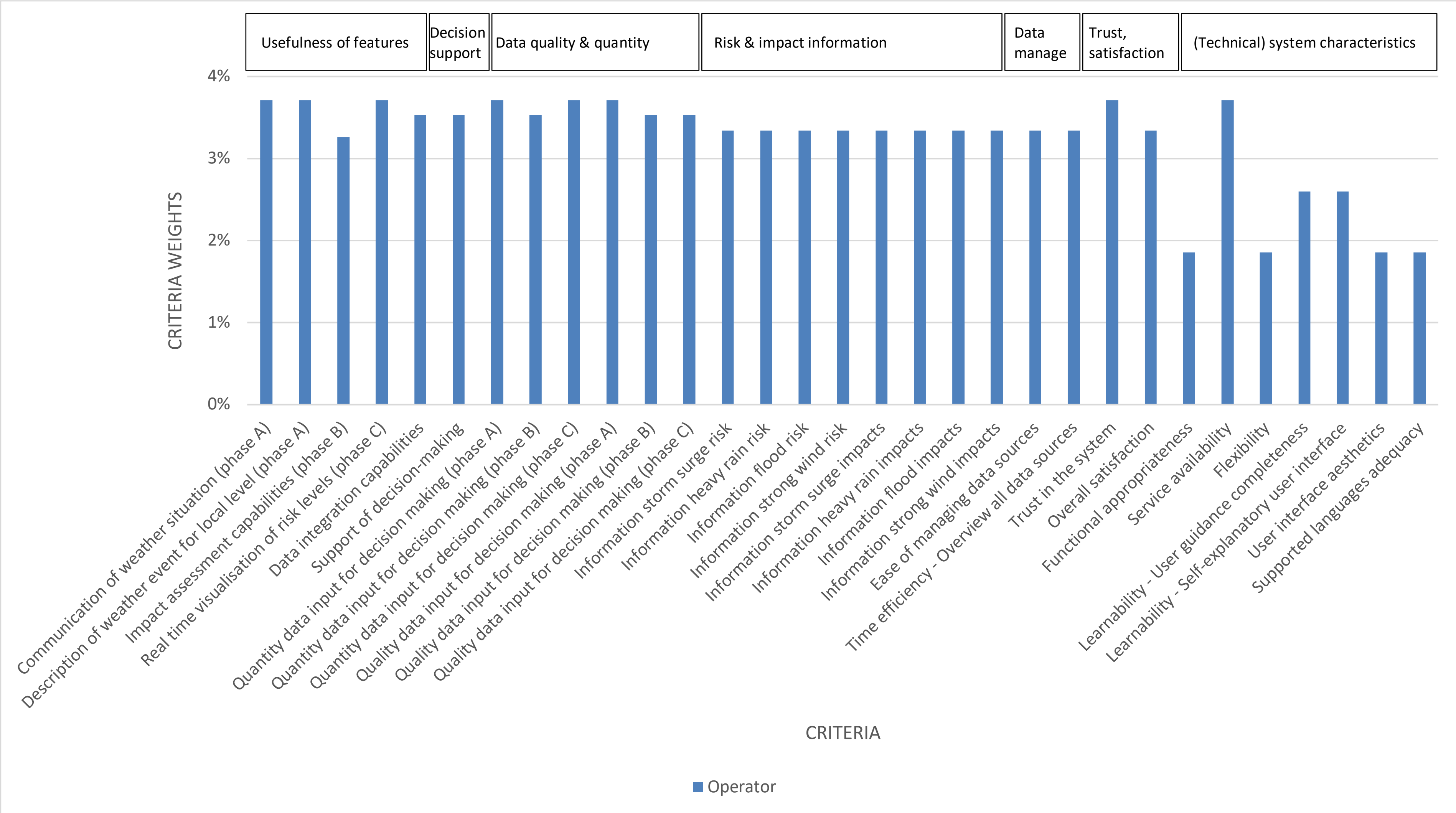


Figure 58: A4NOR - Relative importance of all co-evaluation criteria

7.21 A4COR - Relative importance of all co-evaluation criteria

