VARIABILITY OF REMOTE SENSING SPECTRAL INDICES IN BOREAL LAKE BASINS

Taina Hakala[1] , Ilkka Pölönen[1], Eija Honkavaara[2], Roope Näsi[2], Teemu Hakala[2], Antti Lindfors[3]

[1] Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland - (taina.hakala, ilkka.polonen)@jyu.fi
[2] Finnish Geospatial Research Institute FGI, Kirkkonummi, Finland - (eija.honkavaara, roope.nasi, teemu.hakala)@nls.fi
[3] Luode Consulting Ltd, Espoo, Finland - antti.lindfors@luode.net

ICWG II/III

KEY WORDS: Water quality, Optically complex waters, Remote sensing, Hyperspectral imaging, Spectral indices

ABSTRACT:

Remotely sensed hyperspectral data has widely been used to determine water quality parameters in oceanic waters. However in freshwater basins the dependence between the hyperspectral data and the parameters is more complicated. In this work some ideas are presented concerning the study of this dependence. The data used in this study were collected from the lake Hiidenvesi in southern Finland. The hyperspectral data consists of reflectances in 36 bands in the wavelength area 508...878 nm and the separately measured water quality parameters are turbidity, blue-green algae, chlorophyll, pH and dissolved oxygen. Hyperspectral data was used as bare band reflectances, but also in the form of two simple spectral indices: ratio A/B and difference A-B, where A and B go through all the bands. The correlations of the indices with the parameters were presented visually as 1- or 2-dimensional arrays. To examine the significance on the results of different variables, the data was classified in two different ways: the natural basins and the values of the water quality parameters. It was noticed that the variability of the correlation arrays was particularly strong among different basins in both the magnitude of correlation and the best performing indices. Further studies are needed to clarify which features of the basins are of most importance in predicting the shapes of the correlation arrays.

1. INTRODUCTION

Boreal lakes are one of the worlds largest fresh water storages. Worldwide fresh water is a limited natural resource. In Nordic countries state and county authorities have the responsibility to monitor the condition of water sources. This can be done by in-situ measurements, but there is need for remotely sensed monitoring methods, which can be utilized also in cloudy weather. Hyperspectral imaging using manned small planes or unmanned aerial vehicle, can help in this challenge. The aim of this work is to study how hyperspectral data could be used to investigate water quality parameters in inland waters.

Open ocean waters (often referred as Case 1 waters) have been successfully studied with multispectral cameras carried by satellites. Multispectral data is somewhat similar to hyperspectral data, but the number of the bands is smaller and the bands are wider. In the case of inland and near- coastal waters (Case 2 waters) multispectral data has not proved as functional since waters of this type are optically more complex (Odermatt et al., 2012).

When hyperspectral data is used to investigate water areas, the aim is to associate the characteristics of the hyperspectral image to the values of certain water quality parameters. In the simplest form this means that from the collected hyperspectral and parameter value data it is observed that a certain hyperspectral band correlates well with some parameter value. In principle, we can then use a similar hyperspectral image to predict the values of the parameter in somewhere else.

Instead of single bands, hyperspectral data can also be used in the form of spectral indices, which are arithmetical operations among different bands. The idea behind indices is that they often provide more explicit correlations with the water quality parameters than bare band reflectances. The correlation of an index with a water quality parameter depends on how the spectral signature (see 3.1) changes when the parameter value increases.

Traditionally spectral indices have been used with multispectral data for certain fixed bands. A familiar example is the normalized difference vegetation index (NDVI), which is defined as (NIR - Red)/(NIR + Red), where NIR and Red are the near infrared and red bands respectively. Hyperspectral data consists of a larger number of narrow bands, so for a given index type it is reasonable to ask which bands work best in estimating a certain parameter. To this end the indices of the given type are calculated with all the band combinations and further the correlations of the obtained indices with the parameter are determined. The correlation results can then be presented in an array with dimension equal to the number of independent bands in the index. Indices have been used in this way for example in estimating biomass in (Näsi et al., 2017).

This study applies these methods in water quality analysis and also aims to make the concepts more precise. In section 2.3 index families and their dimensions along with visualization methods are introduced. Practical examples of these visualizations are seen in section 4.

