A Novel Method on Hydrographic Survey Technology Selection Based on the Decision Tree Supervised Learning

Ivana Golub Medvešek *, Igor Vujović, Joško Šoda and Maja Krčum

Faculty of Maritime Studies, University of Split, Rudera Boškovića 37, 21000 Split, Croatia; ivujovic@pfst.hr (I.V.); jسود@pfst.hr (J.Š.); mkrcum@pfst.hr (M.K.)
* Correspondence: igolub@pfst.hr

Featured Application: If one wants to perform a hydrographic survey, many available types of equipment are presenting different technologies, which could (or not) be applicable in specific environments and conditions. The main question is how to choose the optimal available technology. In this paper, we develop a novel method for choosing the best-suited technology for a specific survey region, and it is applied on Kaštela Bay as an application example. The example shows the step-by-step process of identifying appropriate technology.

Abstract: Hydrographic survey or seabed mapping plays an important role in achieving better maritime safety, especially in coastal waters. Due to advances in survey technologies, it becomes important to choose well-suited technology for a specific area. Moreover, various technologies have various ranges of equipment and manufacturers, as well as characteristics. Therefore, in this paper, a novel method of a hydrographic survey, i.e., identifying the appropriate technology, has been developed. The method is based on a reduced elimination matrix, decision tree supervised learning, and multicriteria decision methods. The available technologies were: remotely operated underwater vehicle (ROV), unmanned aerial vehicle (UAV), light detection and ranging (LIDAR), autonomous underwater vehicle (AUV), satellite-derived bathymetry (SDB), and multibeam echosounder (MBES), and they are applied as a case study of Kaštela Bay. Results show, considering the specifics of the survey area, that UAV is the best-suited technology to be used for a hydrographic survey. However, some other technologies, such as SDB come close and can be considered an alternative for hydrographic surveys.

Keywords: supervised learning; decision tree; hydrographic survey; weighted sum model

1. Introduction

Hydrographic survey technologies are rapidly advancing nowadays. Generally, survey technology encompasses traditional and modern techniques. The traditional technique uses marine technology support as a platform. Vessels, ROV (remotely operated underwater vehicle), UAV (unmanned aerial vehicle), etc., are the most commonly used platforms. These platforms have an acoustic sensor that can be divided into SSS (side scan sonar), MBES (multibeam echo sounder), SBES (single beam echo sounder), etc. On the other hand, the modern technique uses air or satellite platforms and has optical sensors. Light detection and ranging (LIDAR), aerial multispectral and hyperspectral sensors, satellite or aerial panchromatic imaging are often used optical sensors for modern hydrographic survey techniques. The equipment of both techniques represents a whole range of devices whose performance is constantly upgraded and improved.

Examples of such technology applications are published in many references. Sand beach topography surveying by drones and photogrammetry was reported in [1]. Analysis of state-of-the-art hydrographic survey technologies is presented in [2]. UAV was used in [3] for multispectral bathymetry. In addition, 3D multibeam sonar data were used...
to make the depth area comparable with depth areas of electronic charts [4]. Geospatial modeling of the tombolo phenomenon was addressed in [5], and measurements were carried out with an unmanned surface vehicle (USV). Measurements of the tombolo oceanographic phenomenon using UAV and USV were presented in [6]. The paper [7] presented the synergistic use of a small commercial UAV and a USV for beach rock mapping. A methodological proposal to generate a topobathymetric model, using low-cost unmanned platforms in a very shallow/shallow and turbid tidal environment was presented in [8]. Some references prefer ellipsoidal referenced surveying techniques (e.g., [9]) or satellite bathymetry methods [10].

Current technical and technological solutions in the implementation of a hydrographic survey show a high degree of development. It is reasonable to assume that the cost is one of the main factors affecting the implementation of a hydrographic survey. The problem of costs in conducting a hydrographic survey has already been recognized in previous studies. Cost-effectiveness research has focused on new technologies and the time required for the implementation of the hydrographic survey [11–14]. The impact of new technologies and computational techniques such as artificial intelligence is analyzed in [15,16]. Satellite observation has been recognized to increase coverage by reducing the survey in shallow water [17].

In such a variety of technologies, it is important to choose the right one for a specific area based on demands by international institutions, such as the International Hydrographic Organization (IHO). Such choice should include characteristics of the terrain, equipment, and internationally recognized demands. Requirements are not uniform, so the choice needs to be adjusted to the conditions. Interest in automatizing the learning process and non-manual decision tree was presented in [18]. Evolutionary algorithms for decision tree induction were reviewed. Evolutionary decision tree methods can be classified as full tree methods and methods that use tree components. The first uses classification regression (regression or model trees) and the former hyperplanes, pruning, and others. The genetic algorithms in combination with multicriteria decision (MCD) analysis were used in [19]. A benchmarking methodology for the optimal selection model was presented in [20].

According to the available literature, no one has researched a complete approach in making the optimal survey technique solution. This paper aims to select the optimal technique while maintaining the required level of the hydrographic survey quality. The subject of this paper is to propose the model of identifying the appropriate technology. The model is based on the formulation of a binary elimination matrix, which serves as an input variable for the decision tree model creation. Finally, applying the weighted sum model (WSM) for optimization produces the optimal technology for the hydrographic survey.