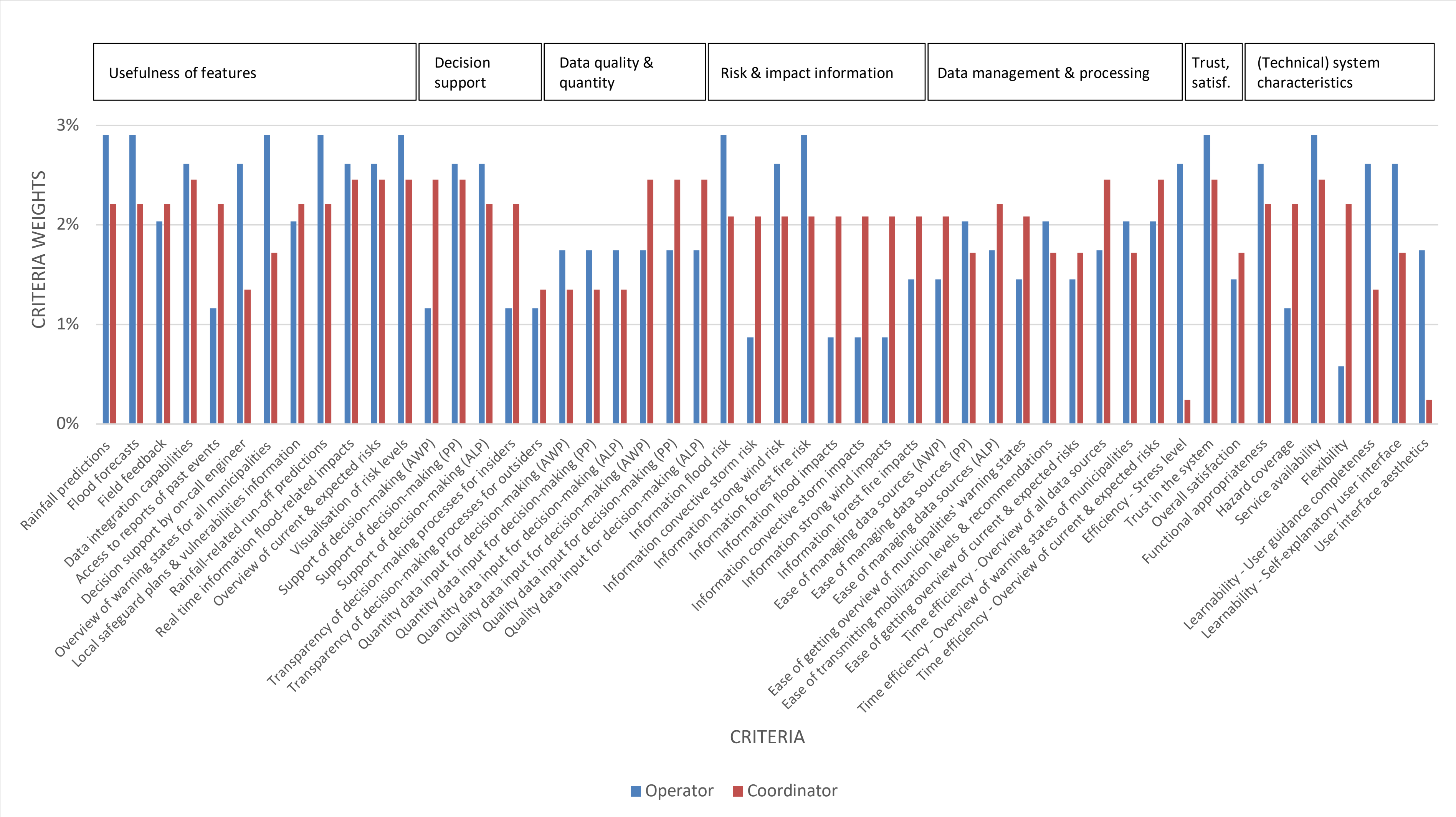


Figure 59: A4COR - Relative importance of all co-evaluation criteria

7.22 A4COR - Relative importance of criteria comparing operator's and coordinator's perspectives

