Decision support for offshore asset construction using expert judgments for supply disruptions risk

Leontaris, George; Morales Napoles, Oswaldo; Dewan, Ashish ; Wolfert, Rogier

DOI
10.1016/j.autcon.2019.102903

Publication date
2019

Document Version
Accepted author manuscript

Published in
Automation in Construction

Decision support for offshore asset construction using expert judgments for supply disruptions risk

Georgios Leontaris[a,*], Oswaldo Morales-Nápoles[b], Ashish Dewan[c,1], A.R.M. (Rogier) Wolfert[a]

[a]*Engineering Asset Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628 CN Delft, The Netherlands*
[b]*Hydraulic Structures and Flood Risk, Faculty Civil Engineering and Geosciences, Delft University of Technology, 2628 CN Delft, The Netherlands*
[c]*R&D department of offshore wind energy, ECN part of TNO, 1755 LE Petten, The Netherlands*

Abstract

Offshore asset construction is a complex and costly process that is subject to various uncertainties within the entire supply chain. Hence, both the construction management optimization and the reduction of deployment expenditures should be supported by automated decision support models which include proper representations of predominant uncertainties. One of these is the supply disruption risk that is often ignored in existing models. Therefore, this article proposes a methodology to properly take this construction risk into account. An algorithm to model this risk was developed and a study was conducted to obtain the required probability distributions of disruption delays using real data and expert judgments for an offshore wind farm construction application. The simulation of a realistic test case with an appropriately modified stochastic simulation tool showed that it is important to consider this risk in order to make optimal decisions for different offshore wind farm construction strategies.

Keywords: Probabilistic decision support; Supply disruptions risk; Construction logistics; Offshore wind construction process; Simulation model; Structured expert judgment

*Corresponding author
Email addresses: G.Leontaris@tudelft.nl (Georgios Leontaris), O.MoralesNapoles@tudelft.nl (Oswaldo Morales-Nápoles), ashish.dewan@nuon.com (Ashish Dewan), R.Wolfert@tudelft.nl (A.R.M. (Rogier) Wolfert)
[1]Current address: Project Support Onshore BA Wind, Vattenfall, Hoekenrode 8, 1102 BR Amsterdam Zuidoost, The Netherlands

1. Introduction

It is a fact that large and complex construction projects are subject to various uncertainties, which may hinder the construction processes. Therefore, models and tools, which support the management of the construction activities, must allow a proper representation of these uncertainties. Supply disruptions are an important issue that may occur in every large construction project and cause significant delays, resulting in budget and ~~or~~ schedule overruns. Although a lot of research has been conducted in the past regarding the supply chain risk (for a thorough review the reader is referred to Snyder et al. (2016); Fahimnia et al. (2015); Heckmann et al. (2015); Shi et al. (2016); Arashpour et al. (2017)), this has not been investigated for the construction of offshore assets and the fast-growing offshore wind industry in particular.

Although the previously high costs of offshore wind industry have been already partly reduced (Lacal-Arántegui et al., 2018), the proper description of uncertainties in all different phases of the offshore wind farms (OWFs) life cycle is needed to make offshore wind even more financially competitive. Especially, certain aspects related to construction management of OWFs should be improved to tackle the arising logistical challenges (Poulsen & Lema, 2017). This will be particularly crucial in the coming years, in order to cope with the increasing challenges due to the necessity to move farther offshore. Furthermore, service providers have already identified "installation and logistics of OWFs as one of the main construction risks" due to capital intensive construction activities (European Wind Energy Association, 2014).

Construction activities of OWFs are not only expensive, but also complex. Their complexity stems from the fact that these are subject to various uncertainties such as environmental offshore conditions, supply disruptions and failures which may occur during the construction process. All ~~of~~ these uncertainties should be taken into account especially during the planning phase. Otherwise, planning based on inaccurate estimates, may lead to decisions which will cause significant schedule and budget overruns during the construction phase. To avoid these undesirable outcomes, probabilistic decision support tools should be utilized in the planning phase to support optimal construction management given these uncertainties. Thus, reliable tools that take into account various uncertainties during the entire OWF supply chain, would be essential for achieving cost reduction.

For the aforementioned reasons, during the last years, various models have been developed concerning different aspects of OWFs decision support. A thorough review of the developed models until 2011 is presented in (Hofmann, 2011). The majority of these models were focused on the maintenance strategies. ~~Research regarding the construction process did not use a detailed approach to model logistics with the aforementioned uncertainties.~~ Since then, more studies were conducted and various models concerning the construction (or installation) process of OWFs were developed. ~~However, these did not take into account the aforementioned uncertainties.~~ For example, Kaiser & Snyder (2013) developed a model of the installation costs of offshore wind projects on the U.S. Outer Continental Shelf. While Sarker & Faiz (2017) proposed a method to identify the characteristics of OWFs installation processes that minimize the total time requirement for transportation and installation, without taking into account the uncertainties. Moreover, most of the developed models use a simulation-based approach and focus on developing different approaches to better describe the environmental condition uncertainty. In particular, Vis & Ursavas (2016) developed a simulation-based decision support tool to investigate different logistical approaches within the installation phase of OWFs while taking into account the external influence of weather by the use of a Markov chain with three states. Paterson et al. (2018) developed a software tool that relies on Monte Carlo methods to simulate multiple independent scenarios of the defined installation strategy for an offshore wind farm, while considers the risk imposed by adverse weather conditions by using a hidden Markov model (HMM).Morandeau et al. (2013) presented the MERMAID (Marine Economic Risk Management Aid) simulation software package that was used for the analysis and optimization of marine energy installations and the investigation of a vessel designed for installation of OWFs. Leontaris et al. (2016) proposed a methodology to produce realistic synthetic time series of wind speed and wave height in order to incorporate the environmental risk into the estimates of the duration of cable installation of OWFs. Also, Guo et al. (2017) proposed a fuzzy duration forecast model for the construction of onshore wind turbines which are only subject to the impact of wind uncertainty.

Other researchers focused on investigating optimization techniques concerning the installation of OWFs. Irawan et al. (2017) developed an integer linear programming (ILP) model to determine the optimal installation schedule considering constraints regarding weather conditions and the availability of vessels. Kerkhove & Vanhoucke (2017) proposed a Markovian model to describe the weather component and an approach that uses both general meta-heuristic optimization approaches and dedicated heuristics to optimize the project planning. Ursavas (2017) proposed a two-stage stochastic integer program that considers disruptions arising from uncertain weather conditions and the solution approach of the planning problem of wind farms is based on partial Benders decomposition strategy. Barlow et al. (2018) proposed a decision-support tool in a combined framework of an optimization and simulation model which improves the capabilities of both models to provide a mechanism to address current OWF installation projects while taking into account the seasonal uncertainties. Scholz-Reiter et al. (2011) have developed an optimization model for OWF installation scheduling using mixed-integer linear programming (MILP). Particularly, Scholz-Reiter et al. (2011) recommended to develop a simulation model that takes into account possible supply disruptions and to integrate this with their model, in order to have a robust design for planning of offshore installation. Thus, none of the aforementioned models addressed the risk of supply disruptions during the installation process of OWFs. So far, only Mogre et al. (2016) and Leontaris et al. (2017) partly considered this. Leontaris et al. (2017) did this for a limited scope regarding only the cable unavailability; whereas, Mogre et al. (2016) proposed a model using expert judgments which were elicited using pairwise-comparison matrices, which cannot lead to numerical estimates but only to preferences.

Therefore, this article proposes a methodology to model this supply disruption risk of the required components based on real data and expert judgments (Section 2.1). These expert judgments were elicited, evaluated and combined by using Cooke's classical model (Cooke, 1991). The theoretical background of the classical model and ~~its application details~~ the details for applying it in this study can be found in Section 2.2. For the analysis and synthesis of the expert judgments, a newly developed and recently released open-source MATLAB toolbox ~~called~~ named ANDURIL (Leontaris & Morales-Nápoles, 2018), was used. The results of the expert judgments elicitation can be found in Section 3.

The current study covers the supply disruption risk of the major components which are most commonly used in OWFs. Namely, the monopiles (MPs), the transition pieces (TPs), the towers, the nacelle and the rotor that consists of 3 blades. These components of an offshore wind turbine generator (WTG) are illustrated in Figure 1. To quantify the supply disruptions risk an existing stochastic decision support tool (ECN Install, see Section 4.1) was modified with the proposed algorithm (Section 2.1). This allows to simulate different scenarios and compare these in terms of the cumulative probability distributions (CDFs) of the estimated duration and cost of the installation accounting for the risk of supply disruptions (Section 4). Hence, the proposed method can be used to investigate the impact of supply disruptions during the construction process allowing the comparison of different scenarios and support decision makers in planning and choosing the “optimal” scenario.

Finally, it is worth mentioning that the proposed methodology within this article can not only be applied to the optimization of offshore but also to inland construction management processes such as for example the road construction supply chain with its uncertainties within the entire supply chain, from the asphalt production plant until the final paving.

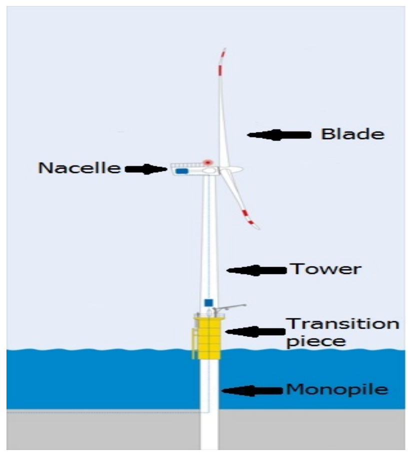


Figure 1: Different components of an offshore WTG.

2. Methodology

A methodology is proposed to include the supply disruptions risk into probabilistic models for construction of offshore assets. As it was mentioned before, the existing decision support models for scheduling of offshore constrcuction activities focus mostly on describing uncertainties regarding the environmental conditions which hinder the construction operations. Hence, the proposed methodology is particularly relevant for offshore assets. However, it could also be applied to different construction projects which are subject to various uncertainties and their planning requires probabilistic models. Assuming such a probabilistic model on hand, one who is interested in including the supply disruptions risk, should first recognise whether this is relevant for the application under investigation. If so, then the existing model should be appropriately modified in order to take into account the events that may lead to delays due to supply disruptions. Afterwards, one should investigate whether sufficient relevant data are available and perfom statistical analyses to obtain the probability distributions which are required as inputs to the modified model. If relevant data are scarce, then the application of a structured expert judgment (SEJ) method such as Cooke's classical model (Cooke, 1991) is recommended (see Section 2.2.1). The proposed methodology is depicted in the form of a flowchart diagram in Figure 2.

