Method Article

Water column compensation workflow for hyperspectral imaging data

Deep Inamdar[a], Gillian S.L. Rowan[a], Margaret Kalacska[a,*], J. Pablo Arroyo-Mora[b]

[a] *Applied Remote Sensing Laboratory, Department of Geography, McGill University, Montréal, QC H3A 0B9, Canada*
[b] *Flight Research Laboratory, National Research Council of Canada, Ottawa, ON K1A 0R6, Canada*

ABSTRACT

Our article describes a data processing workflow for hyperspectral imaging data to compensate for the water column in shallow, *clear* to *moderate* optical water types. We provide a MATLAB script that can be readily used to implement the described workflow. We break down each code segment of this script so that it is more approachable for use and modification by end users and data providers. The workflow initially implements the method for water column compensation described in Lyzenga (1978) and Lyzenga (1981), generating depth invariant indices from spectral band pairs. Given the high dimensionality of hyperspectral imaging data, an overwhelming number of depth invariant indices are generated in the workflow. As such, a correlation based feature selection methodology is applied to remove redundant depth invariant indices. In a post-processing step, a principal component transformation is applied, extracting features that account for a substantial amount of the variance from the non-redundant depth invariant indices while reducing dimensionality. To fully showcase the developed methodology and its potential for extracting bottom type information, we provide an example output of the water column compensation workflow using hyperspectral imaging data collected over the coast of Philpott's Island in Long Sault Parkway provincial park, Ontario, Canada.

- Workflow calculates depth invariant indices for hyperspectral imaging data to compensate for the water column in shallow, *clear* to *moderate* optical water types.
- The applied principal component transformation generates features that account for a substantial amount of the variance from the depth invariant indices while reducing dimensionality.
- The output (both depth invariant index image and principal component image) allows for the analysis of bottom type in shallow, *clear* to *moderate* optical water types.

---

* Corresponding author.
  E-mail address: Margaret.kalacska@mcgill.ca (M. Kalacska).

Specifications table

Subject Area:	Environmental Science
More specific subject area:	*Remote Sensing of Optically Shallow Water Aquatic Ecosystems*
Method name:	*Hyperspectral Depth invariant index*
Name and reference of original method:	*Lyzenga, D.R. (1978). Passive remote sensing techniques for mapping water depth and bottom features. Applied Optics, 17, 379–383*
	Lyzenga, D.R. (1981). Remote sensing of bottom reflectance and water attenuation parameters in shallow water using aircraft and Landsat data. International Journal of Remote Sensing, 2, 71–82
Resource availability:	*Required: MATLAB 2020B with Hyperspectral Imaging Library from Image Processing Toolbox*

*Method details

Background

Over the past four decades, hyperspectral imaging (HSI) has provided rich spectral-spatial information that has led to a variety of applications for Earth observation [3]. Although most of these efforts have concentrated on analyzing the Earth's land surface, the benthic zones of aquatic ecosystems (both marine and fresh water) have also been successfully studied [2,4,18]. Such applications are the most feasible for optically shallow waters where the reflected electromagnetic energy from the bottom of the water body contributes to the water-leaving signal [10]. Optically shallow waters are typically found in inland and marine coastal environments. In these environments, the waters are optically complex as their light-matter interactions can be highly variable and are primarily influenced by sources such as phytoplankton (Case 1 waters) and mineral particles and dissolved organic matter (Case 2 waters) [11,12]. To extract information about the benthic zones of shallow water ecosystems from reflected electromagnetic radiation, it is critical to compensate for the complex effects of absorption and scattering that occur in the overlying water column [13]. A recently proposed optical water type classification scheme for complex waters (such as those inland) organizes water types into five categories, from clear to brown, according to their bio-optical properties [21]. Due to the low concentrations of optically significant constituents, it is the most feasible to analyze the benthic zone of aquatic ecosystems with *clear* and *moderate* optical water types using HSI [21].

As electromagnetic radiation traverses the water column in *clear* to *moderate* optical water types, its intensity decreases close to exponentially with increasing depth [7]. The degree to which the intensity of electromagnetic radiation is attenuated varies by wavelength; as the wavelength of electromagnetic radiation increases, more energy is attenuated by the overlying water column. For instance, Gordon and McCluney [5] showed that 90 % of electromagnetic radiation at 600–700 nm is attenuated in high-clarity sea water at a depth of 3.8 m while Green et al. [6] demonstrated that the electromagnetic radiation at 800–1100 nm is completely attenuated at the same depth. The overall effects of the water column can be described by Eq. (1):

$$L_i = L_{si} + a \cdot R_i \cdot e^{-2k_i \cdot z} \tag{1}$$

where R_i is the bottom type reflectance at band i that is independent of the water column, L_i is the radiance observed by the sensor (including the water column) at band i, z is the height of the water column, k_i is the water attenuation coefficient (wavelength dependent) at band i, a is a constant (representative of the solar irradiance, the transmittance of the atmosphere and the water surface and the reduction of the radiance due to refraction at the water surface) and L_{si} is the radiance observed over deep water (due to external reflection from the water surface and atmospheric contributions) at
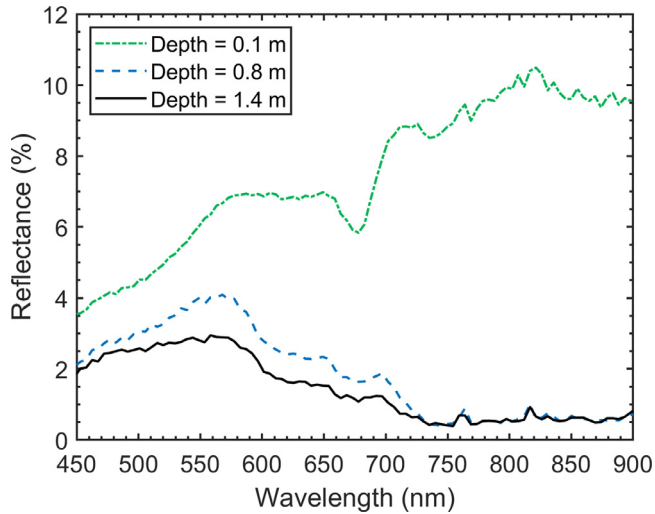
Fig. 1. Reflectance spectra from Hyperspectral Imaging data observed over a flooded cement road at the Long Sault Parkway in Ontario, Canada at various water depths. As the water depth increases, the observed reflectance is attenuated.

band *i*. Considering Eq. (1), the influence of the water column on the signal recorded by a sensor can create confusion when analyzing bottom type. For instance, the observed reflectance of a substrate at a depth of 0.1 m will look considerably different than the observed reflectance of the same substrate at 1.5 m, especially since the decay rate of electromagnetic radiation is not constant from wavelength to wavelength (see Fig. 1 for an example of reflectance spectra from a submerged cement road at various water depths). In principle, the effects of the water column could be calculated and removed if the water depth and water optical properties were known throughout the scene. However, this is not practically feasible since water depth [7] and optical properties [21] are highly variable in space.