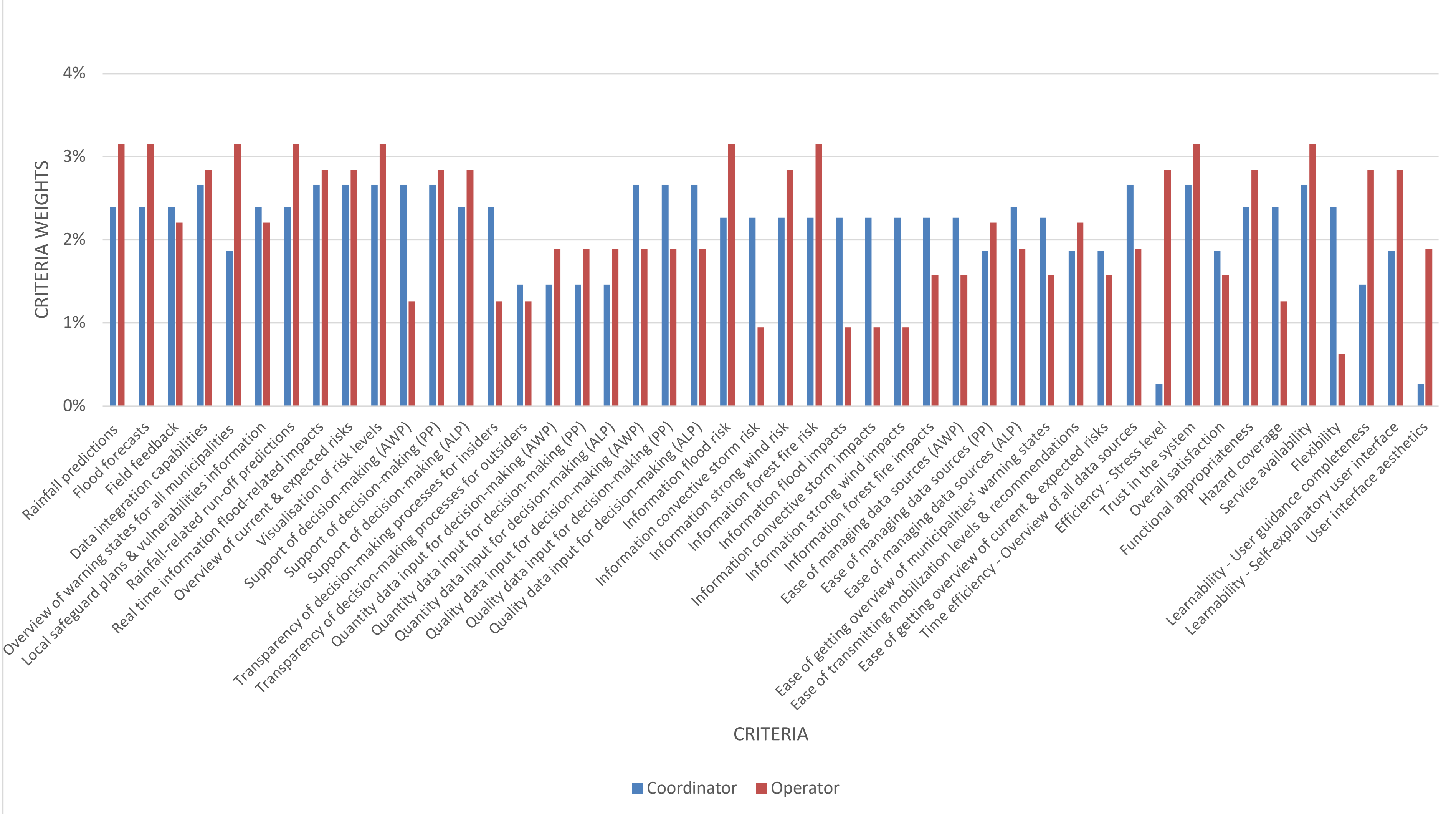


Figure 60: A4COR - Relative importance of criteria for comparison of operator's and coordinator's perspectives

