data

MDPI

Data Descriptor

A High-Accuracy GNSS Dataset of Ground Truth Points Collected within Îles-de-Boucherville National Park, Quebec, Canada

Kathryn Elmer and Margaret Kalacska *

Applied Remote Sensing Lab, Department of Geography, McGill University, Montreal, QC H3A-0B9, Canada; kathryn.elmer@mail.mcgill.ca

* Correspondence: Margaret.kalacska@mcgill.ca

Abstract: A new ground truth dataset generated with high-accuracy Global Navigation Satellite Systems (GNSS) positional data of the invasive reed *Phragmites australis* subsp. *australis* within Îles-de-Boucherville National Park (Quebec, Canada) is described. The park is one of five study sites for the Canadian Airborne Biodiversity Observatory (CABO) and has stands of invasive *P. australis* spread throughout the park. Previously, within the context of CABO, no ground truth data had been collected within the park consolidating the locations of *P. australis*. This dataset was collected to serve as training and validation data for CABO airborne hyperspectral imagery acquired in 2019 to assist with the detection and mapping of *P. australis*. The locations of the ground truth points were found to be accurate within one pixel of the hyperspectral imagery. Overall, 320 ground truth points were collected, representing 158 locations where *P. australis* was present and 162 locations where it was absent. Auxiliary data includes field photographs and digitized field notes that provide context for each point.

Dataset: http://doi.org/10.5281/zenodo.4504922 (accessed on 13 March 2021).

Keywords: GNSS; ground control points; invasive species

Citation: Elmer, K.; Kalacska, M. A High-Accuracy GNSS Dataset of Ground Truth Points Collected within Îles-de-Boucherville National Park, Quebec, Canada. *Data* **2021**, *6*, 32. https://doi.org/10.3390/data6030032

Academic Editor: Richard Ross Shaker

Received: 5 February 2021
Accepted: 12 March 2021
Published: 14 March 2021

1. Summary

Phragmites australis subs. *Australis* (hereafter *Phragmites*) is considered one of the most aggressive invasive plant species in eastern North America. It can be found in all 49 of the mainland states in the United States, as well as throughout the southern portions of six Canadian provinces, in a variety of dry and wet soil habitats (both freshwater and brackish conditions) [1,2]. *Phragmites* is highly aggressive and is considered an indicator of ecosystem disturbance due to its tendency to grow in dense monotypic stands that prevent the growth of other surrounding vegetation, thereby greatly reducing plant biodiversity [3]. These dense monotypic stands can grow up to 3 m tall and have extensive networks of stolons and rhizomes that make removal of the plant difficult, especially once it has spread over a large extent. *Phragmites* is able to survive under a wide range of conditions, although it has demonstrated an affinity for wetlands, especially those that have been enriched with nitrogen from agricultural or other residential sources [4]. Management and eradication of *Phragmites* populations is a difficult and ongoing process. Most removal methods require extensive physical labor and field work, over vast areas. Physical methods of removal must be carried out with great caution to avoid accidental transfer of *Phragmites* or the stimulation of further growth [5,6]. Therefore, it is critical that land managers are able to locate stands of *Phragmites* while they are still relatively young and small in extent so

that they can more easily be eradicated or prevented from spreading using sensible and appropriate methods.

As part of the Canadian Airborne Biodiversity Observatory, airborne hyperspectral imagery (HSI) was acquired with the Compact Airborne Spectrographic Imager-1500 (CASI-1500) over the entire extent of Îles-de-Boucherville National Park in 2019. One of the purposes of the imagery was to identify and map invasive *Phragmites* across the entire park. For this approach, a ground truth dataset is needed to both train target detection algorithms and assess their resulting accuracies. The airborne HSI is geo-corrected with a resampled pixel size of 2 m $\times$ 2 m. The reported positional accuracy of the HSI is an average error in easting and northing of 0.91 m and 0.67 m, respectively (Table 1). However, as described by [7], only 55.5% of the signal from a given pixel originates from materials within its spatial boundaries, while the remaining signal is from the materials within neighboring pixels. For the CASI-1500 HSI, the net point spread function (PSF) extends approximately 3 m in the along-track direction and 2 m in the cross-track direction. Based on the characteristics of the imagery, for this application, ground truth points with less than 2.25 m error in both northing and easting were required. The requirement for precision of the ground truth points had been set to less than one half standard deviation of the net PSF in the cross track (i.e., narrower) direction (16.5 cm). Therefore, ground truth data needed to be collected using a high-precision GNSS (Global Navigation Satellite Systems) instrument, in order to ensure that the points are geolocated correctly within the HSI.

Table 1. The reported mean and standard deviation in positional accuracy of the CASI-1500 HSI acquired over Îles-de-Boucherville National Park in 2019.

	Easting	**Northing**
Mean error (m)	0.91	0.67
Standard deviation (m)	$\pm$0.98	$\pm$0.56

2. Data Description

The dataset available for download is comprised of 320 ground truth data points, as well as digitized field notes and field photographs. The dataset is available as a comma separated value file (.csv) (Figure 1) and as an ESRI shapefile (.shp) with associated metadata. The field notes are available as an Excel workbook (.xlsx) (Figure 2). The field photographs are provided in Portable Network Graphics format (.png) (Figure 3). The ground truth points cover the extent of Îles-de-Boucherville National Park. The positions are reported in the NAD83-CSRS (epoch 2010.01.01) coordinate frame and ellipsoidal height for elevation (see Section 3.2). The user must apply a conversion of the GNSS ellipsoidal height to orthometric height (height above mean sea level) by applying a gravimetric geoid model. The Natural Resources Canada GPS-H tool [8] is recommended for applying the Canadian Geodetic Vertical Datum of 2013 (CGVD2013).

The ground truth dataset contains 15 fields for each of the 320 points collected, and these fields correspond to the fields within the shapefile's attribute table. The field names and a brief description are given in Table 2. The accompanying field notes file contains auxiliary information about the points collected, and consists of 6 fields for each of the 320 points collected. The field names and a brief description are given in Table 3.