In Wezernak and Lyzenga [23], the variable effects of water column were compensated by calculating the ratio between two spectral bands with attenuation coefficients that were assumed to be equal. Although showing some success, this water column compensation (WCC) methodology was limited as substrates with similarly shaped reflectance spectra (e.g., mud, sand) have nearly equal reflectance ratios [7]. Lyzenga [7] expanded on the simple ratio model, developing a WCC methodology that combines the information from pairs of spectral bands to produce depth invariant indices (DII) that are representative of bottom type. This approach does not require the used spectral bands to have identical attenuation coefficients and thus can produce DIIs that are representative of various bottom types. The only parameter required to generate a DII is the ratio of the water attenuation coefficients between the pair of spectral bands used. The ratios of the water attenuation coefficients can be solved following the method described in Lyzenga [8].

The DII calculations from Lyzenga [7] were originally developed for multispectral imaging data (e.g., [1,9,14,15]), however, have since been applied to HSI data (e.g. [16,19,25]) without fundamental changes to the original formulation. For instance, Rowan et al. [19] calculated DIIs from HSI data to accurately detect submerged aquatic vegetation in a freshwater ecosystem. As shown in this work, various modifications are necessary when compensating for the water column in HSI data by calculating DIIs. For instance, an overwhelming number of indices are generated when calculating DIIs from HSI data as there are an abundance of band pair combinations (e.g. a HSI cube with 100 bands would result in 4950 DIIs). For this reason, feature selection and extraction methodologies help to reduce the dimensionality of the dataset to a more manageable size. Given the intricacies in compensating for the water column in HSI data, an open access WCC tool would help end users to apply HSI data over shallow, *clear* to *moderate* optical water types.

Our work herein describes a data processing workflow that was successfully implemented in Rowan et al. [19] to compensate for the water column in HSI data collected over shallow, *clear* to

moderate optical water types. Before HSI data can be input into the WCC workflow it must be pre-processed. Given the low signal levels of aquatic environments, conventional radiometric correction processes often result in exceptionally low or negative values of radiance at wavelengths < 450 nm. As such, a two-part in-flight radiometric refinement (IFRR) as described by Soffer et al. [20] is recommended as an initial pre-processing step. Additionally, before being input into the workflow (with or without an IFRR), the HSI data must be atmospherically compensated. In aquatic applications, spectral measurements have contributions from surface reflected radiance (L_{SR}), which is incident radiance reflected off the water's surface into the HSI sensor's field of view (e.g. glint). If L_{SR} is being accounted for, L_{SR} removal techniques (see Windle and Silsbe [24] for examples) should also be applied prior to the WCC workflow. Additionally, land pixels must be masked from the imagery.

In the WCC workflow, high frequency noise is first removed using a Savitzky–Golay smoothing filter. After, the effects of the water column are compensated for by calculating a DII from every pair of spectral bands following the methods described in Lyzenga [7] and Lyzenga [8]. A correlation based feature selection methodology is then applied to remove redundant DIIs. In an additional post-processing step, a principal component transformation is applied, extracting informative features from the DIIs while further reducing dimensionality. In this work, we provide a MATLAB script that can be readily used to implement the described workflow, breaking down each code segment so that it is more approachable for use and modification by end users and data providers. To showcase the effectiveness of the developed methodology, the WCC workflow was applied to HSI data collected over the coast of Philpott's Island in Long Sault Parkway provincial park, Ontario, Canada. In the final section, we briefly discuss the application that the WCC compensation was used for in Rowan et al. [19]. Overall, by modifying traditional WCC methodologies (through the inclusion of the Savitzky–Golay smoothing filter, the correlation based feature selection algorithm and the principal component transformation), HSI data can be effectively applied to study aquatic ecosystems with shallow, *clear* to *moderate* optical water types.

Method workflow

The WCC workflow for HSI data is shown in Fig. 2. It is important to note that the input HSI data to the WCC must be accurately radiometrically corrected, preferably with an IFRR [20], atmospherically compensated and L_{SR} removed (if necessary). Additionally, land pixels must be masked from the imagery.

Even after proper pre-processing methodologies have been applied to HSI data over *clear* to *moderate* optical water types, the water-leaving signal tends to be exceptionally low in comparison to terrestrial imagery [20]. Spectral filtering techniques are often used to remove high frequency noise in low signal data [22]. Accordingly, our WCC first applies a Savitzky-Golay filter of a user specified order. Afterwards, the workflow calculates DIIs following Lyzenga [7].

In the DII calculations, the relationship between water column depth and radiance at a given band, *i*, is first linearized through the following transformation (Eq. (2)):

$$X_i = \ln (L_i - L_{si}) \tag{2}$$

where X_i is a variable that is linearly dependent on water column depth and is representative of L_i. In practice, the inputs to Eq. (2) (L_i and L_{si}) are not in units of radiance, but reflectance as the imagery must be atmospherically compensated prior to the WCC workflow. The *y*-intercept of a bi-plot of X_i and X_j (the output of Eq. (2) for given band *j* that is different than band *i*) is representative of the bottom surface reflectance independent of the water column and thus a DII [7]. To solve for the DII, Eq. (3) is used:

