A Random Forest Model Based on Lidar and Field Measurements for Parameterizing Surface Roughness in Coastal Modeling

Stephen C. Medeiros, Scott C. Hagen, and John F. Weishampel

***Abstract*—A novel technique for parameterizing surface roughness in coastal inundation models using airborne laser scanning (lidar) data is presented. Two important parameters to coastal overland flow dynamics, Manning's n (bottom friction) and effective aerodynamic roughness length (wind speed reduction), are computed based on a random forest (RM) regression model trained using field measurements from 24 sites in Florida fused with georegistered lidar point cloud data. The lidar point cloud for each test site is separated into ground and nonground classes and the z-dimensional (height or elevation) variance from the least squares regression plane is computed, along with the height of the nonground regression plane. These statistics serve as the predictor variables in the parameterization model. The model is then tested using a bootstrap subsampling procedure consisting of removal without replacement of one record and using the surviving records to train the model and predict the surface roughness parameter of the removed record. When compared with the industry standard technique of assigning surface roughness parameters based on published land use/land cover type, the RM regression models reduce the parameterization error by 93% (0.086–0.006) and 53% (1.299–0.610 m) for Manning's n and effective aerodynamic roughness length, respectively. These improvements will improve water level and velocity predictions in coastal models.**

***Index Terms*—Aerodynamic roughness, land cover, lidar, Manning's n, random forest (RM).**

I. Introduction

TIDAL AND hurricane storm surge modeling provide resource and emergency managers with actionable intelligence that allows them to protect people, property, and the environment. These elements are sensitive to changing conditions on a variety of time scales and, in this context, are almost always located in the nearshore and upland areas. Therefore, tidal and/or storm surge models must have sufficient resolution and be parameterized correctly to compute accurate water levels and velocities. This paper focuses on the latter requirement, in particular, the parameterization of surface roughness.

The framework for this investigation is based on two-dimensional (2-D), depth integrated, long-wave shallow water flow models, such as the advanced circulation model for oceanic, coastal, and estuarine waters (ADCIRC), [1], [2] and TELEMAC-2D [3], [4]. Recent studies using these models in the context of simulating astronomic tides and hurricane storm surge have been conducted by Coughlan *et al.* [5], Westerink *et al.* [6], Jones and Davies [7], Jones *et al.* [8], Bunya *et al.* [9], and Dietrich *et al.* [10]. In terms of developing models of this type, surface roughness is the most important input parameter for inundation behavior after topography [11] as it influences both wind velocity and overland flow [12]. The surface roughness parameters most often used in contemporary tidal and storm surge modeling are Manning's n (bottom friction), surface canopy closure (reduction or elimination of vertical wind effects), and effective aerodynamic roughness length (also known as z_0, reduction of horizontal wind effects). The current methodology for specifying surface roughness parameters across large scale models, especially in the United States, relies on published land use/land cover (LULC) data such as coastal change analysis project (C-CAP), [13] and the national land cover dataset [14], [15]. Surface roughness characteristics are then derived from these LULC maps based on established "look-up" tables for each land cover type and roughness parameter [6], [9]. This method is useful because it is easily automated, straightforward to apply, and scientifically defensible; these are important attributes as model scope and resolution increase. However, Medeiros *et al.* determined that this method is insufficient due to the variability of surface roughness within each LULC class, misclassification errors within the LULC data, and errors arising from parameterizing a continuous variable (roughness) using discrete look-up tables [16]. Therefore, an improved method for computing these parameters, especially bottom friction coefficient and effective aerodynamic roughness length, is warranted.

Past research on the translation of field conditions into numerical model parameters is extensive. While they have made excellent contributions to our understanding of roughness within natural flow fields, microscale studies investigating the drag forces and flow resistance in and around individual roughness elements, or very small patches of roughness elements, such as Stephan and Gutknecht [17] and Wilson *et al.* [18], are

Manuscript received July 22, 2014; revised October 31, 2014; accepted March 23, 2015. Date of publication April 19, 2015; date of current version May 26, 2015. This work was supported in part by the U.S. National Aeronautics and Space Administration (NASA) under Grant NNX09AT44G, in part by the U.S. National Oceanic and Atmospheric Administration (NOAA) under Grant NA10NOS4780146, in part by BakerAECOM LLC, and in part by the Louisiana Sea Grant Laborde Chair Endowment.

S. C. Medeiros is with the Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL 32816 USA (e-mail: Stephen.Medeiros@ucf.edu).

S. C. Hagen is with the Department of Civil and Environmental Engineering, Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803 USA (e-mail: shagen@lsu.edu).

J. F. Weishampel is with the Department of Biology, University of Central Florida, Orlando, FL 32816 USA (e-mail: John.Weishampel@ucf.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2015.2419817

1939-1404 © 2015 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

not considered here in order to focus on regional scale parameterization techniques, mainly those that rely on remotely sensed data. In particular, studies that investigated methods for using remotely sensed data to describe above-ground vegetation are especially relevant.

Remotely sensed data, particularly those acquired via satellite, have proved to be effective in describing the properties of the ground surface and vegetation. In a hybrid approach, Straatsma and Baptist [19] fused lidar with multispectral data to enhance the parameterization of spatially distributed Chézy bottom friction coefficients (an approximate analog for Manning's n, the parameter used herein) in a floodplain. Three-dimensional (3-D) lidar point cloud data have been shown to contain abundant descriptive information about the landscape [20]. In fact, Hyyppä and Hyyppä [21] demonstrated that lidar point cloud data outperformed various optical remotely sensed data (e.g., SPOT and Landsat Thematic Mapper) for extracting measureable attributes from vegetation [22], [23]. Therefore, 3-D lidar point cloud data will be used in the parameterization model proposed here. To further constrain the scope, large footprint or full waveform lidar studies, such as Drake *et al.* [24] and Hollaus *et al.* [25], are not considered because small footprint lidar is more prevalent in tidal and storm surge modeling due to its ability to describe topography. Since lidar data have been integrated into nearly every modern flood mapping study, it is advantageous to further exploit those data.