7.23 A4COR - Performance comparison with Legacy system (operator)

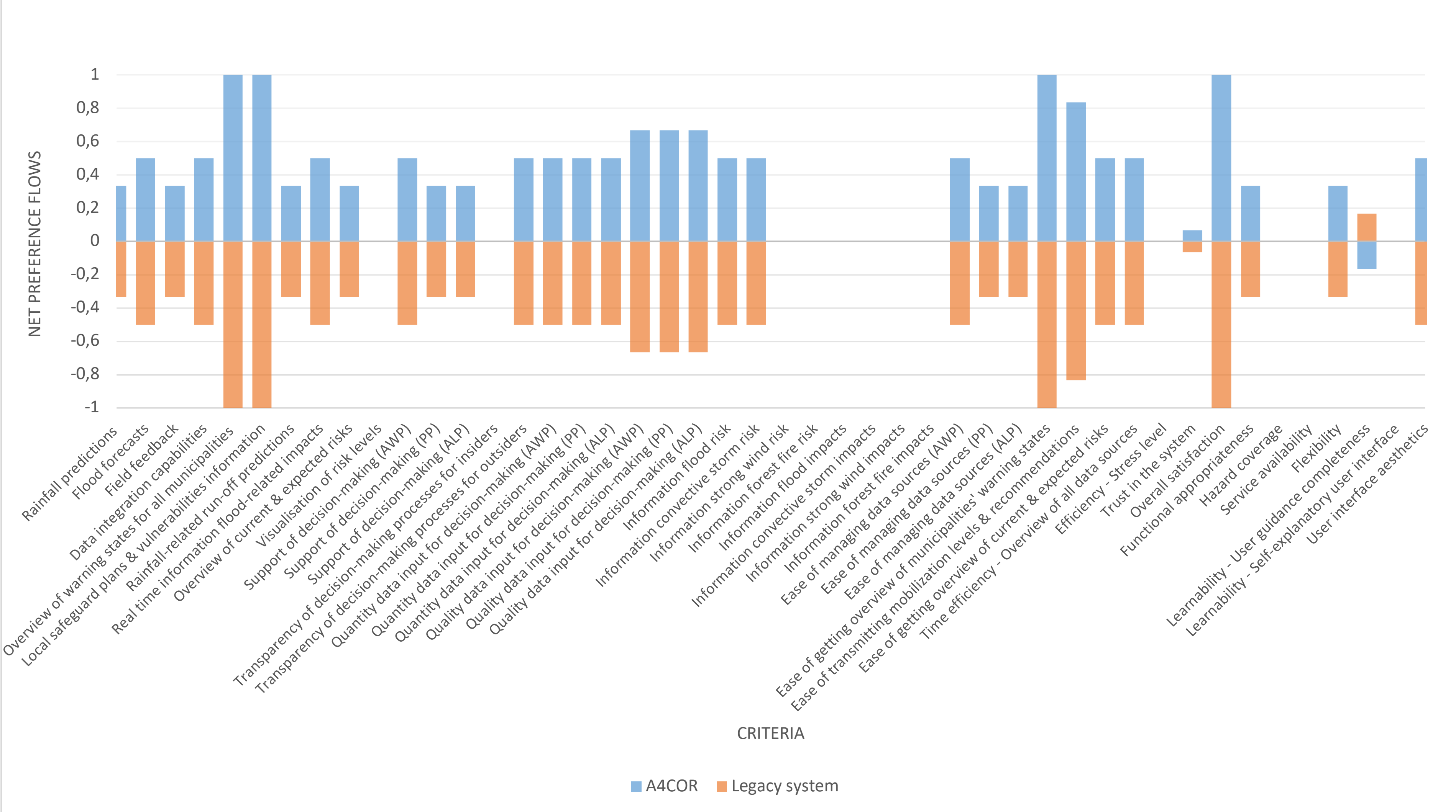


Figure 61: A4COR - Performance comparison with Legacy system (operator)