To investigate which features are mainly responsible for the variation of the correlation patterns, the data is divided into classes in two ways (see 2.2). First classification is done by the lakes natural basins to see how the basin type affects the results. Secondly the classification is done separately for each parameter by the parameters value, which helps to see how the results depend purely on the parameters. To connect the variation in the correlation results with the properties of the data, the notions of spectral reflectance signature and parameter profile are defined in section

Corresponding author

https://doi.org/10.5194/isprs-archives-XLII-2-411-2018 | © Authors 2018. CC BY 4.0 License.

depend on the equipment used to collect the hyperspectral data. Two sets of bands with similar central wavelengths but different widths can give rise to different correlation arrays for the same index family.

3. DATA CHARACTERISTICS

3.1 Spectral reflectance signature

To characterize the spectral properties of the data, a histogram was made for each band by dividing the reflectance axis into bins of equal size and for each bin finding the number of points for which the reflectance in this band falls into the bin. Histograms for all bands have been depicted in one figure by organizing them vertically side by side and presenting the numeric values (bin sizes) with a colormap for which lighter colors indicate larger numbers. In the figures obtained this way the lighter parts form a curve, which in a way represents the ’average’ reflectance spectrum of this dataset. These figures are called *spectral reflectance signatures*, sometimes abbreviated as spectra.

Figure 2 depicts the spectral reflectance signatures of the data from distinct basins of Hiidenvesi. The general shape of these spectra with a peak around the band 10 = R_{575} [1] is typical to some turbid waters (Doraxan et al., 2002). From figure 2 it can be seen that in basins 2 and 3 the peak is clearly higher compared to the other basins.

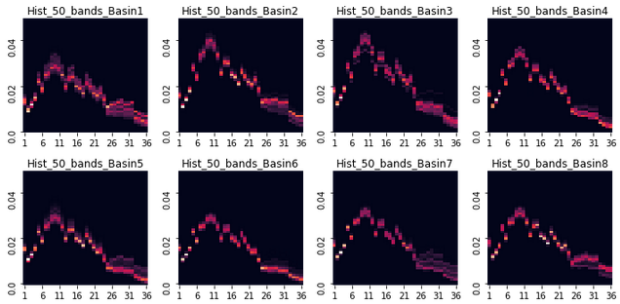

Figure 2. Spectral reflectance signatures of the basins

Figure 3 shows the spectral reflectance signatures for the low-high classes of the parameters BGA and ODO respectively. For BGA the spectral signature of the low- and high-classes differ notably: the spectrum of the class BGA_low has lower peak and is lucid, whereas the spectrum of BGA_high has higher averages and is more spread out. On the other hand for ODO the two spectra of the classes ODO_low and ODO_high look rather similar. This may indicate that the reflectance spectrum of the water is more sensitive to the values of the parameter BGA than those of the parameter ODO. It can also be noted that the peaks of the spectral signatures of the basins 2 and 3 in figure 2 are much clearer than the spread-out one of the class BGA_high.

The other parameters Turbid, CHL and pH behave in many respects similarly to BGA, so in here and what follows their corresponding figures have been left out. All the figures used in this study can be found at (Hakala, 2018).

[1] This notation indicates that band 10 corresponds to reflectance at the wavelength 575 nm.

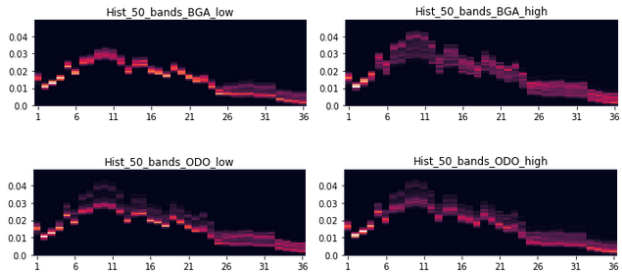

Figure 3. Spectral reflectance signatures for the low- and high-classes of the parameters BGA and ODO

3.2 Parameter profiles

Figures similar to the spectral reflectance signatures can also be made using the parameter values instead of the band reflectances. These figures are not as generic as the spectral reflectance signatures, since the set of water quality parameters used is not standardized nor have they any natural ordering. With a fixed parameter set however these figures can be useful in understanding the essence of the parameters behavior in different classes of the data. This kind of figures are called *parameter profiles*.

Figure 4 shows the parameter profiles of the distinct basins. Note that since the units and scales of the parameters vary, the y-axis cannot be given any reasonable numeric scale. For this reason the y-axis here shows only the common bin-numbers for the histograms of different parameters.

From figure 4 it can be seen that the basins 6, 7 and 8 have quite similar parameter profiles, notably they have very low turbidity and BGA-CHL-content. Basins 2 and 3 are most turbid and basin 2 has exceptionally high BGA and CHL content. Comparing figures 2 and 4 it can be hypothesized that the sharpness of the peak of the spectrum could be related mainly to the turbidity of the water.