It is appropriate to note that the term airborne laser scanning, or ALS, is synonymous with both laser altimetry and lidar [26]. While the extraction of object surfaces is one of the most beneficial uses of the 3-D lidar point cloud [27], in terms of computing hydrodynamic model roughness parameters, measureable vegetation characteristics such as height and frontal area (area occupied by vegetation as opposed to empty space when looking parallel to the ground) are essential [28]. In fact, much of the earliest work in this field focused on establishing that lidar could be used to measure vegetation and forest characteristics accurately and efficiently [29]. Nelson *et al.* [30] presented one of the earliest studies using airborne laser scanning (ALS) to determine vegetation properties, in particular, the characteristics of a forest canopy. That study was able to establish a relationship between canopy closure (a relevant surface roughness parameter) and the penetration capability of the airborne laser. Menenti and Ritchie [31] also used ALS to estimate a surface roughness parameter. In that study, an empirical relationship between the mean vegetation height and standard deviation of vegetation height was able to estimate the effective roughness length z_0 at the watershed scale. De Vries *et al.* [32] applied a similar and slightly more advanced technique to compute z_0 based on the relationship among obstruction height, frontal area, and planimetric area (area of observation plot when looking from down from above) developed by Lettau [33]. The standard deviation of lidar elevations was also used by Davenport *et al.* [34] and Hopkinson *et al.* [35] to estimate measured vegetation height; this methodology was adapted to parameterize roughness for river flood modeling by Cobby *et al.* [36]. Weltz *et al.* [37] used ALS not only to approximate vegetation heights, but also to distinguish among different plant communities. Ritchie [12] further explored the application of this technology to hydrologic studies, including the measurement of surface roughness. In contrast, Straatsma and Middelkoop [38] present some shortcomings of lidar data in terms of computing hydrodynamic roughness including the lack of stem stiffness information and dependence of the empirical relationships on land cover. These shortcomings are important to note because this paper intends to take a step toward disconnecting the parameterization of surface roughness from categorical LULC types as well as plant species.

To be practically applicable, methods for computing surface roughness parameters for regional scale coastal hydrodynamic models must also be automatable. Two primary criteria must be satisfied in order to meet this requirement. First, it must be relatively straightforward to translate the mathematics behind the method into computer code. This motivates the researcher to use the simplest possible mathematical model to describe the empirical relationships between the lidar point cloud and the ground truth data [39]. Second, the source data required by the method must be available over the applicable areas of the model domain. When these source data are remotely sensed, it should provide adequate spatial coverage of the domain as well as temporal relevance.

This paper presents a method that uses classified 3-D lidar point cloud data to enhance the currently employed method for parameterizing surface roughness. While comparing the vegetation characteristics to lidar data at the tile or plot level (i.e., on a defined patch of terrain) has been done in the past (see Kato *et al.* [40] for a listing of previous studies), the research presented here applies a roughness scheme proved in geologic lidar-based roughness assessments [41], [42] to a setting where vegetation and terrain are the primary contributors to surface roughness. The technique presented here utilizes least squares regression planes to develop roughness values for the terrain (ground points) and the overlying vegetation (non-ground points). These roughness values are integrated with the LULC-based parameterizations along with the field measured surface roughness parameters of 24 field test sites in Florida from Medeiros *et al.* [16] to derive empirical relationships for Manning's n and effective roughness length through multiple linear regression (MLR) and random forest (RF) regression approaches. Note that surface canopy closure is not examined here because it is an already mature application from a remote sensing standpoint (due in large part to its use in a wide variety of scientific areas such as forestry) and its incorporation as a surface roughness parameter in coastal models is, at this time, understudied.

II. Methods

The general methodology is as follows. First, we tabulated the test sites with which lidar point cloud data, LULC classifications, and surface roughness values computed based on field measurements were available. The latter of these data represent the "true" surface roughness parameters associated with each field test site and serve as the response variables in the training/test protocol described below. Second, we preprocessed the point cloud data for efficient data handling and

Fig. 1. Location of field measurement sites.

computed statistics including the variance between the points and the least squares regression planes. These data, explained in more detail below, serve as the predictor variables. Then, due to the limited amount of data, a bootstrapping algorithm is employed, where the data from all but one test site is used to train the parameterization model, which is then used to predict the remaining value. This is repeated for all 24 field test sites. MLR and RF [43] approaches were used to determine the empirical relationship between the measurable characteristics of the plot and the surface roughness parameters.

A. Data Collection

1) Test Sites: This study uses the lidar data and field measured surface roughness terms associated with 24 test sites in Florida. These sites are located in Volusia, Lake, and Franklin counties and are all encompassed within land cover classes that are prevalent in the coastal, storm surge prone regions of Florida. A map showing the location of the test sites is included as Fig. 1 and a summary of the test sites along with their land cover classifications is presented in Table I. A detailed description of the field measurement process and surface roughness calculations are presented by Medeiros *et al.* [16]; only relevant details are reiterated here.

The sites are rectangular, measuring 30 m in the east–west direction and 15 m in the north–south, following Arcement and Schneider [44]. Some of these sites (e.g., those in Lake County) are inland and not at risk of inundation by coastal surges, but are included because their roughness conditions are representative of coastal regions that are surge prone. The number of lidar points on each site is shown in Table I.

2) Lidar Point Cloud: All lidar data were received in LAS format, classified by the data acquisition vendor. The approach used in this study relies only on the 3-D (xyz) position and classification of the point as ground or nonground. All points classified as open water were deleted from the dataset. For the purposes of this study, all horizontal positions were projected into UTM coordinates (meters) and elevations transformed to the North American Vertical Datum of 1988 (NAVD88) and converted to meters.

The data from Franklin County were acquired for the Florida Division of Emergency Management (FDEM) in the summer of 2007 by Aero-Metric, Inc., and Terrapoint USA using Optech 3100 EA sensors. The horizontal accuracy is stated as 1 m at the 95% confidence level. The vertical accuracy is stated as 0.53 feet (0.16 m) in open terrain and 0.61 feet (0.19 m) in all land cover categories, both at the 95% confidence interval. Positions were delivered in the State Plane Coordinate System, Florida North Zone (feet) referenced to the North American Datum of 1983—High Accuracy Reference Network (NAD83 HARN), Geodetic Reference System 80 ellipsoid (GRS 80). Elevations were delivered in feet referenced to the NAVD88.

The data from Volusia County were acquired for the Volusia County Public Works Department in March of 2006 by Woolpert, Inc., using a Leica ALS50 sensor. No accuracies were reported in the metadata. Positions were delivered in the State Plane Coordinate System, Florida East Zone (feet) referenced to NAD83 HARN, GRS 80. Elevations were delivered in feet referenced to NAVD88.