7.24 A4COR - Performance comparison with Legacy system (coordinator)

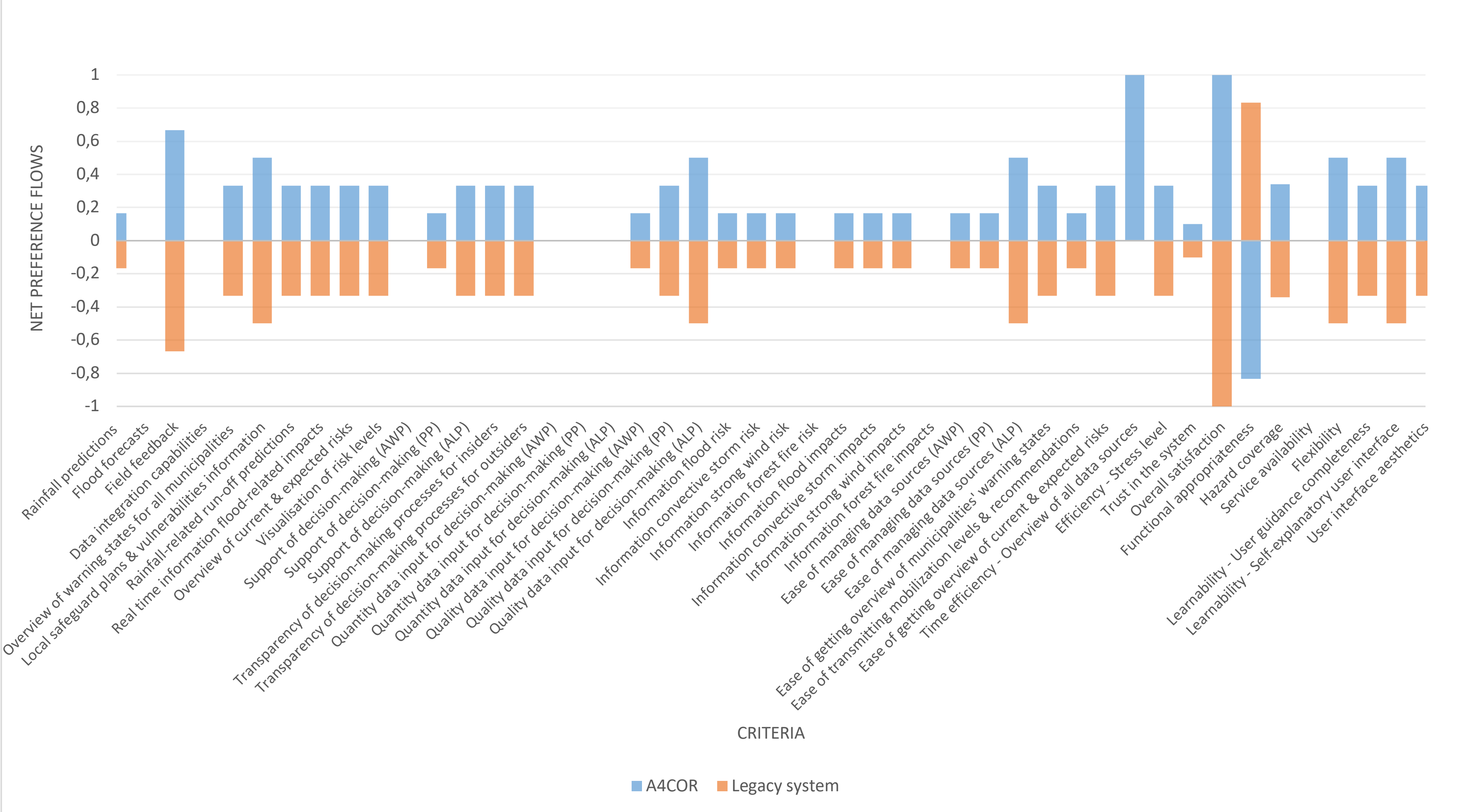


Figure 62: A4COR - Performance comparison with Legacy system (coordinator)

Table 3: List of papers in peer-reviewed scientific journals by ANYWHERE partners