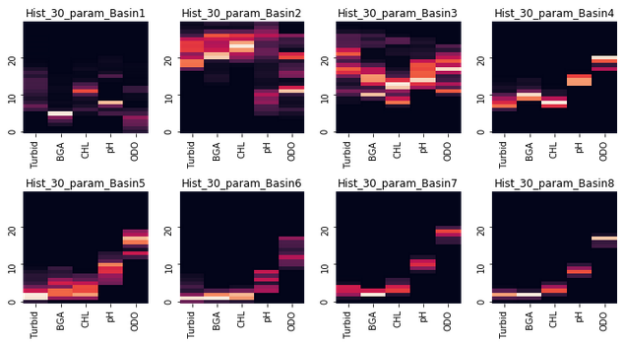

Figure 4. Parameter profiles of the basins

Parameter profiles were also constructed for the different parameter value classes and figure 5 shows these for the parameters BGA and ODO. Here the low- and high-classes are defined separately for each parameter, which can be verified from the figures by observing that the column of the parameter in question has low (high) values in the corresponding low (high) class.

The profiles of the BGA low- and high-classes (top row of figure 5) are very unlike. We see that the values of Turbid and CHL (partly also pH) are low in the class BGA_low and high in the

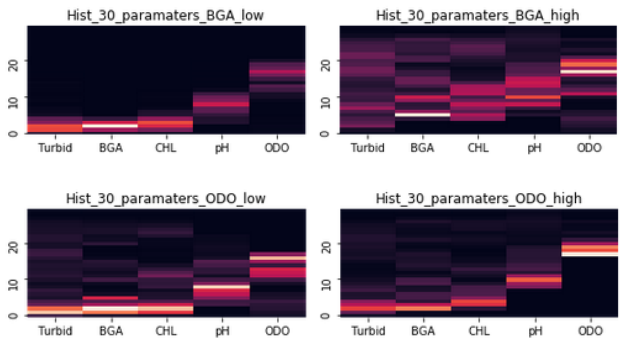

Figure 5. Parameter profiles for the low- and high-classes of the parameters BGA and ODO

class BGA_high, and also the deviations are much higher in the latter class. This indicates that Turbid, CHL (and pH) are mutually correlated to BGA and that the median used as the division point is close to the lower end of the scale.

For the parameters Turbid, CHL and pH the behavior of the spectral reflectance signatures and the parameter profiles of the low- and high-classes is similar to BGA. The large deviation in the spectral reflectance signature of the class BGA_high in figure 3 seems to stem from the deviation of the values of these correlating parameters.

The parameters profiles of the ODO low- and high-classes (bottom row of figure 5) are rather similar (except of course the ODO-columns which show the division to low- and high-classes). This indicates that the parameter ODO is quite independent of the other parameters.

4. INDEX CORRELATIONS

4.1 Band correlations

As explained in section 2.3, the unique 1-dimensional index family consists of the hyperspectral bands. The correlation array of this index family with a given water quality parameter can be displayed as a 1-dimensional array of boxes such that each box represents a band and its color corresponds to the correlation of the parameter with the reflectance in the band.

The upper part of figure 6 shows the correlation arrays of the parameter BGA with the 1-dimensional index family sorted by basins. For comparison, the last row depicts similar correlation array for the whole dataset. The lower part shows the equivalent correlation arrays for the parameter ODO. Note that in the colormap of these figures the sign of the correlation matters, i. e. significant correlations are represented by either dark red or dark blue.

If the horizontal rows in figure 6 (representing the basins) were similar to each other, one could easily determine the bands which generally correlate best with the parameters BGA and ODO respectively. In reality however, the rows are notably different, which means that for different basins the best correlations with the parameters are given by different bands.

Figure 7 also shows the correlation arrays of the 1-dimensional index family with the parameters BGA and ODO, but here the

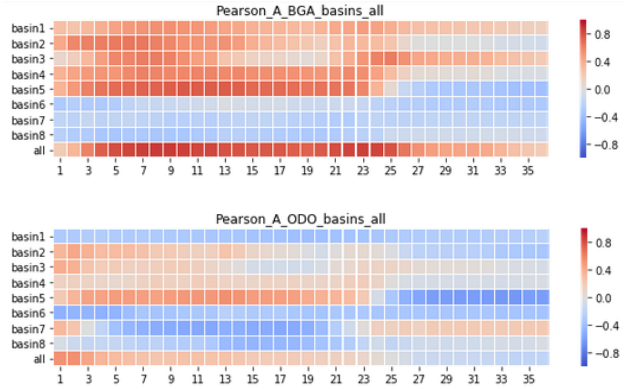

Figure 6. Band correlations for the parameters BGA and ODO, basins separately and the whole data