The data from Lake County were acquired for the Lake County Board of County Commissioners, Department of Information Technology, GIS Division in 2007 by Kucera International Inc., using a Leica ALS50 sensor. The horizontal accuracy is stated as 1 m at the 95% confidence level. The vertical root-mean-square error (RMSE) is stated as 0.205 feet (0.062 m) that translates to a vertical accuracy of 0.402 feet (0.122 m) at the 95% confidence interval. Positions were delivered in the State Plane Coordinate System, Florida North Zone (feet) referenced to the North American Datum of 1983—High Accuracy Reference Network (NAD83 HARN), Geodetic Reference System 80 ellipsoid (GRS 80). Elevations were delivered in feet referenced to the North American Vertical Datum of 1988 (NAVD88).

B. Parameterization Model Development

In order to adequately explain the variability of the surface roughness based on lidar point cloud data, two distinct statistical frameworks were considered: multiple regression and RM. Using statistics derived from the xyz position and classification of the lidar points, predictive parameterization models were constructed as described below.

1) Lidar Point Cloud Statistics: Since it is not our intention to run either of the parameterization models in an unsupervised manner, we have selected specific descriptive statistics that have an intuitive physical connection to surface roughness. In particular, we employ regression planes and the associated square root of variances as an indicator of "roughness."

In order to mimic the development of surface roughness parameters, we divide the lidar point cloud into two groups based on their point classification: ground and nonground. Since both the terrain itself along with the obstacles lying on it contribute to surface roughness, this protocol is warranted. We then construct individual regression planes for ground and

TABLE I
FIELD MEASUREMENT SITE LAND COVER TYPES AND LIDAR METADATA

Site	2001 NLCD classification	Total points	Total point density (pts/m^2)	Ground points	Ground point density (pts/m^2)	Nonground points	Nonground point density (pts/m^2)
1	24 — Developed, high intensity (67%) 31 — Barren land (33%)	2271	5.05	560	1.24	1711	3.80
2	31 — Barren land	1567	3.48	646	1.44	921	2.05
3	52 — Shrub/scrub (58%) 71 — Grassland/herbaceous (42%)	1663	3.70	270	0.60	1393	3.10
4	24 — Developed, high intensity	1573	3.50	464	1.03	1109	2.46
5	90 — Woody wetlands	2486	5.52	266	0.59	2220	4.93
6	31 — Barren land	1540	3.42	528	1.17	1012	2.25
7	52 — Shrub/scrub (74%) 90 — Woody wetlands (26%)	2368	5.26	363	0.81	2005	4.46
8	31 — Barren land	1793	3.98	385	0.86	1408	3.13
9	31 — Barren land	1520	3.38	711	1.58	809	1.80
10	95 — Emergent herbaceous wetlands	2193	4.87	306	0.68	1887	4.19
11	90 — Woody wetlands	1306	2.90	292	0.65	1014	2.25
12	90 — Woody wetlands	697	1.55	15	0.03	682	1.52
13	90 — Woody wetlands	2059	4.58	59	0.13	2000	4.44
14	90 — Woody wetlands	556	1.24	154	0.34	402	0.89
15	21 — Developed, open space	342	0.76	184	0.41	158	0.35
16	21 — Developed, open space	358	0.80	72	0.16	286	0.64
17	90 — Woody wetlands (92%) 95 — Emergent herbaceous wetlands (8%)	335	0.74	73	0.16	262	0.58
18	52 — Shrub/scrub	1971	4.38	569	1.26	1402	3.12
19	52 — Shrub/scrub (70%) 42 — Evergreen forest (30%)	2194	4.88	143	0.32	2051	4.56
20	52 — Shrub/scrub	578	1.28	168	0.37	410	0.91
21	81 — Pasture hay (93%) 42 — Evergreen forest (7%)	652	1.45	317	0.70	335	0.74
22	52 — Shrub/scrub	648	1.44	236	0.52	412	0.92
23	71 — Grassland/herbaceous (61%) 52 — Shrub/scrub (39%)	2439	5.42	401	0.89	2038	4.53
24	42 — Evergreen forest	2366	5.26	414	0.92	1952	4.34

nonground points separately as shown in Fig. 2 and (1) as follows:

$$\begin{bmatrix} \sum 1 & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \\ \sum y_i & \sum x_i y_i & \sum y_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \sum z_i \\ \sum x_i z_i \\ \sum y_i z_i \end{bmatrix} \tag{1}$$

where x_i, y_i, and z_i are the coordinates of the lidar points and β_0, β_1, and β_2 are the coefficients of the regression plane equation such that

$$z = \beta_0 + \beta_1 x + \beta_2 y. \tag{2}$$

The relevant statistic is the square root of the variance (for each class, ground, and nonground points), computed as follows:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [z_i - (\beta_0 + \beta_1 x_i + \beta_2 y_i)]} \tag{3}$$

where n is the number of lidar points in each class. It is important to note that the regression planes are computed based on raw elevations for the ground points and detrended heights for the nonground points. As a result, we have two quantitative measures of surface roughness derived from the lidar data, i.e., square root of ground point elevation variance (σ_g) and square root of nonground point height variance (σ_{ng}).

Another important contributor to surface roughness is their overall height. This is particularly true for effective aerodynamic roughness length, as the mean height of the obstacles is an explicit term in the calculation [33]. The distance between the center of the ground and nonground regression planes is used as an indicator for this contribution. First, we determine the center point of the ground plane by computing the mean x and y coordinates, then using the ground regression plane equation to compute the z coordinate

$$x_g = \frac{1}{n_g} \sum_{i=1}^{n_g} x_i \tag{4}$$

$$y_g = \frac{1}{n_g} \sum_{i=1}^{n_g} y_i \tag{5}$$

$$z_g = \beta_{0g} + \beta_{1g} x_g + \beta_{2g} y_g \tag{6}$$

where x_g, y_g, and z_g are the coordinates of the center point of the ground point regression plane; n_g is the number of ground points in the cloud; and β_{0g}, β_{1g}, and β_{2g} are the coefficients of