The proposed novel methodology is subject to complement as needed by hydrographer knowledge. Nonunified parameters, such as hydrographic survey area characteristics and technology performance, can be complemented. Additionally, the criteria that go into selecting optimal survey technologies are variable and depend on the required quality. Therefore, it is essential to emphasize how the proposed methodology is applicable in any survey area. In this paper, the proposed novel methodology is applied as a case study of Kaštela Bay. The results of the study case, the advantages of the proposed methodology, and future research directions are detailed in the Results, Discussion, and Conclusion sections.

2. Methodology

This paper proposes a novel method of identifying suitable and optimal technology for hydrographic surveys. Figure 1 shows the flow chart of the proposed method.
Figure 1. The process of identifying the appropriate technology.

From Figure 1, it can be seen that the process is divided into two parts:

1. **Stage 1**—problem formulation.
2. **Stage 2**—optimization.

In the first stage, the decision tree model is built. This process begins with an examination of the survey area characteristics. Then, the reduced binary elimination matrix is formed, which serves as an input variable for the decision tree model. Additionally, a decision tree has been made based on the matrix. Furthermore, the advantage of the decision tree lies in the fact that it needs to be created only once, assuming that there have been no changes in the elimination matrix, which implies that the performances of the technologies have remained the same. Hence, these performances can change only if new technologies or upgrades of existing technologies become available. The second stage starts with the evaluation of the built model by setting criteria for hydrographic survey quality. Then, it follows quantifying and ranking criteria. Additionally, it has to be noted, determining the criteria values ensures adequate quality in the optimization process.
Finally, the last part of this stage is the result of optimization, i.e., ranking technologies from top to bottom. On the top is the optimal technology and on the bottom is the least optimal technology.

2.1. Stage 1—Problem Formulation Stage

The problem formulation stage begins with the diversity of the areas in which the hydrographic survey is carried out. Therefore, identifying the area’s characteristics is of paramount importance. To facilitate this process, the proposed distribution of characteristics and sub-characteristics of an area is represented in Table 1.

Table 1. Proposed distribution of characteristics and sub-characteristics of the survey area.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sub-characteristics</th>
</tr>
</thead>
</table>
| A—survey area coverage per day | \( a_1 \leq 1 \text{ km}^2/\text{day} \)  
| | \( a_2 \ 2-25 \text{ km}^2/\text{day} \)  
| | \( a_3 \ 26-65 \text{ km}^2/\text{day} \)  
| | \( a_4 > 66 \text{ km}^2/\text{day} \)  |
| B—the minimum depth to be surveyed | \( b_1 \leq 1 \text{ m} \)  
| | \( b_2 \ 2-5 \text{ m} \)  
| | \( b_3 \ 6-20 \text{ m} \)  |
| C—the maximum depth of a survey area | \( c_1 \leq 25 \text{ m} \)  
| | \( c_2 \ 26-45 \text{ m} \)  
| | \( c_3 > 45 \text{ m} \)  |
| D—exists a possibility of hazards to operation | \( d_1 \) exists a possibility of hazards to operation to surface navigation  
| | \( d_2 \) exists a possibility of hazards to operation to underwater navigation  
| | \( d_3 \) exists a possibility of hazards to surface and underwater navigation  |
| E—seabed type | \( e_1 \) rock/coral  
| | \( e_2 \) soft (mud/sand)  
| | \( e_3 \) heavy vegetation |

Table 1 is made on the basis of processed literature [21–43]. It represents the framework that the hydrographer or marine industry employees uses. The literature references are used to obtain the minimum set of survey area data. The derived minimum survey area data set together with survey technologies from the basis for binary reduction matrix elements. Each identified characteristic has been divided by a single division into sub characteristics that are mutually exclusive. From Table 1, it can be observed that the five (5) distribution characteristics are proposed to create the basics framework of the survey area and as such, form the basic elements for the binary reduction matrix. They are sorted into A, B, C, D, and E characteristics. Additionally, as it can be seen, each characteristic contains sub characteristics. Some characteristics have numerical values (A, B, and C), and some characteristics are descriptive (D and E). The proposed distribution of characteristics and sub characteristics can be applied to any area and easily supplemented with new data as needed. It has to be pointed out that not all areas are available for all technologies. Distribution A (survey area coverage per day) refers to the urgency of the hydrographic survey. For example, distribution A is divided into four (4) categories/variables: \( a_1 \), which covers a survey area of less than 1 \text{ km}^2/\text{day}; variable \( a_2 \), the survey area coverage between 2 and 25 \text{ km}^2/\text{day}; variable \( a_3 \), the survey area between 26 and 65 \text{ km}^2/\text{day}, and with variable \( a_4 \) covers the survey area coverage greater than 66 \text{ km}^2/\text{day}. The distribution B (the minimum depth to be surveyed), C (the maximum depth of a survey area), and E (seabed type) refer to the different survey techniques’ performance depending on the depth and seabed type. For example, distribution B is divided into three (3) categories: variable \( b_1 \) concerning those technologies which can give satisfactory results of work at depths less than or equal to 1 m, variable \( b_2 \) represents minimum survey depths between 2 and 5 m, and the last variable \( b_3 \) denotes the minimum depth to be surveyed between 6 and 20 m. Distribution D (exists a possibility of hazards to operation) refers to some technologies’ potential to disable survey operation and should also be considered. It contains three (3)
descriptive sub characteristics/variables. Variable $d_1$ describes the possibility of hazards to operation to surface navigation, variable $d_2$ describes the existence of a possibility of hazards to surface and underwater navigation, and variable $d_3$ describes the possibility of hazards to surface and underwater navigation. Additionally, available technologies were: remotely operated underwater vehicle (ROV), unmanned aerial vehicle (UAV), light detection and ranging (LIDAR), autonomous underwater vehicle (AUV), satellite-derived bathymetry (SDB), and multibeam echosounder (MBES). They have a set of specific characteristics which the manufacturer determines.