FID	longitude	latitude	elevation	collection_start	collection_end	solution_status	antenna_height	sample_count	SD_r	SD_x	SD_y	SD_z	ACC_r	ACC_z
0	−73.46764958	45.59749003	−19.12149649	2019-09-05 16:35:27 UTC	2019-09-05 16:35:38 UTC	FIX	2.065	57	0.0155	0.0149	0.0046	0.011	1.7576	1.9816
1	−73.47089424	45.59712047	−19.94672308	2019-09-05 16:40:03 UTC	2019-09-05 16:40:11 UTC	FLOAT	2.065	39	0.0072	0.0058	0.0041	0.0187	1.7433	1.9967
2	−73.47162333	45.59688045	−19.72333542	2019-09-05 17:03:58 UTC	2019-09-05 17:04:07 UTC	FLOAT	2.065	48	0.0097	0.0087	0.0042	0.0127	1.7476	1.9849
3	−73.47177396	45.59690193	−19.74514808	2019-09-05 17:06:12 UTC	2019-09-05 17:06:22 UTC	FLOAT	2.065	52	0.0089	0.0064	0.0062	0.0374	1.7462	2.0333
4	−73.47188672	45.59541675	−20.14565	2019-09-05 17:31:23 UTC	2019-09-05 17:31:35 UTC	FLOAT	2.065	8	0.0035	0.0031	0.0016	0.0036	1.7369	1.9671
5	−73.47421156	45.59604689	−21.95235	2019-09-05 17:39:03 UTC	2019-09-05 17:39:09 UTC	FLOAT	2.065	2	0.0053	0.0013	0.0052	0.0052	1.74	1.9702
6	−73.47529771	45.5965223	−20.2982625	2019-09-05 17:45:50 UTC	2019-09-05 17:45:59 UTC	FIX	2.065	16	0.0041	0.0028	0.003	0.0064	1.7379	1.9725
7	−73.47528697	45.59595133	−21.25508333	2019-09-05 17:48:40 UTC	2019-09-05 17:48:51 UTC	FLOAT	2.065	6	0.0039	0.002	0.0034	0.0021	1.7376	1.9641
8	−73.47653296	45.59596773	−21.86551351	2019-09-05 17:55:23 UTC	2019-09-05 17:55:33 UTC	FIX	2.065	37	0.0064	0.006	0.0024	0.0063	1.7419	1.9723
9	−73.4766108	45.5959221	−21.99684286	2019-09-05 17:56:50 UTC	2019-09-05 17:56:59 UTC	FIX	2.065	28	0.0063	0.0057	0.0027	0.0027	1.7417	1.9653
10	−73.47670231	45.59625065	−22.338364	2019-09-05 17:59:07 UTC	2019-09-05 17:59:17 UTC	FIX	2.065	50	0.0067	0.0056	0.0037	0.0034	1.7424	1.9667
11	−73.47699087	45.59634457	−21.64103611	2019-09-05 18:02:53 UTC	2019-09-05 18:03:07 UTC	FIX	2.065	72	0.0131	0.01	0.0085	0.3698	1.7535	2.6848
12	−73.47607822	45.59645495	−20.96758889	2019-09-05 18:06:31 UTC	2019-09-05 18:06:45 UTC	FIX	2.065	9	0.0021	0.0017	0.0012	0.0025	1.7344	1.9649
13	−73.47617696	45.59638931	−19.5086	2019-09-05 18:08:10 UTC	2019-09-05 18:08:21 UTC	FLOAT	2.065	5	0.0149	0.0015	0.0148	0.0251	1.7566	2.0092
14	−73.47789349	45.59539676	−20.44686667	2019-09-05 18:17:21 UTC	2019-09-05 18:17:32 UTC	FIX	2.065	54	0.0066	0.004	0.0052	0.004	1.7422	1.9678
15	−73.4792733	45.59491967	−19.79860769	2019-09-05 18:19:56 UTC	2019-09-05 18:20:08 UTC	FIX	2.065	13	0.0059	0.0056	0.0019	0.004	1.741	1.9678
16	−73.47796724	45.59495678	−20.92404706	2019-09-05 18:27:59 UTC	2019-09-05 18:28:09 UTC	FIX	2.065	51	0.0043	0.0037	0.0023	0.005	1.7382	1.9698
17	−73.47746017	45.59510279	−20.73361667	2019-09-05 18:38:21 UTC	2019-09-05 18:38:34 UTC	FIX	2.065	36	0.0163	0.0135	0.0091	0.009	1.759	1.9776
18	−73.47668789	45.59496357	−21.31206923	2019-09-05 18:41:09 UTC	2019-09-05 18:41:22 UTC	FIX	2.065	39	0.0074	0.0053	0.0052	0.0052	1.7436	1.9702
19	−73.47652043	45.59475379	−20.77343095	2019-09-05 18:43:04 UTC	2019-09-05 18:43:12 UTC	FIX	2.065	42	0.011	0.0056	0.0095	0.0072	1.7498	1.9741
20	−73.47642281	45.5945857	−20.61734651	2019-09-05 18:44:00 UTC	2019-09-05 18:44:16 UTC	FIX	2.065	43	0.0068	0.0041	0.0055	0.0052	1.7426	1.9702

Figure 1. A subset of the 320 ground truth data points listed in the .csv file. The fields contain the FID, longitude, latitude, elevation, the date and time of the beginning/end of collection, the rover's solution status, and antenna height above the ground. Additionally, there are fields for the calculated precision (i.e., standard deviation, SD) in the x, y, and z dimensions, as well as the radial standard deviation (SD_r). The approximate horizontal and vertical accuracy at the 95% confidence level is also reported. Antenna height, SD, and approximate accuracy at the 95% confidence level are reported in meters.