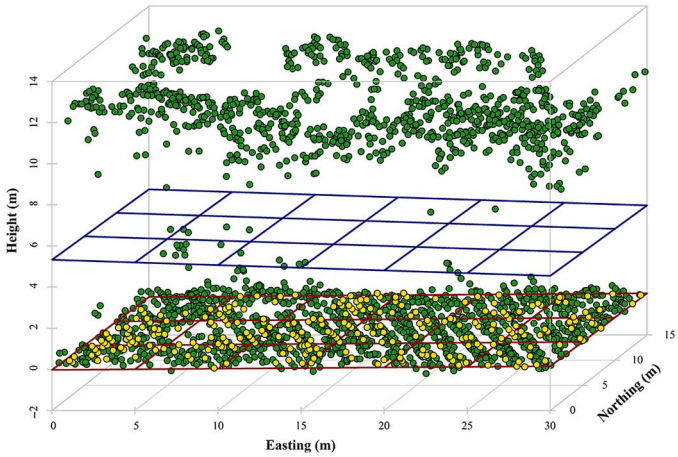


Fig. 2. Lidar point cloud and regression planes for a typical field measurement site.

the ground point regression plane equation. Next, we compute the distance from that point to the nonground regression plane

$$H_{ng} = \frac{|\beta_{1ng}x_g + \beta_{2ng}y_g - z_g + \beta_{0ng}|}{\sqrt{\beta_{1ng}^2 + \beta_{2ng}^2 + (-1)^2}} \tag{7}$$

where H_{ng} is the distance between the ground and nonground regression planes and β_{0ng}, β_{1ng}, and β_{2ng} are the coefficients of the nonground point regression plane equation.

Lastly, in order to preserve the scale of the surface roughness parameters and ensure that the predicted values are within realistic ranges, the parameters derived from LULC will also be included as predictor variables. LULC classes for the test sites are derived from the 2001 National Land Cover Dataset (NLCD) [15] and the associated surface roughness parameters are taken from Bunya *et al.* [9]. This paper is an application of hurricane storm surge modeling that utilizes the subject of surface roughness parameters and is representative of the state-of-the-art. Although it uses a parameterization scheme based on the 1992 NLCD [14], the surface roughness parameters associated with each class are similar, and throughout the literature on this topic they are only modified in select cases where justified by local conditions.

C. Surface Roughness Parameter Prediction

Two frameworks were considered for predicting surface roughness parameters based on lidar point cloud statistics: MLR and RM. Due to the limited amount of data from which to build a predictive model, we employed a recursive bootstrap subsampling procedure. In each iteration, a single data record was removed from the training set and used as the test value. The model was trained using the remaining 23 data records and used to predict the values associated with the removed record. This was repeated a total of 24 times in order to remove each record once. The performance of the model is assessed by aggregating the prediction error achieved during each iteration.

The MLR and RF models were implemented using the statistical programming software R version 3.0.2 [45]. The construction of the MLR model was straightforward using the three variables derived above as predictors (independent variables) and the surface roughness parameter as the response (dependent variable). The RF model used a freely available R package *randomForest* [46]. The only parameter changed from the default when constructing the RF model was the number of trees, which was set to 401 in order to ensure that enough trees were generated to produce a rigorous and thorough regression. No sensitivity tests were conducted to determine the optimal number of trees and the random number seed was set manually at the start of each run to ensure repeatability of results.

TABLE II
STATISTICS DERIVED FROM LIDAR POINT CLOUD AT EACH SITE

Site	σ_g (m)	σ_{ng} (m)	H_{ng} (m)
1	0.049	0.164	0.103
2	0.082	0.068	0.018
3	0.166	2.753	0.943
4	0.033	0.069	0.049
5	0.095	5.338	4.820
6	0.287	0.106	0.013
7	0.093	5.118	4.331
8	0.086	4.223	2.375
9	0.127	0.046	0.052
10	0.067	7.129	5.163
11	0.028	0.755	0.838
12	0.082	6.913	6.166
13	0.019	2.258	2.622
14	0.023	5.910	8.329
15	0.017	0.960	0.430
16	0.152	1.763	1.903
17	0.129	1.959	2.451
18	0.078	2.677	0.892
19	0.169	4.081	3.684
20	0.053	0.446	0.460
21	0.060	3.013	1.895
22	0.170	3.823	1.626
23	0.036	3.389	2.614
24	0.064	2.599	1.176

III. RESULTS

The results of this study demonstrate that the lidar point cloud produces surface roughness parameters that are much closer to measured values than those derived from LULC data.

A. Generation of Lidar Statistics

As stated previously, the specific lidar statistics that serve as predictor variables for the surface roughness parameters are: square root of ground point elevation variance (σ_g), square root of nonground point height variance (σ_{ng}), and height to nonground regression plane (H_{ng}). The values of these variables for the test sites make up the training dataset and are shown in Table II.

Using these predictor variables in the bootstrap sampling procedure, the MLR and RF models both produced surface roughness parameters that are more accurate than LULC, with RF being the most accurate. These results are shown in Figs. 3 and 4.

Both MLR and RF techniques significantly outperform the LULC lookup technique and produce errors that are an order of magnitude smaller. In both MLR and RF cases, the usage of the LULC derived parameter as a predictor variable resulted in

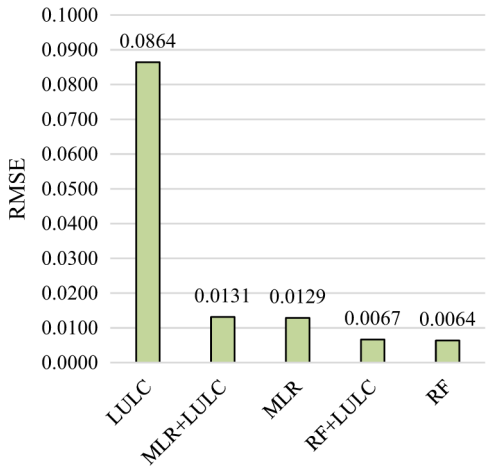


Fig. 3. Manning's n parameterization errors.

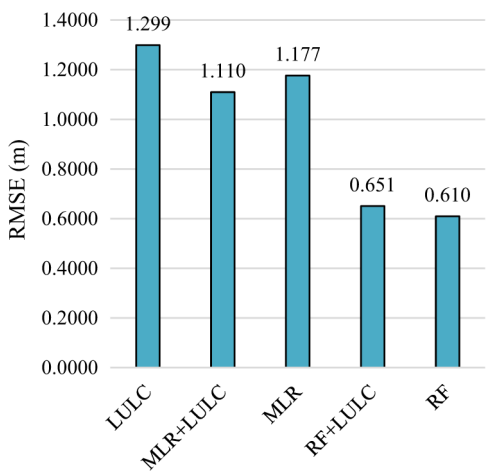


Fig. 4. Effective aerodynamic roughness length parameterization errors.

slightly increased error. This is further reinforced by the residuals shown in Fig. 5. In almost every case, both MLR and RF techniques have smaller residuals than LULC. This figure also highlights the very slight improvement achieved by omitting the LULC derived parameters from the MLR and RF techniques.