Once survey area data are obtained and the available technology is known, a reduced elimination binary matrix can be created. A hydrographer fills the elimination matrix based on survey area data and performances of available technologies. Regarding the knowledge about survey area data, these are data that the hydrographer must be aware of for planning surveying. Additionally, as far as knowledge of technologies’ performances, these data vary depending on the model, and they are available from the manufacturer. Therefore, filling the elimination matrix is a simple process, and it is performed by a hydrographer based on available data. Since all available technologies are not suitable for all survey areas, the elimination matrix is used for rapid and transparent elimination of inappropriate technologies. Hence, the values from the elimination matrix represent the input in the decision tree model.

The basic reduced binary elimination matrix is shown in Figure 2.

![Reduced binary elimination matrix](image-url)

**Figure 2.** Reduced binary elimination matrix with corresponding submatrices.
The rows in Figure 2 represent characteristics and sub characteristics of the survey area (described in Figure 1), and the columns $A_{1k}, A_{2k}, A_{3k}, \ldots, A_{kk}$ represent available technologies. The technologies are represented as alternatives that come into consideration. The reduced binary elimination matrix’s element $x_{kj}$ indicates the binary value (0 or 1) of $k$-th alternatives $A_k$ with respect to the $z$-th characteristic survey area $Ch_z$. Expressions (1) and (2):

$$A_l = \{ A_{1l}, A_{2l}, A_{3l}, \ldots, A_{kl} \}$$

$$Ch = \{ A_1, B_2, C_3, \ldots, Z_z \}$$

A binary filling is proposed to make the process simpler and faster. A value of 0 indicates that a particular alternative is not suitable for the specified area. Conversely, a value of 1 indicates that the observed alternative is suitable for the survey area characteristic. General characteristics of the survey area, to make the selection more precise, are divided into several sub characteristics, Expression (3):

$$Ch = \{ (a_1, a_2, \cdots, a_j), (b_1, b_2, \cdots, b_m), \ldots, (z_1, z_2, \cdots, z_z) \}$$

where $(a_1, a_2, \cdots, a_j)$ are sub characteristics of $A_1$, $(b_1, b_2, \cdots, b_m)$ are sub characteristics of $B_2$, and $(z_1, z_2, \cdots, z_z)$ are sub characteristics of $Z_z$.

In practice, the number of characteristics in the matrix can be smaller than is suggested in Figure 2. If, for example, there is no possibility of hazard (characteristic D), then it can be dropped from the matrix and not considered. Each characteristic under consideration usually has only one sub characteristic. For example, if characteristic C is observed, the maximum depth is generally known, and the characteristic C is represented with only one sub characteristic. In contrast, the characteristic E (seabed type) of the survey area does not have to be unambiguously defined. For example, the bottom part may be rocky, while the other part may be heavy vegetation. In such a case, two sub characteristics of characteristic E are considered. All available alternatives and sub characteristics defined for a particular characteristic of the survey area define submatrices. The total number of submatrices depends on the total number of characteristics of the hydrographic survey area. Each submatrix must have a minimum of one non-zero column. Otherwise, the alternative to which the null column belongs for further analysis will be discarded. Such an outcome means that this alternative is not appropriate for the specified survey area. Additionally, it is important to note that if a technology is no longer available, the corresponding column from the matrix needs to be deleted. This procedure will not be reflected in other values in the elimination matrix.

The flowchart in Figure 3 represents the proposed approach.

The flowchart in Figure 3 is based on fuzzy rules. The process of identifying suitable technology with regards to the characteristics of the hydrographic survey area can be explained by the Expression (4):

If submatrix $A_{s1}$ contains null columns then exclude that alternative and start again else if submatrix $A_{s2}$ contains null columns then exclude that alternative and start again else if submatrix $A_{sN}$ contains null columns then exclude that alternative and start again else indented alternatives for further analysis

$$\text{If submatrix } A_{s1} \text{ contains null columns then exclude that alternative and start again else if submatrix } A_{s2} \text{ contains null columns then exclude that alternative and start again else if submatrix } A_{sN} \text{ contains null columns then exclude that alternative and start again else indented alternatives for further analysis}$$
The structural description does not necessarily need to be represented as rules. Hence, decision trees, which specify sequences of decisions that need to be made along with the resulting recommendation, are another popular means of expression [44].

It is known that decision tree learning is a supervised machine learning technique for inducing a decision from training data [45,46]. It represents one of the most intuitive and frequently used data science techniques [47,48]. Additionally, in decision tree theory, the idea is to split the dataset based on the homogeneity of data [49]. Many measures can be used to determine the best way to split the records. Hence, these measures are defined in terms of the class distribution of the attributes before and after splitting. They are called attribute selection measures (ASM). It has to be pointed out that the most popular ASM in the classification problem is the Gini index and information gain ratio [50].