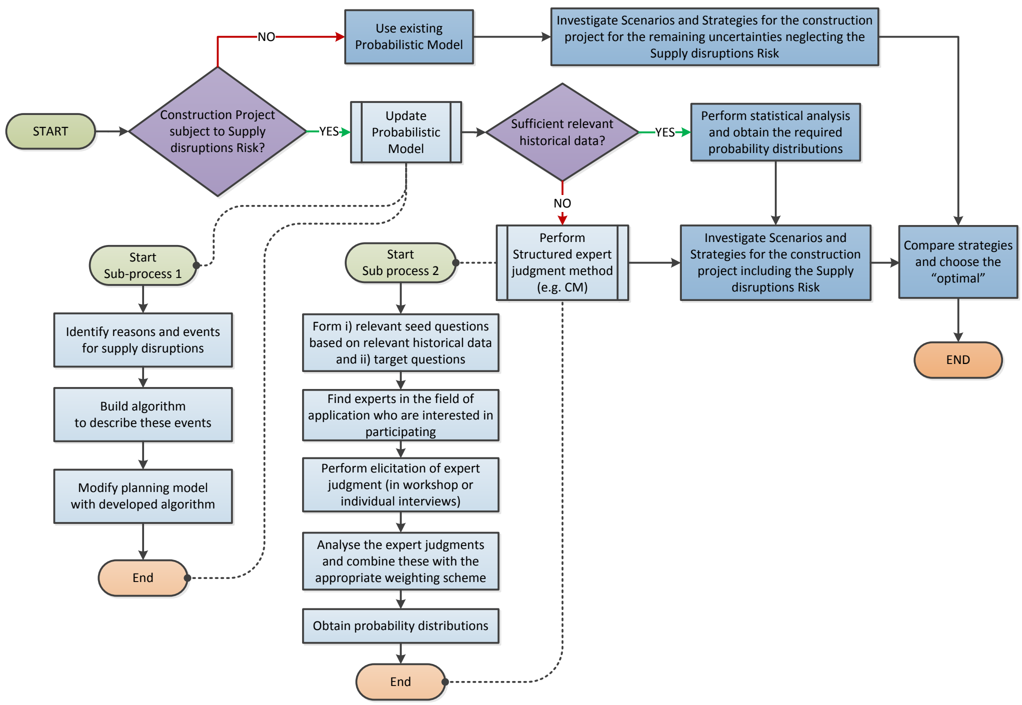


Figure 2: Flowchart diagram of actions one should take to include the risk of supply disruptions into a probabilistic decision support model.

~~In this section~~ In the following subsections, the proposed ~~method~~ methodology is applied to model the supply disruptions and quantify this risk during the installation of OWFs is presented. First, in subsection 2.1 the reader can find a detailed description of the algorithm that was developed to model this risk. Then the application of Cooke's classical model for the quantification of the uncertainty regarding supply disruptions as well as the required theoretical background is presented in subsection 2.2.

2.1. Algorithm for modeling supply disruptions

Usually, different types of components of OWF such as monopiles (MPs), transition pieces (TPs), towers, blades and nacelles are transported from manufacturers to the feeder (or installation) port.

Subsequently, transport vessels or installation vessels transport these components to the OWF site.

There are various reasons why the required components might not be available to be loaded to the transport vessel when the vessel is at the port. Some of these reasons can be delays, damages of components and equipment or miscommunication between the parties involved in the installation process; ~~which~~ these may occur during transport from the manufacturers to the feeder port or even within the port. Therefore, this delay can vary significantly.

In this study, the supply risk is modeled as an event at the loading port that is caused in case the required components are not available when needed. This event is described by:

i a *probability of occurrence* P_{U_C}, where U_C denotes the event of unavailability of the required components for every type of component $C \in (1, 2..., n)$ of the total n different types of components

ii the *delay* D_C (in units of time) which denotes the waiting time until the required components of type C are available for loading. Hence, the waiting time D_C is a random variable with a cumulative probability distribution F_{D_C}.

Figure 3 presents the flowchart of the developed algorithm that can be used to take into account the risk of supply disruptions. When the vessel is at the port and a new loading operation of components is about to start at time t, it should be checked if the stock level $S_{C,t}$ of the component C is sufficient to proceed. So, the first step of the algorithm is to examine the amount of components $S_{C,t-1}$ that was available in the port when the previous loading operation of this type of components C has been completed at time $t-1$. Usually, the entire installation process starts only when there is the required amount of components S'_C, concerning every type of component C, at the installation port. Thus, as far as the first loading operation ($t = 1$) is concerned, the stock $S_{C,t=0} = S'_C$. If the stock $S_{C,t-1}$ is not sufficient (i.e. there are less than the required components R_C of component type C) then a pseudo-random number $X \sim U(0, 1)$ is generated. If X has a value smaller or equal than the *probability of occurrence* P_{U_C} for this particular type of component C, then the event of components' unavailability U_C occurs. Subsequently, the *delay* (or waiting time) D_C until these are ready for loading is sampled from F_{D_C}. Otherwise, if $X > P_{U_C}$, the stock S_C was replenished in the meantime with an amount equal to a user-defined replenish strategy Rp_C for the particular component C. Then, the algorithm proceeds with loading the vessel and updating the stock level S_C.

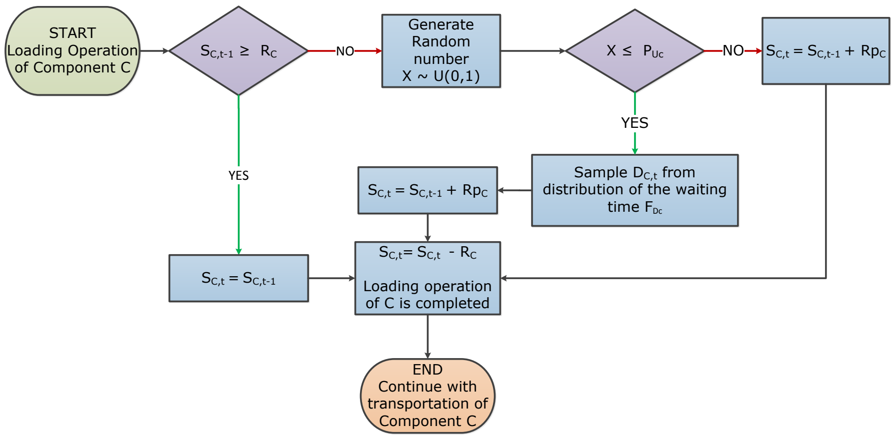


Figure 3: Flowchart of the algorithm that is used to include the risk of supply disruptions.

2.2. *Method to quantify risk of supply disruption*

Detailed data concerning OWFs construction projects are scarce. Moreover, these information are commercially sensitive and it is challenging to acquire a sufficient amount of relevant data. Hence, it is not possible to do a statistical analysis to obtain representative values concerning the probability of occurrence and the distribution of the waiting time which are required to include the risk of supply disruptions (see Figure 3). Therefore, the best alternative is to use expert judgments to assess these uncertain variables.

In order to be able to use the judgments of different experts as inputs of the developed stochastic simulation model, these should be combined for every variable of interest. There are different approaches and methods for combining the assessments of the experts. The subject of treating expert judgments as an alternative source of data has been extensively discussed (Cooke, 1991; O'Hagan et al., 2006; Dias et al., 2018). In this study a performance-based method was used to aggregate the expert judgments, namely the Cooke's classical model for structured expert judgment. This mathematical aggregation method was chosen because it has been shown that such a mathematical approach has an advantage over behavioral approaches (Clemen & Winkler, 1999). It is also worth mentioning that Cooke's classical model was used in the past for a different application on the offshore wind field. Zitrou et al. (2015) proposed a structured expert judgment elicitation process to quantify the required parameters associated with epistemic uncertainties and develop an availability growth model that takes into account the sources of systemic risk to provide a more accurate estimation of farm performance over early life.

2.2.1. Cooke's classical model

Cooke's classical model (CM) for structured expert judgment has been used in numerous studies (databases of these studies can be found in Colson & Cooke (2017); Cooke & Goossens (2008)) and it is widely accepted as a risk analysis tool for quantifying uncertainty and building rational consensus. Although, several good descriptions of this SEJ method can be found in literature, the main concepts of Cooke's classical model are summarized below. This with the purpose of making available to the reader the main elements of the CM ~~method~~. For details and extensive discussion the reader is referred to (Cooke, 1991) and the supplementary information for (Colson & Cooke, 2017).

In CM method, one or more elicitation sessions take place, where the group of experts provide their assessments, answering a questionnaire individually. This questionnaire consists of two parts; the first part comprises the *calibration* or *seed questions*, which concern quantities whose value is known (or will be known within the time frame of the research) to the analysts but it is not known to the experts at the moment of the elicitation, while the second part comprises the *target questions*, concerning the variables of interest (or target variables).

The experts are asked to provide assessments concerning (usually) the 5^{th}, 50^{th} and 95^{th} percentiles of their uncertainty distribution regarding each variable. The performance of each expert in judging uncertainty is measured in terms of *statistical accuracy* (or *calibration score*) and *informativeness*; These are denoted respectively with $C(e)$ and $I(e)$ for every expert e participating in the pool with E experts. A statistically accurate expert is the expert whose assessments capture the true values of the *seed questions* with the long run correct relative frequencies. In other words, *seed variables* are used to ensure empirical control of experts' uncertainty assessments. In order to be informative, the expert should provide narrow distributions. However, it should be noted that statistical accuracy is more important than informativeness with regard to the overall performance. Therefore, experts usually provide wide distributions if these assessments reflect their real uncertainty. Finally, based on the aforementioned measures, the weights of the experts are computed and the expert opinions can be combined. The measures of performance in judging uncertainty are presented in the Appendix A, while the aggregation of the expert judgments based on the performance of every expert is briefly presented next.

Aggregation of expert judgments. In ~~SEJ method~~ CM the combination of experts' assessments is called a *Decision Maker* (DM). This is a weighted average of individual estimates (i.e. the probability densities $f_{e,i}$ of every expert e regarding item i). When the weights are determined based on the performance of experts in the seed variables, we speak of *performance-based* DM. The DM probability densities are thus:

$$f_{DM,i} = \frac{\sum_{e=1}^{E} w_\alpha(e) f_{e,i}}{\sum_{e=1}^{E} w_\alpha(e)} \tag{1}$$

The weights in Cooke's model are weakly asymptotically strictly proper. This property ensures that if an expert wishes to maximize her long run expected weight then she should do this by stating her true beliefs as answer to the seed variables (Cooke, 1991). The weights for each expert $w_\alpha(e)$ are given by the product of calibration and information scores when a certain threshold in calibration is attained:

$$w_\alpha(e) = 1_{\{C(e)>\alpha\}} C(e) I(e). \tag{2}$$

where $1_{\{A\}}$ denotes the indicator function for A. Usually values of $\alpha < 0.05$ (or 0.01) would fail to confer the study the required level of confidence.

For the purpose of this study three different types of DMs were obtained using different weighting schemes. The simplest one is equal weighting and hence falls outside of the performance based DMs. The *Global Weights* DM is computed as described above while the *Item Weights* DM computes the scores in equation 2 using the information score per item rather than the average information score (equation A.3 in Appendix A). The difference between DMs will be discussed further in section 3.2.

2.2.2. Seed questions

As it was mentioned before, some relevant data from the field of offshore wind construction are needed to formulate seed questions. In this study a Dutch marine contractor (i.e. Van Oord) provided the time register database of four past projects concerning the construction of OWFs. This database contains a detailed archive with the duration of all the activities as well as the incidents that occurred during these projects. Hence, although these data, concerning four projects, were not sufficient to do a statistical analysis and draw safe conclusions, they were valuable to formulate the seed questions which were used to evaluate the performance in judging uncertainty of every participant.

For the purpose of this study, 14 seed questions were formulated. These questions concerned frequencies of occurrence as well as the registered waiting time of the delays due to unavailability of different components. The projects were anonymized but important details of these projects were provided to the experts. An example of the provided information regarding project (construction of an OWF with 32 wind turbines, located in the Irish sea) can be found in Table 1. To illustrate the format of the seed questions that were used, one of these is presented below:

SQ. 1*: Consider an installation project (Project 1) of an OWF consisting of 32 WTG in the Irish sea with the details presented in Table 1. A number of times, the monopiles (MPs) were not available while the vessel was on-site ready to start the installation. What do you believe was the* ***maximum*** *registered delay, until the required MPs were available?*

5%-tile (Q_5)	50%-tile (Q_{50})	95%-tile (Q_{95})
...	...	...

