

Learning from synthetic data: Enhancing refraction correction accuracy for airborne image-based bathymetric mapping of shallow coastal waters

Panagiotis Agrafiotis · Konstantinos Karantzas · Andreas Georgopoulos · Dimitrios Skarlatos

Received: date / Accepted: date

Abstract The increasing need for accurate bathymetric mapping is essential for a plethora of offshore activities. Even though aerial image datasets through Structure from Motion (SfM) and Multi-View Stereo (MVS) techniques can provide a low-cost alternative compared to LiDAR and SONAR, offering additionally, important visual information, water refraction poses significant obstacles in delivering accurate bathymetry. In this article, the generation of manned and unmanned airborne synthetic datasets of dry and water covered areas is presented. These data are used to train models for correcting the geometric effects of refraction on real world image-based point clouds and aerial images. Based on a thorough evaluation, important improvements are presented, indicating the increased accuracy and the reduced noise in the point clouds of the derived bathymetric products, meeting also the International Hydrographic Organization's (IHO) standards.

Keywords Bathymetry · Refraction correction · UAV · Airborne · Synthetic data · Support Vector Regression · Machine Learning · Shallow waters · Coastal Mapping

P. Agrafiotis, K. Karantzas, A. Georgopoulos
School of Rural and Surveying Engineering, National Technical University of Athens, 15780, Athens, Greece
Tel.: 00302107722675
E-mail: pagraf@central.ntua.gr, karank@central.ntua.gr, drag@central.ntua.gr

D. Skarlatos
Civil Engineering and Geomatics Dept., Cyprus University of Technology, 3036 Limassol, Cyprus
E-mail: dimitrios.skarlatos@cut.ac.cy

1 Introduction

Accurate and reliable bathymetric mapping is a key element for offshore activities, coastal engineering applications, study of sedimentary processes, hydrographic surveying, as well as archaeological mapping and biological research. With the advent of modern remote sensing techniques (e.g., seismic reflection profiling, aerial imagery, satellites, LiDAR and SONAR), researchers have now gained the ability to effectively interpret and map large portions of the dynamically changing shallow and deep coastal environments along continental shelf margins.

When it comes to shallow waters, SfM and MVS techniques can provide a low-cost alternative compared to other methods such as aircraft-borne and unmanned aerial vehicle (UAV)-borne LiDAR (Mandlbürger et al., 2020) and shipborne multibeam echosounders, offering as well, important visual information. However, refraction seems to be the main factor adversely affecting the geometry and the radiometry of the imagery and consequently of the products of through-water image-based 3D reconstruction methods by delivering erroneous (apparent) depths.

The work presented in this article describes the generation of synthetic Digital Terrain Models (DTMs) and aerial imagery depicting dry and water covered areas. The produced synthetic data enabled us to investigate additional error sources not related to refraction, in a controlled environment. Most importantly, the synthetic data are used for training unbiased linear Support Vector Regression (SVR) models, facilitating the validation in real world cases and delivering valuable results. To train the SVR models and correct the apparent depths (Z_0), the method presented in Agrafiotis

et al. (2019a,b) is followed, which in this article, will be referred to as Method 1.

The rest of the article is organized as follows: Subsection 1.1 reports the related work while Subsection 1.2 the contribution of the article. Section 2 describes the process for generating the synthetic DTMs and images. Section 3 presents the investigation of additional errors performed, not related to refraction, while Section 4 the learning process. In Section 5, the training, validation, and testing of the SVR models of Method 1 is performed using the synthetic data and the results are evaluated. Following, Section 6 describes the implementation of the trained SVR models directly and through the method presented in Agrafiotis et al. (2020), which hereafter will be referred to as Method 2. Both methods are applied on four different real-world test areas, indicating the increased bathymetric accuracy achieved when the models trained on the synthetic data are used. Section 7 is the discussion and Section 8 concludes the article.

1.1 Related work

1.1.1 Refraction Correction

Refraction effect has driven scholars to suggest several correction models for two-media photogrammetry, most of which are dedicated to specific applications. Two-media photogrammetry is divided into through-water and in-water photogrammetry. The through-water term is used when the camera is above the water surface and the object is underwater, hence part of the ray is traveling through air and part of it through water. It is most commonly used in aerial photogrammetry (Mulsow, 2010; Dietrich, 2017; Skarlatos and Agrafiotis, 2018; Mandlburger, 2018, 2019; Agrafiotis et al., 2019a,b, 2020; Mulsow et al., 2020) and satellite photogrammetry (Cao et al., 2019, 2020) for small and large scale seabed mapping surveys or in close range applications (Butler et al., 2002; Georgopoulos and Agrafiotis, 2012). The in-water term is used when both camera and object are in the water (Menna et al., 2018; Agrafiotis et al., 2018), however this is not the case here.

In the literature, three main approaches can be found for correcting refraction in through-water photogrammetry: analytical, image-based, and machine learning-based. The first is based on the modification of the collinearity equation (Fryer, 1983; Wang, 1990; Shan, 1994; Butler et al., 2002; Maas, 2015; Wimmer, 2016; Dietrich, 2017), the second suggests the re-projection of the original photo to correct the water refraction (Georgopoulos and Agrafiotis, 2012; Skarlatos and Agrafiotis, 2018; Agrafiotis et al., 2020) while the third and most

recent one depends on machine learning models that learn the underestimation of depths and predict the correct depth knowing only the apparent one (Agrafiotis et al., 2019a,b, 2020). Other methods that do not fall into the above categories, are multiplying the apparent depth with a constant number, which in most of the cases is the refraction index of the water (Woodget et al., 2015; Chirayath and Li, 2019). As shown in Agrafiotis (2020), the use of this form of correction might be acceptable in the very shallow waters, however, remarkable errors are expected after 2-3 m depth.

A very recent and in-depth presentation of the current state of the art methods in through-water photogrammetry and Satellite Derived Bathymetry (SDB) can be found in Agrafiotis (2020).

1.1.2 Synthetic Images

According to Luhmann (2016), artificial images (or synthetic images called in this article) can be generated if interior orientation and exterior orientation parameters, a Digital Surface Model (DSM) and a texture image are given. By resampling through object space, the image is filled with color values. These images can then be used to create reference or error-free images of which all parameters are known.

Nowadays, synthetic images are widely used in computer vision and machine learning community for facilitating the error-free training, validation, and testing of complex models requiring many data that are difficult and sometimes expensive to find (Peng et al., 2015; Richardson et al., 2016; Ros et al., 2016; Shrivastava et al., 2017; Sankaranarayanan et al., 2018; Saleh et al., 2018; Barbosa et al., 2018; Hinterstoisser et al., 2018). The use of synthetic images in the photogrammetric applications is quite adopted too, especially for some specific applications and error investigation (Skarlatos and Georgopoulos, 2006; Li et al., 2011; Luhmann, 2016; Sun et al., 2016; Kahmen et al., 2019, 2020).

1.2 Contribution

In this work, the generation of synthetic DTMs and aerial images picturing clear and calm waters is performed in order to: (a) discover additional error sources in the process, not related to refraction and (b) use the synthetic data for training unbiased SVR models following Method 1. The latter one will facilitate the validation of those models in real world cases, demonstrating a noticeable increase in accuracy achieved by the already published methods in Agrafiotis et al. (2019a,b, 2020) where the SVR models are trained with real world data only.

To this end, a periodic mathematical function is used to form two different DTMs, which are coloured by a suitable texture pattern, generating eight synthetic aerial datasets; for the four of them, refraction is added on the imagery while for the remaining four, images without refraction are created, allowing for the investigation of additional error sources in the process. Then, by using these data, SfM-MVS processes are performed and the point clouds created using the refracted imagery are corrected by the refraction effects using Method 1. Consequently, the corrected point clouds are used for correcting the refracted imagery following Method 2. Results are compared with the ones delivered using the SVR models trained on real-world datasets as well as three other state of the art methods.

The advantage of the synthetic data is the accuracy and reliability of the depth information and the exact knowledge of exterior and interior orientations of the cameras used. Moreover, when it comes to seabed imaging, errors and limitations in image matching caused by the visibility restrictions due to the depth, especially in depths more than 10-15 m (Agrafiotis, 2020), and errors introduced by the wavy surface are excluded, so that the only unknown is the refraction effect. Once the synthetic environment is created, it is fast and cheap to produce as much data as needed for training a model. Additionally, synthetic data can have perfectly accurate labels, including labeling that may be very expensive or impossible to obtain by hand. By using a mathematical function to generate and describe the DTMs, incompatibilities and errors that might be transferred to the solution by the true depth data are avoided, leading to independent and objective results. This furthermore allows the generation of high quality training data.

2 Synthetic Data Generation

In this section, the synthetic data generation process is described, given that the interior and exterior orientations of the cameras and a realistic texture are a priori available. Considering this, images depicting the synthetic seabed are generated with and without the presence of water. Generating synthetic images acquired by different heights, different flight paths, different cameras, different platforms (UAV and aircraft) and over different DTMs is also considered of high importance. Radial and tangential distortions are not added to the images since they would insert undesirable additional errors in the models.

2.1 DTM Generation

A continuous function $Z = f(X, Y)$ is adopted in order to know the exact elevation for each horizontal position X, Y of the DTM. This way, when comparing the true and the apparent depths, the interpolation between discrete points is avoided, since it would introduce additional errors in the process, not related to the refraction effect. For our simulation, we use two synthetic DTMs: (a) DTM1 describes in a realistic way a typical seabed anaglyph in shallow water areas, and (b) DTM2 is characterized by changes in the elevation and more intense slopes. In Fig. 1 the DTM areas captured by the simulated UAV flights are shown.

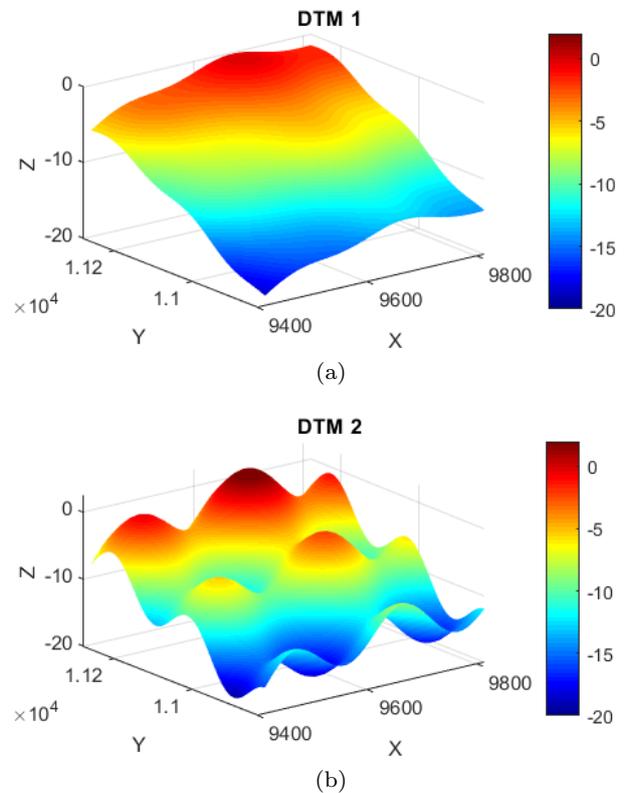


Fig. 1: The two different synthetically generated DTMs. X, Y, Z are in meters. (a) DTM1 and (b) DTM2

Although DTM2 does not represent a realistic seabed anaglyph, this type of complex seabed model is important for investigating the noise in the SfM-MVS elevations, especially when refraction is added. Moreover, this way the robustness or potential limitations of

Method 1 can be discovered. Equation 1 describes the function $Z = f(X, Y)$ adopted for both DTMs:

$$\begin{aligned} Z(X, Y) = & (Z_0 + (X - X_0) \times inc + \\ & (Y - Y_0) \times inc + ap \times \sin((X - X_0) \times ep) - \\ & ap \times \sin((Y + Y_0) \times ep) - as \times \sin((X - X_0) \\ & \times es) - as \times \sin((Y + Y_0) \times es) \end{aligned} \quad (1)$$

where for DTM1: $X_0 = 9312.94$, $Y_0 = 10729.49$, $Z_0 = -19$, the inclination $inc = 0.005$, the amplitude parameters of the two sinusoidal functions $ap = 7$ and $as = 0.5$, and the frequency parameters $ep = 0.00448785722$ and $es = 0.0314150006$. For DTM2, all values remained the same, except of ap and as that are altered to 6 and 3 respectively.