$$DII_{ij} = X_i - \frac{k_i}{k_j} X_j \tag{3}$$

where $\frac{k_i}{k_j}$ is the ratio of attenuation coefficients between band *i* and *j* (slope of the bi-plot of X_i and X_j). Although $\frac{k_i}{k_j}$ can be linearly regressed with training data, the results depend on which bands are chosen as the dependent and independent variables. This problem can be avoided by minimizing
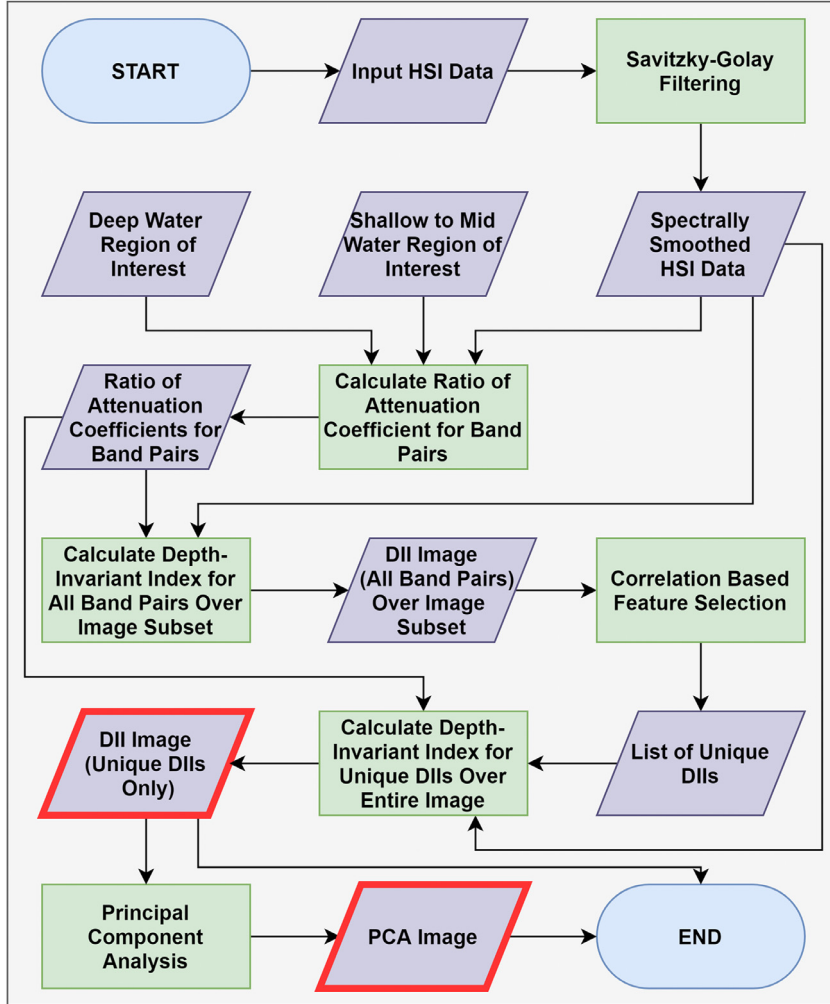
Fig. 2. Flow chart of the water column compensation workflow for hyperspectral imaging (HSI) data. The workflow calculates depth invariant indices (DII) for each band pair from the original HSI data. Due to the high dimensionality of HSI data, the number of calculated DIIs are reduced using a correlation based feature selection algorithm. The data dimensionality is additionally reduced using a principal component transformation (PCA). The DII and PCA end products are highlighted with a red boarder in the workflow.

the mean square deviation perpendicular to the regression line to calculate the ratio of attenuation coefficients [8] (Eq. (4)):

$$\frac{k_i}{k_j} = \frac{\sigma_i - \sigma_j}{2\sigma_{ij}} + \sqrt{\left(\frac{\sigma_i - \sigma_j}{2\sigma_{ij}}\right)^2 + 1} \tag{4}$$

where σ_i is the variance in X_i, σ_j is the variance in X_j and σ_{ij} is the covariance between X_i and X_j. Following Lyzenga [8], $\frac{k_i}{k_j}$ is calculated from spectra over a constant substrate along a transect of varying water column depths.

Given the high dimensionality of HSI data, a large number of DIIs can be generated as there is an abundance of band pair combinations (e.g., a HSI cube with 100 bands would result in 4950 DIIs).

Therefore, it is not computationally feasible to calculate all possible DIIs across a hyperspectral image. As such, dimension reduction through an initial feature selection is required. In this process, DIIs are calculated over a random subset (user specified size) of the imagery for all possible band pairs. A correlation based feature selection algorithm is then applied to determine the most unique DIIs. The feature extraction method operates under the assumption that features with high correlation are more linearly dependent and hence have almost the same effect on the dependent variable during data applications. When the correlation coefficient between two DIIs is above a user specified threshold, one of the DIIs is dropped. By setting a sufficiently strict correlation threshold, it is possible to reduce the DIIs to a manageable number. After applying the correlation based feature selection algorithm, the selected DIIs are calculated over the entire image. This DII image is given as one of the final end products.

In an additional post-processing stage of the workflow, a principal component transformation is applied to the DII image to generate features that account for a substantial portion of the variance from the DIIs while reducing dimensionality. In this transformation, the user specifies the percentage of the total variance that they want to be explained by the components output in the final data product. The PCA image is the other final output of the WCC workflow.

The described WCC workflow is implemented as a single MATLAB function (HSI_WCC.m). In the following section we break down this function, describing the most important code segments.

MATLAB function

The presented MATLAB function (HSI_WCC.m) carries out seven main tasks: (1) smooth spectra with Savitzky-Golay filter; (2) calculate DIIs on image subset; (3) select unique DIIs through correlation based feature selection; (4) generate DII image of selected indices; (5) reduce dimensionality of DII image with principal component transformation, identifying important principal components; (6) generate PCA image of the identified principal components; and (7) output DII and PCA images as ENVI standard data files. The inputs and outputs to the function are outlined in the function description:

```
%%% Input Parameters
%   image_name...           Name of input Hyperspectral Imaging dataset (must be ENVI
standard). Input as character vector(e.g. 'C:') not string (e.g. "C:"). Please do not
include '.dat' extension.
%   image_loc...            Location of folder path of the Hyperspectral Imaging
dataset. Input as character vector(e.g. 'C:') not string (e.g. "C:").
%   dw_roi...               Region of interest with deep water pixels (2 x m matrix with
column and row pixel number from the imagery).
%   sw_mw_roi...            Region of interest with shallow to mid water pixels over
constant substrate (2 x m matrix with column and row pixel number from the imagery).
%   no_data_value...        Value of nodata pixels. Land should be masked before data is
input into the water column compensation workflow.
%   wl_range...             Wavelength range for analysis. Longer wavelengths are
typically more attenuated and thus may produce uninformative depth invariant indices.
%   SG_filt_deg...          Polynomial order for Savitzky-Golay finite impulse response
(FIR) smoothing filter.
%   window_size...          Frame length for Savitzky-Golay finite impulse response
(FIR) smoothing filter.
%   percent_explained...    Percentage of the total variance explained by the principal
components encoded in the PCA image.
%   num_samples...          The number of samples used in the correlation based feature
selection.
%   cor_thresh...           Correlation coefficient threshold for feature selection. If
two DIIs have a correlation coefficient over the specified threshold, one of the DIIs is
dropped.
%
%%% Output Parameters
%   image_DII...            Image with the unique DIIs chosen during the correlation
based feature selection.
%   DII_band_combs...       The bands used to generate each of the DIIs from image_DII
%   image_PCA...            Principal component image generated from the DII image.
```

Before starting task 1, the HSI data needs to be imported as a n (number of image rows) by *m* (number of image columns) by b (number of spectral bands) matrix. Since the built in functions in MATLAB (e.g. sgolayfilt, corrcoef, cov, pca, etc.) require a two-dimensional input, the spatial dimensions of the HSI data matrix also needed to be flattened, creating a n*m by b matrix.

```
%%% Import input hyperspectral imaging data as a n x m x b matrix
Filename=[image_loc image_name '.dat'];
Filename_2=[image_loc image_name '.hdr'];
data=hypercube(Filename,Filename_2);
%%%%Reshape imported image into a n*m x b matrix
size_data_cube=size(data.DataCube);
data_linear=
double(reshape(data.DataCube,size_data_cube(1)*size_data_cube(2),size_data_cube(3
)));
```