2.2.3. Target questions

The purpose of this study is to obtain probability distributions regarding the delays that may occur due to unavailability of different components (such as monopiles, transition pieces, towers, blades, nacelles) during the construction of OWFs. For every different component, two types of target questions were queried, resulting in 10 target questions in total. The first type concerns the relative frequency of the event ~~that~~ in which the required components are not available. The second type concerns the waiting time (i.e. delay) due to the occurrence of such an event. The obtained distributions can be used to support projects which will happen in the near future with certain characteristics. These characteristics are presented in Table 2. To illustrate the two different types of target questions (TQ. 1 and TQ. 2) that were used, an example concerning the monopiles is presented below.

TQ. 1*: Assuming that the operation of loading monopiles (MPs) to the transportation vessel will be performed 1000 times,* ***how many times*** *would you expect that the required MPs will not be available to start the loading operation?*

Components	Monopiles	Transition Pieces
Installation port	Birkenhead, Liverpool	Birkenhead, Liverpool
Manufacturer location	Rostock	Teesport (16 TPs) Aalborg (16 TPs)
Distance of installation port from manufacturer	$\approx 1150NM$	$\approx 750NM$ $\approx 980NM$
Transportation method to installation port	Shipped (vessel speed 15 kn)	Shipped (vessel speed 15 kn)
Estimated transportation duration to installation port	$\approx$ 75 h	$\approx$ 50 h $\approx$ 65 h
Number of trips from manufacturer	8	8
Buffer stock at installation port at the commencement of the installation operation	$\approx$ 20	$\approx$20
Transportation from installation port to OWF site	Tugs towed floating MPs to the installation vessel on-site	Barges transferred TPs to the installation vessel on-site

Table 1: Information provided to the experts regarding Project 1.

5%-tile (Q_5)	50%-tile (Q_{50})	95%-tile (Q_{95})
...	...	...

***TQ. 2**: If the required monopiles (MPs) are not ready for loading while the transportation vessel is in port, what would you expect to be the delay (i.e. waiting time) until the required MPs are available for loading?*

5%-tile (Q_5)	50%-tile (Q_{50})	95%-tile (Q_{95})
...	...	...

Location	North Sea
Number of Wind turbines	More than 50
Distance from manufacturers to the installation port	More than 150 NM
Distance from installation port to the OWF site	More than 20 NM

Table 2: Characteristics of projects which can be supported by the results of this study.

3. Analysis of expert judgments

In this study, 11 experts with experience in the offshore wind field have participated. The participants were affiliated to different types of companies (such as marine contractors, manufacturers, OWF owners, consultancy firms etc.) operating in ~~from~~ 4 different European countries (i.e. the Netherlands, Germany, UK and Belgium) ~~and different types of companies (such as marine contractors, manufacturers, OWF owners, consultancy firms etc.)~~ in order to ensure the elicitation of expert judgments from a pool of experts with diverse expertise. The criteria for the participation of an expert concerned his/her experience in the offshore wind industry and involvement in the construction of OWFs. A list with the names and affiliation of the experts can be found in Appendix B.

The expert judgments were elicited in the period from July 12^{th} until August 15^{th} 2017. Few experts participated in an expert judgment elicitation workshop that took place on July 12^{th} 2017 in Delft while the expert judgments of the remaining participants were elicited by individual (teleconference or in person) interviews with the analyst. The structure of all the elicitation sessions was kept the same, in order to ensure that all the experts were provided with the same information. Each elicitation session consisted of: i) a short presentation of the purpose of this study; ii) an introduction to Cooke's classical model for structure expert judgment; iii) a short example exercise and iv) the filling out of the questionnaire regarding the risk of supply disruptions during the OWF installation process. The vast majority of the SEJ studies conducted in the past made use of the closed-source software EXCALIBUR to perform the analysis and synthesis of expert judgments according to CM. In this study, a newly developed open-source MATLAB toolbox called ANDURIL[1](Leontaris & Morales-Nápoles, 2018) was used. This toolbox provides open access to a code that makes transparent the calculations of performance measures and the aggregation of expert judgments, allowing the current methods to be more accessible and different approaches or extensions to current methods to be further explored. It should be noted that the obtained results were validated using EXCALIBUR software.

3.1. Performance of the experts

As it was mentioned in 2.2.1, the performance of the experts in judging uncertainty is evaluated using two measures, the calibration score (or statistical accuracy) and the information score. A more detailed explanation about these measures can be found in the Appendix A. Based on these measures, it is possible to compute the un-normalized weights using eq. 2 and subsequently the normalized weights.

Table 3 shows the measures of performance in judging uncertainty as well as the un-normalized and normalized weights of every participant. It should be noted that the order of these values concerning the anonymized experts does not correspond to the order of the names of the experts (provided in Appendix B). The presented values of the information score are computed by ANDURIL considering a log-uniform background measure and the weights are computed when considering a significance or cut-off level α equal to zero. It can be seen that experts 2 and 4 are the only experts with calibration score larger than 0.01 (a value that is usually chosen for the significance level α). This leads to larger weights which means that their judgments will mainly constitute the DM's distributions of every queried variable. It is worth mentioning that the information scores of the best calibrated experts are low compared to those of the remaining participants. Although this is not always the case, it is logical since the provided quantiles cover a larger range. In this study most of the remaining experts with low calibration scores, had high information scores. That means that these experts were confident about their assessments and provided narrower distributions for the queried seed variables, which were not statistically accurate.

3.2. Synthesis of DMs

There are different weighting schemes that can be used to aggregate expert opinions for every variable and obtain the so called decision maker (DM) (i.e. the combined opinion). The simplest one is to assign equal weights to every expert, which provides the equal-weight decision maker

[1] ANDURIL toolbox is freely available at `https://github.com/ElsevierSoftwareX/SOFTX_2018_39`. The acronym ANDURIL stands for "ANalysis and Decisions with UnceRtaInty: Learning from expert judgments" and this name was inspired from the universe of Lord of the Rings by J.R.R. Tolkien.

Expert ID	Calibration Score	Information Score (All items)	Information Score (Seed items)	Un-normalized Weights	Normalized Weights excl. DM
Expert 1	0,0002060	0,70675	0,86518	0,00017825	0,0006529
Expert 2	0,011904	0,471358	0,516478	0,0061485	0,022523
Expert 3	6,804e-10	0,93655	1,090144	7,418e-10	2,717e-09
Expert 4	0,569084	0,452635	0,4606201	0,2621317	0,9602551
Expert 5	1,983e-07	1,214771	1,1789866	2,338e-07	8,565e-07
Expert 6	1,314e-05	0,82397	0,712980	9,366e-06	3,431e-05
Expert 7	1,192e-07	1,214716	1,195092	1,425e-07	5,219e-07
Expert 8	0,00036218	1,165429	1,137962	0,00041215	0,00150981
Expert 9	2,547e-11	0,8838053	0,802524	2,044e-11	7,489e-11
Expert 10	1,762e-05	0,8577063	0,897994	1,582e-05	5,795e-05
Expert 11	0,00441208	0,8424475	0,9259006	0,0040851	0,0149649

Table 3: Measures of performance in judging uncertainty and weights for every participant, obtained from the analysis with ANDURIL.

(DM_{equal}) which of course does not take into account the performance of the experts in judging uncertainty. As it was introduced in section 2.2.1, CM method uses the two performance measures (i.e. *statistical accuracy/calibration score* and *informativeness*) to compute the weights based on the performance of the experts in judging uncertainty.

There are two different types of performance-based weighting schemes. The first one is called *global weights*, for which the weight of every expert is computed by taking the product of the calibration score and the average relative information over all the seed items. These global weights are multiplied with the density of every expert for every variable (see equation 1) to obtain the density of the combined opinion (DM_{global}). Hence, DM_{global} was synthesized using the values of the normalized weights presented in Table 3. The second type of performance-based weighting schemes is called *item weights* and the main difference with *global weights* is that the weights are different for every item based on the relative information for this particular item. This means that the opinion of every expert (whose calibration score is larger than the significant level α) has a different weight for every item. By multiplying the non-zero weight of every expert for every item with the experts density of the variable, the density of the combined opinion (DM_{item}) is obtained.

~~These~~ The aforementioned three different weighting schemes were investigated in this study and the performance of the resulting DMs was evaluated in terms of statistical accuracy and relative information (with respect to the log-uniform background measure for every variable). To achieve this, every synthesized DM is treated as a "virtual expert" that enters the pool of experts. Table 4 summarizes the performance measures (statistical accuracy or calibration and mean relative information) for every decision maker obtained using different weighting schemes (DM_{global}, DM_{item} and DM_{equal}). It is interesting to note that both the statistical accuracy and the informativeness of the decision makers which were obtained using performance-based weights (i.e. DM_{global} and DM_{item}) are significantly better than those of the decision maker obtained with equal weights (DM_{equal}). Also, the statistical accuracy of the DM_{equal} is below 0.05 that is the rejection threshold usually set as significance level. Moreover, the statistical accuracy of the performance-based decision makers is higher compared to that of the expert with the higher calibration score (i.e. Expert 4). However, the relative information of the performance based decision makers is lower than that of every individual expert. This means that in general the synthesized decision makers have wider distributions than most experts for the different items. Finally, one can see that there is not a big difference between the measures of performance for the two performance-based decision makers DM_{global} and DM_{item}.

It is important to note that the asymptotic strictly scoring rule property requires a cut-off level beneath which an expert is unweighed. Usually, in practice, the value of α is either set equal to 0.01 (that is more "generous" than the traditional 0.05 for hypothesis testing) or it is chosen to optimize the combined score of the resulting DM. However, in this study, the cut-off level was chosen to be equal to zero. In this way, the judgment of every expert, no matter how small its contribution, is included in the obtained distributions. Although, all the participants in this study were experts in the field of offshore wind energy, they had different functions and were affiliated to different types of companies. Hence, the diversity of the pool of experts could be reflected to the resulting distributions, by setting $alpha = 0$. However, the analysis was also performed when considering $alpha = 0.01$. It was found that this had a minor influence to the obtained distributions and the measures of performance of the resulting performance based DMs (which can be found in Appendix F). This was expected as experts 2 and 4 have significanlty higher statistical accuracy compared to the other experts.

Name	Calibration Score	Information score (total)	Information score (seed items)
DM_{global}	0.96812	0.34834	0.34809
DM_{item}	0.96812	0.35677	0.35466
DM_{equal}	0.03868	0.13560	0.131805

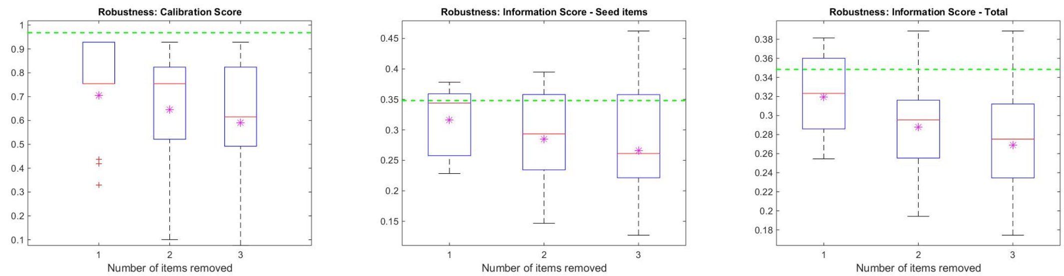
Table 4: Comparison of the three DMs' performance measures.

3.3. Robustness of performance-based DMs using ANDURIL

It is also worth investigating the robustness of the obtained DMs with respect to the seed questions. To achieve this, one or more seed items are removed and the DMs are computed again. Then the resulting DMs are evaluated in terms of statistical accuracy and informativeness.