No.	Title	Authors	Year	Journal, Impact Factor (IF)	doi
2016					
1	What if the 25 October 2011 event that struck Cinque Terre (Liguria) had happened in Genoa, Italy? Flooding scenarios, hazard mapping and damage estimation	Francesco Silvestro, Nicola Rebora, Lauro Rossi, Daniele Dolia, Simone Gabellani, Flavio Pignone, Eva Trasforini, Roberto Rudari, Silvia De Angeli, Cristiano Masciulli	2016	Natural Hazards and Earth System Sciences IF = 2.82	10.5194/nhess-16-1737-2016
2017					
2	An Assessment of the ECMWF Extreme Forecast Index for Water Vapor Transport during Boreal Winter	David A. Lavers, Ervin Zsoter, David S. Richardson, Florian Pappenberger	2017	Weather and Forecasting IF = 2.13	10.1175/waf-d-17-0073.1
3	Improving Predictions of Precipitation Type at the Surface: Description and Verification of Two New Products from the ECMWF Ensemble	Estíbaliz Gascón, Tim Hewson, Thomas Haiden	2017	Weather and Forecasting IF = 2.13	10.1175/WAF-D-17-0114.1
4	Improving Forecasts of Biomass Burning Emissions with the Fire Weather Index	Francesca Di Giuseppe, Samuel Rémy, Florian Pappenberger, Fredrik Wetterhall	2017	Journal of Applied Meteorology and Climatology IF = 2.28	10.1175/jamc-d-16-0405.1
5	Frequently used drought indices reflect different drought conditions on global scale	Niko Wanders, Anne F. Van Loon, Henny A. J. Van Lanen	2017	Hydrology and Earth System Sciences Discussions IF = 4.54	10.5194/hess-2017-512
6	Scale Characterization and Correction of Diurnal Cycle Errors in MAPLE	Aitor Atencia, Isztar Zawadzki, Marc Berenguer	2017	Journal of Applied Meteorology and Climatology IF = 2.28	10.1175/jamc-d-16-0344.1
7	Impact of Rainfall Assimilation on High-Resolution Hydrometeorological Forecasts over Liguria, Italy	Silvio Davolio, Francesco Silvestro, Thomas Gastaldo	2017	Journal of Hydrometeorology IF = 4.20	10.1175/jhm-d-17-0073.1
8	Severe hydrometeorological events in Liguria region: calibration and validation of a meteorological indices-based forecasting operational tool	Maria Laura Poletti, Antonio Parodi, Barbara Turato	2017	Meteorological Applications IF = 1.84	10.1002/met.1653
2018					
9	The Po Delta is restarting progradation: geomorphological evolution based on a 47-years Earth Observation dataset	A. Ninfo, P. Ciavola, P. Billi	2018	Nature Scientific Reports IF = 4.12	10.1038/s41598-018-21928-3

10	Intercomparison of attenuation correction algorithms for single-polarized X-band radars	K. Lengfeld, M. Berenguer, D. Sempere Torres	2018	Atmospheric Research IF = 4.41	10.1016/ j.atmosres.2017. 10.020
11	Characterisation and prediction of meteorological drought using stochastic models in the semi-arid Chéliff–Zahrez basin (Algeria)	Brahim Habibi, Mohamed Meddi, Paul J.J.F. Torfs, Mohamed Remaoun, Henny A.J. Van Lanen	2018	Journal of Hydrology: Regional Studies IF = 3.21	10.1016/j.ejrh.20 18.02.005
12	Diagnosing drought using the downstreamness concept: the effect of reservoir networks on drought evolution	Pieter R. van Oel, Eduardo S. P. R. Martins, Alexandre C. Costa, Niko Wanders, Henny A. J. van Lanen	2018	Hydrological Sciences Journal IF = 2.23	10.1080/ 02626667.2018. 1470632
13	Assessing heat-related health risk in Europe via the Universal Thermal Climate Index (UTCI)	Claudia Di Napoli, Florian Pappenberger, Hannah L. Cloke	2018	International Journal of Biometeorology IF = 2.36	10.1007/s00484 -018-1518-2
14	Estimating the snowfall limit in alpine and pre-alpine valleys: A local evaluation of operational approaches	Michael Fehlmann, Estíbaliz Gascón, Mario Rohrer, Manfred Schwarb, Markus Stoffel	2018	Atmospheric Research IF = 4.41	10.1016/ j.atmosres.2018. 01.016
15	Caliver: An R package for CALIbration and VERification of forest fire gridded model outputs	Claudia Vitolo, Francesca Di Giuseppe, Mirko D'Andrea	2018	PLOS ONE IF = 2.88	10.1371/ journal.pone.018 9419
16	Using the Fire Weather Index (FWI) to improve the estimation of fire emissions from fire radiative power (FRP) observations	Francesca Di Giuseppe, Samuel Rémy, Florian Pappenberger, Fredrik Wetterhall	2018	Atmospheric Chemistry and Physics IF = 5.63	10.5194/acp-18-5359-2018
17	Linking source with consequences of coastal storm impacts for climate change and risk reduction scenarios for Mediterranean sandy beaches	Marc Sanuy, Enrico Duo, Wiebke S. Jäger, Paolo Ciavola, José A. Jiménez	2018	Natural Hazards and Earth System Sciences IF = 2.82	10.5194/nhess-18-1825-2018
18	Analysis of the streamflow extremes and long-term water balance in the Liguria region of Italy using a cloud-permitting grid spacing reanalysis dataset	Francesco Silvestro, Antonio Parodi, Lorenzo Campo, Luca Ferraris	2018	Hydrology and Earth System Sciences IF = 4.54	10.5194/hess-22-5403-2018
19	A surface soil moisture mapping service at national (Italian) scale based on Sentinel-1 data	Luca Pulvirenti, Giuseppe Squicciarino, Luca Cenci, Giorgio Boni, Nazzareno Pierdicca, Marco Chini, Cosimo Versace, Paolo Campanella	2018	Environmental Modelling & Software IF = 4.87	10.1016/ j.envsoft.2017.1 2.022
20	Retrieval of snowflake microphysical properties from multifrequency radar observations	Jussi Leinonen, Matthew D. Lebsock, Simone Tanelli, Ousmane O. Sy, Brenda Dolan, Randy J. Chase, Joseph A. Finlon, Annakaisa von Lerber, Dmitri Moisseev	2018	Atmospheric Measurement Techniques IF = 3.52	10.5194/amt-11-5471-2018