For the UAV-borne datasets (see 2.2), an area of 400×400 m for both the DTMs is covered by the imagery. For this area, the maximum elevation of DTM1 is -0.13 m (Fig. 1a) while the maximum elevation of DTM2 is 2.65 m (Fig. 1b). Their minimum elevation (or maximum depth) is -18.57 m and -19.05 m respectively. For the aircraft-borne datasets (see 2.2), a larger area of 1332×1332 m of the DTM2 is covered by the imagery. The maximum elevation of DTM2 in this area is 4.95 m and the minimum elevation is -40 m.

To texture both DTMs, affine transformations are applied on an already available typical texture of a seabed area in the Mediterranean Sea, depicting rock formations, sandy areas and seagrass covered areas, in order to cover the whole area of the DTM. For the UAV-borne datasets, the Ground Sampling Distance (GSD) of the generated texture is set to 0.03 m \times 0.03 m while for the aircraft-borne dataset the GSD of the generated texture is set to 0.10 m \times 0.10 m, facilitating the synthetic imagery generation in adequate and realistic resolution later. It is considered that the used texture does not affect the SfM-MVS elevations, however, it is a factor that could affect the generated noise in the point clouds.

2.2 Synthetic Image Generation and Processing

For achieving the goals of this work, it is necessary to generate separate blocks of synthetic images having exactly the same exterior orientations and depicting the seabed areas with and without the presence of the water. As such, eight synthetic datasets are created for four representative flying heights; for DTM1, two flights (with and without water) are simulated at a flying height of 150 m and two flights are simulated at a flying height of 200 m, for DTM2 two flights are simulated at a flying height of 200 m and two flights are simulated at a flying height of 2800 m.

For generating the synthetic imagery, the already mentioned texture is treated as an orthoimage. As such, the orthoimage is back projected to create perspective images from the specified camera positions, featuring the specified interior orientation. For the generation of the non-refracted images, the straight-forward approach presented in Skarlatos and Georgopoulos (2006) is implemented, expanding also to generate RGB images. For introducing the refraction effect on the imagery, the inverse approach of the image refraction correction method (Method 2) described in (Skarlatos and Agrafiotis, 2018; Agrafiotis et al., 2020) is followed. The air-water interface is considered flat at 0 m elevation.

Regarding the camera parameters, for the images depicting the textured artificial surface described by DTM1, where only UAV-borne images are simulated, a focal length of 3.60 mm with pixel size of $1.55 \mu\text{m}$ and a typical image size of 4000×3000 pixels is selected. For the UAV-borne images over DTM2, a focal length of 4.50 mm is used, having the same pixel size and the same image size as the images of DTM1, representing a vast majority of the commercial light weight RGB sensors exploited in low altitude UAV image-based mapping (i.e. the DJI FC330, the GoPro Hero 4 etc.).

For the aircraft-borne images over DTM2, a focal length of 100.5 mm is selected with pixel size of $4 \mu\text{m}$ and an image size of 26460×17004 pixels, representing an airborne sensor for large scale image-based mapping (i.e. Vexcel's UltraCam Eagle M3 etc.).

Fig. 2 depicts an example of a generated synthetic image without refraction for the 150 m [DTM1] dataset in the left column of row I, while in the right column, the same image with refraction is shown. The example images of the 200 m [DTM2] and 2800 m [DTM2] datasets are shown in rows II and III respectively. In the same figure, it can be noticed that the images which are affected by refraction are slightly magnified, depicting a smaller area on the seabed. This is in line with the results presented in Agrafiotis and Georgopoulos (2015) where according to the authors, when water is added in the light ray path, the effective camera constant is always larger than the camera constant in air, leading to a reduction of the field of view.

Table 1 presents the details of the four non-refracted datasets and of the subsequent SfM-MVS processing in Agisoft Metashape software. This processing is necessary for the generation of the point clouds to be used in Method 1. The same details apply for the remaining four refracted datasets, except of the *Root Mean Square Errors (RMSEs)*, which for those variants are in parentheses. It is noted that for the refracted datasets, the *RMSEs* of the camera positions exceed the ones resulted from the non-refracted ones and that as an effect

of refraction on the images, overlapping is expected to be slightly smaller.

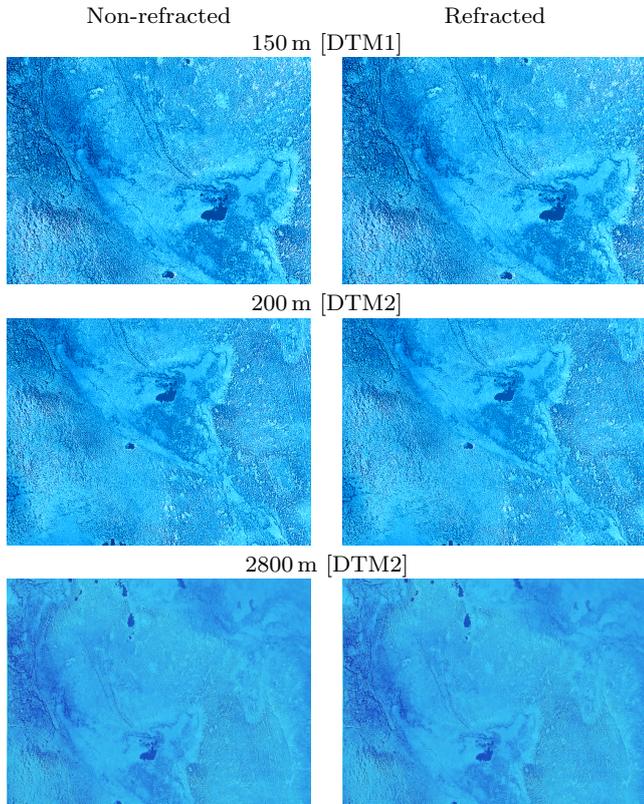


Fig. 2: Indicative synthetic imagery without refraction (left column) and with refraction (right column). Both image variants share exactly the same exterior orientation.

3 Investigating additional errors not related to refraction

To estimate the amount and the type (random or systematic) of the errors in depth determination that are not related to the trained SVR models but are introduced by the SfM-MVS processing and possibly by the synthetic data generation pipeline, the following comparison is performed: The elevations of the X , Y points generated from these processes using the non-refracted synthetic datasets, are compared with the respective true elevations calculated using Equation 1 for exactly the same X , Y points.

Fig. 3 serves as an explanatory figure of the 2D histograms that are presented next in the article. There, the 2D histogram of the comparisons performed for the 150 m [DTM1][n-r] synthetic dataset, which juxtaposes

the elevation and the differences between the true and the calculated elevations, is accompanied with two 1D histograms; the one on top is the histogram of the elevation differences of -6 m (cross section of the 2D histogram) while the one on the right is the elevation histogram of a difference of 0.055 m (longitudinal section of the 2D histogram), which is the average elevation difference for this dataset.

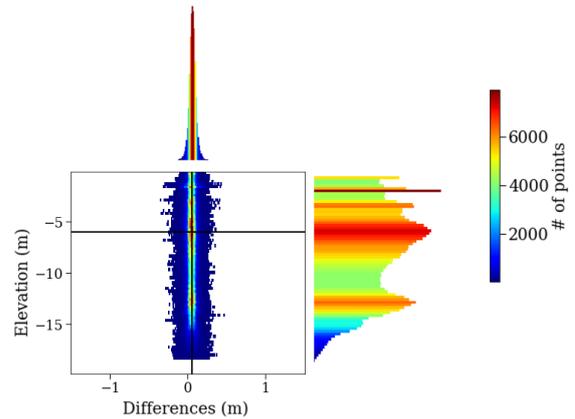


Fig. 3: An indicative 2D histogram of the differences between the true (Z) and the SfM-MVS calculated elevations using the non-refracted imagery. Traces of the cross and longitudinal sections are in black.

Regarding the histogram of the elevations, it is directly related to the DTM used for the synthetic data generation, and thus, for the different DTMs it would be different. On the contrary, the histogram of the differences is approximated by a higher-order Gaussian or super-Gaussian distribution for all the tests performed (see Fig. 4). Results of the comparisons are presented in Table 2 while the histograms of the calculated differences in relation to the true elevation are illustrated in Fig. 4.

Fig. 4, depicts the histograms of the differences between the true and the SfM-MVS calculated elevations using the non-refracted imagery, for all the four non-refracted datasets. There, it can be noticed that SfM-MVS processing, is underestimating the elevations (\bar{x} is positive). This systematic effect is also reported in Smith and Vericat (2015) and Nesbit and Hugenholtz (2019). These systematic errors, are accompanied by larger differences that are scattered almost equally on either side of the mean value, forming the higher-order Gaussian distribution already reported. These larger differences, that according to the literature (Smith and Vericat, 2015; Nesbit and Hugenholtz, 2019) are present

Dataset	150 m	200m	200m	2800m
	[DTM1] [n-r]	[DTM1] [n-r]	[DTM2] [n-r]	[DTM2] [n-r]
# Images (# strips x # images)	24 (4 x 6)	12 (4 x 3)	16 (4 x 4)	12 (4 x 3)
Avg. B/H ratio along strip (n-r)	0.45	0.45	0.31	0.21
Avg B/H ratio across strip (n-r)	0.52	0.52	0.69	0.32
Along strip overlap (n-r)	65%	65%	70%	70%
Across strip overlap (n-r)	70%	70%	50%	70%
Image ground footprint[m] (n-r)	260 x 195	344 x 258	277 x 208	2949 x 1887
GSD [m] (n-r)	0.065	0.086	0.069	0.111
$RMSE_X$ [m]	0.005 (0.006)	0.002 (0.004)	0.004 (0.012)	0.006 (0.008)
$RMSE_Y$ [m]	0.002 (0.014)	0.001 (0.014)	0.002 (0.063)	0.004 (0.016)
$RMSE_Z$ [m]	0.001 (0.015)	0.001 (0.007)	0.001 (0.016)	0.001 (0.029)
Pixel size [μ m]	1.56	1.56	1.56	4
Camera constant c [mm] (n-r)	3.61	3.61	4.5	100.5

Table 1: Details about the simulated aerial campaigns, the derived datasets and relative information. n-r stands for non-refracted and B/H for Baseline to Height. $RMSEs$ are for camera position and $RMSEs$ in parenthesis stand for the refracted datasets.