The reduction in error for the z_0 parameter is not as severe as for Manning's n, although the RF technique still achieves the greatest reduction in error. For this parameter, inclusion of the values derived from LULC data caused an increase in the RMSE for the RF technique and a decrease for the MLR technique. The less significant reduction in error for this parameter in comparison to that for Manning's n is further illustrated in Fig. 5 (lower). It is evident that while the MLR and RF techniques achieve a lower overall RMSE, there are many test sites for which the LULC method results in a lower residual.

IV. Discussion

As shown in the results, deriving Manning's n and effective roughness length from lidar data represents a more locally accurate technique. The increase in accuracy is attributed to the ability of the lidar point cloud, and more technically, the physically connected statistics we have derived from it, to describe the actual *in situ* configuration and density of the terrain and obstructions. In this context, "locally accurate" means that the parameter at each specific location will be closer to its true, or field measured value. By extension, the global parameterization will also be more accurate. However, the significance of the importance of local accuracy cannot be overstated.

With the LULC technique, each pixel of a particular LULC class is considered to be identical and an average representation of the surface roughness of that LULC type. Inherent in this technique is the acceptance of the fact that the continuity and resolution of the parameter is coarsened. Stated another way, by using the LULC technique, a modeler exchanges maximized local parameter accuracy for a global parameterization accuracy that is acceptable when averaged over the entire model domain. In hurricane storm surge modeling, the domain is typically of a geographic scope in which millions of computation points are to be parameterized with surface roughness. In this case, the global mean parameterization error of the LULC technique over the domain would converge to a minimum. This is a direct consequence of assigning the surface roughness parameters that represent the "average" conditions to each LULC class; with a large enough dataset, the "average" conditions emerge and present the illusion of parameterization accuracy.

If a study were able to field measure a sufficiently large sample of test sites, the error generated by the LULC parameterization technique would converge to a minimum as the number of sites increased. However, a technique that accounts for the local structure of the terrain, such as those presented herein, would generate an error distribution that was essentially flat. This is one of the more subtle points evident from careful examination of Fig. 5. The MLR and RF techniques produce a residual graph that is flatter than that of the LULC technique for both the Manning's n and z_0 parameter, although this concept is far more obvious in the case of Manning's n. The lack of flatness in Fig. 5 (lower) is primarily driven by the large residuals (for all methods) at site 16. This site suffers from a particularly severe misclassification in the LULC data. As shown in Table I, site 16 is classified as developed, open space. Fig. 6(a) shows an aerial photograph of the site and it is apparent that the site is heavily wooded. This is confirmed in the ground level site photograph shown in Fig. 6(b). The result is a significant under prediction of z_0 parameter at this site for the LULC method, which the parameterization models presented herein are able to improve by 56.5%.

This leads to the fundamental motivation of this study. The overland flow conditions (i.e., hurricane storm surge or river flooding) are only affected by the local roughness conditions. While the overall hydrodynamics are influenced by even a slight perturbation anywhere in the domain, an incoming surge at a section of coastline is not going to experience any drag force from the roughness conditions 100 km away. The mean parameterization error over the domain is not as important as the local error with respect to overland flow conditions. Therefore, the method with the more consistent and smaller error on a site by site basis should be favored over one with a small mean error over the entire domain. The method proposed herein has the benefit of exceeding the LULC technique on both counts.

While the performance of the proposed technique has clearly demonstrable advantages over the LULC technique, there are obvious limitations that should be carefully considered. While