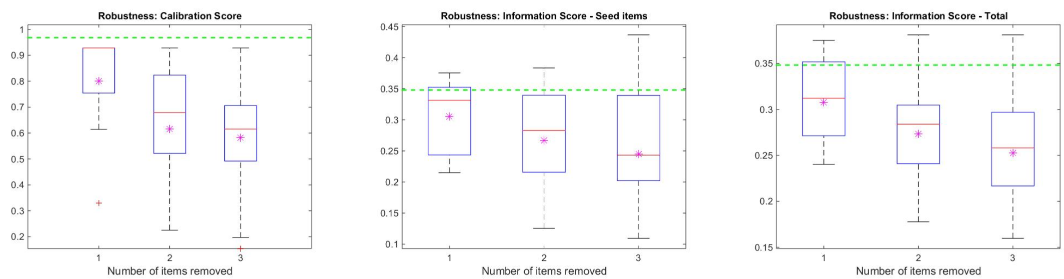
In this study, the robustness analysis has been performed using the newly developed ANDURIL toolbox (Leontaris & Morales-Nápoles, 2018). The reason for this choice is that ANDURIL gives the opportunity to the user to decide the number of the seed items which should be excluded at a time. For the purpose of this study, it was chosen to exclude all the possible combinations up to 3 seed items at a time (i.e. a total of 469 combinations).

The results of the robustness investigation of DM_{item} and DM_{global} are presented in Fig. 4 and Fig. 5 respectively. It can be seen that the calibration score of the DM_{item} can range from 0.7 to 0.92 approximately with a median of ≈ 0.75, when one seed item is excluded at a time. As it was expected the range of the obtained calibration scores increases as more than one seed items are excluded at a time. However, the median of the calibration score when excluding two seed items at a time remains close to the median occurring when only one seed item is excluded. On the contrary, the median of the calibration score is ≈ 0.6 when all combinations of 3 seed items are excluded at a time. Although the resulting calibration scores are not very close to the original calibration score (i.e. green dashed horizontal line), these are significantly larger than 0.05 that is usually set as the threshold at which the study would cast doubts regarding its estimates. Moreover, the resulting information scores of the DM_{item} are less spread and the medians are closer to the original value concerning both the information score over the seed items as well as the one over the target questions. As far as the robustness box-plots of the DM_{global} (Fig. 5) are concerned, similar conclusions can be drawn. Following the same approach, it is also possible to investigate the robustness of the obtained DM with respect to the experts. It was chosen to investigate this when excluding one expert at a time. The performance measures of the resulting performance-based DMs are presented in Table 5. Both measures of performance and especially informativeness decrease significantly when expert 4 is excluded, while only the statistical accuracy is reduced when expert 2 is excluded.



(a) Calibration score (b) Information - Seed (c) Information - Total

Figure 4: Robustness box-plots concerning (a) the calibration score, (b) the information score over the seed items and (c) the information score over all the queried variables of the DM_{item}, when excluding up to three seed items at a time.



(a) Calibration score (b) Information - Seed (c) Information - Total

Figure 5: Robustness box-plots concerning (a) the calibration score, (b) the information score over the seed items and (c) the information score over all the queried variables of the DM_{global}, when excluding up to three seed items at a time.

3.4. Obtained distributions - Inputs for simulation tool

The obtained distributions concern the relative frequency of occurrence of a delay due to unavailability of required components and the waiting time distribution in between availability of the major components required for the installation of offshore WTGs. These components are: the

Excluded Expert	DM_{global} Calibration Score	DM_{global} Information score (total)	DM_{global} Information score (seed items)	DM_{item} Calibration Score	DM_{item} Information score (total)	DM_{item} Information score (seed items)
Expert 1	0.96812	0.33982	0.34875	0.96812	0.34771	0.35533
Expert 2	0.43118	0.37786	0.40134	0.43118	0.38028	0.39926
Expert 3	0.96812	0.34834	0.34809	0.96812	0.35677	0.35466
Expert 4	0.65873	0.15863	0.13098	0.65873	0.18163	0.17062
Expert 5	0.96812	0.34834	0.34810	0.96812	0.35677	0.35466
Expert 6	0.96812	0.34837	0.34816	0.96812	0.35680	0.35472
Expert 7	0.96812	0.34834	0.34809	0.96812	0.35677	0.35466
Expert 8	0.96812	0.35075	0.35146	0.96812	0.35927	0.35782
Expert 9	0.96812	0.33866	0.34091	0.96812	0.34699	0.34724
Expert 10	0.96812	0.34842	0.34818	0.96812	0.35686	0.35474
Expert 11	0.96812	0.35697	0.33835	0.96812	0.36722	0.35152

Table 5: Robustness of performance-based DMs with respect to experts.

monopiles (MPs) the transition pieces (TPs), the towers, the blades and the nacelles. Figure 1 shows the different main components of a typical offshore WTG that were taken into account in this study.

Among the different combinations of the experts' judgments presented in section 3.2, it would be a sensible decision to use as inputs for the stochastic simulation model, the distributions obtained using item weights, (i.e. DM_{item}). The main reasons for this choice are based on the evaluation of the performance of the obtained DMs (treated as "virtual experts") in assessing uncertainty. In Table 3 and Table 4, it can be seen that DM_{item} has higher statistical accuracy compared to every individual expert and higher relative information (which is a slowly varying function) compared to DM_{global}.

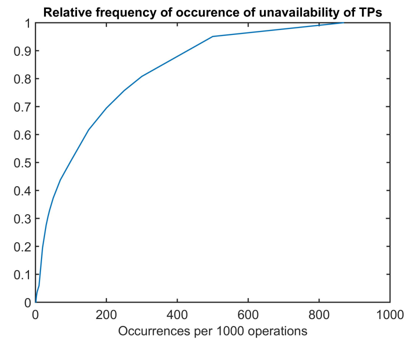
To illustrate the obtained distributions using the item weights weighting scheme, Figure 6 presents the distributions concerning the relative frequency of occurrence of unavailability of required TPs and required blades, as well as the waiting time due to this. The distributions, obtained with performance based weighting schemes, presented small differences across the different types of components. Moreover, these distributions can be characterized as significantly wide (e.g. the waiting time until the required TPs become available can range from minutes up to several days). This can be explained because the queried variables express large variance since these are influenced by many different aspects. Also, the obtained distributions can be used to support projects with various characteristics, rather than one project that partly express the large variance of the output distributions. In Appendix D and E, one can find the figures that present the distributions of every individual expert as well as the obtained distributions of every DM concerning the seed and target variables respectively.

4. Model implementation

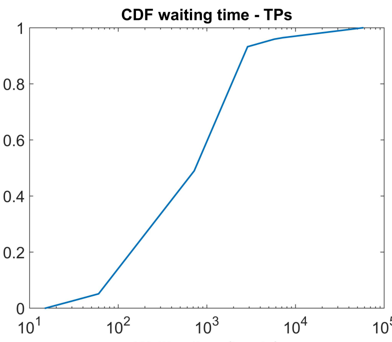
4.1. ECN Install

In order to investigate the impact of including/neglecting the uncertainties regarding the supply disruptions during the installation process of OWFs, a realistic test case should be simulated. For this purpose, a modified version of a software developed from ECN (ECN Install v. 2.1) (Dewan et al., 2015) was used to simulate different scenarios and provide the user with cumulative distributions of time and cost of each scenario. ECN Install is a time driven decision support tool that simulates an installation scenario for a different number of historical environmental time series and provides the user with the estimates regarding the duration and the cost. The tool provides excellent flexibility in the hands of the user to model the desired planning and export the cost and time outputs for any project. Due to the high reliance on the user-defined inputs, the outputs are profoundly dependent on the quality of input data. Figure 7 presents the start screen and the user interface of the software.

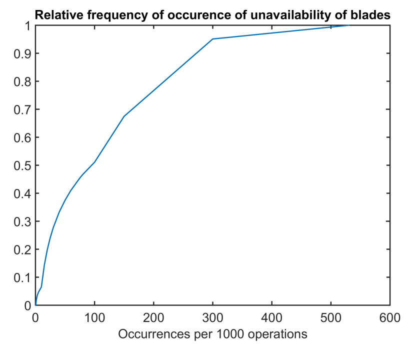
ECN Install highlights the barriers during the installation activities and supports in eliminating project risks. ECN Install is designed to test various conceptual installation strategies for accelerating the knowledge transfer between different actors involved. It leads towards efficient resource management to minimize the possible delays and overall costs for simulated schedules. The ECN Install simulation tool is in existence from early 2014, where over the years it has seen systematic improvements. The latest commercial tool available is based on version 2.1. This version of the



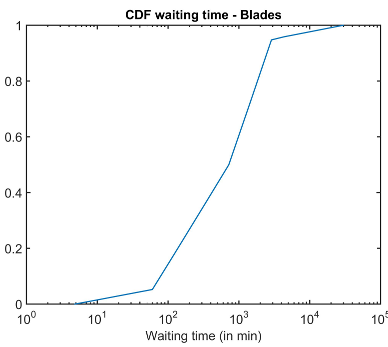
(a) Relative Frequency of occurrence - TPs



(b) Waiting time due to unavailability of TPs



(c) Relative Frequency of occurrence - Blades



(d) Waiting time due to unavailability of required blades

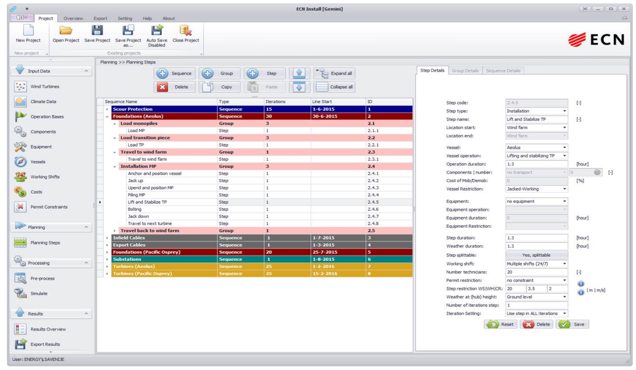
Figure 6: Obtained distributions concerning the relative frequency of delays due to unavailability of the required TPs (a), blades (c) and the waiting time due to unavailability of the required TPs (b), blades (d).

tool was modified with the algorithm presented in subsection 2.1. A realistic test case that was developed from ECN was simulated using the modified ECN Install tool.

4.2. Test case details

The realistic test case that was developed from ECN and presented in (Katsouris, 2015), concerns an OWF consisting of 150 wind turbines and simulates scour protection activities and the installation operations of the support structures (i.e. monopiles and transition pieces), the wind turbines, the (infield and export) cables and the offshore substation. The installation plan of the simulated test case is split in the following phases:

(a) Start screen



(b) Example of user interface

Figure 7: ECN Install software

1. Scour protection: Rock dumping is performed for each of the 150 monopiles as scour protection method. It is assumed that rock loading to the vessel takes place in the port of Eemshaven.

2. Foundations: The foundation of each wind turbine consist of a monopile and a transition piece. Two vessels were employed for the installation of foundations. Both vessels carry three complete foundations at a time. The installation of 90 foundations is performed by vessel 1 while 60 of them by vessel 2. Loading of foundations to the vessels is carried out in Orange Blue Terminal of Eemshaven port. Last, it is noted that an extra restriction concerning the piling is applied in the Netherlands from January 1st until July 1st.

3. Export cables: Two AC 220 kV export cables of over 100 km each, weighting 90 kg/m are installed. Their installation is a challenging and complex engineering process because of several restricted areas. Briefly, the installation of export cables is split in shallow waters, near shore, deep waters and connector cable installation.

4. Substations: Two 300 MW substations are installed on top of two jacket structures. A heavy lift vessel installs first the two jackets and subsequently positions the substations. The jackets are towed by tugboats from Bow Terminal of Vlissingen port whereas the substation are transported from Hoboken, Antwerp.