After this initial pre-processing, the imported HSI data are spectrally smoothed with a Savitzky-Golay filter, completing task 1.

```
%Apply Savitzky-Golay smoothing filter
data_linear_temp=sgolayfilt(data_linear',SG_filt_deg,window_size)';
data_linear=data_linear_temp;
```

To calculate the DIIs, the spectra from a user specified deep water region of interest (ROI) need to be extracted. Similarly, the workflow must extract the spectra from a user specified ROI composed of pixels over a constant substrate at varying water column depths.

```
%Import Deep water and shallow/medium mid-water sand ROIs.
Deep=dw_roi;
Shallowmed=sw_mw_roi;

%Convert subscripts (row, columns)for input ROIs to a linear index for m*n by b
HSI data matrix
deep_index=sub2ind(size_data_cube(1:2),Deep(:,2),Deep(:,1));
Shal_Med_index=sub2ind(size_data_cube(1:2),Shallowmed(:,2),Shallowmed(:,1));

%Extract Spectra from Deep water ROI
DW_ref=double(data_linear(deep_index,:));
%Extract Spectra from Shallow water ROI
S_M_ref=double(data_linear(Shal_Med_index,:));
```

Next, L_{si} needs to be calculated. In this WCC workflow, L_{si} is approximated with the mean and standard deviation of the spectra from a user specified deep water ROI. Specifically, L_{si} is equal to the mean deep water spectra minus two standard deviations. If this results in any negative values, L_{si} is set to the mean deep water spectra minus one standard deviation. If this still results in negative values, an error appears, and a new deep water ROI must be selected.

```
%Calculate mean of deep water
DW_mean=mean(DW_ref);

%Calculate standard deviation of deep water
DW_SD=std(DW_ref);

%Shift mean by a multiple of the standard deviation
DW_mean_shift=DW_mean-DW_SD*2;
if sum(DW_mean_shift<0)>0
    DW_mean_shift=DW_mean-DW_SD
end
if sum(DW_mean_shift<0)>0
    error("Error: Please reselect the deep water region of interest.")
end
%Define LSI
LSI=DW_mean_shift;
```

With L_{si} calculated, the workflow linearizes the effect of depth on the spectra from the user specified ROI over a constant substrate along a transect of varying water column depths.

```
%Linearize water column depth dependence
S_M_log=log(S_M_ref-LSI);
```

After, the transformed spectra are used to calculate $\frac{k_i}{k_j}$. In this process, we first calculate $a_{ij} = \frac{\sigma_i - \sigma_j}{2\sigma_{ij}}$ for all possible band combinations. We then use a_{ij} to calculate $\frac{k_i}{k_j}$.

```
%Calculate covariance matrix of S_M_log
S_M_log_covar_matrix=cov(S_M_log);

%Calculate all possible band combinations
band_combs=nchoosek(1:sum(~bad_bands_index), 2);
size_band_combs=size(band_combs);


%Calculate aij
a_i_j=zeros(size_band_combs(1),1);

for k=1:size_band_combs(1)
a_i_j(k)=(S_M_log_covar_matrix(band_combs(k,1),band_combs(k,1)) -
S_M_log_covar_matrix(band_combs(k,2),band_combs(k,2))
)/2/S_M_log_covar_matrix(band_combs(k,1),band_combs(k,2));
end

%filter out positive values
neg_index=a_i_j<0;
a_i_j_filt=a_i_j(neg_index);
band_combs_filt=band_combs(neg_index,:);

%Calculate ki/kj
k_i_j=a_i_j_filt+sqrt(a_i_j_filt.^2+1);
```

Using $\frac{k_i}{k_j}$, DIIs can be calculated on a subset of the HSI data, completing task 2.

```
%Identify and remove nodata pixels and pixels for which data_linear-LSI<0 (pixels
that will generate invalid DIIs after log transformation)

data_linear_sub_index=~all(data_linear==no_data_value,2)& (min((data_linear-
LSI)')>0)';
data_linear_sub=data_linear(data_linear_sub_index,:);
sub_size=size(data_linear_sub);

%Subset the imagery to a user specified number of randomly selected pixels for
%the purposes of features selection
number_samples=num_samples;

if sub_size(1)>number_samples
    rand_vec=randi(sub_size(1),number_samples, 1);
    data_linear_sub=data_linear_sub(rand_vec,:);
end

%Calculate depth invariant indices on the selected subset
data_linear_2=zeros(number_samples,length(k_i_j));

for k=1:length(k_i_j)
data_linear_2(:,k)=(log(double(data_linear_sub(:,band_combs_filt(k,1)))-
LSI(band_combs_filt(k,1))))-(log(double(data_linear_sub(:,band_combs_filt(k,2)))-
LSI(band_combs_filt(k,2))))*k_i_j(k);
end
```

Task 3 is completed by applying a custom correlation based feature selection algorithm to the DIIs calculated on a subset of the imagery. The size of this subset is user specified and should be large enough so that the mean and standard deviation of each DII is representative of the site. The feature selection algorithm first calculates the correlation coefficient between all possible DII pairs. After, the script generates a list of potentially redundant DII pairs (correlation coefficient > then user specified threshold). The script goes through this list one DII pair at a time, identifying the first DII in the pair as redundant. Once a DII is identified as redundant, all of the potentially redundant DII pairs that use that identified DII are removed. By repeating this process until there are no longer any potentially redundant DII pairs, a list of redundant DIIs is generated. The end product of the described feature selection is a logical of redundant DIIs, completing task 3.

```
%Feature selection based on correlation coefficient thresholding between DIIs

%Calculate correlation matrix( correlation coefficient between each set of DIIs)
R = corrcoef(data_linear_2);
size_R=size(R);

%Make a list of DII pairs that are highly correlated
R_tri=triu(R,1);
thresh=cor_thresh;
[idx_R_row, idx_R_column] = find(abs(R_tri)>thresh); %Apply correlation threshold
idx_R_row2=idx_R_row;
idx_R_column2=idx_R_column;

%Loop through all DIIs one at a time. If a DII is part of a pair of DIIs that
%are highly correlated, mark it as redundant and remove it from the list
%generated in the previous section.
redundant_DII=zeros(size_R(1),1);

for i=1:size_R(1)
temp_index=or(idx_R_row2==i,idx_R_column2==i);
idx_R_row2=idx_R_row2(~temp_index);
idx_R_column2=idx_R_column2(~temp_index);

if sum(temp_index)>0
redundant_DII(i)=1;
end
end

%Logical of redundant DIIs
redundant_DII_log=logical(redundant_DII);
```