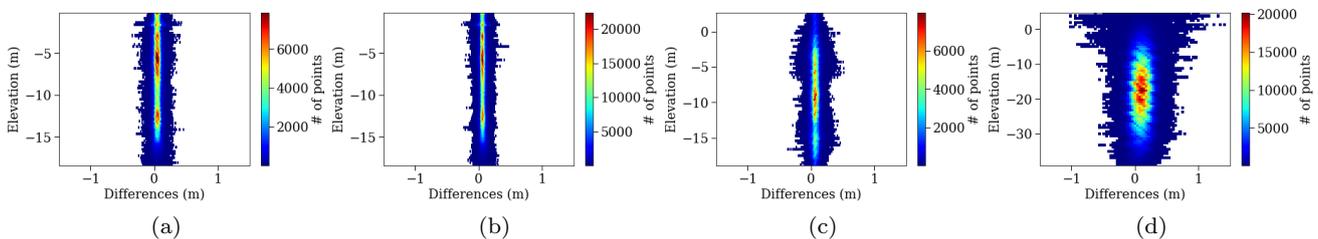


Fig. 4: The 2D histograms of the differences between the true elevations and: (a) the SfM-MVS calculated elevations using the imagery of the 150 m [DTM1][n-r] dataset, (b) the 200 m [DTM1][n-r] dataset, (c) the 200 m [DTM2][n-r] dataset, and (d) the 2800 m [DTM2][n-r] dataset.

Dataset [n-r]	\bar{x} [m]	σ [m]	$RMSE_Z$ [m]	Expected $RMSE_Z$ according to Smith and Vericat (2015) [m]	Evaluated Points
150 m [DTM1]	0.055	0.043	0.070	0.23	2.507.667
200 m [DTM1]	0.069	0.030	0.078	0.31	5.789.766
200 m [DTM2]	0.071	0.047	0.085	0.31	2.162.523
2800 m [DTM2]	0.121	0.110	0.164	not applicable	4.398.822

Table 2: Differences between the SfM-MVS and the true elevations of the point clouds generated using the non-refracted datasets. \bar{x} is the average distance of the point cloud from the true elevations and σ the standard deviation.

in most of the nadir image blocks, are justifying the calculated standard deviations of Table 2 and they are indicating the noise in the elevations of the point cloud, introduced mainly by the SfM-MVS process.

It is expected that this noise will be more intense when refraction is added on the images, primarily due to the erroneous key point matching of points at the edges of the images which are severely affected by the larger incidence angles and secondarily by the usage of RGB imagery, since the amount of refraction is different for each wavelength (Agrafiotis et al., 2020; Agrafiotis, 2020).

4 Learning from Synthetic Data

This section presents the methodology for learning from the generated synthetic datasets. Initially, the additional errors in the point cloud generation process, not related to image ray refraction are investigated. Consequently, this section discusses the training, validation, and testing schemes using the synthetic data.

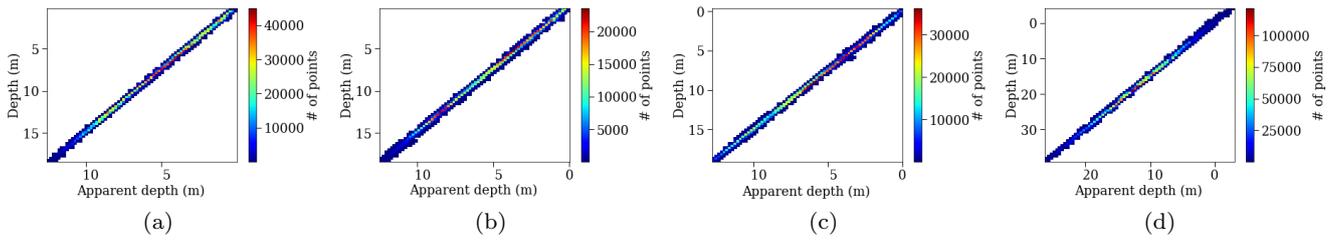


Fig. 5: The 2D histograms of the juxtaposed real and apparent depths and the respective point densities for (a) 150 m [DTM1], (b) 200 m [DTM1], (c) 200 m [DTM2] and (d) 2800 m [DTM2] refracted datasets.

4.1 Training, Validation, and Testing Methodology

Having investigated the additional, non-refraction related errors introduced in the point cloud generation process, Method 1 is implemented to process the already described synthetic datasets being affected by the refraction. In this method, SVR models are trained using the apparent (from SfM-MVS) and the true depths of point clouds, facilitating the prediction of the correct depths of unseen data. In order to demonstrate the potential for applications over different areas and shallow waters, training and cross-testing is performed in different synthetic sites, including also the synthetic aircraft-borne imagery. As such, a model trained using the apparent and true depths of the 3D point cloud of a synthetic test site is tested by predicting the correct depths on the rest of the synthetic test sites, where the true depths are also available but used only for evaluation. Later, in Section 6, these trained models will be used to predict the correct depths over real-world datasets as well delivering important results.

For the experiments performed here, four different sets are formed for training, validation, and testing. These sets are depicted in Fig. 5. There, the 2D histograms of the juxtaposed true and apparent elevations and the respective point densities are presented, showing also their clear linear relationship.

4.1.1 Training and Validation Scheme

SVR has a great potential to deliver robust models, even using sparse data for training (Bishop, 2006; Awad and Khanna, 2015; Vapnik and Chervonenkis, 2015). Due to the large size of the available data for correlation in all the four synthetic datasets, the use of percentages of the total data for training is investigated. To that direction, models are trained using 80%, 30% and 5% of the data. Results suggested that for all three percentages, models with comparable accuracy are derived. However, expectedly training on 80% of the data required much more computational resources compared

to 30% and 5% of the data. Considering this, training is performed using 5% of the available points.

This way, samples are first shuffled and then split into 5%-95% blocks. Splits are created by preserving the same percentage for each group of ordered indexes in the complete set (Pedregosa et al., 2011), ensuring a representation of all depths. For all four datasets, 5% of the points are used for training and the remaining 95% for validation.

4.1.2 Testing Scheme

Testing is performed only on unseen data i.e. different datasets. Using the model trained on the 5% of the 150 m [DTM1] dataset, the corrected depth over the 100% of the 200 m [DTM1], the 200 m [DTM2], and the 2800 m [DTM2] datasets is predicted. The respective testing procedure is repeated for models trained on the remaining three datasets.

5 Experimental Results and Validation on the Synthetic datasets

In this section, the results of the training, validation and testing of the synthetic data processed by Method 1 are presented, delivering valuable additional results. These trained models are used for correcting the geometric effects of refraction on real world image-based point clouds later in Section 6.

5.1 Training Results

To facilitate the visual comparison between the trained models, Fig. 6 demonstrates zoomed details of the four models predicting the depths of the 150 m [DTM1] dataset. In this figure, the model trained on the 150 m [DTM1] dataset is also plotted to show the clear overlap with the remaining two models trained on the UAV-borne datasets.

It is obvious that the models trained on the UAV-borne datasets are overlapping across the entire range of

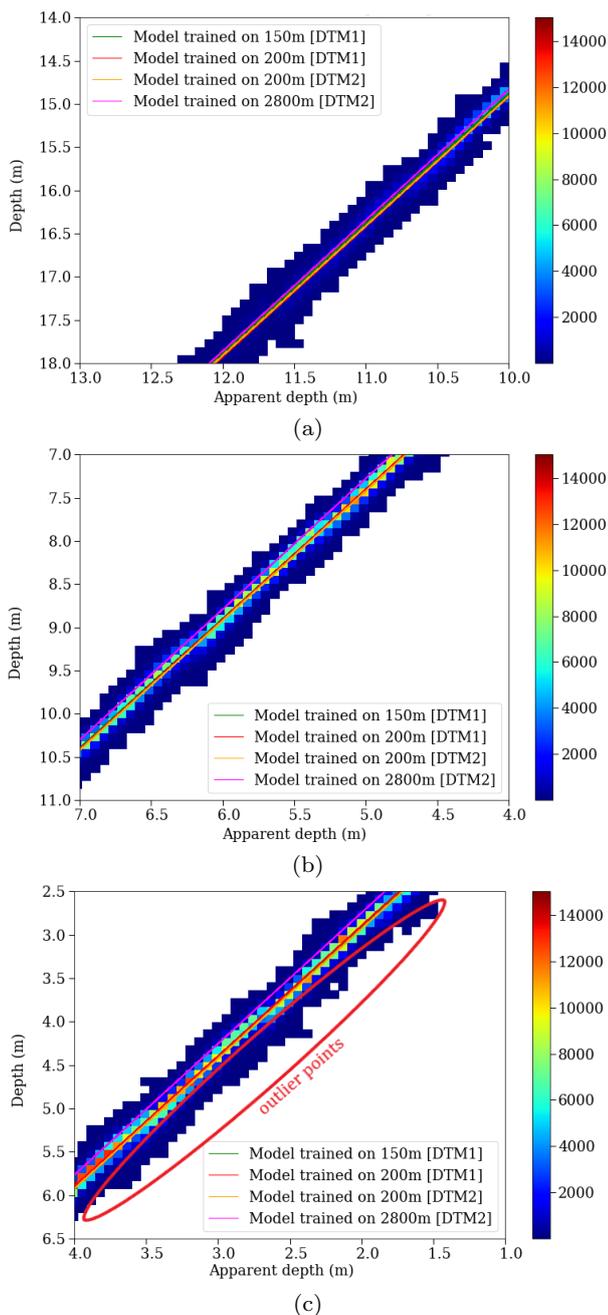


Fig. 6: Zoomed details of the juxtaposed true and apparent depths of the 150 m [DTM1] dataset, the point densities, and the respective linear SVR models.

the 2D histogram, although they are trained in datasets having different flying heights, different interior and exterior orientations, different B/H ratios, and different seabed anaglyph. Slight differences are observed between the UAV and the aircraft-borne model, being in the range of 0.002-0.031 m in the deeper areas (Fig. 6a). This deviation is less than the half of the GSD size of

the imagery used. On the contrary, in the shallower areas of the 150 m [DTM1] dataset (Fig. 6b, c), this deviation is slightly smaller than the pixel size, i.e. 0.07-0.09 m. All the models, especially the ones trained on the UAV-borne datasets, succeed in following the Z - Z_0 distribution of the large percentage of the points (see also Fig. 5). Compared with the models trained on real world datasets presented in Agrafiotis et al. (2019b,a), these results are very promising regarding the generalization applicability of the models.

In Fig. 6, a number of outlier points appear to lie away from the predicted models (points in dark blue that are indicatively marked in Fig. 6c with the red ellipse). However, as can be seen in Figure 7, these points are less than 5% of the total juxtaposed points and they affect the final accuracy of the results only marginally. In the next paragraphs, the above models are evaluated in terms of accuracy highlighting the high performance of this proposed method, discussing also issues and differences observed between the predicted and true depths calculated by Equation 1.