5. Inter-array cables: Approximately 140 km of inter-array cables are required to interconnect the wind turbines with an average weight of 30-40 kg/m. Inter-array cable laying is carried

out by a cable laying vessel whereas the post-burying is performed by a multi purpose vessel and a remotely operated vehicle (ROV).

6. Wind turbines: Wind turbines installation is performed by 2 vessels (75 turbines each) while components are loaded in the port of Esbjerg. For both vessels, one loading includes three complete wind turbines and for each turbine, the tower comes at one piece while nacelle and hub are pre-assembled. Hence, five lifts are required for the installation of one turbine.

The details of the simulated test case are summarized in Table 6.

Number of WTG:	150
Location:	North Sea
Starting Date:	1 June
Installation operations:	Scour protection, Support structures, Wind turbines, Cables and Offshore Substations
Environmental conditions:	20 years of observations in the OWF site
Strategy 1: Initial stock at the installation port in the commencement of project	10 units of each component (MPs, TPs, Towers, Nacelles, Rotors)
Strategy 2: Initial stock at the installation port in the commencement of project	20 units of each component (MPs, TPs, Towers, Nacelles, Rotors)

Table 6: Details of the simulated test case.

The presented installation test case was simulated for three different approaches, which represent the risk aversion of the user regarding the supply disruptions. In *Approach 1 (or Base)*: the risk of supply disruptions is neglected, in *Approach 2 (or Neutral)*: the risk of supply disruptions is described with a "moderate" frequency of occurrence and in *Approach 3 (or Pessimistic)*: the risk of supply disruptions is included with an "high" frequency of occurrence. The terms "moderate" and "high" represent the median (50^{th} percentile) and the 95^{th} percentile of the obtained probability distributions regarding the frequency of occurrence, respectively.

4.3. Results

The realistic test case was simulated 1000 times in total (i.e. 50 simulations for every available year of environmental time series) to properly incorporate the risk of supply disruptions. Following this approach, different strategies can be compared in order to support decisions with a certain confidence level. For illustration purposes, two different strategies concerning the initial stock in the commencement of the installation project were simulated (see also Table 6). The results of the test case when *Strategy 1* and *Strategy 2* are considered, are presented below. It should be mentioned that for both strategies the replenished amount of every component (as presented in Figure 3) was equal to 5 units and the stock in the harbor could not exceed the initial stock of every strategy.

4.3.1. Initial stock: Strategy 1

In Figure 8 and 9 are presented the duration of the installation (in days) and the total cost of the installation (when the cost of the components is not included) respectively. It can be seen that when the risk of supply disruption is included with a "moderate" relative frequency of occurrence (i.e. the 50^{th} percentile of the CDF of every component), the cumulative distributions of the duration and the cost of the installation present small differences compared to the case where this risk is neglected. More precisely, concerning the 80^{th} percentile (or P80 value), there is a difference of approximately equal to $\approx$ 4 days and $\approx$ 0.42 M€ for the duration and the cost of the installation respectively. However, these values differ significantly in the pessimistic case, where the risk of supply disruptions is included with a "high" relative frequency of occurrence (i.e. the 95^{th} percentile of the CDF of every component). In this case, the difference compared to the *Base* approach is $\approx$ 15 days for the duration and $\approx$ 3.9 M€, for the cost of the installation.

An important remark concerning the resulting cost CDF of the *Neutral* approach is that the total cost of the installation sometimes appears to be smaller than the *Base* approach. Although this may be counter-intuitive, it shows a more realistic modeling of the total cost calculation. The total cost of the installation is rigorously calculated based on the resources utilization. One can find a short description of the cost calculation in Appendix C and a more detailed description of the cost calculation that is supported by ECN Install in (Katsouris, 2015). ~~This~~ The aforementioned remark can be explained by the difference in the cost of the vessels depending on their state (i.e. remaining idle, traveling or performing an activity); the weather windows that may happen to be longer and the fact that the vessels are hired for the entire day. Hence, there are a few cases where although a small delay in the loading operation results in a delay in the total duration, the total cost of the installation is lower compared to the case where this delay has not occurred.

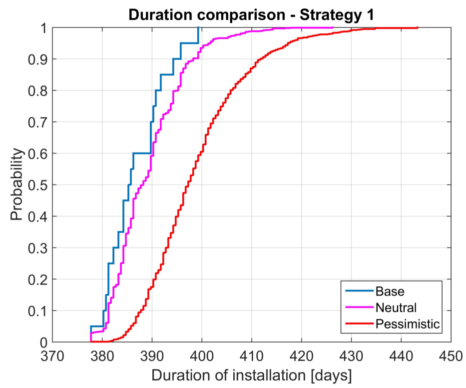


Figure 8: Duration of Strategy 1 for different risk aversions.

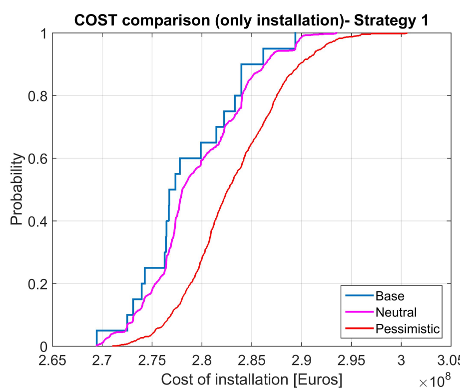


Figure 9: Cost of Strategy 1 for different risk aversions.

4.3.2. Initial stock: Strategy 2

A similar trend concerning the difference between the CDFs obtained for different risk aversion approaches, can be seen when *Strategy 2* is considered for the initial stock at the port. In this scenario, all the details of the test case were kept the same, except the cost of the harbor. Since in *Strategy 2* there are 20 (instead of 10) units of every component in the commencement of the project, it was assumed that the cost of the harbor will be twice as high as the harbor cost in

Strategy 1. Figures 10 and 11 show the resulting CDFs of the duration and cost of the installation respectively. When *Strategy 2* is considered, the difference of the P80 value between the *Base* and *Neutral* case is ≈ 3.5 days for the duration and ≈ 0.52 M€, for the cost of the installation. Whereas the difference between the *Base* and the *Pessimistic* approach is ≈ 14.5 days for the duration and ≈ 4.23 M€, for the cost.

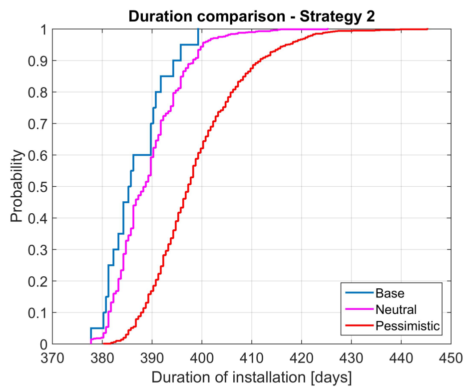


Figure 10: Duration of Strategy 2 for different risk aversions.

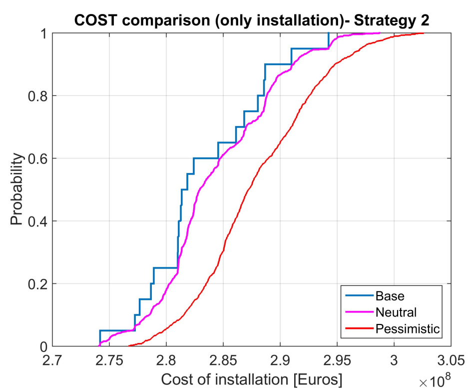


Figure 11: Cost of Strategy 2 for different risk aversions.

4.3.3. Comparison of strategies

It was shown that there is a significant impact into the estimates of the duration and cost of the installation, when the risk of supply disruptions is included with a *Pessimistic* approach. However, it is also interesting to investigate which of the two simulated strategies, regarding the initial stock, is the most economical when this risk is taken into account. Figure 12 shows the CDFs of the cost of the installation for the *Strategy 1* and *Strategy 2*. It can be clearly seen that *Strategy 1* should be preferred over *Strategy 2* in terms of cost. However, it must be mentioned that this result is heavily influenced by the cost of the harbor, which was considered twice as high for *Strategy 2*. Also, as it can be seen from Figures 8 and 10, the estimated duration present minor differences for both strategies. This would differ depending on the amount of initial stock as well as the replenished amount of components.

~~More~~ Numerous scenarios ~~could also~~ can be investigated. For example, an extreme scenario where it is assumed that all the required components for the installation of an OWF are available at the port in the commencement of the project, will result in a CDF of the duration identical to the *Base* approach. However, the CDF of the cost will be much different, showing a significantly higher total cost. To investigate such a scenario more realistically, the user could also add a monetary penalty above a defined threshold for the duration. These and many more different scenarios and strategies can be investigated and compared with the developed approach. In this way, it is possible for the decision makers to choose the most economical scenario with a certain confidence level.

To further explore alternative scenarios for the presented realistic test case, four different strategies regarding the initial stock of different components in the commencement of the installation process were also simulated. These strategies differ in an important characteristic of the project (i.e. initial stock of the components) that can be intuitively used as a mitigation measure for the risk of supply disruptions. The investigation of these additional strategies would show whether difference in the initial stock of particular component(s) would result in important differences in the estimated cost and time of the test case. It was found that the cumulative distribution of the total duration had insignificant differences for the simulated strategies. This can be explained by the small differences in the obtained distributions regarding the relative frequency of occurrence and the waiting time for different components. On the other hand the CDFs of the cost varied proportionally to the required area in the port. The details of all simulated strategies (including *Strategy 1* and *2*) as well as comparison of the obtained CDFs of the duration and cost for the pessimistic approach can be found in Appendix G.

5. Discussion

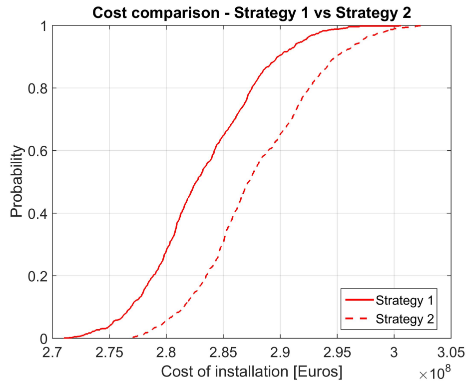


Figure 12: Cost comparison of different strategies for the pessimistic approach.

The proposed methodology was depicted in Figure 2 and it has some challenges and limitations that should be discussed. To begin with, considering the first sub-process of the proposed methodology, past construction projects from the field of application should be investigated in order to identify potential events which may cause delays due to supply disruptions. These are used for creating algorithm(s) in order to update the probabilistic model. This can be particularly challenging because in case unforeseen events are not modeled properly or are completely neglected, this might lead to underestimation of the supply disruptions risk. Furthermore, regarding the second sub-process of the proposed methodology, a number of challenges or limitations of applying a SEJ method such as the CM are listed below:

1. the data are often scarce and commercially sensitive for studies which are similar to the presented study. However, these are crucial for forming a sufficient number of seed questions based on which the performance of the experts is evaluated.

2. it can be challenging to find experts in the field of application who are interested in participating and available for several hours.

3. the analyst should have been trained in CM, to ensure the proper elicitation of the expert judgments in a structured way.

There are also limitations concerning the obtained probability distributions from applying the proposed methodology to the presented case. These probability distributions (i.e. the relative frequency of occurrence and the waiting time due to unavailability of components during the installation of

OWFs) are compatible with future projects which comply with the characteristics presented in Table 2. These characteristics were deliberately set as such to support a wide range of future projects. However, there might be the case that in the long-term future, significant changes to the associated technology and the offshore construction practice will render the obtained distributions invalid. In this case, it should be mentioned that the proposed methodology could be applied again to update the probability distributions.