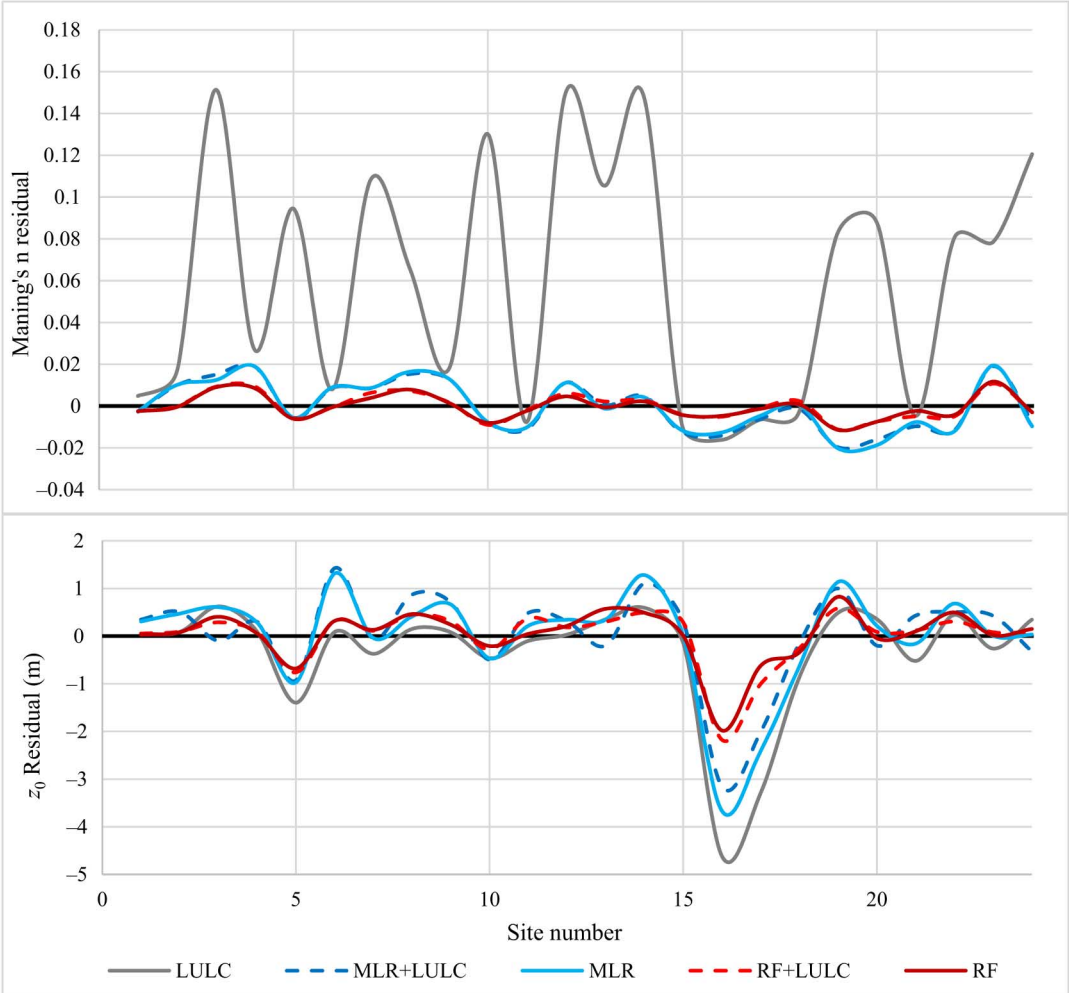


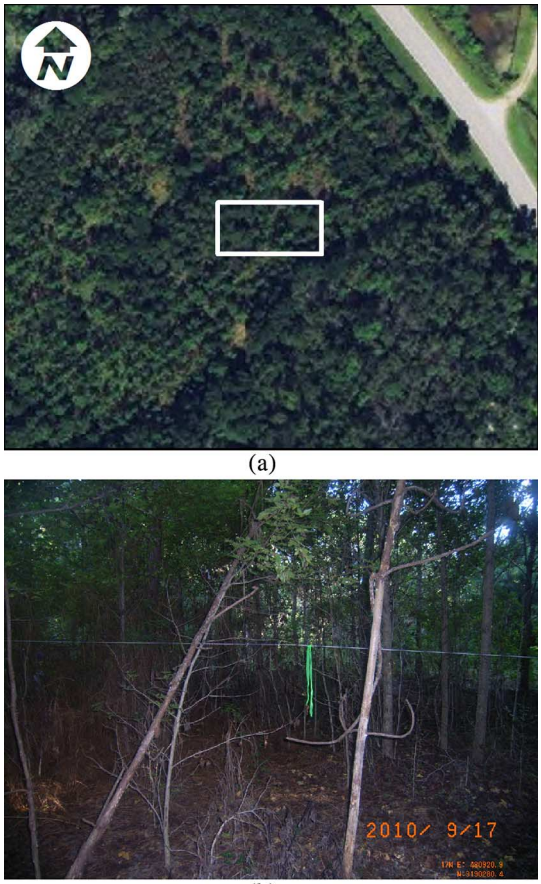
Fig. 5. Residual error for prediction of Manning's n (upper) and effective aerodynamic roughness length (lower).

Medeiros *et al.* [16] emphasize that every effort was made to select field sites that generally captured the range of terrain conditions in hurricane storm surge prone regions of Florida, the computed surface roughness parameters fall within a much narrower range than traditional parameterizations. This is especially true of Manning's n (see Fig. 5). This could be attributed to two factors: 1) error in field measurements and 2) subsequent computations; or errors in the Manning's n look-up tables associated with various LULC classification systems. Random errors in field measurements cannot be accurately quantified in this case because the intensive labor required in the field makes repeat measurements impractical. Systematic errors were minimized by utilizing established protocols along with averaging of multiple independent estimations when required. It is more likely that the Manning's n results falling in a narrow band reinforces the thesis of this study, namely that look-up tables based on LULC data account for almost none of the *in situ* variability of surface roughness. Instead, it is apparent that the LULC technique insists on imposing the classical parameter range.

Another well-established feature of the Manning's n roughness coefficient is that it is dependent on depth. Models that are typically used for simulating hurricane storm surge are deterministic and modify the shear stress due to bottom friction as a quadratic function of depth during run-time. Since the technique proposed here is directed at this specific application, no treatment of the depth dependence is conducted; the Manning's n roughness coefficients computed using this technique serve as the "base" Manning's n that will be adjusted for depth internally within any particular numerical model.

Also of concern with the parameterization of Manning's n and effective roughness length in this context is the temporal discontinuity between the field measurements and the lidar data acquisition. The technique proposed here assumes that the lidar data accurately reflect the field measured conditions. While we can state that there were no major developments in the field site areas in the time between the lidar data acquisition (circa 2007) and the field measurements (circa 2010), it is possible that natural processes may have altered the landscape in the interim. This would lead to errors in the parameterization that would be evident in the residual graphs shown in Fig. 5. Since there are none apparent, we assume that this factor had minimal impact on the training and performance of the parameterization models. In particular, the impact is further minimized by the single site removal bootstrap subsampling procedure employed to train the parameterization models.

Lastly, there is a notable omission in the field measurements that could limit the applicability of the method in production surge studies, namely the lack of urban or developed sites.



(a)

(b)

Fig. 6. (a) Aerial image and (b) ground level photograph of field site 16.

Storm surge is the major driver of damage during tropical storms and for this reason, it is essential to understand the physics of flow through built environments. This is especially important for the effective roughness length parameter. The wind has a significant influence on the behavior of overland storm surge; this is compounded in the highly anisotropic built environment since buildings often block wind in one direction while in another the spaces between them permit and often accelerate free wind flow. The techniques developed in this study are applicable to urban settings with no adjustment to the method other than expanding the training data to include a sufficient number of sites representing the range of urban development conditions in surge prone regions. This is a primary objective of future work.

V. Conclusion

A novel technique for parameterizing surface roughness in coastal models using airborne lidar data is presented. Two important parameters to coastal overland flow dynamics, Manning's n (bottom friction) and effective aerodynamic roughness length (wind speed reduction), were computed based on an RM regression model trained using field measurements from 24 sites in Florida fused with georegistered lidar point cloud data. The lidar point clouds for each test site were separated into ground and nonground classes and the height, or elevation, variance from the least squares regression plane was computed, along with the height of the nonground regression plane. These statistics served as predictor variables in the parameterization model that was tested using a bootstrap subsampling procedure consisting of removal without replacement of one record and using the surviving records to train the model and predict the surface roughness parameter of the removed record. When compared with the industry standard technique of assigning surface roughness parameters based on published land use/land cover type, the RM regression models reduced the parameterization error by 93% (0.086–0.006) and 53% (1.299–0.610 m) for Manning's n and effective aerodynamic roughness length, respectively.

This work represents the first step in what is hopefully a significant advance in coastal modeling. We now have the capability to remotely sense the actual *in situ* structure of the terrain quickly and accurately, which enables coastal modelers to more correctly parameterize surface roughness. It is important that future work clarify the role of surface canopy closure and rigorously connect that parameter to the governing equations for wind stress on the water surface. Models that incorporate these types of data represent substantial increases in local parameterization accuracy, which is key to obtaining accurate model results.