5.1.1 Validation and Testing Results

As a first step towards the evaluation of the trained models and their predicted depths, the amount of the errors that is introduced by the refraction effect is demonstrated by comparing the initial (uncorrected or apparent) depths (Z_0) of the synthetic datasets with the true depths (Z) calculated by the Z function (Equation 1). In Fig. 7, it can be clearly observed that the differences between the true depths and the uncorrected depths are increasing proportionally to the depth, reaching a mean difference of 5.5 m at the depth of 15 m. This means that the mean difference is in the order of 36.7% of the true depth. In all the cases demonstrated in Fig. 7, the average value of the differences is significant, being between 2.48 m to 2.96 m with a standard deviation (1σ) of 1.47 m and 1.54 m, and $RMSEs_Z$ of 2.87 m and 3.34 m, respectively. For the 2800 m [DTM2] synthetic dataset, these values are larger, confirming that refraction cannot be neglected, even when images are captured from higher altitudes (Skarlatos and Agrafiotis, 2018).

As can be observed in the same figure, the majority of the differences in all of the histograms are far from the red dashed lines. These lines, also appear in some of the next figures represent the maximum allowable total vertical uncertainty (TVU) dictated by IHO's Special Order (S44) (IHO, 2020). To compute the maximum TVU at the 95% confidence level, the following formula is used:

$$TVU_{\max}(d) = \sqrt{a^2 + (b \times d)^2} \quad (2)$$

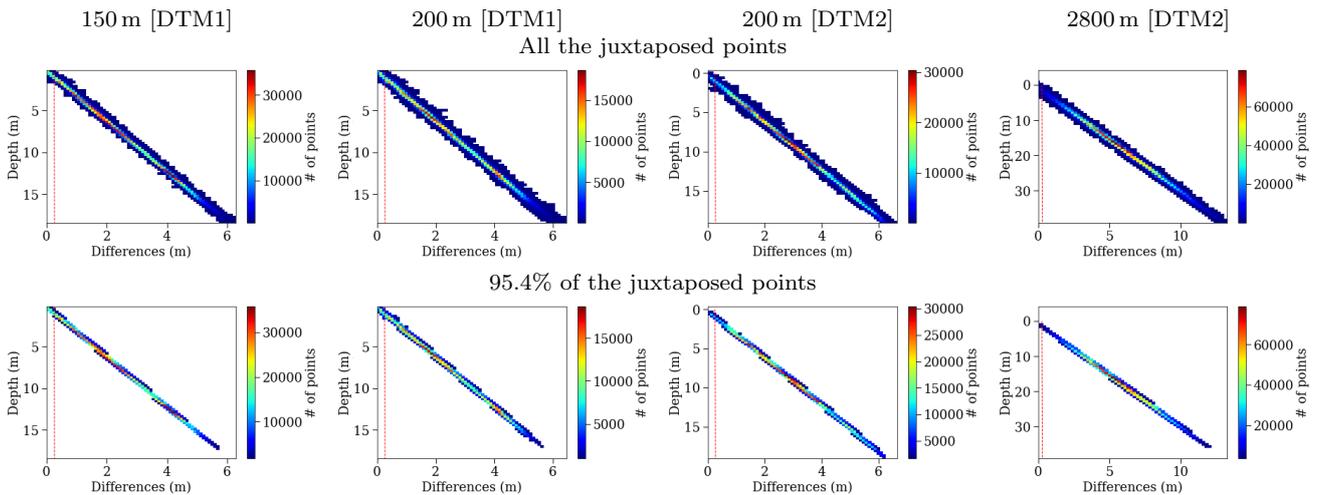


Fig. 7: The 2D histograms of the differences between the true and the uncorrected SfM-MVS depths for the 150 m [DTM1] dataset (column I), the 200 m [DTM1] dataset (column II), the 200 m [DTM2] dataset (column III) and the 2800 m [DTM2] dataset (column IV).

where for the Exclusive Order, $a=0.15$ m, $b=0.0075$, the Special Order, $a=0.25$ m, $b=0.0075$, the Order 1a, $a=0.50$ m, $b=0.013$, while for the Order 2, $a=1$ m, $b=0.023$, and d is the depth. The 95% confidence level for depths is defined as $1.96 \times \sigma$.

Fig. 8 presents the 2D histograms of the differences between the true depths and the corrected depths produced from the predicted model trained on each synthetic dataset. In the odd rows of Fig. 8, the 2D histograms of the juxtaposed depths and differences of the synthetic datasets and the respective point densities of 100% of the points are presented while in the even rows of Fig. 8, the 2D histograms of the 95.4% of the juxtaposed points are presented. It is observed that the amount of points that surround the core of the 2D histograms and exceed the maximum TVU of the Special Order (IHO, 2020) is insignificant, representing less than the 4.6% of the total points.

Contrary to Fig. 4, for all the cases presented in Fig. 8, the majority of the points present mean differences close to 0 m. These results, confirm that, as anticipated previously, the adopted method for correcting the geometric effects of refraction based on trained SVR models, can absorb the systematic effects observed in the elevations of the non-refracted datasets (Table 2). On the contrary, the noise in the elevations of the point clouds, introduced mainly by the SfM-MVS process is still present in Fig. 8 and, as expected, is more intense due to the refraction effect (Agrafiotis, 2020). However, this noise is limited to less than the 4.6% of the compared points.

A further important observation on the results derived from Fig. 8 is that in some of the 2D histograms,

an inclination of the high density core is detected, deviating from the normal. This inclination seems to be directly related to the depth. This effect is more evident in the column IV, where the models trained using the 2800 m [DTM2] are used and evaluated. This is also the case when testing the UAV models on the 2800 m [DTM2] dataset (last two rows in Fig. 8). However, for the majority of the compared points, this deviation is no more than 0.09 m at the depth of 0 and 30 m, as also shown in Fig. 6. This effect is systematically attributed to the 2800 m [DTM2] dataset, indicated by a slight rotation around the Z axis of the figures between the models at the depth of 20 m.

To explain this indicated systematic depth-dependent effect, experiments including more high-altitude aircraft-borne datasets should be performed. However, a deeper look at the SfM processing of the images depicting both dry and water-covered areas should provide a preliminary insight into the causes.

Indeed, in both synthetic and real-world cases, it is observed that when low-altitude UAV images are used, image matching is heavily dependent on the percentages of the depicted emerged and submerged areas. Those images are mainly being matched and consequently aligned using the key-points found in the prevailing area, dry or water-covered. Candidate matches in this area are dominating the epipolar geometry, which slightly differs depending on whether the area is submerged or not, due to the effects of refraction. This way, matches that are not following the epipolar constraint of the prevailing area are filtered out.

Due to the lower altitude of the UAV, incidence angles are larger, compared to the higher-altitude im-

Training site

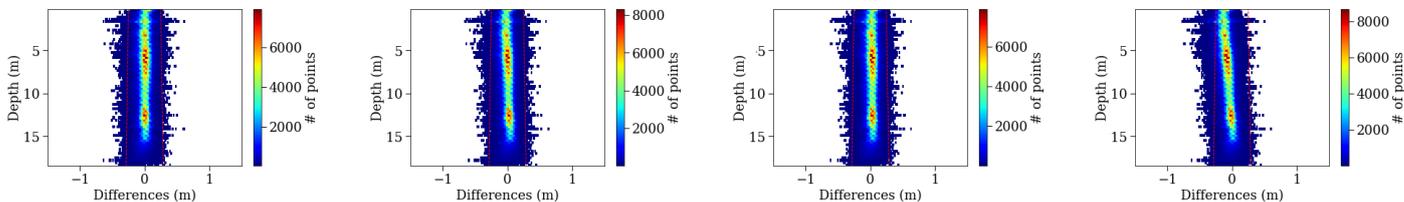
150 m [DTM1]

200 m [DTM1]

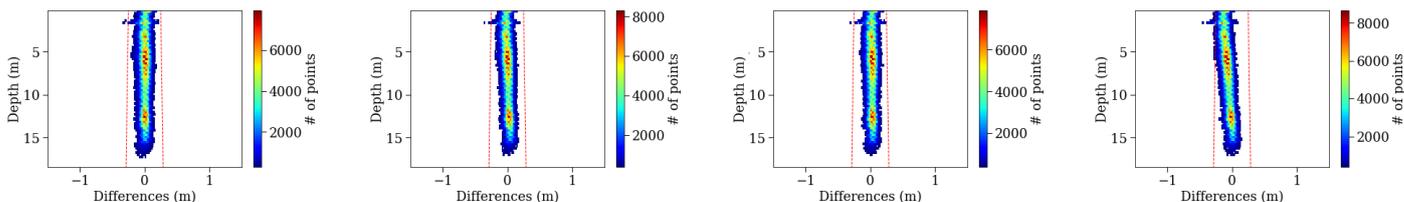
200 m [DTM2]

2800 m [DTM2]

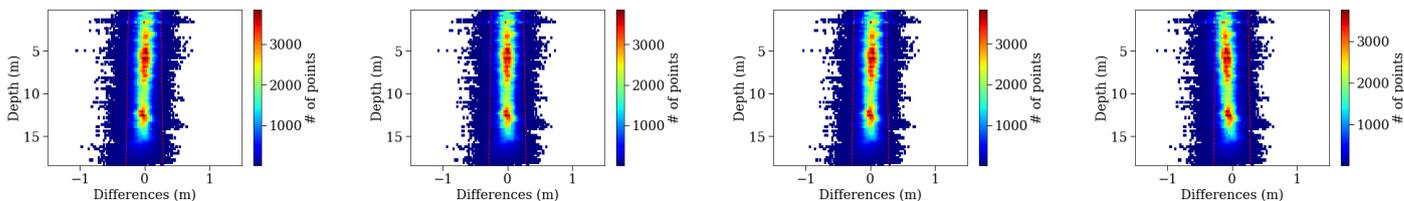
Testing on 150 m [DTM1] - Plotting the 100% of the tested points



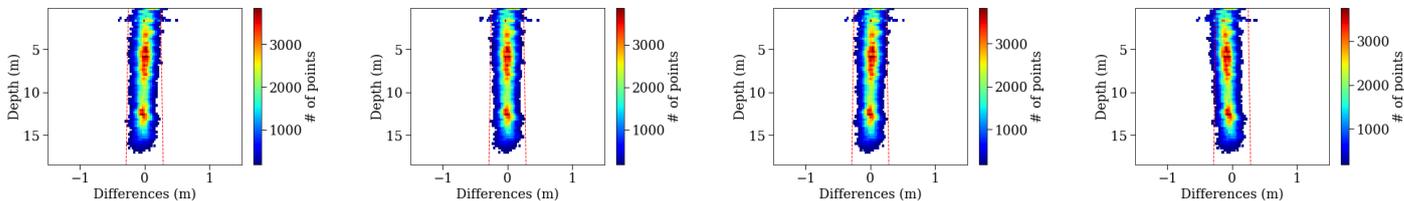
Testing on 150 m [DTM1] - Plotting the 95.4% of the tested points