The Gini index determines the purity of a specific class after splitting along a particular attribute. The best split increases the purity of the sets resulting from the split. If a data set \( L \) contains examples from \( j \) classes, the Gini index, \( gini(L) \), is defined as in Equation (5) [51,52]:

\[
gini(L) = 1 - \sum_{i=1}^{j} p_i^2
\]  

(5)
where \( p_i \) is the relative frequency of class \( i \) in \( L \).

If the dataset is split on an attribute \( A \) into two subsets \( L_1 \) and \( L_2 \), with sizes \( N_1 \) and \( N_2 \), respectively, the Gini index can be calculated as in Equation (6) [51]:

\[
gini_A(L) = \frac{N_1}{N} gini(L_1) + \frac{N_2}{N} gini(L_2) \tag{6}
\]

Reduction in impurity is calculated as in expression Equation (7) [51]:

\[
\Delta gini(A) = gini(L) - gini_A(L) \tag{7}
\]

The information gain ratio is a ratio of information gain to entropy. Information gain can be expressed as in Equation (8):

\[
IG(L, f) = entropy(L) - \sum_{i=1}^{V} \frac{|L^V|}{|L|} \left( entropy(L^V) \right) \tag{8}
\]

where \( f \) is a feature, \( V \) is different values for a feature, \(|L^V|\) is the subset of \( L \) with \( f = v \).

Therefore, entropy is defined as the probability sum of each label times the log probability of that same label, as is shown by Equation (9) [51,52]:

\[
entropy(L) = -\sum_{i=1}^{i} p_i \log_2(p_i) \tag{9}
\]

### 2.2. Stage 2—Optimization Stage

After the technology suitable for the required hydrographic survey area has been selected and visually substantiated by the decision tree, multi-criterion optimization is applied. Multicriteria decision making (MCDM) is a procedure that combines the performance of decision alternatives across several contradicting, qualitative, and/or quantitative criteria and yields in a compromise solution [53,54]. Relevant methods are frequently applicable, implacable, or explicable, in numerous real-life problems. They are widely used where sets of decision alternatives are evaluated against conflicting criteria [54,55]. The most commonly used multicriteria decision method approach is the weighted sum method (WSM) [56–58]. Figure 4 shows MCDM and WSM procedures.

![Figure 4. Working process of multicriteria decision and weighted sum methods.](image-url)
From Figure 4, it can be observed there are three major steps. The first step starts with the construction of a pair-wise comparison matrix $C = (c_{ij})$, type $(nxn)$, between sets of criteria as is shown by Equation (10) [59]:

$$
C = \begin{bmatrix}
1 & c_{12} & \cdots & c_{1n} \\
\frac{1}{c_{12}} & 1 & \cdots & c_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{c_{1n}} & \frac{1}{c_{2n}} & \cdots & 1 \\
\end{bmatrix}
$$

(10)

where is $c_{ij} = \frac{1}{c_{ji}}$, $i \neq j$ criteria of positive reciprocal matrix $C$.

Further, the pair-wise comparison matrix between criteria is obtained using Saaty’s scale from Table 2, [60–62].

Table 2. Saaty’s scale of relative importance.

<table>
<thead>
<tr>
<th>Intensity of Importance</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Same importance</td>
</tr>
<tr>
<td>2</td>
<td>Intermediate values</td>
</tr>
<tr>
<td>3</td>
<td>Relatively more important</td>
</tr>
<tr>
<td>4</td>
<td>Intermediate values</td>
</tr>
<tr>
<td>5</td>
<td>Some more important</td>
</tr>
<tr>
<td>6</td>
<td>Intermediate values</td>
</tr>
<tr>
<td>7</td>
<td>Very more important</td>
</tr>
<tr>
<td>8</td>
<td>Intermediate values</td>
</tr>
<tr>
<td>9</td>
<td>Enormously important</td>
</tr>
</tbody>
</table>

For quantifying qualitative data, Saaty suggested a scale of relative importance, as shown in Table 2. The scale has five degrees of intensity (1, 3, 5, 7, and 9) and four intermediate levels (2, 4, 6, and 8). Thus, a value judgment corresponds to how many times one criterion is more important than another to each of them. Hence, the values used for any given pair of criteria vary from one (1) where they have equal importance to nine (9) where one is enormously more important than the other.

Once a comparison matrix has been made, it is necessary to sum the values in each column of the pair-wise matrix, compute each element of the matrix by its column total, and calculate the weight vector by finding the row average as in Equation (11) [60,63]:

$$
W = \begin{bmatrix}
w_1 \\
w_2 \\
\vdots \\
w_n
\end{bmatrix} = \begin{bmatrix}
\frac{\sum_{j=1}^{n} c_{1j} \cdot w_j}{w_1} \\
\frac{\sum_{j=1}^{n} c_{2j} \cdot w_j}{w_2} \\
\vdots \\
\frac{\sum_{j=1}^{n} c_{nj} \cdot w_j}{w_n}
\end{bmatrix}
$$

(11)

The weighted sum matrix is found by multiplying the pair-wise comparison matrix and the weight vector. Then, dividing all the elements of the weighted sum matrix by their respective priority vector element, a consistency vector ($CV$) is obtained as by Equation (12), [60–63]:

$$
CV = \begin{bmatrix}
\frac{\sum_{j=1}^{n} c_{1j} \cdot w_j}{w_1} \\
\frac{\sum_{j=1}^{n} c_{2j} \cdot w_j}{w_2} \\
\vdots \\
\frac{\sum_{j=1}^{n} c_{nj} \cdot w_j}{w_n}
\end{bmatrix}
$$

(12)
Step 1 ends with performing the consistency index (CI) and consistency ration (CR). CI measures the degree of inconsistency, as is shown by Equation (13) [60–63]:

\[
CI = \frac{\lambda_{\text{max}} - n}{n - 1}
\]  

where \( n \) is matrix size and \( \lambda_{\text{max}} \) is the largest principal eigenvalue of the positive reciprocal pair-wise comparison matrix.