GPS Survey Date	FID	Phrag? (Y/N)	Site Characteristics	Photo ID	Other notes
Sept. 5, 2019	0	Y	monostand	FID-0	near parking lot
Sept. 5, 2019	1	Y	large phrag stand, some mixed	FID-1	near parking lot
Sept. 5, 2019	2	N	goldenrod, phalaris	FID-2	open field, trees on edge
Sept. 5, 2019	3	Y	large phrag stand	FID-3	open field, trees on edge
Sept. 5, 2019	4	N	grass field, strips of phalaris	FID-4	lawn w/ rows of reeds (overflow parking lot)
Sept. 5, 2019	5	N	grass, phalaris	FID-5	meadow with reeds and grasses
Sept. 5, 2019	6	N	golden flowers	FID-6	golden flowers, milkweed
Sept. 5, 2019	7	N	dry grasses and reeds	FID-7	meadow with dry reeds and grasses
Sept. 5, 2019	8	Y	border between phrag and phalaris	FID-8	large phrag stand, edge mixed with others
Sept. 5, 2019	9	N	meadow of reeds, grasses, goldenrod	FID-9	grasses, reeds
Sept. 5, 2019	10	Y	phrag mixed with golden	FID-10	phrag stand mixed with goldenrod
Sept. 5, 2019	11	Y	phrag near water	FID-11	near water edge
Sept. 5, 2019	12	N	goldenrod monofield	FID-12	goldenrod field
Sept. 5, 2019	13	N	patchy dirt with sparse vegetation	FID-13	dirt patch
Sept. 5, 2019	14	N	reeds, phalaris	FID-14	large field with reeds, phalaris
Sept. 5, 2019	15	Y	phrag stand	FID-15	near trail
Sept. 5, 2019	16	Y	phrag stand	FID-16	at edge of field

Figure 2. A subset of the digitized field notes for the ground truth dataset. For each point collected, there is auxiliary information detailing the survey date, FID, presence/absence of *Phragmites*, the site characteristics, the corresponding field photograph ID, and any other notes. The full set of digitized field notes is available for download.

Figure 3. A subset of the field photographs taken for each measured location. Each photograph has an identifier that corresponds to a specific point's FID, as well as in the accompanying metadata containing location and copyright information. The full set of photos are available for download.

Table 2. A list and descriptions of the fields in the ground truth dataset.

Field Name	Description
FID	The point's numerical identifier (ObjectID of the ESRI shapefile)
longitude	The measured longitude of the point
latitude	The measured latitude of the point
elevation	The measured elevation of the point (ellipsoidal height)
collection_start	The date and time that collection of the point began
collection_end	The date and time that collection of the point ended
solution_status	The solution status at the time the point was collected
antenna_height	The height of the antenna above the ground (m)
sample_count	The number of measurements that were used in the calculation of the position
SD_r *	The radial standard deviation (m)
SD_x *	The standard deviation in the x dimension (m)
SD_y *	The standard deviation in the y dimension (m)
SD_z *	The standard deviation in the z dimension (m)
ACC_r	The approximate horizontal accuracy at the 95% confidence level (m)
ACC_z	The approximate vertical accuracy at the 95% confidence level (m)

* See Section 3.3 for a discussion on the SD fields as reported by the Reach RS+ unit.

Table 3. A list and descriptions of the fields in the auxiliary field notes file.

Field Name	Description
GPS Survey Date	The date that the points were collected
FID	The point's numerical identifier (ObjectID of the ESRI shapefile)
Phrag? (Y/N)	"Y" for "yes" or "N" for "no", representing whether Phragmites was present at the specified point
Site Characteristics	A brief description of the immediate area surrounding the measured point
Photo ID	The identifier for the corresponding field photograph
Other notes	Any other notes/comments regarding the measured point, surrounding area, or other characteristics

3. Methods

3.1. Study Area

Îles-de-Boucherville National Park is located in the Saint Lawrence River between the Island of Montreal and the municipality of Boucherville in Quebec, Canada (Figure 4). The park covers an area of approximately 8 km^2 and consists of five main islands, the largest of which is Île Grosbois (244 ha) [9]. At the end of the 17th century, the first settlers arrived at the islands and began a history of intensive agricultural activity [10], which is still continued today on two of the islands (Île de la Commune and Île Grosbois). The park is characterized by a continental climate, and is relatively flat with an average elevation of 12 m above sea level. Îles-de-Boucherville National Park contains more than 450 plant species in terrestrial, aquatic, and semi-aquatic ecosystems [11]. The majority of the park is covered by young herbaceous and shrub vegetation following the abandonment of several agricultural fields when the area was designated as a protected area by the Quebec Ministry of Forests, Wildlife and Parks in 1984 [9]. Examples include *Populus deltoids* (eastern cottonwood), *Solidago gigantea* (giant goldenrod), *Solidago altissima* (late goldenrod), and *Phalaris arundinacea* (reed canary grass). Some scattered forested areas are present, with the only mature forest located on Île Grosbois, covering an area of 18 ha [9] and consisting of *Fraxinus pennsylvanica* (green ash), *Tilia Americana* (basswood), *Acer saccharinum* (silver maple), *Ulmus americana* (American elm), *Carya cordiformis* (bitternut hickory), and *Quercus macrocarpa* (bur oak) [11,12]. The invasive common reed *Phragmites australis* subsp. *australis* is found throughout the park, within river channels and on riverbanks, along hiking trails, and in agricultural ditches, as *Phragmites* readily colonizes disturbed ecosystems such as former agricultural fields and farmland [4]. Additionally, the largest stand of invasive

Phragmites in the province of Quebec is located in the Courant Channel along the western edge of the park.