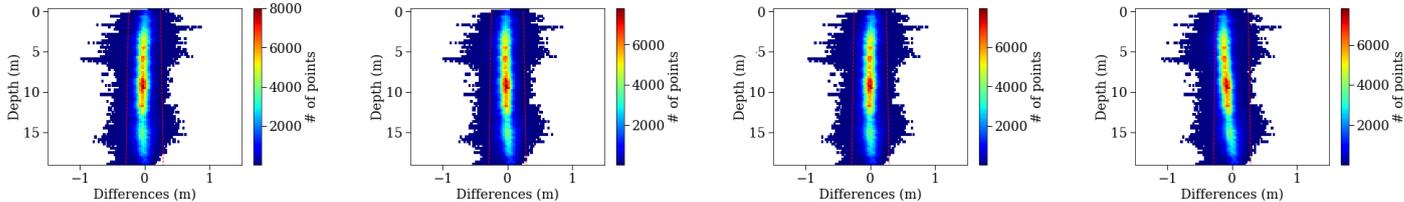
Testing on 200 m [DTM1] - Plotting the 100% of the tested points



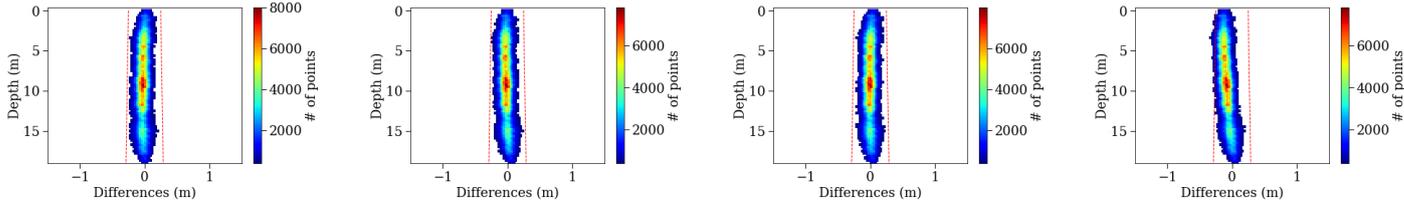
Testing on 200 m [DTM1] - Plotting the 95.4% of the tested points



Testing on 200 m [DTM2] - Plotting the 100% of the tested points



Testing on 200 m [DTM2] - Plotting the 95.4% of the tested points



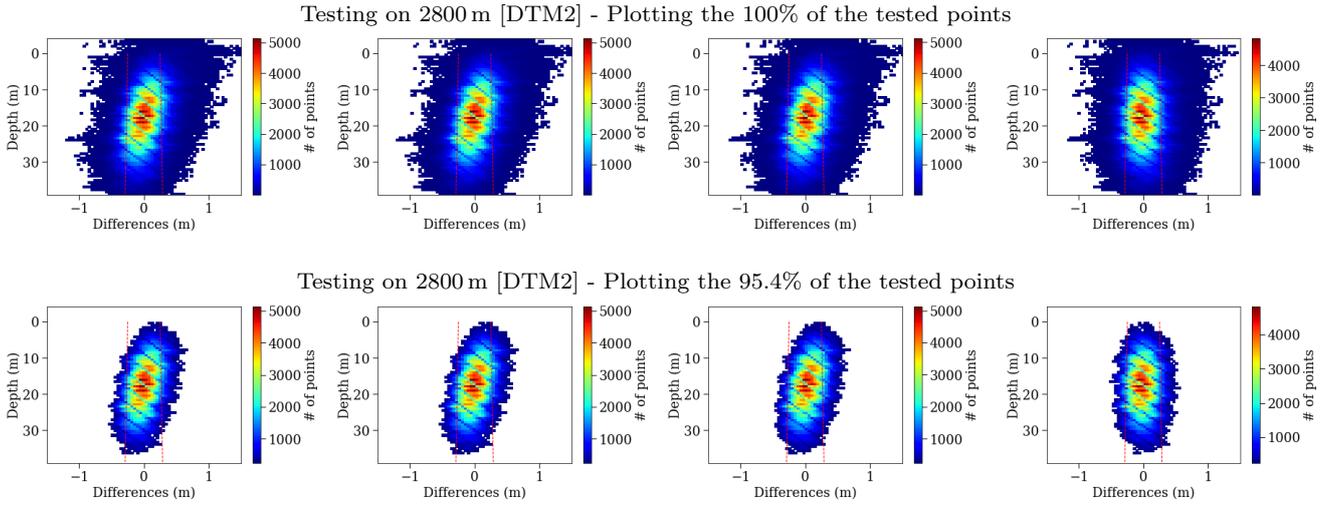


Fig. 8: The 2D histograms of the differences between the true depths and the corrected depths after the application of Method 1, in relation to the real depth. The red dashed lines represent the maximum allowable TVU dictated by the IHO's Special Order (S44) (IHO, 2020).

Dataset [5%]	Train. points	R^2	Eval. Site [95%]	Max /Min depth of test site	Eval. points	Uncorrected data			Corrected data		
						\bar{x} [m]	σ [m]	$RMSE_Z$ [m]	\bar{x} [m]	σ [m]	$RMSE_Z$ [m]
150 m [DTM1]	129.552	0.9997	150 m [DTM1]	18.7/0	2.461.488	2.48	1.47	2.88	0.01	0.07	0.07
150 m [DTM1]	129.552	0.9993	200 m [DTM1]	18.7/0	1.409.673	2.49	1.49	2.88	0	0.11	0.11
150 m [DTM1]	129.552	0.9996	200 m [DTM2]	19.05/0	2.097.713	2.96	1.54	3.34	-0.02	0.09	0.09
150 m [DTM1]	129.552	0.9996	2800 m [DTM2]	40/0	4.071.219	5.86	2.52	6.38	0.03	0.22	0.23
200 m [DTM1]	74.193	0.9997	150 m [DTM1]	18.7/0	2.461.488	2.48	1.47	2.87	0.01	0.07	0.08
200 m [DTM1]	74.193	0.9994	200 m [DTM1]	18.7/0	1.409.673	2.49	1.49	2.88	0	0.11	0.11
200 m [DTM1]	74.193	0.9996	200 m [DTM2]	19.05/0	2.097.713	2.96	1.54	3.34	-0.01	0.09	0.09
200 m [DTM1]	129.552	0.9996	2800 m [DTM2]	40/0	4.071.219	5.86	2.52	6.38	0.03	0.21	0.23
200 m [DTM2]	110.405	0.9996	200 m [DTM2]	19.05/0	2.097.713	2.96	1.54	3.34	0	0.09	0.08
200 m [DTM2]	110.405	0.9997	150 m [DTM1]	18.7/0	2.097.713	2.48	1.47	2.87	0.03	0.07	0.11
200 m [DTM2]	110.405	0.9994	200 m [DTM1]	18.7/0	1.409.673	2.49	1.49	2.88	0.02	0.11	0.09
200 m [DTM2]	129.552	0.9996	2800 m [DTM2]	40/0	4.071.219	5.86	2.52	6.38	0.04	0.20	0.23
2800 m [DTM2]	214.274	0.9996	2800 m [DTM2]	40/0	4.071.219	5.86	2.52	6.38	0.01	0.19	0.20
2800 m [DTM2]	214.274	0.9996	150 m [DTM1]	18.7/0	2.461.488	2.48	1.47	2.87	-0.05	0.09	0.10
2800 m [DTM2]	214.274	0.9996	200 m [DTM1]	18.7/0	1.409.673	2.49	1.49	2.88	-0.03	0.11	0.12
2800 m [DTM2]	214.274	0.9996	200 m [DTM2]	19.05/0	2.097.713	2.96	1.54	3.34	-0.06	0.10	0.11

Table 3: Quantitative evaluation results. Comparing the results derived from all performed experiments.

agery, intensifying the effect and consequently increasing the differences between the calculated epipolar geometry of both the areas. Apparently, this is restricted to a limited area of the image block, being parallel to the coastline, and depicting the shallower areas of the site. It leads to the generation of different apparent depths depending on whether the images used are matched using emerged or submerged key-points. Sometimes, these depths do not follow the strict linear model that applies to the rest of the submerged areas of the block, but they are subject to the flight plan and the image overlap in this area. As a measure to mitigate this problem, a cross hatch pattern of variable flying altitudes is suggested in order to facilitate reliable and evenly distributed matches.

However, this is not the case for the high-altitude aircraft-borne imagery that covers larger portions of emerged and submerged areas in the same image. There, due to the higher altitude, smaller differences in epipolar geometry are found and matches are kept for both the areas. This difference in the SfM derivatives between the UAV and the aircraft-borne datasets results in a slight differentiation of the trained models (see Fig. 6) leading to the discussed rotation. The application of Method 2 presented in Agrafiotis et al. (2020) overcomes the problem by correcting the imagery by the refraction effect (see Fig. 9), facilitating the evenly distributed matches in both the emerged and submerged areas of the corrected images. The effect of refraction on the key-point matching on the deeper areas of the block is discussed in Subsection 6.3.1.

Table 3 presents the results of each one of the 4 validation and 12 testing approaches. Both in Fig. 8 and Table 3, an impressive improvement in depth accuracy is observed. More specifically, for all simulated UAV synthetic datasets used, the initial depth deviation being in the range of 2.48 m to 2.96 m is reduced to the range of 0 m to 0.027 m, the standard deviation of 1.47 m to 1.58 m is reduced to the range of 0.07 m to 0.11 m, and the $RMSEs_Z$ decreased from initially 2.87 m to 3.34 m to 0.07 m to 0.11 m. When real datasets are under testing, this deviation is expected to increase in depths of more than 10 m, since errors and limitations caused by the progressive, depth-related blurring would be introduced (Agrafiotis, 2020). For the aircraft-borne synthetic test sites, the statistical indexes are doubled, as a result of the rotation reported previously and the larger GSD size.

To categorize the results according to IHO’s S44 (IHO, 2020), the 95% TVU is calculated as $1.96 \times \sigma$, for all the tests performed. When the models trained on the 200 m [DTM1] and 200 m [DTM2] datasets are used to correct the depths of the 150 m [DTM1] dataset,

results are satisfying the TVU limits of the Exclusive Order while for the remaining simulated UAV datasets, the maximum TVU of the Special Order is met. It is highlighted that the latter applies also for the models trained on the 2800 m [DTM2] dataset when applied on all the UAV datasets. However, when the models trained on the UAV datasets are applied on the 2800 m [DTM2] dataset, resulted accuracy is within the TVU of Order 1a. Moreover, considering IHO’s white paper on the regulations for international charts and charts specifications (IHO, 2019), when using the synthetic data, the proposed method is qualified to the maximum Zone of Confidence (ZOC) level A1, for which, the minimum depth accuracy is calculated as $0.50+1\% \times d$, where d is again the depth.

It is also important to note that the large distances between the point clouds observed in Fig. 7 disappeared. This improvement is observed in every test performed, proving that the trained SVR models on synthetic data achieve a great reduction of the geometric effects caused by the refraction to the bottom point clouds, and thus successfully eliminate the errors in depth determination.