By removing the $\frac{k_i}{k_j}$ values associated with redundant DIIs, it is possible to calculate all the unique DIIs across the entire HSI dataset, completing task 4.

```
% Generate selected DII for entire image
data_linear_sub_index=~all(data_linear==no_data_value,2);
size_temp=size(data_linear);
data_linear_3=ones(size_temp(1),sum(~redundant_DII_log))*-10000;
k_i_j=k_i_j(~redundant_DII_log);
band_combs_filt=band_combs_filt(~redundant_DII_log,:);

for k=1:length(k_i_j)
data_linear_3(data_linear_sub_index,k)=(log(abs_2(double(data_linear(data_linear_
sub_index,band_combs_filt(k,1)))-LSI(band_combs_filt(k,1)))))-
(log(abs_2(double(data_linear(data_linear_sub_index,band_combs_filt(k,2)))-
LSI(band_combs_filt(k,2)))))*k_i_j(k);
end
```

To complete task 5, a principal component transformation is applied to the image generated in task 4. The workflow then identifies the number of principal components that explain a user specified percentage of the total variance.

```
% Principal component transformation feature extraction. Identify the
% number of principal components that explain a user specified percentage of
% the variance in the DII image.
[coeff,score,latent,tsquared,explained,mu]=pca(data_linear_3(data_linear_sub_inde
x,:));
cumulative_explained=cumsum(explained);
[M,I]=min(abs(cumulative_explained-percent_explained));
```

Task 6 is then completed by generating an image from the selected principal components.

```
%Generate from selected principal components
data_linear_4=ones(size_temp(1),I)*-10000;
data_linear_4(data_linear_sub_index,:)=score(:,1:I);
```

The final task is completed by writing out the generated DII and PCA images. The band names of the DII image are given by the wavelengths used to generate each DII. The band names of the PCA image are given by the component number. Nodata pixels are given a value of -10000.

```
%Output DII image
image_DII=reshape(data_linear_3,size_data_cube(1),size_data_cube(2),length(k_i_j)
);

meta2=data.Metadata;
meta2.Bands=length(k_i_j);
wl_encode=1:length(k_i_j);
wl_1=num2str(good_bands(band_combs_filt(:,1)));
wl_2=num2str(good_bands(band_combs_filt(:,2)));
plus=repmat(" + ",length(k_i_j),1);
newStr = join([wl_1 plus wl_2]);
meta2.BandNames=newStr;
meta2.DataType="double";
meta2.FWHM=[];
meta2.DataIgnoreValue=-10000

DII_band_combs=newStr;

Filename=[image_loc image_name '_DII'];
newhcube = hypercube(image_DII,wl_encode,meta2);
enviwrite(newhcube,Filename);

%Output PCA image
image_PCA=reshape(data_linear_4,size_data_cube(1),size_data_cube(2),I);

meta3=data.Metadata;
meta3.Bands=I;
wl_encode=1:I;
plus=repmat("PC ",I,1);
newStr = join([plus num2str(wl_encode')]);
meta3.BandNames=newStr;
meta3.DataType="double";
meta3.FWHM=[];
meta3.DataIgnoreValue=-10000

Filename=[image_loc image_name '_DII_PCA'];
newhcube = hypercube(image_PCA,wl_encode,meta3);
enviwrite(newhcube,Filename);
```

Below, we provide an example MATLAB code that can be used to call the MATLAB function and generate the DII and PCA images. In our implementation, the two ROIs input into the WCC workflow were drawn in ENVI, exported as comma separated value text files and then imported into MATLAB (see example dataset for more details).

```matlab
%Input Parameters
image_name='Water_only'
image_loc='D:\Water_attenuation_model\'
no_data_value=0
wl_range=[400 950]
SG_filt_deg=3
window_size=11
num_samples=10000
cor_thresh=0.9
percent_explained=95

%Run Function
[image_DII,DII_band_combs,image_PCA] = HSI_WCC(image_name,...
        image_loc, dw_roi,sw_mw_roi,no_data_value,wl_range,SG_filt_deg,...
        window_size,num_samples,cor_thresh, percent_explained);
```

Example dataset

Here we provide an example output from the WCC workflow applied to HSI data collected over the Long Sault Parkway near Cornwall, Ontario, Canada. The parkway is a connecting group of eleven islands created by flooding the Long Sault rapids during the construction of the St. Lawrence Seaway, which connects the North American Great Lakes to the North Atlantic. The specific example study site is a shallow bay just west of Philpott's Island in Long Sault Parkway provincial park (Fig. 3). The bay has a wide diversity in bottom cover and vegetation.

The HSI data used in the WCC workflow (see Fig. 4) was collected by the CASI-1500 hyperspectral imager (ITRES, Calgary, AB, Canada) operated by the National Research Council of Canada – Flight Research Laboratory. The CASI-1500 is a grating-based pushbroom imager with a 39.8° field of view that collects spectral information from 366 to 1053 nm across up to 288 bands with a silicon-based charged coupled device detector. To boost signal levels while capturing spectral information at a sufficiently fine spatial scale ($\sim$1 m) , the spectral bands were summed by a factor of 2 on-chip, resulting in imagery with 144 spectral bands. The imagery was radiometrically corrected with an IFRR applied, atmospherically compensated and geometrically corrected with a standard nearest neighbor resampling. The radiometric correction was carried out with proprietary software provided by the sensor manufacturer. Due to differences between the observed and anticipated at-sensor radiance, the IFRR (i.e., linear correction for mid to high radiance levels, non-linear correction for low radiance levels) developed in Soffer et al. [20] was implemented. The atmospheric compensation was completed in ATCOR4 v7.3.0 (ReSe Applications GmbH, Wil, Switzerland) [17]. The HSI data did not need to be de-glinted as the flight plan and calm atmospheric conditions resulted in no apparent sun glint in the imagery with reflectance of deep water in the NIR $<$ 1% across the field of view. It is important to note that a L_{SR} removal methodology should be applied if there is sun glint or other substantial water surface reflectance artifacts in the imagery. The geometric correction and spatial resampling steps were carried out by software provided by the sensor manufacturer, resulting in a raster HSI dataset with a spatial resolution of 1.16 m. The input parameters to the WCC workflow are shown in Table 1. The location of the two ROIs input into the WCC workflow are shown in Fig. 4; the green ROI is composed of deep water pixels while the red ROI consists of pixels over a constant substrate (cement road) along a transect of varying water column depths from 0.10 to 1.5 m. These ROIs were drawn in ENVI, exported as text files (see DW_roi.txt and sw_mw.txt in supplemental materials). Spectral bands with wavelengths above 950 nm were excluded from the analysis as they were almost completely attenuated by the water column, even at shallow depths ($<$ 10 cm) [19]. Additionally, spectral bands with wavelengths below 400 nm were removed due to low

Fig. 3. Photograph of Philpott's Island in Long Sault Parkway provincial park. The photograph was captured with the X5s camera aboard a DJI Inspire 2 remotely piloted aerial system.