If the pair-wise comparisons are perfectly consistent, then \( \lambda_{\text{max}} \) is equal to the size of the matrix and the \( CI = 0 \). The larger the inconsistency between comparisons is a consequence of the larger \( \lambda_{\text{max}} \) and the larger \( CI \). Then, the consistency ration \( CR \), a metric is calculated as in Equation (14) [58,64]:

\[
CR = \frac{CI}{RI}
\]

where \( RI \) is the random consistency index obtained from a randomly generated pair-wise comparison matrix.

If \( CR \leq 0.1 \), then the comparisons are acceptable. However, if \( CR > 0.1 \), then the values of the ratio are indicative of inconsistent judgments. In such a case, the judgments should be reconsidered and revised [58,60–63,65]. The last part of the step is to compute consistency vector, index, and ratio.

In Step 2, a decision matrix \((mxn)\) of alternatives on each criterion is created as in Equation (15):

\[
R = \begin{bmatrix}
  r_{11} & r_{12} & \cdots & r_{1m} \\
  r_{21} & r_{22} & \cdots & r_{2m} \\
  \vdots & \vdots & \ddots & \vdots \\
  r_{m1} & r_{m2} & \cdots & r_{mm}
\end{bmatrix}
\]  

The next part of the second step includes calculating the normalized decision matrix with positive attributes as in Equation (16) [58,64]:

\[
r_{ij} = \frac{x_{ij}}{x_{i}^{\text{max}}}
\]

where \( r_{ij} \) is the score of the \( i \)-th alternative concerning the \( j \)-th criterion and \( x_{i}^{\text{max}} \) a maximum number of the \( r \) in the column of \( j \).

From Equations (15) and (16), the weighted normalized decision matrix \( R' \) is created and shown by Equation (17) [58]:

\[
R' = \begin{bmatrix}
  \omega_1 \cdot r_{11} & \omega_2 \cdot r_{12} & \cdots & \omega_n \cdot r_{1m} \\
  \omega_1 \cdot r_{21} & \omega_2 \cdot r_{22} & \cdots & \omega_n \cdot r_{2m} \\
  \vdots & \vdots & \ddots & \vdots \\
  \omega_1 \cdot r_{m1} & \omega_2 \cdot r_{m2} & \cdots & \omega_n \cdot r_{mm}
\end{bmatrix}; \sum_{j=1}^{n} \omega_j = 1
\]

Finally, step 3 in WSM, including the optimum solution of each alternative, is obtained by the following Equation (18) [58,63]:

\[
S_i^{\text{WSM}} = \sum_{j=1}^{n} r_{ij} \omega_j; \ i = 1, 2, 3, \ldots, m
\]

where \( S_i^{\text{WSM}} \) represents the weighted sum score and \( \omega_j \) is the weight of the \( j \)-th criterion.

3. Case Study Results

In order to show the applicability and how to implement the novel method, a case study is performed. The study site is Kaštela Bay. It is a semi-enclosed coast and the largest
bay situated in the Middle Adriatic. The bay is located at 43°32′00″ N and 16°21′00″ E [66]. It is 14.8 km long, approximately 6 km wide, and the total area is 61 km² [67,68]. The bay is relatively shallow, with an average depth of 23 m. Maximum depth is 45 m at the bay’s inlet. Heavy traffic across this area is under the Port Authority Split jurisdiction, consisting of passenger ports, trade ports, and a large number of ports and moorings for small boats [66]. Marina Kaštela, in Kaštel Gomilica with 420 moorings, is the large port of nautical tourism. It significantly contributes to the development of nautical tourism and density of maritime transport [69]. The seafloor consists of many different substrate types, including sand, rock, mud, and gravel [70].

In the proposed study case, the choice of alternatives is based on the technologies available by the Hydrographic Institute of the Republic of Croatia. Hence, those technologies would possibly be all available in the future if their purchase or lease proves profitable, as is shown in Table 3.

Table 3. Available technologies.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>ROV (remotely operated underwater vehicle)</td>
</tr>
<tr>
<td>A2</td>
<td>UAV (unmanned aerial vehicle)</td>
</tr>
<tr>
<td>A3</td>
<td>LIDAR (light detection and ranging)</td>
</tr>
<tr>
<td>A4</td>
<td>AUV (autonomous underwater vehicle)</td>
</tr>
<tr>
<td>A5</td>
<td>SDB (satellite derived bathymetry)</td>
</tr>
<tr>
<td>A6</td>
<td>MBES (multibeam echosounder)</td>
</tr>
</tbody>
</table>

Available technologies from Table 2 are considered with respect to the proposed distribution of characteristics and sub characteristics of the survey area from Table 1. Tables 1 and 3 show that the reduced binary elimination matrix is formed and expressed as shown in Table 4.

Table 4. Reduced binary elimination matrix for the study case.