To compare the achieved results with the differences found by the comparison between the uncorrected and the true depths (Fig. 7), it is reminded that at the depth of 15 m, the difference reaches the 5.5 m which relates to 36.7% of the depth. On the contrary, when comparing the corrected depths resulted by the proposed method with the true depths, differences at the depth of 15 m are within the range of -0.20 m to 0.20 m. A majority of the deviations even range from -0.05 m to 0.05 m, i.e. corresponding to 1.33% and 0.33% respectively of the true depth for all the tests performed.

An additional measure to evaluate the predicted models used is the computation of fitting score R^2 , which is defined as:

$$R^2 = 1 - \frac{\sum (Z_{true} - Z_{predicted})^2}{\sum (Z_{true} - Z_{true.mean})^2} \quad (3)$$

The best possible score is 1.0 and it may also be negative (Pedregosa et al., 2011). Z_{true} is the true value of the depth of the points not used for training, while $Z_{predicted}$ is the predicted depth for these points, using the model trained on an independent set of points. Results in Table 3 highlight the great potential of the SVR models when using synthetic data. As can be seen there, the fitting score achieved for all the trained models indicates that they describe the data in a very precise way.

6 Experimental Results and Validation on the real-world datasets

In this section, Method 1 and Method 2 are applied to predict the depths of real world datasets and correct the refracted imagery, respectively, using the SVR models trained on the synthetic datasets and the SVR model trained on a real-world dataset (see 6.2). Additionally, the real world data are corrected using the 3D space correction approaches presented in Dietrich (2017) (Method 3) and in Woodget et al. (2015) (Method 4) and the image-based refraction correction method described in Skarlatos and Agrafiotis (2018) (Method 5). Implementation details and results of these methods are originally presented in Agrafiotis et al. (2020), however, it is considered useful to demonstrate the results also in this article, highlighting the advantages offered by the use of the synthetic data.

6.1 Description of Datasets and Processing

The real world data used for this purpose consist of two shallow water areas in Cyprus; Amathounta and Agia Napa, and two in Greece; Cyclades-1 and 2. All the sites, represent typical examples of seabed and water column characteristics in the Eastern Mediterranean Sea. Both UAV data and true depth data are available for these areas.

In Amathounta, the maximum depth is 5.57 m. A Swinglet CAM fixed-wing UAV with a Canon IXUS 220HS camera with a focal length of 4.3 mm was used for image acquisition, and a total of 182 images were acquired, from an average flying height of 103 m. In Agia Napa, the maximum depth is 14.8 m and the flight is executed with the same system. In total, 383 images were acquired, from an average flying height of 209 m. For both the sites, LiDAR data are used for validation. In Cyclades-1, the maximum depth is 6.9 m. A Phantom 4 UAV with an FC330 camera with a focal length of 3.61 mm is used for data acquisition. In total, 449 images were acquired from three different average flying heights of 88 m, 70 m, and 35 m. Finally in Cyclades-2, the maximum depth is 4.05 m. The same UAV was used and in total, 203 images were acquired from 75 m and 33 m height. For those two test sites, topographic points measured with a total station are used for validation.

A standard SfM-MVS approach is followed in order to obtain the required data for applying the proposed refraction correction methods, i.e. the interior and exterior orientation of the cameras and the initial dense point cloud. For the approaches presented here, Agisoft Metashape is used. Available data and processing are described in detail in Agrafiotis et al. (2019a,b, 2020).

6.2 Refraction Correction

Considering that all the models trained on the UAV-borne synthetic data are similar (Fig. 6), only the trained model over the 150 m [DTM1] dataset is used for correcting the point clouds of the real world areas. To compare the performance of this model with that of a model trained on real world data, both Method 1 and 2 are also applied using an SVR model trained on the Dekelia dataset, described in detail in Agrafiotis et al. (2020). The main aim here is to demonstrate the improved accuracy offered by the synthetic data in the trained SVR models and that these models have significant potential of generalization for applications over different areas and shallow waters. The investigation and evaluation of the use of synthetic data is performed in both the aforementioned methods in order to demonstrate the increased accuracy offered for both the cases.

6.3 Comparative analysis of the results

Having available all the necessary data, the possible improvements on the accuracy of the results achieved when exploiting the models trained on the synthetic data, are investigated and evaluated.

For evaluating the improvements using Method 1 the corrected depths of the point clouds are compared with the true depths available for each dataset. For Method 2, the evaluation of the improvements on the corrected imagery is performed through the evaluation of the SfM results achieved using the imagery. For this method, the dense point cloud regeneration is not necessary after the correction of the imagery since the dense point cloud is already corrected by the intermediate step of applying the first method and used for the image correction. Also, there is no ground truth for the exterior orientations of the cameras. As such, the sparse point clouds of the updated SfM solutions are evaluated to demonstrate the improvements. Indeed, in the literature, a well-established methodology for evaluating SfM performance is to compare the sparse point cloud to some ground truth with the same data representation (Bianco et al., 2018). For reasons of consistency, the evaluation of the first method is also performed on the sparse point clouds.

To this end, the comparison of the depths of the respective point clouds with the LiDAR data is performed for the Amathounta and Agia Napa test sites. For the Cyclades-1 and Cyclades-2 test areas, comparisons are performed using checkpoints measured with a geodetic total station on-site and acquired together with the image data, since LiDAR data are not available.

6.3.1 Quantitative Results and Evaluation

The results of the vertical differences among the true depth point clouds of the four test areas and the respective sparse point clouds resulted from the tested refraction correction approaches are presented in Fig. 9 and Table 4.

In Fig. 9 (row I), it can be observed that the vertical distances between the reference data and the original (uncorrected) image-based point clouds increase in proportion to the depth. Rows II, III, and IV illustrate the vertical distances in relation to the depth between the true and the sparse point clouds produced by Method 4 and Method 3. In this method, the corrected depths resulted from all the refraction angles are given in III while the filtered corrected depths, calculated using a limited set of refraction angles, in this case those less than 35° off-nadir, are given in IV. Row V illustrates the vertical distances between the true and the sparse point clouds produced by Method 1 trained on the Dekelia dataset while row VI presents the same distances, however when the SVR model is trained on the 150 m [DTM1] dataset. All the above compared approaches are meant to correct the geometric effects of refraction in the 3D space and specifically in the sparse or dense point cloud.

Row VII illustrates the vertical distances in relation to the depth between the true and the sparse point clouds produced by Method 5 while the VIII and the IX row present the 2D histograms of the vertical distances between the true and the sparse point clouds produced by the corrected imagery of Method 2, using the SVR models trained on the Dekelia and the 150 m [DTM1] datasets respectively. The latter three compared approaches are correcting the geometric effects of refraction in the image space.

Fig. 9 reveals the improved performance of both of the tested Method 1 and 2 for correcting the geometric effects of refraction when the SVR model trained on the synthetic data is used. A very important outcome of this figure is that, the resulted sparse point cloud is characterized by less noise in the elevations' differences, when the initial imagery is corrected. This is evident by comparing the 2D histograms of Method 1 with the 2D histograms resulted for all the methods that are applied in the 3D point clouds directly. The 2D histograms resulting from this image correction method are narrower in the X axis, indicating less dispersion of apparent elevations. This is even more evident when the SVR model trained on the synthetic data is used (ninth row of Fig. 9), demonstrating another positive effect of the use of the synthetic data for training. Table 4 presents the results of all the tests performed.

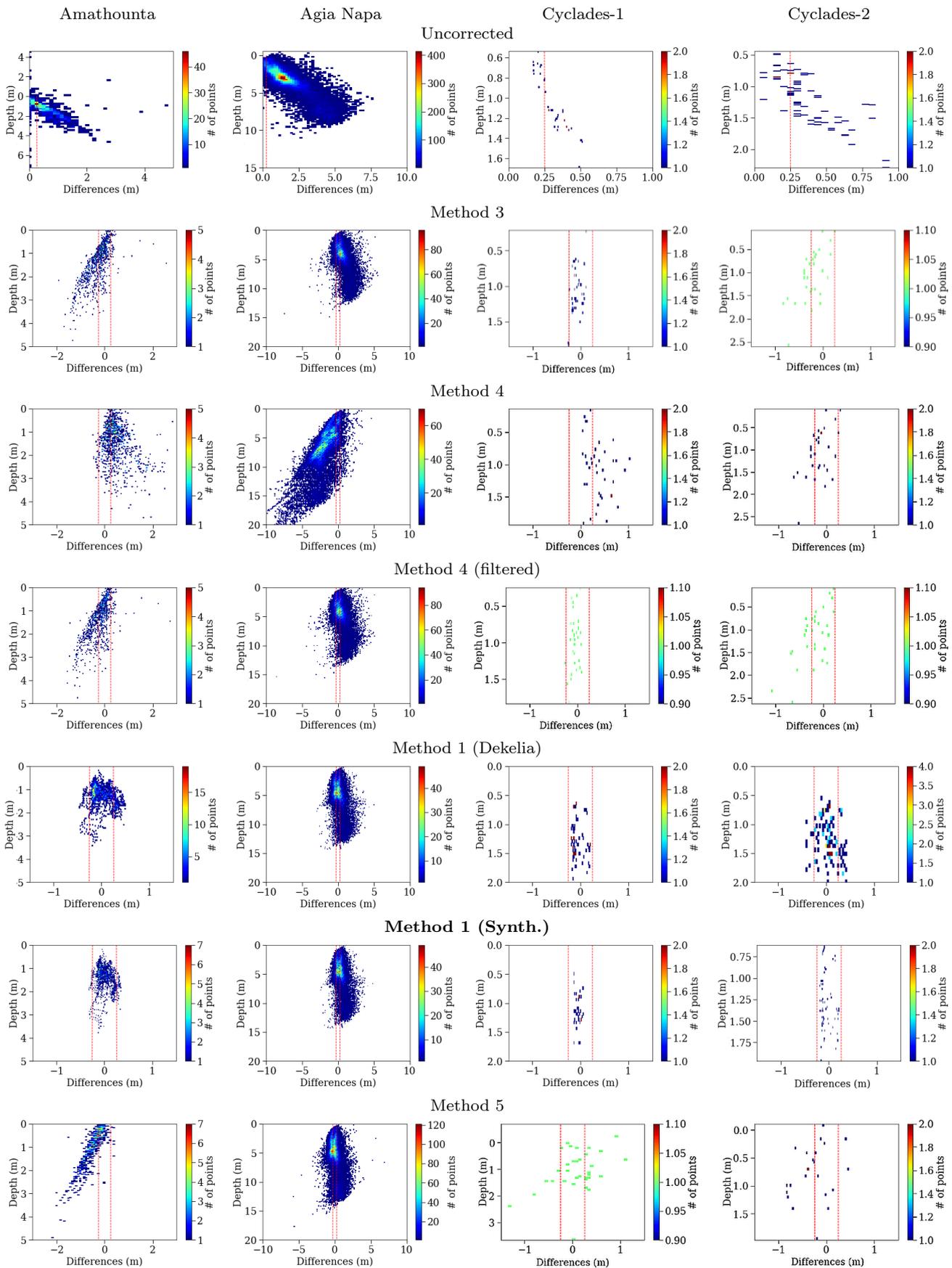
Both in Fig. 9 and Table 4, a great improvement in the depth accuracy of the sparse SfM point clouds is achieved by both Method 1 and 2 when using the synthetic data. As expected, when exploiting the SVR model trained on the 150 m [DTM1] dataset, the calculated statistical indices are better, confirming the higher performance of the model trained on the synthetic dataset compared to the real world areas.