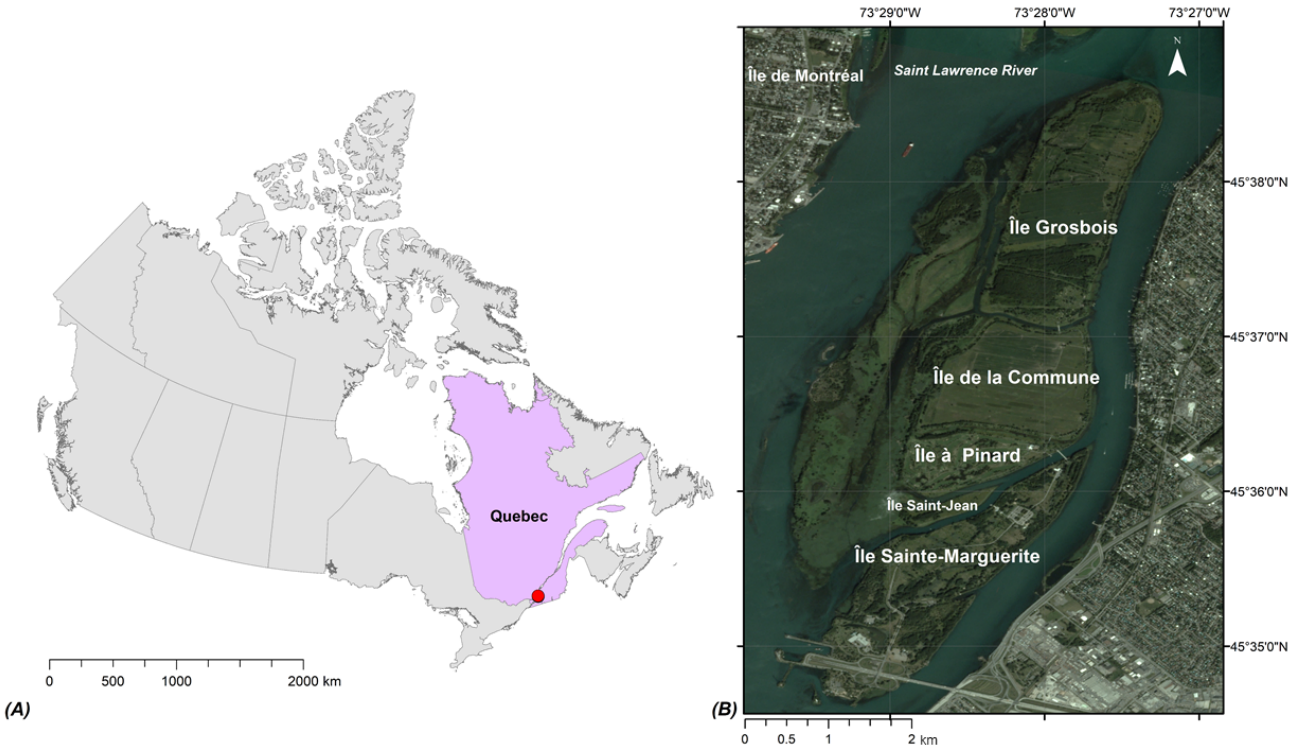

Figure 4. (**A**) The location of Îles-de-Boucherville National Park (red circle) within the Canadian province of Quebec. Projection: Lambert Conformal Conic Projection, NAD83 (EPSG:102002)). (**B**) The five major islands that make up the extent of the park, shown in the St. Lawrence River near the Island of Montreal. Basemap is a PlanetScope Dove, image acquired on September 5, 2019. Projection: NAD83(CSRS)v7, EPSG:8254.

3.2. Ground Truth Data Collection

The dataset is comprised of 320 points collected over 6 days between September 5 and 17, 2019 (Figure 5). A Reach RS+ (EMLID, St. Petersburg, Russia) single-band real-time-kinematic (RTK)-capable GNSS receiver was used. It is capable of receiving signals from several satellite constellations, including GPS, GLONASS, Galileo, QZSS, and Beidou. The Reach RS+ was mounted to the top of a fully extended surveying monopole with an attached bubble level to ensure the system would remain stable during data collection. The antenna height was measured as 2.065 m, which includes the height of the surveying monopole (2 m) and the phase center offset, which is the given distance from the bottom of the system housing to the position of the antenna within the housing (0.065 m). The ground truth data points were collected following an established protocol for the unit [13] and utilized incoming corrections from a SmartNet North America (Norcross, GA, USA) base station, located in Longueuil, Quebec (Station QCLO), with a baseline distance that ranged up to 16 km (depending on where in the park collection was taking place at the time). Corrections were received via NTRIP (Networked Transport of Radio Technical Commission for Maritime Services (RCTM) via Internet Protocol) [14] on a RTCM3-iMAX (individualized master-auxiliary) mount point. The RCTM3-iMAX utilizes real reference base stations to send the network corrections, which provides traceability and consistency for incoming corrections received by the GNSS unit [15]. Both GPS and GLONASS constellations with an update rate of 5 Hz were used. The manufacturer reported approximate accuracy of the Reach RS+ with an NTRIP stream baseline of <10 km with a FIX position is 7 mm + 1 mm/km and 1 m for an NTRIP stream baseline <30 km for FLOAT positions [16]. The necessary wireless hotspot was creating using a Novatel

Wireless MiFi-7000, and an iPhone Xs running the EMLID Reach application controlled the Reach RS+ rover. In order to respect the Société des établissements de plein air du Québec (SEPAQ) park management team's request to limit the use of automobiles within the park so that park visitors would not be disturbed, a bicycle was used as the mode of transportation on the roads and trails, with the remainder of the sampling off-trail conducted on foot.