21	Local-scale post-event assessments with GPS and UAV-based quick-response surveys: a pilot case from the Emilia–Romagna (Italy) coast	Enrico Duo, Arthur Chris Trembanis, Stephanie Dohner, Edoardo Grottoli, Paolo Ciavola	2018	Natural Hazards and Earth System Sciences IF = 2.82	10.5194/nhess-18-2969-2018
22	How Does Riming Affect Dual-Polarization Radar Observations and Snowflake Shape?	Haoran Li, Dmitri Moisseev, Annakaisa von Lerber	2018	Journal of Geophysical Research: Atmospheres IF = 3.86	10.1029/2017JD028186
23	Snowfall retrieval at X, Ka and W bands: consistency of backscattering and microphysical properties using BAECC ground-based measurements	Marta Tecla Falconi, Annakaisa von Lerber, Davide Ori, Frank Silvio Marzano, Dmitri Moisseev	2018	Atmospheric Measurement Techniques IF = 3.52	10.5194/amt-11-3059-2018
24	Estimating the snowfall limit in alpine and pre-alpine valleys: A local evaluation of operational approaches	Fehlmann, Michael; Gascón, Estíbaliz; Rohrer, Mario; Schwarb, Manfred; Stoffel, Markus	2018	Atmospheric Research IF = 4.41	10.1016/ j.atmosres.2018.01.016
2019					
25	How to improve attribution of changes in drought and flood impacts	Heidi Kreibich, Veit Blauhut, Jeroen C.J.H. Aerts, Laurens M. Bouwer, Henny A.J. Van Lanen, Alfonso Mejia, Marjolein Mens, Anne F. Van Loon	2019	Hydrological Sciences Journal IF = 2.23	10.1080/ 02626667.2018.1558367
26	Towards robust pan-European storm surge forecasting	T. Fernández-Montblanc, M.I. Vousdoukas, P. Ciavola, E. Voukouvalas, L Mentaschi, G. Breyiannis, L. Feyen, P. Salamon	2019	Ocean Modelling IF = 3.15	10.1016/ j.ocemod.2018.12.001
27	Using nowcasting technique and data assimilation in a meteorological model to improve very short range hydrological forecasts	Maria Laura Poletti, Francesco Silvestro, Silvio Davolio, Flavio Pignone, Nicola Rebora	2019	Hydrology and Earth System Sciences Discussions IF = 4.54	10.5194/hess-2019-75
28	Comparison of two early warning systems for regional flash flood hazard forecasting	Carles Corral, Marc Berenguer, Daniel Sempere-Torres, Laura Poletti, Francesco Silvestro, Nicola Rebora	2019	Journal of Hydrology IF = 4.54	10.1016/ j.jhydrol.2019.03.026
29	Quantifying Positive and Negative Human-Modified Droughts in the Anthropocene: Illustration with Two Iranian Catchments.	Kakaei, E., Moradi, H.R., Moghaddam Nia, A.R. and Van Lanen, H.A.J.	2019	Water IF = 2.60	10.3390/w11050884
30	Improving medium-range forecasts of rain-on-snow events in pre-alpine areas	Michael Fehlmann, Estíbaliz Gascón, Mario Rohrer, Manfred Schwarb, Markus Stoffel	2019	Water Resources Research IF = 4.27	10.1029/2018WR024644
31	Anticipating cascading effects of extreme precipitation with pathway schemes -	Simone Schauwecker, Estíbaliz Gascón, Shinju Park, Virginia Ruiz-Villanueva, Manfred	2019	Environment International IF=8.17	10.1016/ j.envint.2019.02.072