<table>
<thead>
<tr>
<th></th>
<th>ROV</th>
<th>AUV</th>
<th>LIDAR</th>
<th>SDB</th>
<th>UAV</th>
<th>MBES</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>a₂</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>a₃</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a₄</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>b₁</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>b₂</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>b₃</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>c₁</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>c₂</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>c₃</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>d₁</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>d₂</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>d₃</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>e₁</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>e₂</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>e₃</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4 is filled with elements based on the previous section methodology. The mentioned literature was used to analyze the performance of the alternatives. The performances of hydrographic survey technologies are not uniform, but vary depending on the model and manufacturer. Therefore, in practice, these are data that are available from the manufacturer. Table 4 covers the relationship between available technologies’ performances for the case study related to all possible survey area data. From Table 4, for example, in the case of technology UAV, it is evident survey coverage per day (characteristic A) is to a
maximum of 25 m (sub characteristics $a_1$ and $a_2$). In terms of characteristic B, it can also be used in extremely shallow water (sub characteristics $b_1$, $b_2$, and $b_3$). The maximum depths (characteristic C) at which the UAV shows good performance are up to 45 m. Since its functionality does not include surface or underwater navigation, it can also be used when there is a possibility of hazards to sea navigation. Hence, a value of 1 belongs to all sub characteristic ($d_1$, $d_2$, $d_3$) of characteristic D. Analogously, UAV can work even when the seabed is rocky or coral, soft or covered with vegetation (sub characteristic $e_1$, $e_2$, and $e_3$).

It has to be noted that the reduced binary elimination matrix represents the input variable for decision tree creation. Hence, using the matrix from Table 4 represents a set of training data in the decision tree working process. Based on these data, a general decision tree model was created with the software package RapidMiner version 9.8. It is one of the most used open-source predictive analytics platforms utilized for data analysis. It is accessible as a stand-alone application for information investigation and a data mining engine to integrate into its products. RapidMiner provides an integrated environment for data mining and machine learning procedures [71].

It has to be pointed out that the reduced binary elimination matrix represents the input variable for decision tree creation. Hence, using the matrix from Table 4 represents a set of training data in the decision tree working process. Based on these data, a general decision tree model was created with the software package RapidMiner version 9.8. It is one of the most used open-source predictive analytics platforms utilized for data analysis. It is accessible as a stand-alone application for information investigation and a data mining engine to integrate into its products. RapidMiner provides an integrated environment for data mining and machine learning procedures [71].

It has to be pointed out that the decision tree represents all possible combinations with used technologies. Figure 5 shows a decision tree model using the information gain ratio algorithm for a considered study case based on Table 4.

From Figure 5, it has to be noted that sometimes individual characteristics and sub-characteristics denote only one technology, and sometimes they refer to several of them. Additionally, from Figure 5, it can be seen that multicolors strips denote several alternatives that are possible at the same instance. Further, the name of the terminal node technology depends on the data used for tree training and the probability of that technology as an output value. Hence, multicolors strips of nodes explain the possibility of choosing several technologies. For example, according to Table 1 and Figure 5, if the maximum depth of survey area (characteristic C) from less than or equal to 25 m (sub characteristic $c_1$) and the minimum depth to be surveyed (characteristic B) less than or equal to 1 m (sub characteristic $b_1$), the optimal technologies are LIDAR and UAV.

Characteristics of the considered area—Kaštela Bay are given in Table 5, including survey area (parameter A), depth (parameters B and C), and seabed type (parameter E).
Table 5. Characteristic and sub characteristic of the hydrographic survey area.

<table>
<thead>
<tr>
<th>Characteristic and Sub Characteristic of the Kaštela Bay</th>
</tr>
</thead>
<tbody>
<tr>
<td>A—survey area coverage per day</td>
</tr>
<tr>
<td>B—the minimum depth to be survey</td>
</tr>
<tr>
<td>C—the maximum depth of survey area</td>
</tr>
<tr>
<td>E—seabed type of area</td>
</tr>
<tr>
<td>$a_1$—≤ 1 km$^2$/day</td>
</tr>
<tr>
<td>$b_3$—6–20 m</td>
</tr>
<tr>
<td>$c_3$—45 m</td>
</tr>
<tr>
<td>$e_1 + e_2$—mixed type</td>
</tr>
</tbody>
</table>

It has to be pointed out, the parameter D is excluded from the consideration since it is known that in the Kaštela Bay area, there is no possibility of hazards to operation. It is worthwhile to mention that this study’s area of interest was obtained from the general tree in Figure 5 and is shown in Figure 6a,b.

Figure 6. A tree decision model for Kaštela Bay study case: (a) using information gain ratio algorithm; (b) using Gini algorithm.

Figure 6a,b confirms the state from the reduced binary elimination matrix that all observed technologies are suitable for hydrographic surveys in Kaštela Bay. Therefore, they are taken as available alternatives for choosing the optimal solution.
Furthermore, to perform optimization and determine the best-suited technology for the Kaštela Bay case study, three criteria are performed in multicriteria decision-making: cost ($C_1$), urgency ($C_2$), and accuracy ($C_3$). Hence, the WSM method’s procedure was conducted. Table 6 shows the relationship between the given criteria, weights, and consistency vector.