6. Conclusions & Recommendations

In order to further facilitate the transition from conventional energy sources to sustainable energy technologies, such as offshore wind, it is crucial to improve the management of the installation of OWFs. The installation of OWFs is a complex process that is influenced by various uncertainties. One important uncertainty that is usually overlooked in practice and in scientific literature, is the risk of unavailability of the required components, when needed, due to supply disruptions. To fill this gap, a methodology was proposed to model the risk of supply disruptions.

Due to the absence of sufficient relevant historical data, expert judgments were used to quantify the risk of supply disruptions for different components. For this purpose, 11 experts from different companies and countries provided their assessments. Their judgments were analyzed by using Cooke's classical model and were combined using different weighting schemes. It was found that the combined opinion using item weights (DM_{item}) performed better in terms of statistical accuracy (i.e. calibration score) and informativeness. It is worth mentioning that the DM_{item} had a much higher calibration score compared to every individual expert as well as the equal weight DM (DM_{equal}), whose score was below 0.05 that is usually the threshold below which the study would cast doubts regarding its estimates.

Afterwards, a test case was simulated when following different risk aversion approaches. It was found that disregarding the supply disruptions from the estimated duration and cost may cause significant schedule and budget overrun. More precisely, in the *Pessimistic* case, neglecting this risk may lead to an underestimation of $\approx$ 4 M€, for the P80 value.

Furthermore, ~~two~~ six different strategies concerning the initial stock at the installation port were investigated. It was found that despite the CDFs of the duration were similar for all strategies, a large difference, proportional to the required area in the port, was observed in the CDFs of the total cost of the installation for these ~~two~~ strategies. Of course following the proposed methodology, it is possible to incorporate the risk of supply disruptions to various scenarios. Hence, it was shown that applying the proposed methodology allows to ~~to~~ incorporate the risk of supply disruptions to the cost estimates of various scenarios. This can assist in comparing scenarios and making optimal decisions regarding: 1. Schedule of installation; 2. Buffer stock; and 3. Selection of vessels and installation port

Ultimately, it must be mentioned that the proposed methodology and the obtained distributions can be used in future projects (with particular characteristics) to properly incorporate the risk of supply disruptions. However, since the obtained distributions can support multiple projects which share the characteristics presented in Table 2, ~~it was~~ these were expected to have a wide range. To improve this in future studies, two recommendations are given. First, the expert judgments could be elicited for a particular project of which the details will be given to the experts to reduce their uncertainty. Second, a dependence model, similar to those presented in (Werner et al., 2017), can be developed, where the dependence between the obtained distributions and different project characteristics will be described. Computing conditional distributions of interest for particular projects could be an application of such a model.

Acknowledgements

This research is part of the EUROS research programme, which is supported by NWO domain Applied and Engineering Sciences and partly funded by the Dutch Ministry of Economic Affairs. The authors would like to thank the internationally operating marine contractor Van Oord for providing essential data for this study, Mr. Georgios Katsouris for his useful advice regarding the modification of ECN Install and Dr. Tina Nane for her insightful remarks and support for the elicitation of expert judgments.

References

Arashpour, M., Bai, Y., Aranda-mena, G., Bab-Hadiashar, A., Hosseini, R., & Kalutara, P. (2017). Optimizing decisions in advanced manufacturing of prefabricated products: Theorizing supply chain configurations in off-site construction. *Automation in Construction*, *84*, 146 – 153. doi:10.1016/j.autcon.2017.08.032.

Barlow, E., Tezcaner Öztürk, D., Revie, M., Akartunal, K., Day, A. H., & Boulougouris, E. (2018). A mixed-method optimisation and simulation framework for supporting logistical decisions during offshore wind farm installations. *European Journal of Operational Research*, *264*, 894–906. doi:10.1016/j.ejor.2017.05.043.

Clemen, R. T., & Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, *19*, 187–203. doi:10.1111/j.1539-6924.1999.tb00399.x.

Colson, A. R., & Cooke, R. M. (2017). Cross validation for the classical model of structured expert judgment. *Reliability Engineering & System Safety*, *163*, 109 – 120. doi:10.1016/j.ress.2017.02.003.

Cooke, R. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Environmental ethics and science policy. Oxford University Press. ISBN 9780195064650.

Cooke, R. M., & Goossens, L. L. (2008). TU Delft expert judgment data base. *Reliability Engineering & System Safety*, *93*, 657 – 674. doi:10.1016/j.ress.2007.03.005.

Dewan, A., Asgarpour, M., & Savenije, R. (2015). *Commercial Proof of Innovative Offshore Wind Installation Concepts using ECN Install Tool*. Technical Report March. Retrieved from https://publications.ecn.nl/WIN/2015/ECN-E–15-007.

Dias, L. C., Morton, A., & Quigley, J. (Eds.) (2018). *Elicitation: The Science and Art of Structuring Judgement* volume 261 of *International Series in Operations Research & Management Science*. (1st ed.). Springer International Publishing. ISBN 978-3-319-65051-7

European Wind Energy Technology Platform (2014). Strategic Research Agenda / Market Deployment Strategy. Technical Report. Retrieved from https://windeurope.org/about-wind/reports/strategic-research-agenda-market-deployment-strategy/.

Fahimnia, B., Tang, C. S., Davarzani, H., & Sarkis, J. (2015). Quantitative models for managing supply chain risks: A review. *European Journal of Operational Research*, *247*, 1 – 15. doi:10.1016/j.ejor.2015.04.034.

Guo, S.-J., Chen, J.-H., & Chiu, C.-H. (2017). Fuzzy duration forecast model for wind turbine construction project subject to the impact of wind uncertainty. *Automation in Construction*, *81*, 401 – 410. doi:10.1016/j.autcon.2017.03.009.

Heckmann, I., Comes, T., & Nickel, S. (2015). A critical review on supply chain risk - Definition, measure and modeling. *Omega (United Kingdom)*, *52*, 119–132. doi:10.1016/j.omega.2014.10.004.

Hofmann, M. (2011). A Review of Decision Support Models for Offshore Wind Farms with an Emphasis on Operation and Maintenance Strategies. *Wind Engineering*, *35*, 1–15. doi:10.1260/0309-524X.35.1.1.

Irawan, C. A., Jones, D., & Ouelhadj, D. (2017). Bi-objective optimisation model for installation scheduling in offshore wind farms. *Computers and Operations Research*, *78*, 393–407. doi:10.1016/j.cor.2015.09.010.

Kaiser, M. J., & Snyder, B. F. (2013). Modeling offshore wind installation costs on the U . S . Outer Continental Shelf. *Renewable Energy*, *50*, 676–691. doi:10.1016/j.renene.2012.07.042.

Katsouris, G. (2015). Cost Modeling of the Installation of Offshore Wind Farms. Technical Report. Retrieved from https://www.slideshare.net/GeorgiosKatsouris/internship-reportgeorgios-katsouris-64880193.

Kerkhove, L. P., & Vanhoucke, M. (2017). Optimised scheduling for weather sensitive offshore construction projects. *Omega (United Kingdom)*, *66*, 58–78. doi:10.1016/j.omega.2016.01.011.

Lacal-Arántegui, R., Yusta, J. M., & Domínguez-Navarro, J. A. (2018). Offshore wind installation: Analysing the evidence behind improvements in installation time. *Renewable and Sustainable Energy Reviews*, *92*, 133–145. doi:10.1016/j.rser.2018.04.044.

Leontaris, G., & Morales-Nápoles, O. (2018). ANDURIL - A MATLAB toolbox for ANalysis and Decisions with UnceRtaInty: Learning from expert judgments. *SoftwareX*, *7*, 313 – 317. doi:10.1016/j.softx.2018.07.001.

Leontaris, G., Morales-Nápoles, O., & Wolfert, A. R. M. (2016). Probabilistic scheduling of offshore operations using copula based environmental time series An application for cable installation management for offshore wind farms. *Ocean Engineering*, *125*, 328–341. doi:10.1016/j.oceaneng.2016.08.029.

Leontaris, G., Morales-Nápoles, O., & Wolfert, A. R. M. (2017). Planning cable installation activities for offshore wind farms including risk of supply delays. In T. B. Lesley Walls, Matthew Revie (Ed.), *Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016 (Glasgow, Scotland, 25-29 September 2016)* (pp. 660–666). CRC Press. ISBN 9781138029972.

Mogre, R., Talluri, S. S., & Damico, F. (2016). A decision framework to mitigate supply chain risks: An application in the offshore-wind industry. *IEEE Transactions on Engineering Management*, *63*, 316–325. doi:10.1109/TEM.2016.2567539.

Morandeau, M., Walker, R. T., Argall, R., & Nicholls-Lee, R. F. (2013). Optimisation of marine energy installation operations. *International Journal of Marine Energy*, *3-4*, 14–26. doi:10.1016/j.ijome.2013.11.002.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., & Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons, Ltd. doi:10.1002/0470033312.

Paterson, J., D'Amico, F., Thies, P. R., Kurt, R. E., & Harrison, G. (2018). Offshore wind installation vessels A comparative assessment for UK offshore rounds 1 and 2. *Ocean Engineering*, *148*, 637–649. doi:10.1016/j.oceaneng.2017.08.008.

Poulsen, T., & Lema, R. (2017). Is the supply chain ready for the green transformation? The case of offshore wind logistics. *Renewable and Sustainable Energy Reviews*, *73*, 758–771. doi:10.1016/j.rser.2017.01.181.

Sarker, B. R., & Faiz, T. I. (2017). Minimizing transportation and installation costs for turbines in offshore wind farms. *Renewable Energy*, *101*, 667–679. doi:10.1016/j.renene.2016.09.014.

Scholz-Reiter, B., Heger, J., Lütjen, M., & Schweizer, A. (2011). A MILP for installation scheduling of offshore wind farms. *International Journal of Mathematical Models and Methods in Applied Sciences*, *5*, 371–378. Retrieved from http://www.naun.org/main/NAUN/ijmmas/19-793.pdf.

Shi, Q., Ding, X., Zuo, J., & Zillante, G. (2016). Mobile internet based construction supply chain management: A critical review. *Automation in Construction*, *72*, 143 – 154. doi:10.1016/j.autcon.2016.08.020.

Snyder, L. V., Atan, Z., Peng, P., Rong, Y., Schmitt, A. J., & Sinsoysal, B. (2016). OR / MS Models for Supply Chain Disruptions : A Review. *IIE Transactions*, *48*, 89–109. doi:10.1080/0740817X.2015.1067735.

Ursavas, E. (2017). A benders decomposition approach for solving the offshore wind farm installation planning at the North Sea. *European Journal of Operational Research*, *258*, 703–714. doi:10.1016/j.ejor.2016.08.057.

Vis, I. F., & Ursavas, E. (2016). Assessment approaches to logistics for offshore wind energy installation. *Sustainable Energy Technologies and Assessments*, *14*, 80–91. doi:10.1016/j.seta.2016.02.001.

Werner, C., Bedford, T., Cooke, R. M., Hanea, A. M., & Morales-Nápoles, O. (2017). Expert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions. *European Journal of Operational Research*, *258*, 801 – 819. doi:10.1016/j.ejor.2016.10.018.

Zitrou, A., Bedford, T., & Walls, L. (2015). Quantification and modelling of epistemic uncertainties for availability risk of future offshore wind farms using expert judgment. *Safety and Reliability: Methodology and Applications - Proceedings of the European Safety and Reliability Conference, ESREL 2014*, (pp. 805–812). ISBN 9781138026810.