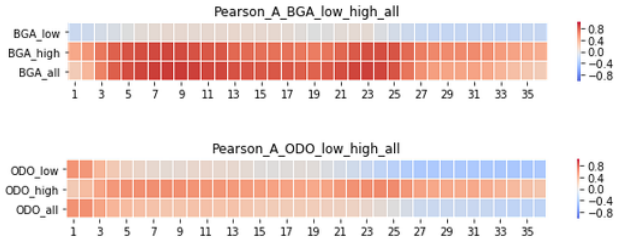

Figure 7. Band correlations for the parameters BGA and ODO, parameter classes separately and the whole data

classification is done by the parameter classes instead of basins (again the last row represents whole data).

From figure 7 it can be seen that for BGA the correlations in the high-class are much stronger than those in the low-class and that the correlation array of the whole data is very similar to that of the high-class. This could be related to the fact that the deviation in the high-class of BGA was clearly greater in both the spectral reflectance signature and parameter profile (see figures 3 and 5). For the parameter ODO the difference between the correlation arrays of the low- and high-classes is striking and here the array of the low-class looks more similar to the one concerning the whole data.

4.2 Correlations with index family A/B

For the 2-dimensional index family A/B the correlation array with a water quality parameter is 2-dimensional and can be displayed as a pixelated square of size 36*36. Each pixel corresponds to a unique ordered pair (A, B) of bands and its color represents the correlation of the index A/B with the parameter in question. The correlation figures of the index family A/B are close to skew symmetric, because the indices reflected by the main diagonal are multiplicative inverses. The deviations from this symmetry occur mainly in the weakly correlating areas of the array. In all the subsequent 2-dimensional correlation figures the band A is on the vertical axis and the band B on the horizontal axis.

Figure 8 shows the correlation arrays of the index family A/B with the parameters BGA and ODO for the whole data. The index family A/B behaves very differently for the different parameters BGA and ODO. For the parameters Turbid, CHL and pH however

the corresponding figures are quite similar to the one for BGA, likely due to the high mutual correlation of these parameters.

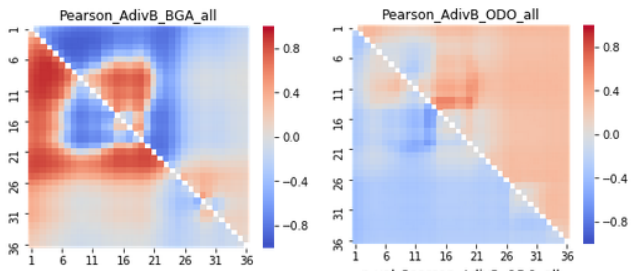

Figure 8. A/B-correlations for the parameters BGA and ODO in the whole data

In figures 9 and 10 the correlation of the index family A/B with the paramaters BGA and ODO is depicted with the data divided into basins. These figures show that significant variation on the shape of the correlation arrays occurs from one basin to another: e.g. the correlations for ODO look almost opposite for the basins 5 and 7.

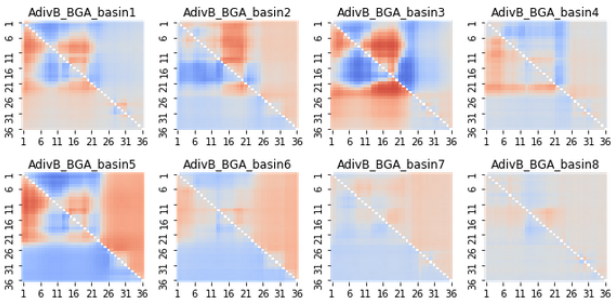

Figure 9. A/B-correlations for the parameter BGA classified by basins

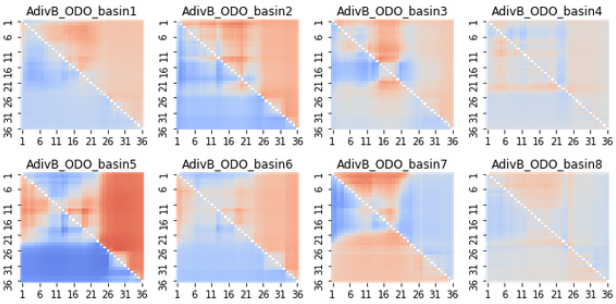

Figure 10. A/B-correlations for the parameter ODO classified by basins

For some basins the overall correlation is weak. Comparing figure 9 to figure 4 it can be observed that for the parameter BGA the overall strength of the correlation figure seems to be related to the deviation of the parameter in distinct basins so that weak correlation corresponds to small deviation. The same observation in fact holds also for the parameters Turbid, CHL and pH. For the parameter ODO this does not hold, since e.g. the basins 1 and 6 have weak correlation figures but rather large deviations. For the parameter ODO the correlations in the basin 5 are actually stronger than those in the whole data, whereas for BGA (also Turbid, CHL and pH) the correlations in the individual basins are generally lower than in the whole data.