Table 6. Normalized pair-wise comparison matrix, weights, and consistency vector.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$\Sigma$</th>
<th>Weights Vector</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>2.0795</td>
<td>0.6932</td>
<td>2.9998</td>
</tr>
<tr>
<td>$C_2$</td>
<td>0.33</td>
<td>1</td>
<td>3</td>
<td>0.6907</td>
<td>0.2302</td>
<td>3.0004</td>
</tr>
<tr>
<td>$C_3$</td>
<td>0.11</td>
<td>0.33</td>
<td>1</td>
<td>0.2294</td>
<td>0.0764</td>
<td>3.0026</td>
</tr>
</tbody>
</table>

Table 6 is derived using Equations (10)–(12) from the methodology section. Finally, the relation between practical alternatives and given criteria is graded, and the correlation matrix is obtained as shown in Table 7.

Table 7. The normalized decision matrix.

<table>
<thead>
<tr>
<th>Technology</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROV</td>
<td>0.8</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>UAV</td>
<td>1</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>LIDAR</td>
<td>0.6</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>AUV</td>
<td>0.8</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>SDB</td>
<td>1</td>
<td>0.6</td>
<td>1</td>
</tr>
<tr>
<td>MBES</td>
<td>0.4</td>
<td>1</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 7 is derived using Equations (15)–(17), and it shows correlations between technologies and three criteria. For example, the correlation between the ROV with the cost ($C_1$) criteria is 0.8, which indicates a high correlation. Also, the ROV has an ideal correlation with urgency ($C_2$) criteria. Finally, it can be observed, the ROV technology has a weak correlation (0.4) with the accuracy ($C_3$) criteria. It has to be pointed out that a high correlation between technology and criteria indicated suitability between the criterion and technology.

Next, the final step in WSM includes the optimum solution of each alternative is obtained using the Equation (18). Therefore, ranked technologies for the survey area are represented in Table 8.

Table 8. Ranked technologies for the survey area.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>UAV</td>
<td>0.92</td>
</tr>
<tr>
<td>SDB</td>
<td>0.90</td>
</tr>
<tr>
<td>AUV</td>
<td>0.83</td>
</tr>
<tr>
<td>ROV</td>
<td>0.81</td>
</tr>
<tr>
<td>MBES</td>
<td>0.70</td>
</tr>
<tr>
<td>LIDAR</td>
<td>0.53</td>
</tr>
</tbody>
</table>

From Table 8, it can be observed that according to the WSM method, the optimal technology following the given criteria for the proposed Kaštela Bay study case is UAV technology. Further, the second-best alternative with a score of 0.90 is SDB technology. It can be observed that there is a small difference between the two best options candidates for the Kaštela Bay hydrographic survey. These numerical results suggest that both alternatives can be considered because they meet the given criteria almost equally. On the other hand, the MBES technology has the lowest score (0.53). Therefore, it can be concluded that the technology is least desirable to use for the proposed hydrographic survey area.
4. Discussion and Conclusions

The main contribution of this paper is the development of the hydrographic survey optimal technologies’ selection model. First, it has to be pointed out that hydrographic surveys are performed in different survey areas. Further, each hydrographic survey area has an abundance of features to be considered. Additionally, all technologies are not suitable for every area. Hence, their performances are not universal and depend on the manufacturer, model, components, etc. Otherwise, the overall budget, urgency, and accuracy play an essential role. For example, some areas require an urgent hydrographic survey, then most often, the increase in implementation speed decreases the accuracy. Therefore, in addition to finding suitable technology for the survey area, the selected technology must satisfy the required hydrographic survey criteria.

Since the paper proposes a novel methodology in selecting the optimal survey technology, it has to be emphasized that the method can be applied to all survey areas and all available technologies. The first step involves the obtained hydrographic survey area data. These are the primary data that the surveyor needs before surveying (such as min/max depth, seabed type, etc.). A reduced elimination binary matrix contains the distribution of characteristics and sub-characteristics of a hydrographic survey area to facilitate data management. It has to be pointed out that this matrix is the novelty of the proposed methodology and represents the input in the model. However, it is generally applicable and can be supplemented with new data as needed and assessed by a hydrographer. It has to be highlighted that the purpose of the matrix is to select suitable technologies that may be taken for further consideration with respect to the characteristics of the survey area. The performances of the technologies are available from the manufacturer. Given all the access data, the hydrographer has a simple task of binary filling the matrix. The most significant purpose is to make a decision tree based on binary data from the matrix.

The decision tree’s primary function is to create a model that can predict all possible alternatives depicted in Figure 5 as multicolor strips. Then, suitable technologies are further eliminated and graduated given their performance and required hydrographic survey quality.

From the above discussion about the elimination matrix and a decision tree model, it is evident that the created decision tree model is done once. Additionally, if a new technology becomes available and has different performances, it requires a new tree to be made.

Applicability and validation of the proposed method is shown in the proposed study case. Kaštela Bay represented the hydrographic area of interest. Binary values in the elimination matrix indicate the suitability of using all observed technologies concerning hydrographic survey characteristics of Kaštela Bay. Based on the results of a matrix, ROV, UAV, LIDAR, AUV, SDB, and MBES were the appropriate technologies, as it is shown in Table 4.

The binary data from the matrix represents an input for a decision tree model. The decision tree model is made using the following metrics the information gain ratio and Gini algorithms (Figure 6a,b). It has to be pointed out, using both metrics, the obtained decision tree models differ from each other with respect to the root node, i.e., tree depth, but they produce identical results. Therefore, the significant dissimilarity between the decision tree leave depths seeks to provide equal choice. Additionally, using a Gini algorithm for the training tree leads to a less complex and transparent decision tree (Figure 6b). Hence, all observed technologies are taken as available alternatives for choosing the optimal technological solution for Kaštela Bay.