Table 1

Parameters used in the water column compensation workflow applied to the CASI-1500 imagery collected over the Long Sault Parkway.

Water Column Compensation Workflow Parameter	Value
Wavelength Range	400–950 nm
Polynomial order for Savitzky-Golay Smoothing Filter	3
Window Size for Savitzky-Golay Smoothing Filter	11
Number of Pixels used in the Correlation Based Feature selection	10,000
Threshold used in the Correlation Based Feature selection	0.9
Percentage of the Total Variance Explained by the Principal Components Encoded in the PCA Image	95 %

signal levels. To avoid errors in the DII calculations during log transformation, bands with negative reflectance values were also removed from the analysis. In these bands, negative reflectance values have no physical significance and were retrieved during atmospheric compensation due to low signal-to-noise ratios. With the remaining 106 bands, the WCC workflow generated a total of 5565 DIIs. These DIIs were reduced to 124 using the correlation based feature extraction with a threshold value of 0.9. The remaining 124 DIIs were reduced to 31 features using the principal component analysis, generating components that explained 95% of the data variability.

Examples of the DII and PCA images are shown in Fig. 5B and Fig. 5C, respectively. In contrast to the original HSI data shown in Fig. 5A, the DII and PCA images show more variability across the site. For instance, in the true color image, the only easily identifiable shallow water substrates are the cement road and rocky bottom along the edges of each island (Fig. 5A), despite the known diversity in bottom cover at the site. In the DII and PCA images, however, various bottom types are clearly distinguished. In the DII image, dense stands of *Vallisneria americana* and *Potamogeton* sp. are visible
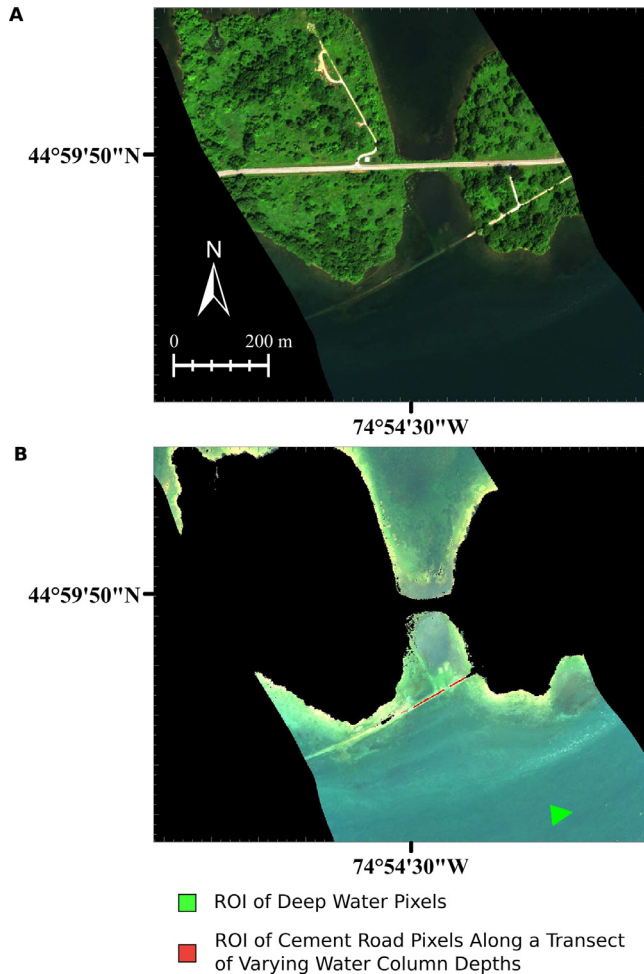
Fig. 4. (A) The CASI hyperspectral imaging data collected over the Long Sault Parkway (red = 639.6 nm, green = 548.7 nm, blue = 472.1 nm, linearly stretched between 0 and 13%). Before being input into the water column compensation workflow, the land pixels needed to be masked from the imagery. (B) The CASI hyperspectral imaging data (red = 639.6 nm, green = 548.7 nm, blue = 472.1 nm, linearly stretched between 0 and 3%) after masking out land pixels. The regions of interest (ROIs) required by the workflow are shown in the figure in green (deep water pixels) and red (pixels along a transect of varying water column depths over a constant substrate).

in addition to the cement road and rocky substrates (Fig. 5B). The PCA image can similarly be used to detect the variability in bottom cover types (e.g., *V. americana, Potamogeton* sp., rocky substrate seen in Fig. 5C).