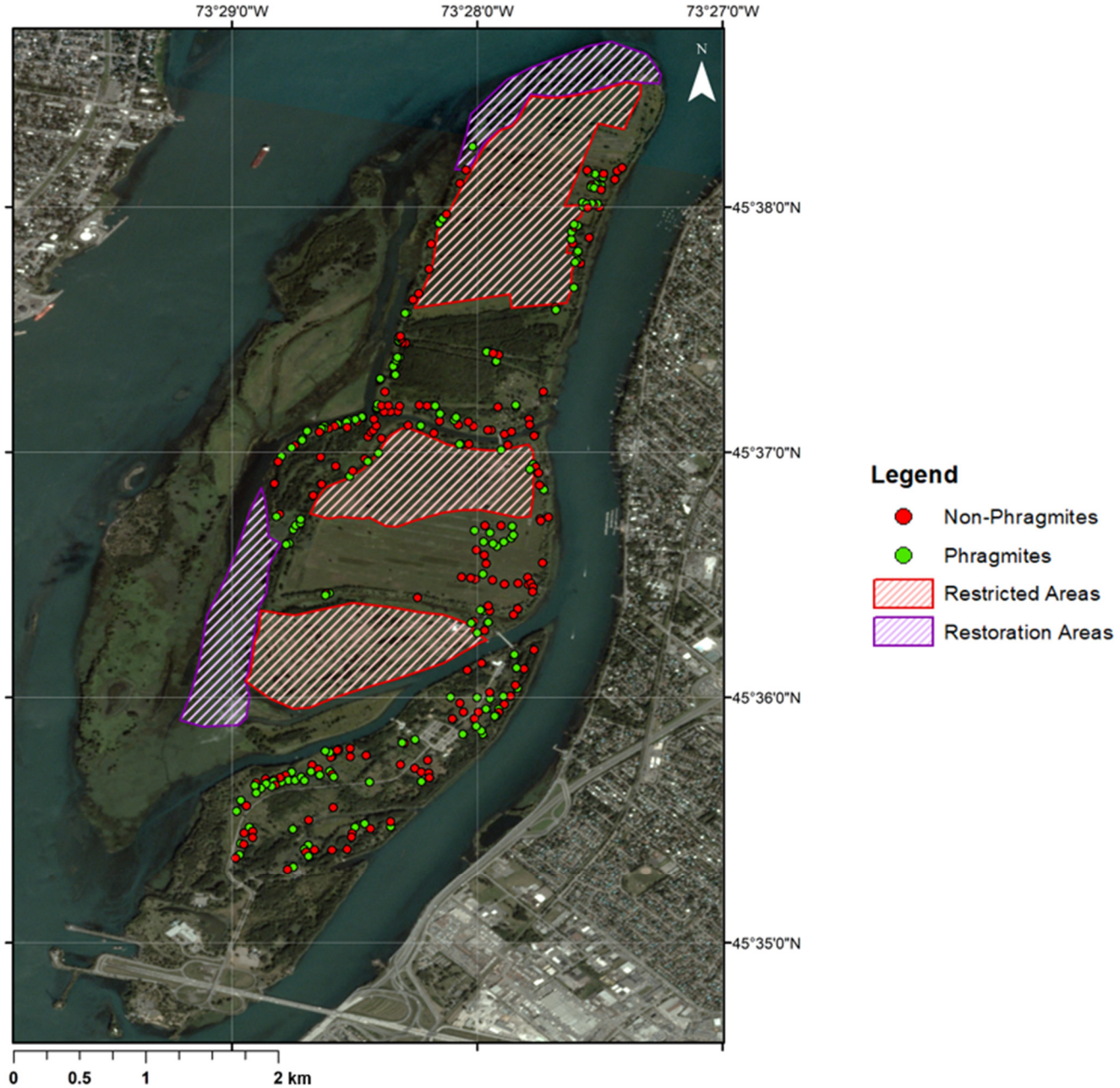

Figure 5. Map of Îles-de-Boucherville National Park showing the 320 ground truth points. Red points indicate no *Phragmites* was present at the measured location, and green points indicate that *Phragmites* was present at the measured location. Areas where sampling was restricted due to active restoration efforts (purple) or private land (red) are also shown. Basemap is a PlanetScope Dove, image acquired on 5 September 2019. Projection: NAD83(CSRS)v7, EPSG:8254.

In addition to the ground truth points, detailed field notes were also collected for each sampling point. These field notes included a site photo of the measurement point (along with the appropriate photo ID number), whether invasive *Phragmites* was present or not,

basic site characteristics such as the type(s) of vegetation present, and more detailed notes regarding the location or characteristics of the site. The data underwent quality assessment and any improperly collected points were removed from the dataset, such as duplicate points (due to operator error) or points that reported horizontal standard deviation of 0 m. Ultimately 320 points were retained: 158 points where *Phragmites* was present and 162 points where *Phragmites* was absent. The 320 points were distributed over as much of the accessible extent of the park as possible (Figure 5). In order to avoid a bias of points only collected near trails or other easily accessible areas (such as near beaches or parking lots), care was taken to go off-trail into harder-to-reach areas of the park in order to capture the natural variation and distribution of vegetation species. This off-trail sampling was done with permission from the SEPAQ park management team, and no sampling took place in any of the active restoration areas of the park (Figure 5).

3.3. Approximate Positional Accuracy

The Reach RS+ rover internally calculates the position of sample points relative to a continuously updating average position [17]. The deviation from the average position is reported by the unit both on-screen and in the output files as the root mean squared error (RMSE). The RMSE is a measure of the difference between values that are predicted and those that are observed. However, because there is no 'observed' known location to which the collected sample point locations are compared, the reported values are the standard deviation. In this dataset, the RMSE field names have been changed to SD to reflect what the values represent.

For this application, we report SD_x, SD_y, SD_z, and the horizontal linear SD in the radial direction (SD_r) from Equation (1) [18]. These are referred to incorrectly as $RMSE_x$, $RMSE_y$, $RMSE_z$, and $RMSE_r$ in the original data files output by the unit.

$$SD_r = \sqrt{SD_x{}^2 + SD_y{}^2}, \quad (1)$$

The SD are reported in the ground truth dataset file, as well as the attribute table of the shapefile. An approximate horizontal ($\sim ACC_r$) and vertical ($\sim ACC_z$) accuracy were calculated from Equations (2) and (3).

$$\sim ACC_r = 1.7308 \times (AA_r + \delta_r), \quad (2)$$

$$\sim ACC_z = 1.9600 \times (AA_z + \delta_z), \quad (3)$$

where δ_r is the uncertainty in the radial dimension (SD_r) and δz is the uncertainty in the z dimension (SD_z). Because the SD values represent precision, rather than accuracy compared to a known position, the manufacturer-stated approximate accuracy (AA) of 1 m for baselines < 30 km is adopted here. Equation (2) is modified from the National Standard for Spatial Data Accuracy (NSSDA) horizontal accuracy (ACC_r) at the 95% confidence level [18]. However, because ACC_r as described by [18] is calculated using the error compared to a known position, the NSSDA ACC_r could not be determined for this dataset, and an approximate horizontal accuracy at the 95% confidence level is reported. Equation (3) is modified from the ASPRS Accuracy Standards for Digital Geospatial Data, which notes that the vertical error in vegetated terrain does not typically follow a normal distribution [18]. Because ACC_z as described by [18] is calculated using the error compared to a known position, the ASPRS ACC_z could not be determined for this dataset and an approximate vegetated vertical accuracy at the 95th percentile is reported. The same EMLID Reach RS+ unit with incoming corrections from SmartNet North America via NTRIP had been previously verified for accuracy using a Natural Resources Canada High Precision 3D Geodetic Network station (#95K0003) [19]. The positional accuracy was assessed to be <2.5 cm (x and y) and <3 cm (z) [20].

A brief summary of the average, minimum, maximum, and standard deviation of SD in each dimension, $\sim ACC_r$, and $\sim ACC_z$ for all 320 points is given in Table 4. Tables 5 and 6

contain the average, minimum, and maximum SD in each dimension, $\sim ACC_r$, and $\sim ACC_z$ for the ground truth points with a solution status of FIX (143 points) or FLOAT (177 points), respectively. All 320 points had an $\sim ACC_r < 2.25$ m and a $SD_r < 16.5$ cm, which meet the requirements of the project as determined by the characteristics of the HSI.