In more detail, when the model trained on the 150 m [DTM1] dataset is used in Method 1, a large decrease in the mean distances calculated from the true data is observed; in Amathounta, reduced from 0.67 m to -0.04 m, in Agia Napa from 1.71 m to 0.06 m, in Cyclades-1, from 0.32 m to -0.05 m, while in Cyclades-2, from 0.54 m to -0.05 m too. Comparing with the results achieved by the same method, but using the SVR model trained on the Dekelia dataset, these are -0.09 m, -0.13 m, 0.02 m and -0.01 m respectively. Although the Cyclades-1 and Cyclades-2 datasets feature higher mean distances, it is important to note that the standard deviation (σ) and the *RMSE* are reduced for all the cases.

In the same context, when the model trained using the 150 m [DTM1] dataset is used in Method 2, the initial mean distance in Amathounta is reduced to -0.04 m, in Agia Napa, to -0.04 m, in Cyclades-1, to 0 m, while in Cyclades-2, to -0.05 m. Comparing with the results achieved by the same method, when exploiting the SVR model trained on the Dekelia dataset, these are -0.19 m, -0.31 m, -0.02 m and -0.06 m respectively. Again here, a great reduction is observed for the remaining statistical indices (σ and *RMSE*).

Actually, the reduction when using Method 2 is larger than previously and can be partially attributed to the correction of the imagery instead of the correction of the point clouds (Agrafiotis et al., 2020), since a decrease is apparent when the model trained on the Dekelia dataset is used too. However, a higher reduction can be achieved using the model trained on the synthetic dataset, proving the positive effect of using these models. This reduction also indicates the generation of point clouds with reduced noise.

This is a very important advantage of Method 2, indicating that when correcting the initial imagery from refraction, SfM delivers more accurate and more reliable results with less noise in both the sparse and dense point clouds. This is explained by the fact that the geometric effects of refraction in the images are affecting feature matching to an important degree; matchable key points are not matched together or they are matched with the wrong key points, because due to the refraction effect, each image is distorted in a different way. This effect drives feature matching algorithms to



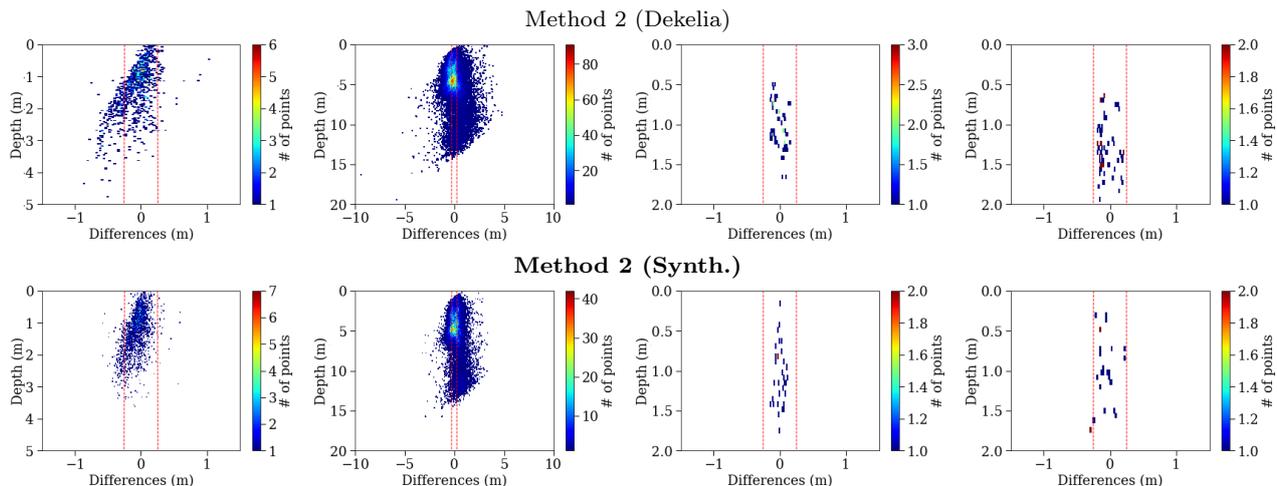


Fig. 9: The 2D histograms of the differences between the true and the uncorrected and corrected image-based sparse point clouds derived from the SfM for Amathounta (column I), Agia Napa (column II), Cyclades-1 (column III) and Cyclades-2 (column IV) test sites respectively. The red dashed lines represent the maximum allowable TVU dictated by the IHO’s Special Order (S44) (IHO, 2020). ”Synth.” stands for synthetic data and specifically the 150 m [DTM1] dataset.

	Test Site											
	Amathounta			Agia Napa			Cyclades-1			Cyclades-2		
Check points	1K			75K			23			34		
Max/Min depth [m]	5.57/0.10			14.8/0.20			6.9/0.0			4.05/0.0		
Point clouds from different methods	Statistical Analysis [m]											
	\bar{x}	σ	$RMSE_Z$	\bar{x}	σ	$RMSE_Z$	\bar{x}	σ	$RMSE_Z$	\bar{x}	σ	$RMSE_Z$
Uncorrected images	0.67	2.19	2.28	1.71	1.18	2.08	0.32	0.10	0.33	0.54	0.29	0.62
Method 3	-0.27	0.40	0.49	0.63	1.02	0.98	-0.08	0.10	0.12	-0.23	0.26	0.34
Method 4	0.49	0.54	0.73	-1.55	1.49	1.75	0.38	0.25	0.46	-0.15	0.24	0.28
Method 4 (filt.)	-0.22	0.40	0.45	0.43	0.72	0.84	-0.06	0.09	0.10	-0.20	-0.30	0.36
Method 1 (Dekelia)	-0.09	0.18	0.28	-0.13	0.51	0.55	0.02	0.09	0.09	-0.01	0.21	0.21
Method 1 (Synth.)	-0.04	0.13	0.14	0.06	0.41	0.42	-0.05	0.06	0.07	-0.05	0.12	0.13
Method 5	-0.39	0.88	0.96	-0.05	0.74	0.74	0.15	0.42	0.46	-0.28	0.36	0.46
Method 2 (Dekelia)	-0.19	0.28	0.31	-0.04	0.37	0.38	-0.02	0.09	0.09	-0.06	0.14	0.15
Method 2 (Synth.)	-0.04	0.12	0.13	-0.03	0.21	0.23	0.00	0.06	0.07	-0.05	0.06	0.09

Table 4: Quantitative evaluation results of the comparisons. \bar{x} is the average distance of the point cloud from the true values and σ is its standard deviation. Negative values suggest overestimation of the depth and positive suggest underestimation. Values achieved using the synthetic data are highlighted in bold format. ”Synth.” stands for synthetic data and specifically the 150 m [DTM1] dataset.

match points fulfilling the epipolar geometry only because of the refraction effect.

In all the cases, results include outlier points, such as seagrass, that are not captured in the true depth clouds for all the cases or are caused due to point cloud noise again in areas with seagrass or poor texture.

Compared with the results derived from Method 3 and 4, Method 1 outperformed both. Method 2, com-

pared with the accuracy levels reached by Method 1, managed to achieve slightly better accuracy in all the test sites. However, Method 2, delivered better standard deviations and $RSMEs_Z$, especially when employed SVR models are trained using the synthetic dataset. This way, it is shown that their use facilitates an improvement not only of the accuracy but also of the quality of the generated point clouds by reducing the noise.

Regarding the rest of the compared refraction correction approaches, both Method 3 and 4 are developed for correcting the effects of refraction in very shallow waters riverbeds and they seem to produce quite accurate and reliable results in the shallower areas, especially when filtering is applied in Method 4.

To categorize the results according to IHO's S44, the 95% TVU is calculated as $1.96 \times \sigma$. Results suggest that when the model trained on 150m [DTM1] dataset is used, Method 1 meets the TVU limits of: the Exclusive Order in Cyclades-1, the Sepcial Order in Amathounta and Cyclades-2, and the Order 2 in Agia Napa test sites. Method 2 meets the TVU limits of: the Exclusive Order both in Cyclades-1 and 2, the TVU limits of the Special Order in Amathounta, and the TVU limits of the Order 1a in Agia Napa. As with the synthetic data, performed tests shown that Method 2 is qualified to the maximum ZOC (IHO, 2019) level A1 for all the real world test sites. Method 1, is qualified to A1 for all the test sites, except Agia Napa, the deepest site, that is qualified to A2.

7 Discussion

In this article the generation of synthetic DTMs and of both UAV-borne and aircraft-borne aerial images of clear water-covered areas is described. To the best of authors knowledge, this is the first attempt to exploit synthetic data for correcting the refraction effect on UAV imagery and SfM-MVS derivatives.

The most important aim of this work is the exploitation of synthetic data for training unbiased SVR models as in Agraftotis et al. (2019a,b) facilitating the validation in real world cases and increasing the accuracy achieved compared to models trained in real world data. To achieve this goal, four synthetic datasets with refraction and four without, are generated and processed accordingly.

Compared with the real world data, the advantage of the synthetic data is the accuracy and the reliability of the depth information. More importantly, when it comes to seabed imaging, errors and limitations caused by visibility restrictions due to the depth, as well as errors introduced by the wavy surface are excluded, delivering unaffected results. The mathematical function used to generate the DTMs led to independent and objective results by eliminating errors transferred to the training, validation and testing process by asynchronous true data. Indeed, as in any supervised learning approach, the quality of the training data affects the accuracy of the predicted models.

The performed experiments, produced a number of important outcomes regarding the systematic errors and

the noise in the SfM-MVS elevations for both the non-refracted and refracted datasets. In all the tests performed using the non-refracted datasets, it is noticed that the average elevation differences are increasing as the flying height and the complexity of the DTM increases. This is inline with reports in the literature (Smith and Vericat, 2015; Nesbit and Hugenholtz, 2019). Moreover, SfM-MVS processing underestimated the elevations of those datasets.

In the experiments performed using both the refracted and non-refracted datasets, this underestimation, which is almost equal to the GSD size, is accompanied by larger differences, which are not systematic and they are scattered almost equally on either side of the mean value, approximating the high-ordered Gaussian distribution already reported and shown in Fig. 3. These effects, reported also in Smith and Vericat (2015) and Nesbit and Hugenholtz (2019), are apparent in most of the nadir image blocks and they are explaining the calculated standard deviations (σ) in Table 2 and Table 3 since they are indicating that the noise in the elevations of the point cloud are mainly introduced by the SfM-MVS process. However, the use of oblique images for seabed mapping applications is not suggested since this would asymmetrically increase the incidence angles and consequently the refraction and the noise in the point clouds instead of reducing it.

This noise is not related to the proposed method and cannot be avoided by any of the proposed approaches presented in the literature, since it seems to be a fundamental feature of the SfM-MVS process and as expected, it is more intense when refraction is added to the images (Fig. 8). However, this noise is limited to a mere 4.6% of the compared points. The remaining noise due to refraction is reduced when correcting the refraction in image space, using Method 2 (Agraftotis et al., 2020) and especially when the SVR model trained on the synthetic data is used.