As an illustration of the WCC workflow effectiveness, points of pure bottom cover were extracted from the untransformed hyperspectral, the DII and the PCA images across a 22 m by 22 m area (Fig. 6). The points were chosen according to manually defined polygons of bottom cover created using a high resolution RGB orthoimage of the area and in situ field knowledge of the site (Fig. 6A,B). The points extracted from the hyperspectral image were plotted across the 450–900 nm region (Fig. 6D). Points from the DII and PCA images were plotted against three axes (Fig. 6F, H). As the reflectance amplitude of aquatic materials depends heavily on the thickness of the overlying water column, first order statistics such as mean reflectance cannot be used to accurately differentiate between classes
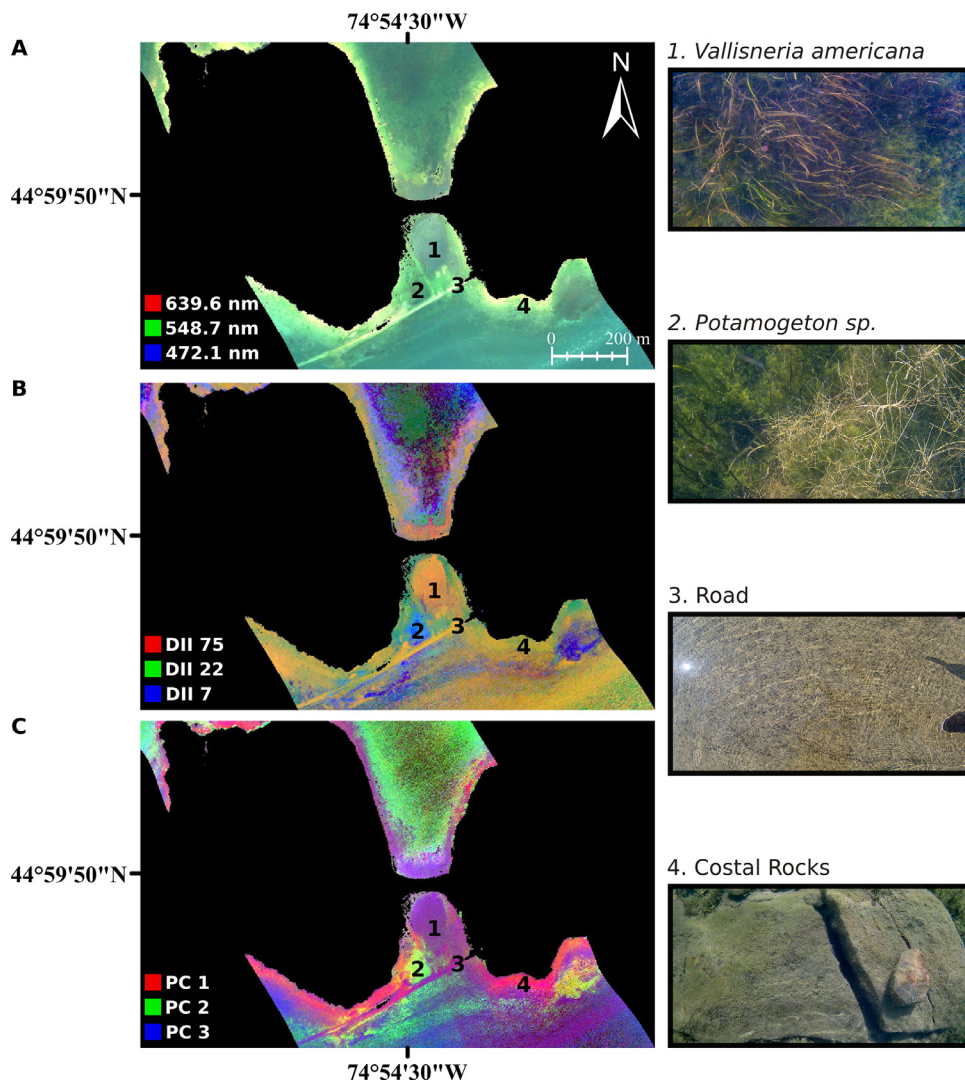
Fig. 5. Example output of the water column compensation workflow. (A) Input CASI hyperspectral imagery over Long Sault Parkway (red = 639.6 nm, green = 548.7 nm, blue = 472.1 nm, linearly stretched between 0 and 3%). (B) DII data product (red = DII 75 (682.7 nm & 701.8 nm), green = DII 22 (553.5 nm & 563.1 nm), blue = DII 7 (424.3 nm & 438.7 nm), linearly stretched from minimum to maximum value on extent). (C) PCA transformed DII data product (red = PC 1, green = PC 2, blue = PC 3, linearly stretched from minimum to maximum value on extent).

in the original HSI data. Furthermore, since there is no WCC, constituents within the water column itself are also contributing to the observed reflectance spectra, complicating applications further. The class plots extracted from the DII and PCA images demonstrate that after compensating for the effect of the water column, class clusters are discernable with as few as three dimensions (Fig. 6F, H).

Effective application of the water column compensation workflow

As previously mentioned, the described WCC workflow was successfully implemented in Rowan et al. [19] to detect submerged aquatic vegetation in the water surrounding the Long Sault Parkway.
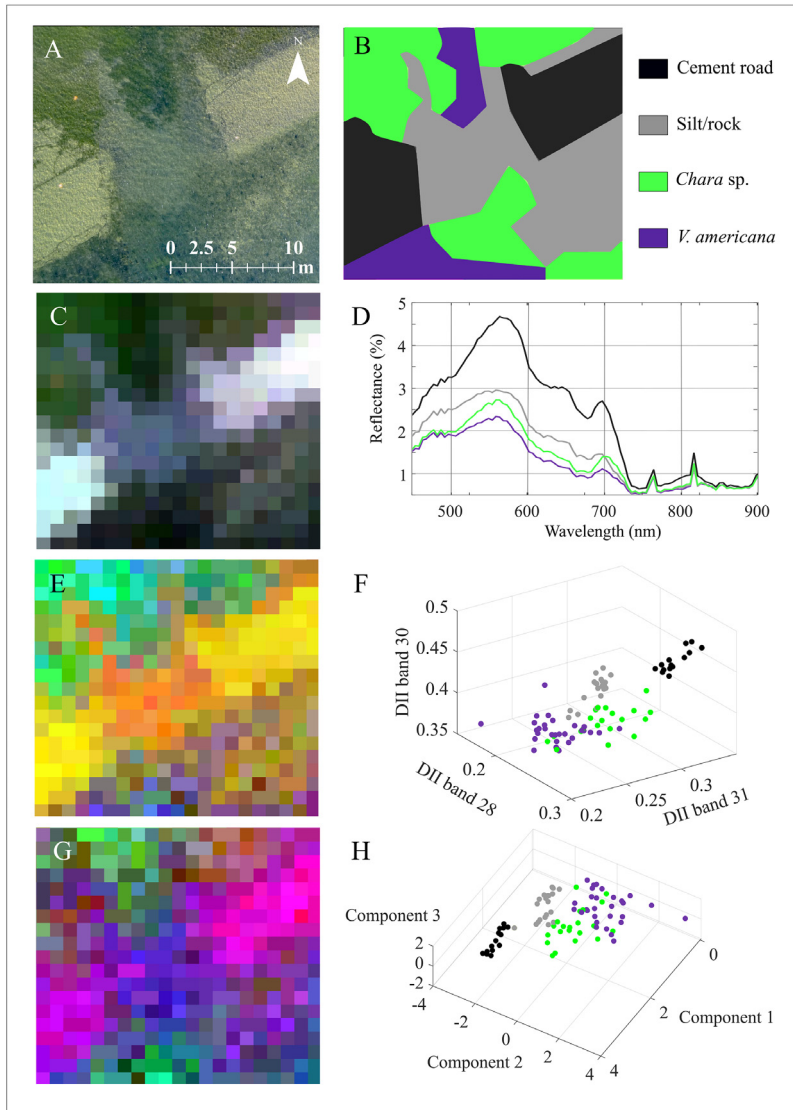
Fig. 6. Example application of the water column compensation workflow and resultant class spectra and plots demonstrating the improved performance of the DII and PCA as compared to original hyperspectral imagery. (A) High resolution RGB photograph of the study area (22 m by 22 m region) captured with an X5s camera aboard a DJI Inspire 2 remotely piloted aerial system. (B) Manual delineation of bottom cover classes defined from the orthomosaic and in situ field knowledge of the site. (C) Input CASI hyperspectral image over Long Sault Parkway (red = 644.4 nm, green = 548.7 nm, blue = 472.1 nm, linearly stretched from minimum to maximum value on extent). (D) Plot of mean class spectra extracted from the airborne hyperspectral image. (E) DII data product (red = DII 75 (682.7 nm & 701.8 nm), green = DII 22 (553.5 nm & 563.1 nm), blue = DII 7 (424.3 nm & 438.7 nm), linearly stretched from minimum to maximum value on extent). (F) Plots of the DII values of each class in the study area for DII bands 28 (572.63 nm + 577.42 nm), 30 (596.56 nm + 601.35 nm) and 31 (606.13 nm + 610.92 nm). The DII bands were selected as they had the greatest separability between the two vegetation classes (absolute difference between the mean DII for each class, normalized by the product of the standard deviation of the DII for each class). (G) PCA transformed DII data product (red = PC 1, green = PC 2, blue = PC 3, linearly stretched from minimum to maximum value on extent). (H) Plots of the PCA values of each class in the study area for the first three principal components. The legend in subplot B is applicable to subplots D, F and H.