Acknowledgment

The authors would like to thank the students in the CHAMPS and GAMES Labs at UCF for their field work.

References

[1] R. A. Luettich, J. J. Westerink, and N. W. Scheffner, "ADCIRC: An advanced three-dimensional circulation model for shelves, coasts, and estuaries, Report 1: Theory and methodology of ADCIRC-2DDI and ADCIRC-3DL," U.S. Army Corps of Engineers Waterways Experiment Station, Vicksburg, MS, USA, Tech. Rep. DRP-92-6, 1992.

[2] R. A. Luettich and J. J. Westerink. (2006). *ADCIRC: A Paralell Advanced Circulation Model for Oceanic, Coastal and Estuarine Waters* [Online]. Available: http://adcirc.org/documentv46/ADCIRC_title_page.html

[3] J.-C. Galland, N. Goutal, and J. M. Hervouet, "TELEMAC: A new numerical model for solving shallow water equations," *Adv. Water Resour.*, vol. 14, pp. 138–148, 1991.

[4] J. M. Hervouet, "TELEMAC modelling system: An overview," *Hydrol. Processes*, vol. 14, pp. 2209–2210, 2000.

[5] C. Coughlan, C. E. Vincent, T. J. Dolphin, and J. M. Rees, "Effects of tidal stage on the wave climate inshore of a sandbank," *J. Coastal Res.*, vol. SI 50, pp. 751–756, 2007.

[6] J. J. Westerink *et al.*, "A basin- to channel-scale unstructured grid hurricane storm surge model applied to southern Louisiana," *Mon. Weather Rev.*, vol. 136, pp. 833–864, 2008.

[7] C. Le Provost, F. Lyard, J. Molines, M. Genco, and F. Rabilloud, "A hydrodynamic ocean tide model improved by assimilating a satellite altimeter derived data set," *J. Geophys. Res. Oceans*, vol. 103, pp. 5513–5529, 1998.

[8] J. E. Jones, P. Hall, and A. M. Davies, "An inter-comparison of tidal solutions computed with a range of unstructured grid models of the Irish and Celtic Sea regions," *Ocean Dyn.*, vol. 59, pp. 997–1023, 2009.

[9] S. Bunya *et al.*, "A high-resolution coupled riverine flow, tide, wind, wind wave and storm surge model for southern Louisiana and Mississippi. Part I: Model development and validation," *Mon. Weather Rev.*, vol. 138, pp. 345–377, 2010.

[10] J. C. Dietrich *et al.*, "Hurricane Gustav (2008) waves and storm surge: Hindcast, synoptic analysis and validation in southern Louisiana," *Mon. Weather Rev.*, vol. 139, pp. 2488–2522, 2011.

[11] M. Straatsma, "3D float tracking: In situ floodplain roughness estimation," *Hydrol. Processes*, vol. 23, pp. 201–212, 2009.

[12] J. C. Ritchie, "Remote sensing applications to hydrology: Airborne laser altimeters," *Hydrol. Sci. J.*, vol. 41, pp. 625–636, Aug. 1996.

[13] C. S. C. National Oceanic and Atmospheric Administration, *The Coastal Change Analysis Program (C-CAP) Regional Land Cover*. Charleston, SC, USA: NOAA Coastal Services Center, 1995.