All ground truth points had an $\sim ACC_z \geq 1$ m, and 89.69% of all ground truth points had an $\sim ACC_z < 2$ m. Thirty-three points (10.31%) had an $\sim ACC_z \geq 2$ m, but no points had an $\sim ACC_z > 3$ m. This indicates the vertical data would not be suitable for geospatial mapping projects that require high vertical accuracy, but would be suitable for visualization and projects without high vertical accuracy requirements.

The empirical cumulative distribution functions (CDF) were determined for SD_r and $\sim ACC_r$ for the full dataset, FLOAT data subset, and FIX data subset, as shown in Figure 6. Additionally, the corresponding frequency distribution histograms are shown within each corresponding CDF plot.

Table 4. The average, minimum, maximum, and standard deviation of the precision and the approximate ACC_r and ACC_z calculated for the entire set of 320 points, reported in centimeters.

Metric	Average	Standard Deviation	Minimum	Maximum
SD_r	0.80	0.68	0.12	9.21
SD_x	0.53	0.59	0.01	8.44
SD_y	0.53	0.43	0.04	3.69
$\sim ACC_r$	174.46	1.17	173.28	189.02
$\sim ACC_z$	198.09	4.47	196.13	268.48

Table 5. The average, minimum, maximum, and standard deviation of the precision and the approximate ACC_r and ACC_z calculated for the subset of 143 points with a solution status as FIX, reported in centimeters.

Metric	Average	Standard Deviation	Minimum	Maximum
SD_r	0.54	0.79	0.13	9.21
SD_x	0.37	0.72	0.06	8.44
SD_y	0.36	0.36	0.04	3.69
$\sim ACC_r$	174.02	1.37	173.31	189.02
$\sim ACC_z$	197.43	6.24	196.27	268.48

Table 6. The average, minimum, maximum, and standard deviation of the precision and the approximate ACC_r and ACC_z calculated for the subset of 177 points with a solution status as FLOAT, reported in centimeters.

Metric	Average	Standard Deviation	Minimum	Maximum
SD_r	1.01	0.48	0.12	2.84
SD_x	0.66	0.40	0.01	1.85
SD_y	0.67	0.44	0.09	2.74
$\sim ACC_r$	174.82	0.82	173.28	177.99
$\sim ACC_z$	198.63	2.02	196.13	206.76

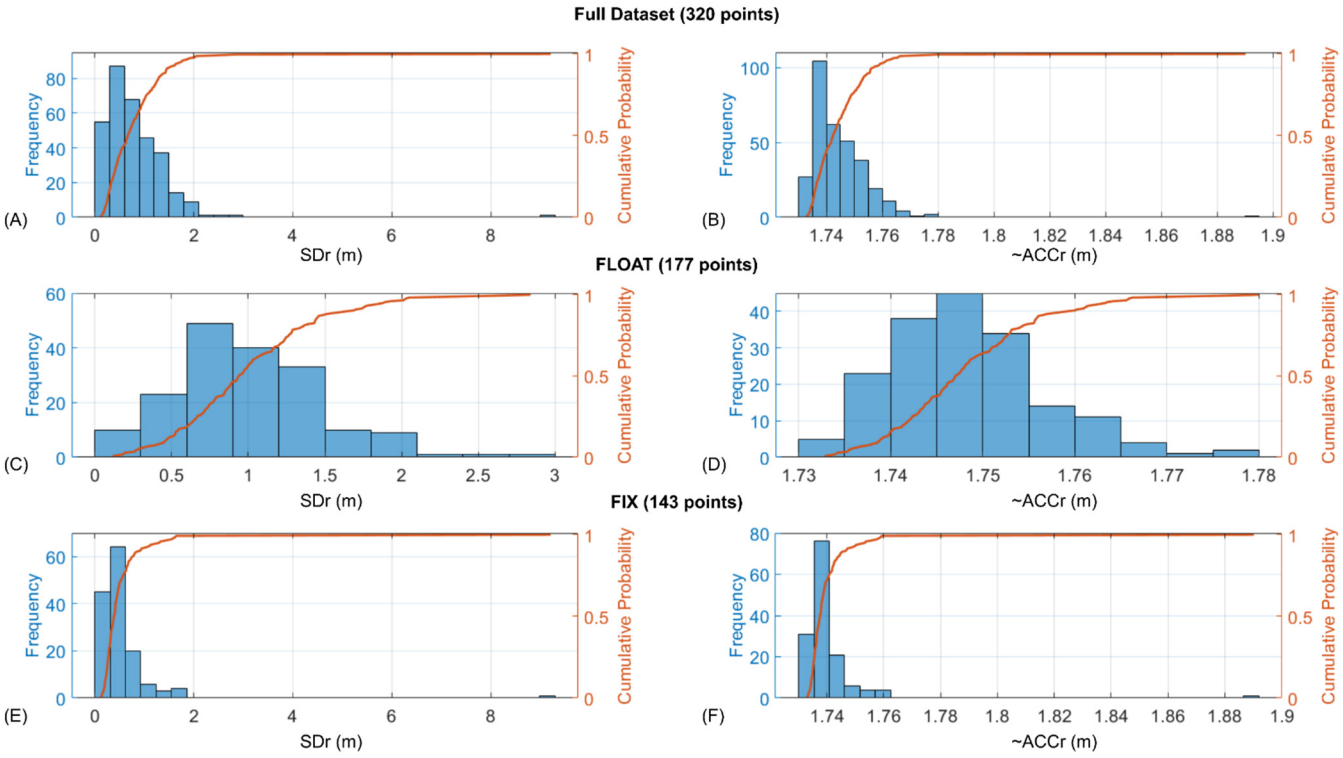

Figure 6. The empirical cumulative distribution functions for different subsets of the data. Subsets (**A**) and (**B**) are the SD_r and $\sim ACC_r$, respectively, for the full dataset of 320 points. Subsets (**C**) and (**D**) are the SD_r and $\sim ACC_r$, respectively, for the subset of 177 points where the solution status was FLOAT. Subsets (**E**) and (**F**) are the SD_r and $\sim ACC_r$, respectively, for the subset of 143 points where the solution status was FIX. Each plot also illustrates the frequency distribution histogram for the associated metric.