Following the above discussion, multicriteria decision-making is performed by three criteria: cost, urgency, and accuracy. First, the cost has been selected as the preferred criteria for the study, then the criterion urgency follows, and the least important measure is accuracy for the Kaštela Bay study case (Table 6). Further, applying WSM, the numerical results, which have ranked technologies for the hydrographic survey, are obtained (Table 8). The optimal technology for the Kaštela Bay study case was the UAV, with a score of 0.92. Then, with a score of 0.90 follows SDB technology as the best alternative to UAV. Additionally, it
has to be noticed that the MBES technology had the lowest score of 0.53 and represents the least favorable alternative to be used for a hydrographic survey of Kaštela Bay. For other considered technologies, no evident difference has been observed. Furthermore, if the costs were the only decision criterion, after applying our algorithm, only the cheapest alternative would be selected as the best technology, which may not be as efficient and effective as other technologies. Hence, such an approach would represent a constraint in practice. Therefore, it is not recommended to select optimal technologies to be guided solely by cost as the only decision criterion. The time factor was also considered an important criterion in a hydrographic survey and presents the second criterion in importance and had to be included in a multi-decision optimization. Based on Table 7 from study case results, it is evident if decision-making is based on two equally important criteria, cost and urgency, then ROV, UAV, and AUV would be selected as optimal technologies. However, by such a decision, the accuracy is completely neglected. For example, suppose the time is not an issue, and the costs and accuracy are equally important in decision making. In that case, SDB will be the optimal choice for represented study case. Likewise to the above, when funds are sufficient, and urgency and accuracy are the only and equally important criteria, the optimal technology choice for Kaštela Bay would be LIDAR.

In the end, it has to be highlighted the following: parameters that represent the input to the model are the reduced binary elimination matrix and the optimization criteria. In addition, the binary matrix is obtained on the basis of available data of the survey area (Table 1) and available technologies (e.g., Table 3). Further, optimization criteria are determined based on the financial budget, the time available, and the desired level of accuracy. Finally, all mentioned elements in the matrix and optimization criterion are filled by a hydrographer using the proposed methodology and available data.

Future research direction includes other multi-decision methods that could be investigated on this subject. Additionally, it is necessary to compare them to conclude on the most appropriate.

Further, the limitations of the proposed method should be mentioned. For example, the financial budget of some countries does not allow consideration of more than two/three technologies. In addition, it has to be noted that, currently, that from the available literature, there are no similar developed algorithms at present.

Author Contributions: Conceptualization, I.G.M., I.V. and J.Š.; methodology, I.G.M.; validation, I.G.M., J.Š. and I.V.; formal analysis, I.G.M. and M.K.; resources, I.G.M.; data curation, I.G.M., I.V. and J.Š.; writing—original draft preparation, I.V., J.Š., I.G.M. and M.K.; visualization, I.G.M.; supervision, J.Š., I.V. and I.G.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References
3. Rossi, L.; Mammi, I.; Pelliccia, F. UAV-derived multispectral bathymetry. Remote Sens. 2020, 12, 3897. [CrossRef]
5. Specht, M.; Specht, C.; Mindykowsk, J.; Dąbrowski, P.; Maśnicki, R.; Makar, A. Geospatial modeling of the tombolo phenomenon in Sopot using integrated geodetic and hydrographic measurement methods. Remote Sens. 2020, 12, 737. [CrossRef]
47. Kotu, V.; Deshpande, B. Classification. In Data Science, 2nd ed.; Elsevier Science: Cambridge, MA, USA, 2019; pp. 65–163. [CrossRef]
57. Briot, J.P.; Almeida, A.; Alvarez, I.; Guessoum, Z.; Marin, O.; Perrot, J.-F. Multi-Criteria Decision Aggregation and Ranking Approaches for Combining Multiple Information Strategies to Decide Which Agents Should Be Replicated. Deliverable, FT-CAT Project, EuroControl INO CARE III Programme, Research Report, France, 9 November 2007. Available online: https://d1wqtxts1xzle7.cloudfront.net/30607528/FTATC_8-with-cover-page.pdf?Expires=1622172669&Signature=aULkBzXHnog--IjD0ppl6R3xYtg45s-wMzhlIdVkvPvozJboKp0Nxg-hvFpg5eqQlAMF2y5Kw4oo0TJyjDzh-ssiOOVLP-MfCrQHassj0kpwAjt2TVtldZ0RyDMQPHZNjxh--[yp5pKiyqyqCZ7g7PO5hWjbNw;MeFryGb2-]--[0njb9Dp-auvwNH8zZo0LpPYUUVXgz2Hvuan67HVQ3wr--[QZIOm8bIloOwFH5vHvs4--[d8ryj3lUoTh7nbfSmwrRlkVkcUjVQ8gkKSxTvuv5E8blGwbasWbygZOAuemUkQZRUh4jqqufbd0Dw2Mh0f5Gpnu2C148uavbr7dWpsVqbrWP3yQ_&Key-Pair-Id=APKAJLOHFSGSLRBV4ZA (accessed on 27 May 2021).


59. Wu, Z.; Abdul-Nour, G. Comparison of multi-criteria group decision-making methods for urban sewer network plan selection. CivilEng 2020, 1, 26–48. [CrossRef]