In particular, a set of target detections was performed on water column compensated HSI data for various classes of interest (i.e., five canopy-forming vegetation types, the paved asphalt road, silt/rock, all vegetation combined, and a non-vegetation class). For pure pixels, the overall recall of the target detection was 87.8% across the individual operational taxonomic units and two non-vegetation classes, and 93.6% for the binary vegetation/non-vegetation classes. In these target detections, submerged aquatic vegetation was detected in 5527 m^2 of the 66,148 m^2 of identified water area. Such an approach to monitoring submerged aquatic vegetation could be beneficial in water body management (e.g., prevent the establishment of tall stands of vegetation near water intakes). This application shows how this WCC workflow, and the associated MATLAB script, can easily be used and modified for HSI data applications over shallow, *clear* to *moderate* optical water types.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2021.101601.

References

[1] M.A. Amran, R.A. Rappe, Estimation of seagrass coverage by depth invariant indices on quickbird imagery, Biotropia 17 (2010) 42–50.
[2] H.M. Dierssen, S.G. Ackleson, K.E. Joyce, E.L. Hestir, A. Castagna, S. Lavender, M.A. McManus, Living up to the hype of hyperspectral aquatic remote sensing: science, resources and outlook, Front. Environ. Sci. 9 (2021).
[3] M.T. Eismann, 1.1 Hyperspectral remote sensing, Hyperspectral Remote Sensing, SPIE, Bellingham, 2012.
[4] B. El Mahrad, A. Newton, J.D. Icely, I. Kacimi, S. Abalansa, M. Snoussi, Contribution of remote sensing technologies to a holistic coastal and marine environmental management framework: a review, Remote Sens. 12 (2020) 2313.
[5] H.R. Gordon, W.R. McCluney, Estimation of the depth of sunlight penetration in the sea for remote sensing, Appl. Opt. 14 (1975) 413–416.
[6] E. Green, P. Mumby, A. Edwards, C. Clark, Mapping bathymetry, in: Remote Sensing: Handbook for Tropical Coastal Management, United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris, 2000, pp. 219–234.
[7] D.R. Lyzenga, Passive remote sensing techniques for mapping water depth and bottom features, Appl. Opt. 17 (1978) 379–383.
[8] D.R. Lyzenga, Remote sensing of bottom reflectance and water attenuation parameters in shallow water using aircraft and Landsat data, Int. J. Remote Sens. 2 (1981) 71–82.
[9] M.D. Manessa, A. Kanno, M. Sekine, E.E. Ampou, N. Widagti, A.R. As-syakur, Shallow-water benthic identification using multispectral satellite imagery: investigation on the effects of improving noise correction method and spectral cover, Remote Sens. 6 (2014).
[10] L.I.W. McKinna, P.J. Werdell, Approach for identifying optically shallow pixels when processing ocean-color imagery, Opt. Express 26 (2018) A915–A928.
[11] C. Mobley, Light and Water: Radiative Transfer in Natural Waters, Academic Press, 1994.
[12] A. Morel, L. Prieur, Analysis of variations in ocean color 1, Limnol. Oceanogr. 22 (1977) 709–722.
[13] P.J. Mumby, C.D. Clark, E.P. Green, A.J. Edwards, Benefits of water column correction and contextual editing for mapping coral reefs, Int. J. Remote Sens. 19 (1998) 203–210.
[14] P.J. Mumby, A.J. Edwards, Mapping marine environments with IKONOS imagery: enhanced spatial resolution can deliver greater thematic accuracy, Remote Sens. Environ. 82 (2002) 248–257.
[15] M. Nurlidiasari, S. Budiman, Mapping coral reef habitat with and without water column correction using Quickbird image, Int. J. Remote Sens. Earth Sci. (IJReSES) 2 (2010).
[16] R. Pu, S. Bell, C. Meyer, L. Baggett, Y. Zhao, Mapping and assessing seagrass along the western coast of Florida using Landsat TM and EO-1 ALI/Hyperion imagery, Estuarine Coast. Shelf Sci. 115 (2012) 234–245.
[17] R. Richter, D. Schläpfer, ATCOR-4 User Guide, Version 7.3.0, March 2020, Atmospheric/Topographic Correction for Airborne Imagery, DLR - German Aerospace Center, Wessling, 2020.

[18] G.S.L. Rowan, M. Kalacska, A review of remote sensing of submerged aquatic vegetation for non-specialists, Remote Sens. 13 (2021) 623.

[19] G.S.L. Rowan, M. Kalacska, D. Inamdar, J.P. Arroyo-Mora, R. Soffer, Multi-scale spectral separability of submerged aquatic vegetation species in a freshwater ecosystem, Front. Environ. Sci. 9 (2021).

[20] R. Soffer, G. Ifimov, Y. Pan, S. Belanger, Acquisition and spectroradiometric assessment of the novel WaterSat Imaging Spectrometer Experiment (WISE) sensor for the mapping of optically shallow coastal waters, OSA Optical Sensors and Sensing Congress, OSA Publishing, Virtual Event, 2021.

[21] K. Uudeberg, I. Ansko, G. Põru, A. Ansper, A. Reinart, Using optical water types to monitor changes in optically complex inland and coastal waters, Remote Sens. 11 (2019) 2297.

[22] C. Vaiphasa, Consideration of smoothing techniques for hyperspectral remote sensing, ISPRS J. Photogramm. Remote Sens. 60 (2006) 91–99.

[23] C. Wezernak, D. Lyzenga, Analysis of Cladophora distribution in Lake Ontario using remote sensing, Remote Sens. Environ. 4 (1975) 37–48.

[24] A.E. Windle, G.M. Silsbe, Evaluation of unoccupied aircraft system (UAS) remote sensing reflectance retrievals for water quality monitoring in coastal waters, Front. Environ. Sci. 9 (2021).

[25] C. Zhang, D. Selch, Z. Xie, C. Roberts, H. Cooper, G. Chen, Object-based benthic habitat mapping in the Florida Keys from hyperspectral imagery, Estuarine Coast. Shelf Sci. 134 (2013) 88–97.