[14] J. E. Vogelmann *et al.*, "Completion of the 1990s national land cover data set for the conterminous United States from Landsat thematic mapper data and ancillary data sources," *Photogramm. Eng. Remote Sens.*, vol. 67, pp. 650–652, 2001.
[15] C. Homer *et al.*, "Completion of the 2001 national land cover database for the conterminous United States," *Photogramm. Eng. Remote Sens.*, vol. 73, pp. 337–341, Apr. 2007.
[16] S. C. Medeiros, S. C. Hagen, and J. F. Weishampel, "Comparison of floodplain surface roughness parameters derived from land cover data and field measurements," *J. Hydrol.*, vol. 452–453, pp. 139–149, 2012.
[17] U. Stephan and D. Gutknecht, "Hydraulic resistance of submerged flexible vegetation," *J. Hydrol.*, vol. 269, pp. 27–43, 2002.
[18] C. A. M. E. Wilson, T. Stoesser, P. D. Bates, and A. Batemann Pinzen, "Open channel flow through different forms of submerged flexible vegetation," *J. Hydraul. Eng.*, vol. 129, pp. 847–853, 2003.
[19] M. W. Straatsma and M. J. Baptist, "Floodplain roughness parameterization using airborne laser scaning and spectral remote sensing," *Remote Sens. Environ.*, vol. 112, pp. 1062–1080, 2008.
[20] M. Wang and Y.-H. Tseng, "Automatic segmentation of Lidar data into coplanar point clusters using an octree-based split-and-merge algorithm," *Photogramm. Eng. Remote Sens.*, vol. 76, pp. 407–420, Apr. 2010.
[21] H. Hyyppä and J. Hyyppä, "Comparing the accuracy of laser scanner with other optical remote sensing data sources for stand attribute retrieval," *Photogramm. J. Finland*, vol. 16, pp. 5–15, 1999.
[22] G. T. Raber, J. R. Jensen, S. R. Schill, and K. Schuckman, "Creation of digital terrain models using an adaptive lidar vegetation point removal process," *Photogramm. Eng. Remote Sens.*, vol. 68, pp. 1307–1315, Dec. 2002.
[23] J. T. Fisher *et al.*, "Savanna woody vegetation classification—Now in 3-D," *Appl. Veg. Sci.*, vol. 17, pp. 172–184, 2014.
[24] J. B. Drake *et al.*, "Estimation of tropical forest structural characteristics using large-footprint lidar," *Remote Sens. Environ.*, vol. 79, pp. 305–319, Feb. 2002.
[25] M. Hollaus, C. Aubrecht, B. Höfle, K. Steinnocher, and W. Wagner, "Roughness mapping on various vertical scales based on full-waveform laser scanning data," *Remote Sens.*, vol. 3, pp. 503–523, 2011.
[26] D. W. Coggin, "LiDAR in coastal storm surge modeling: Modeling linear raised features," M.S. Thesis, Univ Central Florida, Orlando, FL, USA, 2008, 140 pp.
[27] G. Vosselman, B. G. H. Gorte, G. Sithole, and T. Rabbani, "Recognising structure in laser scanner point clouds," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 46, pp. 33–38, 2004.
[28] D. C. Mason, D. M. Cobby, M. S. Horritt, and P. D. Bates, "Floodplain friction parameterization in two-dimensional river flood models using vegetation heights derived from airborne scanning laser altimetry," *Hydrol. Processes*, vol. 17, pp. 1711–1732, 2003.
[29] K. Lim, P. Treitz, M. Wulder, B. St-Onge, and M. Flood, "LiDAR remote sensing of forest structure," *Prog. Phys. Geogr.*, vol. 27, pp. 88–106, 2003.
[30] R. Nelson, W. Krabill, and G. Maclean, "Determining forest canopy characteristics using airborne laser data," *Remote Sens. Environ.*, vol. 15, pp. 201–212, 1984.
[31] M. Menenti and J. C. Ritchie, "Estimation of effective aerodynamic roughness of Walnut Gulch watershed with laser altimeter measurements," *Water Resour. Res.*, vol. 30, pp. 1329–1337, 1994.
[32] A. C. De Vries *et al.*, "Effective aerodynamic roughness estimated from airborne laser altimeter measurements of surface features," *Int. J. Remote Sens.*, vol. 24, pp. 1545–1558, 2003.
[33] H. Lettau, "Note on aerodynamic roughness-parameter estimation on the basis of roughness-element description," *J. Appl. Meteorol.*, vol. 8, pp. 828–832, Oct. 1969.
[34] I. J. Davenport *et al.*, "Improving bird population models using airborne remote sensing," *Int. J. Remote Sens.*, vol. 21, pp. 2705–2717, 2000.
[35] C. Hopkinson *et al.*, "Wetland grass to plantation forest—Estimating vegetation height form the standard deviation of lidar frequency distributions," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 36, pp. 288–294, 2004.
[36] D. M. Cobby, D. C. Mason, and I. J. Davenport, "Image processing of airborne scanning laser altimetry data for improved river flood modeling," *ISPRS J. Photogramm. Remote Sens.*, vol. 56, pp. 121–138, 2001.
[37] M. A. Weltz, J. C. Ritchie, and H. D. Fox, "Comparison of laser and field measurements of vegetation height and canopy cover," *Water Resour. Res.*, vol. 30, pp. 1311–1319, May 1994.
[38] M. Straatsma and H. Middelkoop, "Extracting structural characteristics of herbaceous floodplain vegetation under leaf-off conditions using airborne laser scanner data," *Int. J. Remote Sens.*, vol. 28, pp. 2447–2467, Jun. 2007.
[39] K. Kraus and W. Rieger, "Processing of laser scanning data for wooded areas," in *Proc. Photogramm. Week '99*. Heidelberg, Germany, 1999, pp. 221–231.
[40] A. Kato, G. F. Schreuder, D. Calhoun, P. Schiess, and W. Stuetzle, "Digital surface model of tree canopy structure from LiDAR data through implicit surface reconstruction," in presented at the *Amer. Soc. Photogramm. Remote Sens. Annu. Conf. (ASPRS)*, Tampa, FL, USA, 2007.
[41] R. M. Pollyea and J. P. Fairley, "Estimating surface roughness of terrestrial laser scan data using orthogonal distance regression," *Geology*, vol. 39, pp. 623–626, 2011.
[42] R. M. Pollyea and J. P. Fairley, "Experimental evaluation of terrestrial LiDAR-based surface roughness estimates," *Geosphere*, vol. 8, pp. 222–228, 2012.
[43] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[44] G. J. Arcement and V. R. Schneider, *Guide for Selecting Manning's Roughness Coefficients for Natural Channels and Flood Plains*, U.S. Geological Survey, Water-Supply Paper 2339, 1989.
[45] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013.
[46] A. Liaw and M. Weiner, "Classification and regression by randomforest," *R News*, vol. 2, pp. 18–22, 2002.

Stephen C. Medeiros was born in Massachusetts, USA, in 1977. He received the Bachelor's and Master's degrees in civil engineering from Florida Institute of Technology, Melbourne, FL, USA, in 1999 and 2002, and the Ph.D. degree in civil engineering from the University of Central Florida (UCF), Orlando, FL, USA, in 2012.

He is currently a Research Assistant Professor with the Department of Civil, Environmental, and Construction Engineering, UCF, and is also affiliated with the research cluster of excellence on the Coastal Dynamics of Sea Level Rise. His research interests include fusion of remote sensing, field measurements, and computer modeling for natural hazard applications.

Prof. Medeiros is a Registered Professional Engineer in the State of Florida.

Scott C. Hagen was born in Iowa, USA, in 1961. He received the Bachelor's degree in civil engineering from the University of Iowa, Iowa City, IA, USA, in 1993, and the Ph.D. degree in civil engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1998.

He is currently a Professor with the Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, USA, and holds the John P. Laborde Endowed Chair for Sea Grant Research and Technology Transfer. He also has an appointment with the LSU Center for Computation and Technology, Baton Rouge, and is a Fellow of the LSU Coastal Studies Institute, Baton Rouge. He spent the previous +17 years with the University of Central Florida, Orlando, FL, USA, where he established the internationally recognized CHAMPS Laboratory. His research interests include coastal hydroscience and engineering, high performance computational modeling of ocean, coastal, and inland astronomical, and meteorological tides and flows.

Prof. Scott is a Registered Professional Engineer in the State of Florida, and is a Diplomate of both Coastal and Water Resources Engineering. He is a past member of the Board of Governors for the ASCE/Coasts, Oceans, Ports, and Rivers Institute and he was elected a Fellow of the American Society of Civil Engineers in 2014.

John F. Weishampel was born in Pennsylvania, USA, in 1963. He received the Bachelor's degree in biology from Duke University, Durham, NC, USA, in 1985, and the Ph.D. degree in environmental sciences from the University of Virginia, Charlottesville, VA, USA, in 1994.

He is currently a Professor with the Department of Biology, University of Central Florida (UCF), Orlando, FL, USA, and the Director of the Geospatial Analysis and Modeling of Ecological Systems (GAMES) Laboratory. He is also affiliated with the Research Cluster of Excellence on the Coastal Dynamics of Sea Level Rise. His research interests include landscape ecology, i.e., how landscape patterns influence ecological processes and biodiversity with a particular emphasis on remote sensing, field observations, and simulation models.