	Three case studies from Europe	Schwarb, Daniel Sempere-Torres, Markus Stoffel, Claudia Vitolo, Mario Rohrer			
32	Long-term analysis of gauge-adjusted radar rainfall accumulations at European scale	Shinju Park, Marc Berenguer, Daniel Sempere-Torres	2019	Journal of Hydrology IF = 4.54	10.1016/ j.jhydrol.2019.03 .093
33	Mapping combined wildfire and heat stress hazards to improve evidence-based decision making	Claudia Vitolo, Claudia Di Napoli, Francesca Di Giuseppe, Hannah L. Cloke, Florian Pappenberger	2019	Environment International IF = 8.17	10.1016/ j.envint.2019.03. 008
34	Using paired catchments to quantify the human influence on hydrological droughts	Anne F. Van Loon, Sally Rangecroft, Gemma Coxon, José Agustín Breña Naranjo, Floris Van Ogtrop, and Henny A. J. Van Lanen	2019	Hydrology and Earth System Sciences IF = 4.54	10.5194/hess-23-1725-2019
35	ANYCaRE: a role-playing game to investigate crisis decision-making and communication challenges in weather-related hazards	Galateia Terti, Isabelle Ruin, Milan Kalas, Ilona Láng, Arnau Cangròs i Alonso, Tommaso Sabbatini, Valerio Lorini	2019	Natural Hazards and Earth System Sciences IF = 2,82	10.5194/nhess-19-507-2019
36	Growing spatial scales of synchronous river flooding in Europe	Wouter R. Berghuijs, Scott T. Allen, Shaun Harrigan, James W. Kirchner	2019	Geophysical Research Letters IF = 4.74	10.1029/2018gl0 81883
37	Exploration of the importance of physical properties of Indonesian peatlands to assess critical groundwater table depths, associated drought and fire hazard	M. Taufik, A.A. Veldhuizen, J.H.M. Wösten, H.A.J. van Lanen	2019	Geoderma IF = 4.52	10.1016/ j.geoderma.201 9.04.001
38	Verification of Heat Stress Thresholds for a Health-Based Heat-Wave Definition.	Di Napoli, Claudia; Pappenberger, Florian; Cloke, Hannah L.	2019	Journal Applied Meteorology Climatology IF = 2.28	10.1175/JAMC-D-18-0246.1.
39	Moving from drought hazard to impact forecasts.	Sutanto, S.J., van der Weert, M., Wanders, N., Blauhut, V. and Van Lanen, H.A.J.	2019	Nature Communications IF = 11.80	10.1038/s41467 -019-12840-z
40	Evaluating skill and robustness of seasonal meteorological and hydrological drought forecasts at the catchment scale – Case Catalonia (Spain).	Van Hateren, T.C., Sutanto, S.J., Van Lanen, H.A.J.	2019	Environment International IF = 8.17	10.1016/j.envint. 2019.105206.
41	Heatwaves, droughts, and fires: exploring compound and cascading dry hazards at the pan-European scale.	Sutanto, S.J., Vitolo, C., Di Napoli, C. D'Andrea, M., Van Lanen, H.A.J.	2019	Environment International IF = 8.17	10.1016/j.envint. 2019.105276.
42	Potential of pan-European seasonal hydro-meteorological drought forecasts obtained from a Multi-	Sutanto, S.J., Van Lanen, H.A.J., Wetterhall, F., Llort, X.	2019	Bulletin of the American Meteorological Society (accepted) IF = 6.19	

	Hazard Early Warning System.				
43	Characterisation of the dynamics of past droughts.	Diaz, V., Corzo Perez, G.A., Van Lanen, H.A.J., Solomatine, D., Varochakis, E.	2019	Science of the Total Environment (accepted). IF = 5.90	