Appendices

A. Measures of performance in judging uncertainty

Statistical accuracy. Assume that $e = 1, \ldots, E$ experts provided their assessments regarding three quantiles $q_{i,5}$, $q_{i,50}$ and $q_{i,95}$ (for the 5^{th}, 50^{th} and 95^{th} quantiles of each uncertain quantity) of $i =$

$1, \ldots, N$ seed variables and $1, \ldots, T$ variables of interest. There are thus $j = 1, \ldots, 4$ interquantile bins. The procedures described next may be easily extended by assuming more quantiles are assessed from each expert. For each quantity, each expert divides her belief range into four interquantile intervals, for which the j corresponding probabilities of occurrence $p = [p_1, \ldots, p_4]$ are: $p_1 = 0.05$ for a realization value $\leq 5^{th}$ percentile, $p_2 = 0.45$ for a realization value $\in (5^{th}, 50^{th}]$ percentiles, $p_3 = 0.45$ for a realization value $\in (50^{th}, 95^{th}]$ bin, and $p_4 = 0.05$ for a realization value $> 95^{th}$ percentile. The empirical version of $p = (p_1, \ldots, p_4)$ for expert e, is denoted $s(e) = (s_1, \ldots, s_4)$, where $s_j(e)$ is equal to the number of realizations of seed variables falling in the j^{th} interquantile assessed by expert e divided by the total number of seed variables.

$$
\begin{aligned}
s_1(e) &= \frac{\text{Number of realizations} \leq 5^{th} \text{ quantile}}{N} \\
s_2(e) &= \frac{\text{Number of realizations} \in (5^{th}, 50^{th}] \text{ quantile}}{N} \\
s_3(e) &= \frac{\text{Number of realizations} \in (50^{th}, 95^{th}] \text{ quantile}}{N} \\
s_4(e) &= \frac{\text{Number of realizations} > 95^{th} \text{ quantile}}{N}
\end{aligned}
$$

One way to measure the difference between p and $s(e)$ is through relative information or entropy (equation A.1), which is a measure of the disagreement between them.

$$
I(s(e), p) = \sum_{j=1}^{4} s_j(e) \ln \frac{s_j(e)}{p_j} \tag{A.1}
$$

Experts' assessments are treated as statistical hypotheses. Consider for each expert the null hypothesis H_0 : The interquantile interval containing the true value for each variable is drawn independently from the probability vector p. The quantity $2NI(s(e), p)$ where $I(s(e), p)$ is given in equation (A.1) is asymptotically χ^2_3 (where the degrees of freedom are equal to 3 in this case). This quantity can be used to test H_0 and it defines the calibration score (equation A.2). The probability in equation A.2 can be evaluated by a χ^2_3 distribution. The statistical accuracy or calibration score $C(e)$ of every expert is the probability that a deviation at least as large as r could be observed on N realizations if H_0 were true. Where r is the percentile of interest in the χ^2 distribution of interest obtained from evaluating $2NI(s(e), p)$ for the data corresponding to a particular expert. Values of calibration close to zero mean that it is unlikely that the experts' assessments are correct.

$$C(e) = P\{2NI(s(e), p) > r\} \tag{A.2}$$

Informativeness. The informativeness or information score measures the degree to which a distribution is concentrated (or spread out) with respect to a background measure. In the classical model the uniform or log-uniform background measures are used. An intrinsic range is calculated for each expert's density. The intrinsic range is obtained by adding a $k\%$ overshoot to the smallest interval containing all quantiles and realizations (when available), where k is selected by the analyst (typically $k\% = 0.1$). The lowest (l) and highest (h) values for the intrinsic range are $l_i = min\{q_{i,5}(e), v_i\}$ and $h_i = max\{q_{i,95}(e), v_i\}$ where v_i is the realization of interest. Then $q_{l_i} = l_i - k(h_i - l_i)$ and $q_{h_i} = h_i + k(h_i - l_i)$. The *information score* is then computed as shown in A.3. Also notice that in equation A.3 a uniform background measure is applied. For a log-uniform background measure the log of $q_{\cdot,i}$ would be used instead.

$$I(e) = \frac{1}{N}\sum_{i=1}^{N}\left[\ln(q_{h_i} - q_{l_i}) + p_1 \ln\frac{p_1}{q_{5,i} - q_{l,i}} + \ldots + p_4 \ln\frac{p_4}{q_{h,i} - q_{95,i}}\right] \tag{A.3}$$

B. Participants of the study

Name	**Affiliation**	**Country**
Andersen K. J.	Veja Mate	Germany
Claus M.	Siemens	the Netherlands
De Ridder E.	Jan de Nul	Belgium
Engelmann L.	Maritime Technik	Germany
Garrett C.	DNV-GL	United Kingdom
Holy A. L.	Vattenfall	Germany
Knipping D.	Van Oord	the Netherlands
Rabaut D.	DEME	Belgium
Rainey P.	EON	United Kingdom
Robert P.	DAMEN	the Netherlands
Warnaar P.	ECN	the Netherlands

C. Cost calculation - ECN Install

ECN Isntall calculates the cost of the installation process by keeping track of the installation activities in the time domain and hourly save the utilization of resources. For every performed simulation, information is gathered regarding the working time and waiting time (due to weather, supply disruptions, harbour and shift delays) for every vessels and equipment. Then, these are used accordingly to the user-defined costs. More precisely, the total cost of the installation process is computed using equation C.1.

$$Cost_{installation} = c_{vessels} + c_{equipment} + c_{ports} + c_{labour} \tag{C.1}$$

where:

$$c_{vesssels} = \sum_{v \in V} c_{fixed,v} + N_{dr,v} * d_{r,v} + N_{drw,v} * d_{rw,v} + N_{mob/demob,v} * c_{mob/demob,v} + N_{trips,v} * c_{add,v} \tag{C.2}$$

where v and V a vessel and the set of vessels respectively; $c_{fixed,v}$ the fixed cost of vessel v; $N_{dr,v}$ and $N_{drw,v}$ the number of working and waiting days of vessel v respectively; $d_{r,v}$ and $d_{rw,v}$ the day rates for working and waiting respectively; $N_{mob/demob,v}$ and $c_{mob/demob,v}$ number and cost of (de-)mobilizations respectively; $N_{trips,v}$ and $c_{add,v}$ number of trips and additional cost respectively.

$$c_{equipment} = \sum_{e \in E} c_{fixed,e} + N_{dr,e} * d_{r,e} + N_{drw,e} * d_{rw,e} \tag{C.3}$$

where e the equipment and E the set of used equipment; $c_{fixed,e}$ the fixed cost of the equipment; $N_{dr,e}$ and $N_{drw,e}$ the number of working and waiting days of equipment e respectively; $d_{r,e}$ and $d_{rw,e}$ are the day rates while working and waiting respectively.

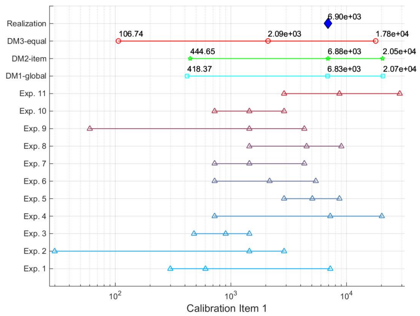
$$c_{ports} = \sum_{p \in P} c_p * D_p \tag{C.4}$$

where p and P denote one port and the set of ports respectively; c_p and D_p are the cost of port p per day and the number of days port p was used.

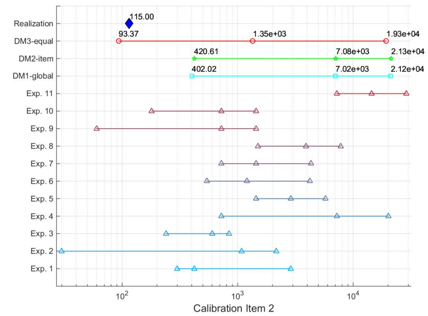
$$c_{labour} = \sum_{d_l \in D_l} c_l \tag{C.5}$$

where d_l and D_l denote one day and the set of days for which labour is performed; c_l is the labour cost per day.

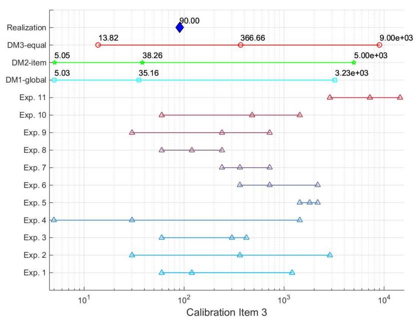
D. Calibration variables



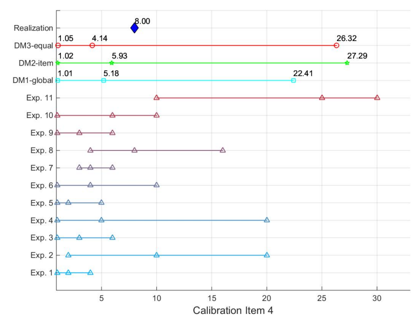
(a) Maximum registered delay because required MPs not available (Prj. 1).



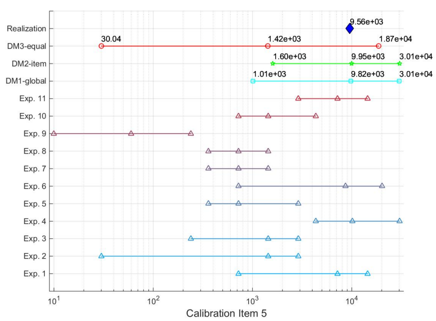
(b) Maximum registered delay because required TPs not available (Prj. 1).



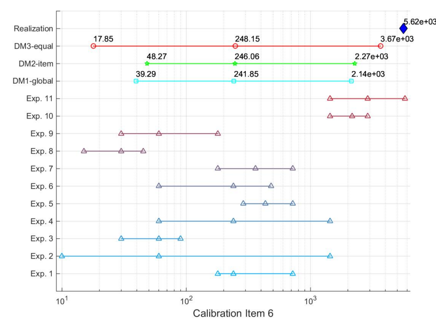
(c) Average of registered delays because required TPs not available (Prj. 1).



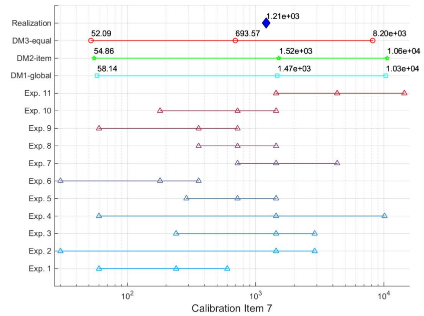
(d) Number of times there was a delay larger or equal to one hour because the required MPs were not available (Prj. 1).



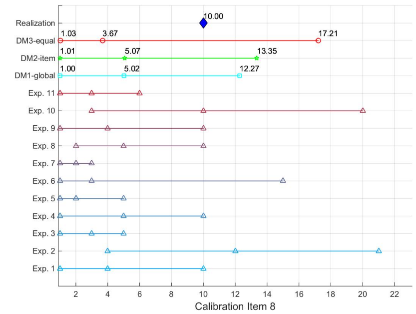
(e) Maximum registered delay because required TOWERS not available (Prj. 2).



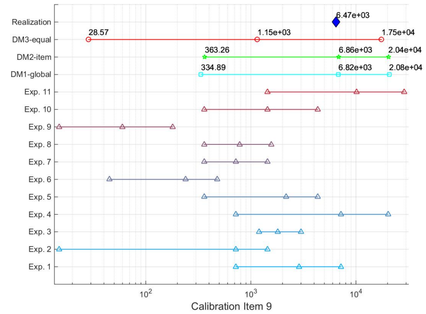
(f) Average of registered delays because required TOWERS not available (Prj. 2).



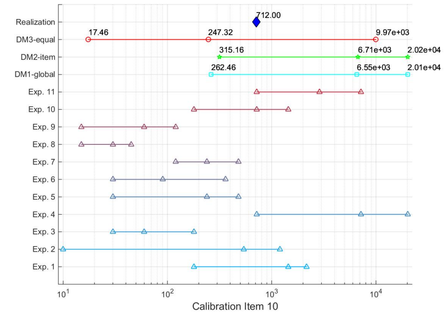
(g) Maximum registered delay because required BLADES not available (Prj. 2).