4. Conclusions

A new ground truth dataset containing detailed positional information regarding the presence of invasive *Phragmites* and other vegetation within Îles-de-Boucherville National Park is described. The dataset was collected using an established protocol and a high-accuracy GNSS system consisting of an EMLID Reach RS+ rover and incoming RTK corrections via internet using the SmartNet North America NTRIP service. The purpose of these data is to provide ground truth for the training and validation of *Phragmites* target detection using airborne HSI acquired over the extent of the park, in order to map the presence of *Phragmites*. The precision and approximate accuracy of the dataset were assessed, and were found to meet the requirements of the application (i.e., <16.5 cm for precision and <2.25 m for accuracy), as determined by the characteristics of the HSI (Figure 7). Auxiliary field notes are provided in order to give context to the measured points, and include brief descriptions of the site characteristics as well as corresponding field photographs for visual reference.

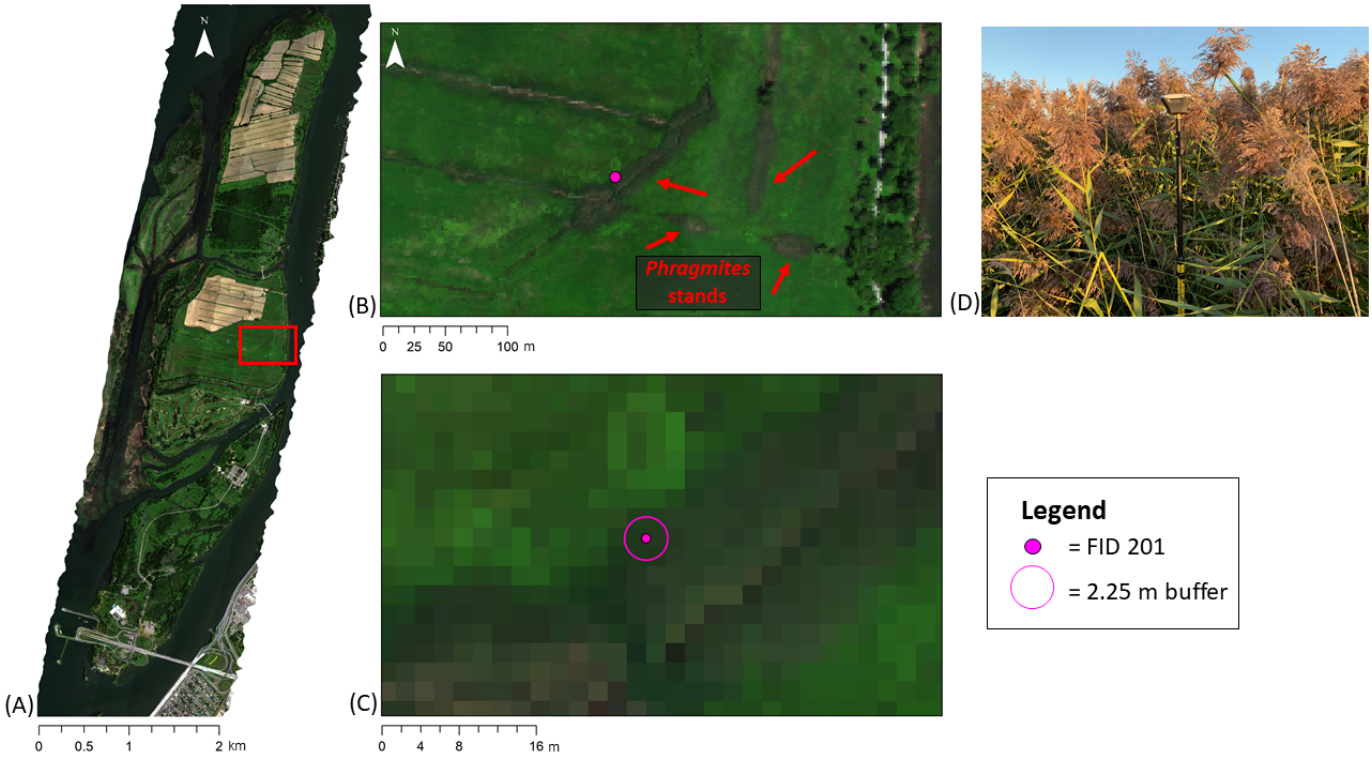

Figure 7. (**A**) A mosaic of the four CASI-1500 flight lines acquired over Îles-de-Boucherville National Park on July 8, 2019 (Projection: NAD83(CSRS)v7, EPSG:8254). The HSI are displayed as a true color RGB composite: 650 nm/548 nm/442 nm. The red rectangle shows the location of subset (**B**), which illustrates the appearance of *Phragmites* stands in the HSI (red arrows) and the location, as an example, of one of the ground truth points (FID 201) collected within a *Phragmites* stand. (**C**) shows a closeup of the HSI with individual pixels visible, and a 2.25 m buffer drawn around FID 201 to demonstrate that all pixels within the buffer are part of the *Phragmites* stand. (**D**) shows the corresponding field photograph for FID 201 to give context to the measured location.

5. User Notes

Users should be sure to note the "Solution Status" field of the dataset, as solutions of "FLOAT" and "FIX" may have implications on the overall accuracy of the collected location. While both "FLOAT" and "FIX" solution statuses indicated that base corrections are included and positioning is relative to the determined base coordinates, "FLOAT" means that the integer ambiguity is not resolved, while "FIX" means that the integer ambiguity is resolved. Precision in float mode is typically at the sub-meter level, while precision in fix mode is at the centimeter level. Points collected in dense vegetation, such as standing in a developed *Phragmites* stand that could measure over 2 m in height, are likely to have lower accuracy. This can be attributed to the fact that such dense vegetation restricted the view of the sky, and therefore restricted the ability of the Reach RS+ to see and acquire sufficient satellites to produce lower errors. To improve the positional accuracy of the ground truth points, a multi-band receiver with incoming corrections over a shorter baseline is recommended. For the Reach RS+ specifically, a local base station (< 10 km baseline) in addition to the incoming NTRIP corrections would be required for higher accuracy.

Due to logistical constraints, it was not possible to remeasure the 320 points in order to recheck them for accuracy. Additionally, there are no previously established geodetic monuments within the park that could serve as a reference point. Because the Reach RS+ unit had been previously assessed for accuracy using a Natural Resources Canada High Precision 3D Geodetic Network station (see Section 3.3), the 320 points were collected in the field with confidence, and their approximate accuracy exceeds the requirements of the project under which they were acquired.