Experimental results over the four synthetic test areas which are characterized by different flying heights, different depths, and different seabed anaglyph, along with the quantitative validation performed, indicated the high potential of the SVR models trained on these synthetic data, even if the image acquisition altitude differed by 2650 m. Additionally, results over four different real-world test areas suggested that the SVR model trained on the synthetic dataset outperforms the models trained on the real-world datasets. This is expected due to the increased quality of the synthetic training data, compared with the real-world data.

The notable performance of the SVR models trained on the synthetic datasets for correcting the refraction effect on the real world ones, demonstrated the cor-

rect modeling of the refraction effect in Skarlatos and Agrafiotis (2018) and Agrafiotis et al. (2020) too. This model is inverted for adding the refraction on the synthetic images.

8 Conclusions

The synthetic data, simulating UAV and aircraft-borne cameras with flying altitudes ranging from 150 m to 2800 m, enabled the training of accurate linear SVR models for the correction of the geometric effects of refraction and specifically for predicting accurate depths, having knowledge only of the apparent ones. This approach, addressed a number of challenges related to inevitable random measurement noise in the training data and the asynchronous image and true data, delivering generalized models.

It is shown that when the models trained on the UAV datasets are applied on different UAV synthetic datasets, results are satisfying IHO's Exclusive or Special Order TVU limits. It is also shown that the same models can be used to accurately correct the effects of refraction on image-based 3D point clouds resulted by an aircraft-borne sensor, flying more than 2500 m higher, satisfying IHO's Order 1a TVU. However, it is important that when the models trained on the aircraft-borne sensor are used to correct the refraction on the UAV-borne datasets, the accuracy is doubled, satisfying IHO's Special Order TVU. These minor differences in the UAV-borne and aircraft-borne models, are subject of further research, including more synthetic aircraft-borne datasets and thorough investigation of the effects of images depicting both emerged and submerged areas on the SfM, as preliminary reported in this article.

By exploiting the synthetic data, Methods 1 and 2 managed to achieve impressive results over real world test sites, reducing the mean distances between the true and the apparent depths in the range of 0 m to 0.05 m, the standard deviation in the range of 0.06 m to 0.24 m and the $RMSE_Z$ in the range of 0.07 m to 0.24 m. Also, they met IHO's Special Order TVU in the majority of the cases. Exclusive Order's TVU is also met in Cyclades-1 and 2 test sites. As a typical example, it is reported that the initial mean difference of 5.5 m, i.e. the 36.7% of the true depth, between the uncorrected and the true depths at the depth of 15 m, is reduced to the range of -0.05 m to 0.05 m, i.e. the 0.33% of the true depth in the vast majority of the points.

Acknowledgements We are grateful to the reviewers for their valuable comments and suggestions. Also, we would like to acknowledge the Dep. of Land and Surveys of Cyprus for

providing the LiDAR reference data, the Cyprus Dep. of Antiquities for permitting the flight over the Amathouda site and commissioning the flight over Agia Napa, and HFF for partially funding the acquisition of Cyprus' datasets.

References

- Agrafiotis P (2020) Image-based bathymetry mapping for shallow waters. PhD thesis, National Technical University of Athens
- Agrafiotis P, Georgopoulos A (2015) Camera constant in the case of two media photogrammetry. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-5/W5:1–6, DOI 10.5194/isprsarchives-XL-5-W5-1-2015
- Agrafiotis P, Skarlatos D, Forbes T, Poullis C, Skamantzari M, Georgopoulos A (2018) Underwater photogrammetry in very shallow waters: main challenges and caustics effect removal. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2:15–22, DOI 10.5194/isprs-archives-XLII-2-15-2018
- Agrafiotis P, Skarlatos D, Georgopoulos A, Karantzalos K (2019a) Depthlearn: Learning to correct the refraction on point clouds derived from aerial imagery for accurate dense shallow water bathymetry based on svms-fusion with lidar point clouds. Remote Sensing 11(19):2225
- Agrafiotis P, Skarlatos D, Georgopoulos A, Karantzalos K (2019b) Shallow water bathymetry mapping from uav imagery based on machine learning. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W10:9–16, DOI 10.5194/isprs-archives-XLII-2-W10-9-2019
- Agrafiotis P, Karantzalos K, Georgopoulos A, Skarlatos D (2020) Correcting image refraction: Towards accurate aerial image-based bathymetry mapping in shallow waters. Remote Sensing 12(2):322
- Awad M, Khanna R (2015) Support vector regression. In: Efficient Learning Machines, Springer, pp 67–80
- Barbosa IB, Cristani M, Caputo B, Rognhaugen A, Theoharis T (2018) Looking beyond appearances: Synthetic training data for deep cnns in re-identification. Computer Vision and Image Understanding 167:50–62
- Bianco S, Ciocca G, Marelli D (2018) Evaluating the performance of structure from motion pipelines. Journal of Imaging 4(8):98
- Bishop CM (2006) Pattern recognition and machine learning. springer

- Butler J, Lane S, Chandler J, Porfiri E (2002) Through-water close range digital photogrammetry in flume and field environments. *The Photogrammetric Record* 17(99):419–439
- Cao B, Fang Y, Jiang Z, Gao L, Hu H (2019) Shallow water bathymetry from worldview-2 stereo imagery using two-media photogrammetry. *European Journal of Remote Sensing* 52(1):506–521
- Cao B, Deng R, Zhu S (2020) Universal algorithm for water depth refraction correction in through-water stereo remote sensing. *International Journal of Applied Earth Observation and Geoinformation* 91:102108
- Chirayath V, Li A (2019) Next-generation optical sensing technologies for exploring ocean worlds-nasa fluidcam, midar, and nemo-net. *Frontiers in Marine Science* 6:521
- Dietrich JT (2017) Bathymetric structure-from-motion: extracting shallow stream bathymetry from multi-view stereo photogrammetry. *Earth Surface Processes and Landforms* 42(2):355–364
- Fryer JG (1983) Photogrammetry through shallow water. *Australian Journal of Geodesy, Photogrammetry and Surveying* 38:25–38
- Georgopoulos A, Agrafiotis P (2012) Documentation of a submerged monument using improved two media techniques. In: 2012 18th International Conference on Virtual Systems and Multimedia, IEEE, pp 173–180
- Hinterstoisser S, Lepetit V, Wohlhart P, Konolige K (2018) On pre-trained image features and synthetic images for deep learning. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 0–0
- IHO (2019) Zones of confidence
- IHO (2020) Iho s-44 edition 6.0.0
- Kahmen O, Rofallski R, Conen N, Luhmann T (2019) On scale definition within calibration of multi-camera systems in multimedia photogrammetry. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*
- Kahmen O, Rofallski R, Luhmann T (2020) Impact of stereo camera calibration to object accuracy in multimedia photogrammetry. *Remote Sensing* 12(12):2057
- Li C, Zhang J, Guo B (2011) New method of landslide monitoring based on close-range photogrammetry. *Jisuanji Gongcheng yu Yingyong(Computer Engineering and Applications)* 47(3):6–8
- Luhmann T (2016) Learning photogrammetry with interactive software tool phox. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 41:39
- Maas HG (2015) On the accuracy potential in underwater/multimedia photogrammetry. *Sensors* 15(8):18140–18152
- Mandlbürger G (2018) A case study on through-water dense image matching. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2:659–666*, DOI 10.5194/isprs-archives-XLII-2-659-2018
- Mandlbürger G (2019) Through-water dense image matching for shallow water bathymetry. *Photogrammetric Engineering & Remote Sensing* 85(6):445–455
- Mandlbürger G, Pfennigbauer M, Schwarz R, Flöry S, Nussbaumer L (2020) Concept and performance evaluation of a novel uav-borne topo-bathymetric lidar sensor. *Remote Sensing* 12(6):986
- Menna F, Nocerino E, Remondino F (2018) Photogrammetric modelling of submerged structures: influence of underwater environment and lens ports on three-dimensional (3d) measurements. *Latest developments in reality-based 3D surveying and modelling* MDPI Basel, Switzerland pp 279–303
- Mulsoy C (2010) A flexible multi-media bundle approach. *Int Arch Photogramm Remote Sens Spat Inf Sci* 38:472–477
- Mulsoy C, Mandlbürger G, Maas HG (2020) Comparison of subaquatic digital elevation models from airborne laser scanning and imagery. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences V-2-2020:671–677*, DOI 10.5194/isprs-annals-V-2-2020-671-2020
- Nesbit PR, Hugenholtz CH (2019) Enhancing uav-sfm 3d model accuracy in high-relief landscapes by incorporating oblique images. *Remote Sensing* 11(3):239
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830
- Peng X, Sun B, Ali K, Saenko K (2015) Learning deep object detectors from 3d models. In: Proceedings of the IEEE International Conference on Computer Vision, pp 1278–1286
- Richardson E, Sela M, Kimmel R (2016) 3d face reconstruction by learning from synthetic data. In: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, pp 460–469
- Ros G, Sellart L, Materzynska J, Vazquez D, Lopez AM (2016) The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3234–3243

- Saleh FS, Aliakbarian MS, Salzmann M, Petersson L, Alvarez JM (2018) Effective use of synthetic data for urban scene semantic segmentation. In: European Conference on Computer Vision, Springer, pp 86–103
- Sankaranarayanan S, Balaji Y, Jain A, Nam Lim S, Chellappa R (2018) Learning from synthetic data: Addressing domain shift for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 3752–3761
- Shan J (1994) Relative orientation for two-media photogrammetry. *The Photogrammetric Record* 14(84):993–999
- Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W, Webb R (2017) Learning from simulated and unsupervised images through adversarial training. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2107–2116
- Skarlatos D, Agrafiotis P (2018) A novel iterative water refraction correction algorithm for use in structure from motion photogrammetric pipeline. *Journal of Marine Science and Engineering* 6(3):77
- Skarlatos D, Georgopoulos A (2006) The method of two overlapping orthoimages for checking the produced dtm. Tech. rep.
- Smith MW, Vericat D (2015) From experimental plots to experimental landscapes: topography, erosion and deposition in sub-humid badlands from structure-from-motion photogrammetry. *Earth Surface Processes and Landforms* 40(12):1656–1671
- Sun Y, Sun H, Yan L, Fan S, Chen R (2016) Rba: Reduced bundle adjustment for oblique aerial photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing* 121:128–142
- Vapnik VN, Chervonenkis AY (2015) On the uniform convergence of relative frequencies of events to their probabilities. In: *Measures of complexity*, Springer, pp 11–30
- Wang Z (1990) *Principles of photogrammetry:(with remote sensing)*. Press of Wuhan Technical University of Surveying and Mapping
- Wimmer M (2016) Comparison of active and passive optical methods for mapping river bathymetry. TU Wien
- Woodget A, Carbonneau P, Visser F, Maddock IP (2015) Quantifying submerged fluvial topography using hyperspatial resolution uas imagery and structure from motion photogrammetry. *Earth Surface Processes and Landforms* 40(1):47–64