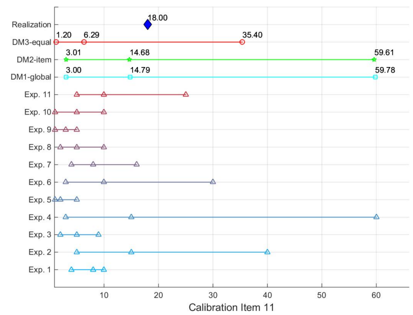
(h) Number of times there was a delay ($\geq 1h$) because the required TOWERS were not available (Prj. 2).



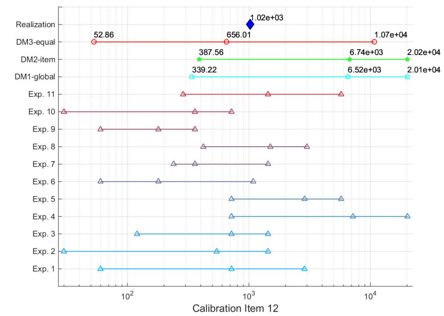
(i) Maximum registered delay because required MPs not available (Prj. 3).



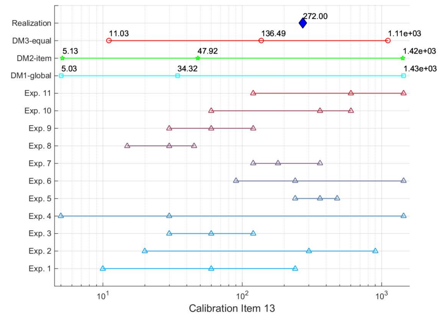
(j) Average of registered delays because required MPs not available (Prj. 3).



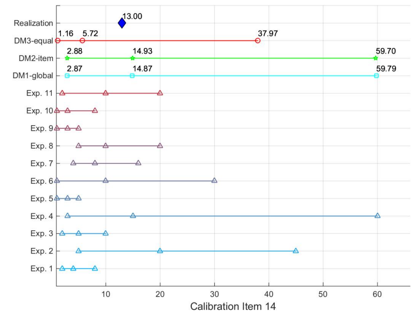
(k) Number of times there was a delay larger or equal to one hour because the required MPs were not available (Prj. 3).



(l) Maximum registered delay because required MPs not available (Prj. 4).

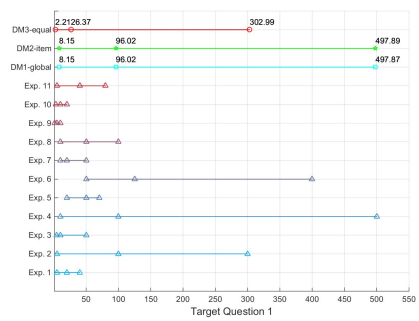


(m) Average of registered delays because required MPs not available (Prj. 4).

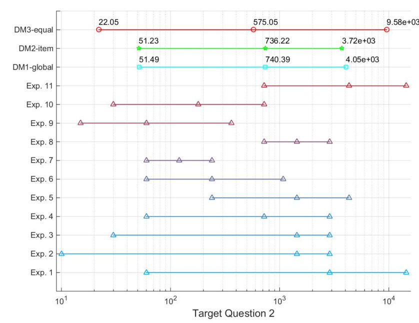


(n) Number of times a delay (>= $1h$) occurred because the required MPs were not available (Prj. 4).

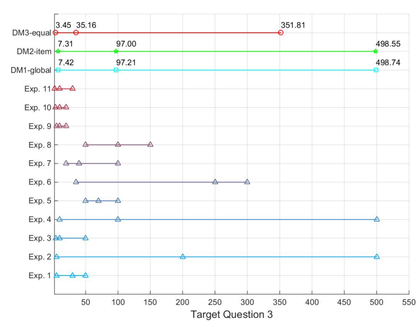
E. Target variables



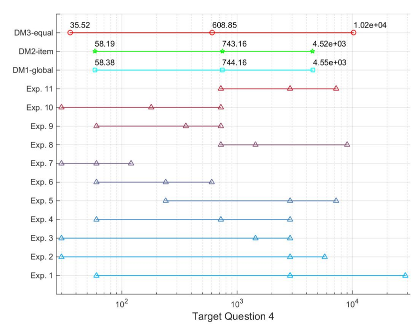
(a) Relative frequency of occurrence (per 1000) of unavailability of required MPs.



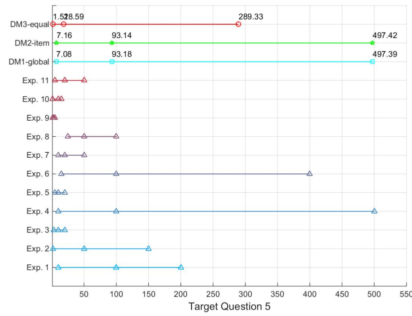
(b) Waiting time (in minutes) because required MPs not available for loading.



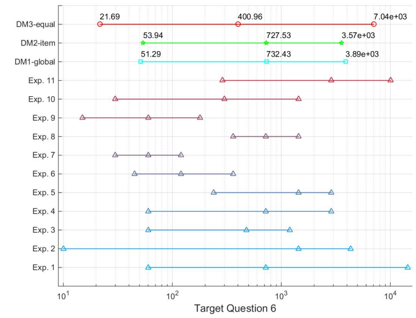
(c) Relative frequency of occurrence (per 1000) of unavailability of required TPs.



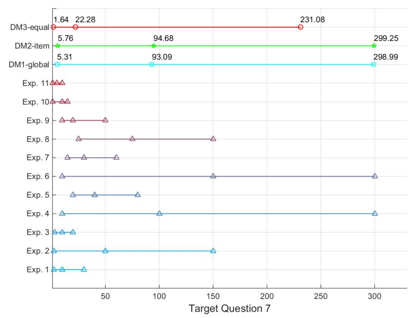
(d) Waiting time (in minutes) because required TPs not available for loading.



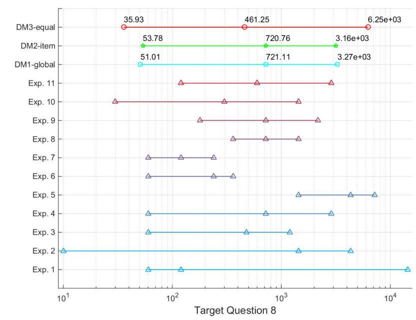
(e) Relative frequency of occurrence (per 1000) of unavailability of required Towers.



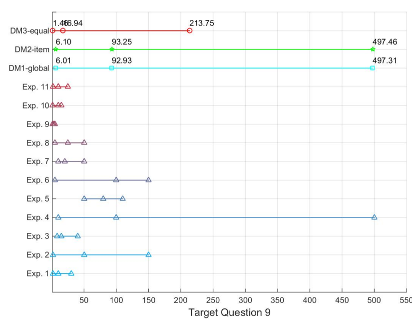
(f) Waiting time (in minutes) because required Towers not available for loading.



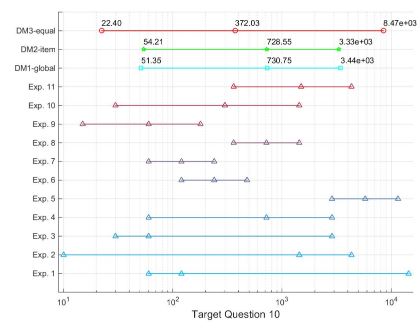
(g) Relative frequency of occurrence (per 1000) of unavailability of required Blades.



(h) Waiting time (in minutes) because required Blades not available for loading.



(i) Relative frequency of occurrence (per 1000) of unavailability of required Nacelles.



(j) Waiting time (in minutes) because required Nacelles not available for loading.

F. Expert judgments analysis with alpha $= 0.01$

Expert ID	Calibration Score	Information Score (All items)	Information Score (Seed items)	Un-normalized Weights	Normalized Weights excl. DM
Expert 1	0,0002060	0,70675	0,86518	0,00	0,00
Expert 2	0,011904	0,471358	0,516478	0,0061485	0.0229
Expert 3	6,804e-10	0,93655	1,090144	0,00	0,00
Expert 4	0,569084	0,452635	0,4606201	0,2621317	0.9771
Expert 5	1,983e-07	1,214771	1,1789866	0,00	0,00
Expert 6	1,314e-05	0,82397	0,712980	0,00	0,00
Expert 7	1,192e-07	1,214716	1,195092	0,00	0,00
Expert 8	0,00036218	1,165429	1,137962	0,00	0,00
Expert 9	2,547e-11	0,8838053	0,802524	0,00	0,00
Expert 10	1,762e-05	0,8577063	0,897994	0,00	0,00
Expert 11	0,00441208	0,8424475	0,9259006	0,00	0,00

Table 7: Measures of performance in judging uncertainty and weights for every participant, obtained from the analysis with ANDURIL using a cut-off level equal to 0.01.

Name	Calibration Score	Information score (total)	Information score (seed items)
DM_{global}	0.96812	0.38179	0.36811
DM_{item}	0.96812	0.39156	0.38038

Table 8: Comparison of the performance measures of two performance-based DMs for *alpha* $= 0.01$.

G. Alternative strategies

Strategy 1: Initial stock at the installation port in the commencement of installation	10 units of each component (MPs, TPs, Towers, Nacelles, Rotors)
Strategy 2: Initial stock at the installation port in the commencement of installation	20 units of each component (MPs, TPs, Towers, Nacelles, Rotors)
Strategy 3: Initial stock at the installation port in the commencement of installation	20 units of foundation components (MPs, TPs) and 10 units of WTGs components (Towers, Nacelles, Rotors)
Strategy 4: Initial stock at the installation port in the commencement of installation	10 units of foundation components (MPs, TPs) and 20 units of WTGs components (Towers, Nacelles, Rotors)
Strategy 5: Initial stock at the installation port in the commencement of installation	20 units of MPs, TPs, Towers and 10 units of Nacelles, Rotors
Strategy 6: Initial stock at the installation port in the commencement of installation	20 units of MPs, TPs, Nacelles, Rotors and 10 units of Towers

Table 9: Details of the different simulated strategies.

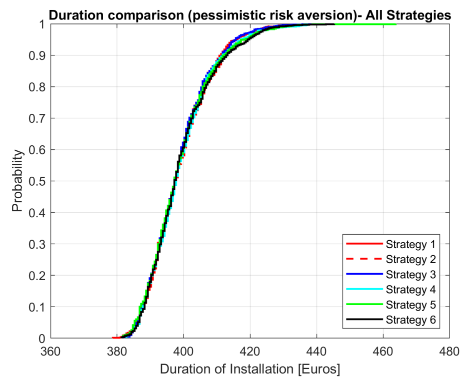


Figure 15: Duration comparison of all different simulated strategies for the pessimistic approach.

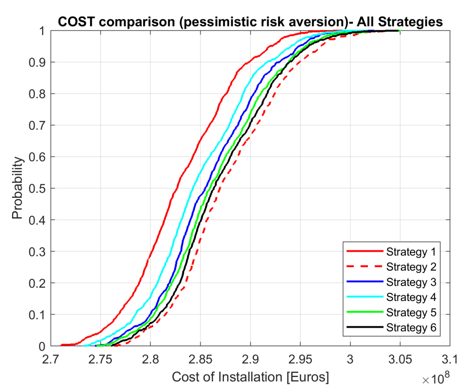


Figure 16: Cost comparison of all different simulated strategies for the pessimistic approach.