The field note information shown in Figure 2 is in a format that can be joined to the attribute table of the .shp file in a GIS. This would allow the user to have the attribute table contain a complete set of descriptions for each point. There is one missing field photograph for the point FID-265. However, a brief description of the immediate area was noted within the auxiliary field notes file.

Author Contributions: Conceptualization, M.K. and K.E.; methodology, M.K. and K.E.; formal analysis, K.E.; data curation, K.E.; writing—original draft preparation, K.E.; writing—review and editing, M.K. and K.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a Natural Sciences and Engineering Research Council (NSERC) Discovery Frontiers grant that supported the Canadian Airborne Biodiversity Observatory (CABO). Additional funding was provided by the McGill Department of Geography Rathlyn Fellowship in Geography (to Elmer). The APC was funded by an NSERC Discovery Grant (to Kalacska).

Data Availability Statement: The dataset is available for download from http://doi.org/10.5281/zenodo.4504922 (accessed on 13 March 2021).

Acknowledgments: We would like to acknowledge the SÉPAQ team at Îles-de-Boucherville National Park, especially Nathalie Rivard, for their logistical support of the data collection campaign. Access to the Planet Scope imagery was provided by the Department of Geography, McGill University. We thank J. Pablo Arroyo-Mora for comments on the manuscript and Deep Inamdar for details about the HSI point spread functions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Chambers, R.M.; Meyerson, L.A.; Saltonstall, K. Expansion of *Phragmites australis* into tidal wetlands of North America. *Aquat. Bot.* **1999**, *64*, 261–273. [CrossRef]
2. NOAA Great Lakes Environmental Research Laboratory. Great Lakes Aquatic Nonindigenous Species Information System: *Phragmites australis australis* (Cav.) Trin. ex Steud. Available online: https://nas.er.usgs.gov/queries/greatlakes/FactSheet.aspx?SpeciesID=2937 (accessed on 8 December 2018).
3. Saltonstall, K. Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2445–2449. [CrossRef] [PubMed]
4. Belzile, F.; Labbé, J.; LeBlanc, M.-C.; Lavoie, C. Seeds contribute strongly to the spread of the invasive genotype of the common reed (*Phragmites australis*). *Biol. Invasions* **2010**, *12*, 2243–2250. [CrossRef]
5. Saltonstall, K. Alien Plant Invaders of Natural Areas. Available online: http://www.nps.gov/plants/alien (accessed on 1 April 2019).
6. Avers, B.; Fahlsing, R.; Kafcas, E.; Schafer, J.; Collin, T.; Esman, L.; Finnell, E.; Lounds, A.; Terry, R.; Hazelman, J. *A Guide to the Control and Management of Invasive Phragmites*; Michigan Department of Environmental Quality: Lansing, MI, USA, 2007.
7. Inamdar, D.; Kalacska, M.; Leblanc, G.; Arroyo-Mora, J.P. Characterizing and Mitigating Sensor Generated Spatial Correlations in Airborne Hyperspectral Imaging Data. *Remote Sens.* **2020**, *12*, 641. [CrossRef]
8. Natural Resources Canada's Canadian Geodetic Survey (CGS). Geodtic Tools and Applications: GPS-H. Available online: https://www.nrcan.gc.ca/maps-tools-publications/maps/tools-applications/10925#gpsh (accessed on 13 February 2021).
9. Laliberté, E.; Cogliastro, A.; Bouchard, A. *Projet pilote de restauration de paysages forestiers au parc national des îles-de-Boucherville. Rapport Final Présenté à la Direction du Parc National des îles-de-Boucherville, Société des Établissements de Plein Air du Québec (SÉPAQ)*; Institut de recherche en biologie végétale: Montréal, QC, Canada, 2006.
10. Giroux, L. *Synthèse et cartographie du potentiel biophysique et humain du Parc des îlesde-Boucherville*; Maîtrise en Sciences de l'environnemen: Montréal, QC, Canada, 1986.
11. SOCIETE des Etablissements de Plein air du Québec. Parc National Des Îles-De-Boucherville: Portait of the park. Available online: https://www.sepaq.com/pq/bou/decouvrir/portrait.dot?language_id=1 (accessed on 17 October 2019).
12. Ross, M. *Étude sur L'évolution de Boisé Grosbois au Parc des Iles-De-Boucherville*; Parcs de la Montérégie: Montérégie, Canada, 1990.
13. Kalacska, M. *Emlid GPS NTRIP Protocol*; Protocols.io: Berkeley, CA, USA, 2018. [CrossRef]
14. Weber, G.; Dettmering, D.; Gebhard, H.; Kalafus, R. Networked transport of RTCM via internet protocol (Ntrip) . . . IP-Streaming for Real-Time GNSS Applications. In Proceedings of Proceedings of the 18th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2005), Long Beach, CA, USA, 13–16 September 2005; pp. 2243–2247.
15. Leica Geosystems. *White Paper: Take it to the MAX! An Introduction to the Philosophy and Technology behind Leica Geosystems' SpiderNET Revolutionary Network RTK Software and Algorithms*; Leica Geosystems AG, Nune: Heerbrugg, Switzerland, 2005.

16. EMLID. Reach RS/RS+ docs: Placing the Base. Available online: https://docs.emlid.com/reachrs/common/tutorials/placing-the-base/ (accessed on 13 February 2021).
17. Fedorov, E.; RMS? [Discussion Post]. EMLID Community Forum, April 14 2018. Available online: https://community.emlid.com/t/rms/6340/5 (accessed on 25 February 2021).
18. American Society for Photogrammetry and Remote Sensing (ASPRS). *ASPRS Positional Accuracy Standards for Digital GEospactial Data (2014)*; ASPRS: Bethesda, MD, USA, 2015; pp. A1–A26.
19. Natural Resources Canada. Station Report: 95K0003. Available online: https://webapp.geod.nrcan.gc.ca/geod/data-donnees/station/report-rapport.php?id=95K0003 (accessed on 13 February 2021).
20. Kalacska, M.; Lucanus, O.; Arroyo-Mora, J.P.; Laliberté, É.; Elmer, K.; Leblanc, G.; Groves, A. Accuracy of 3d landscape reconstruction without ground control points using different uas platforms. *Drones* **2020**, *4*, 13. [CrossRef]