When the data is divided by basins, the correlation figures of the parameters of even the same correlation class (BGA, Turbid, CHL, pH) no longer look similar. On the other hand, also the mutual correlation of these parameters breaks up in this division. (Hakala, 2018)

Figures 11 and 12 show the correlation of the family A/B with the parameters BGA and ODO with the data divided into the low- and high- classes of the parameters BGA and ODO respectively. Figure 11 reveals that for BGA the overall correlation is much stronger in the high-class, which may be associated to the larger deviation of BGA in this class (see figure 5). The parameters Turbid, CHL and pH again behave very similarly.

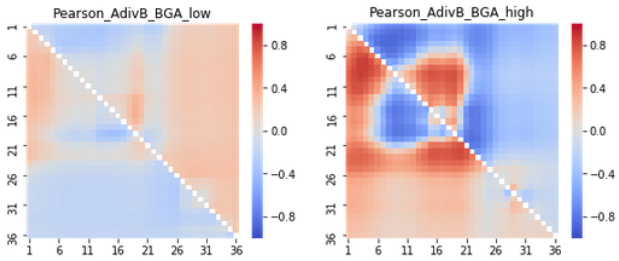

Figure 11. A/B-correlations for the parameter BGA by the parameter classes

For the parameter ODO in figure 12 the correlation strengths in the low- and high-classes are quite similar, although the patterns are very different. This may be surprising as the spectral reflectance signatures of the classes ODO_low and ODO_high (figure 3) look almost the same. Although the behavior of parameter ODO usually deviates from the other parameters, for some reason the correlation figure of the high-class of ODO looks similar to the high-classes of the other parameters.

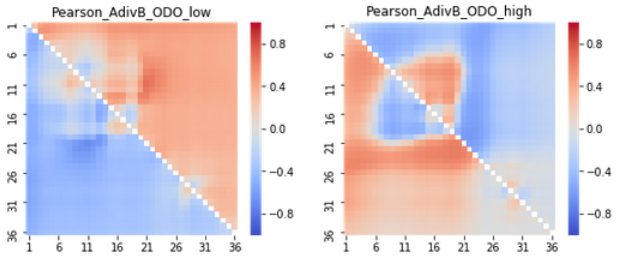

Figure 12. A/B-correlations for the parameter ODO by the parameter classes

4.3 Correlations with index family A-B

The correlation arrays for the 2-dimensional index family $A - B$ are also squares of size 36*36 and can be visualized in the same way as those of the family A / B in section 4.2. For this family the indices reflected by the main diagonal are additive inverses, so the correlation figures are completely skew symmetric.

The behavior of this index family with respect to the different partitions of the data is rather similar to the family A/B, so to avoid repetition the correlation figures for this family will not be presented here in totality. Figure 13 of the $A - B$ -correlation

for BGA and ODO in the whole data serves as an example of the general shape of the correlation array.

Compared to the corresponding figures of the family A / B (see figure 8), the pattern in the figures for A − B is more ’broken’ with vertical and horizontal lines. This makes it a little harder to sort out the distinct areas of strong correlation.

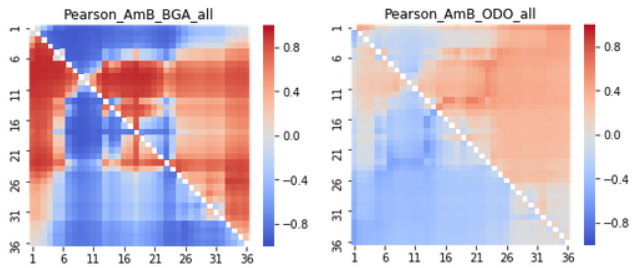

Figure 13. A-B-correlations for the parameters BGA and ODO in the whole data

5. APPLICATIONS

If an index could be found that correlates generally well with a given water quality parameter, then the values of the parameter could be quantitatively predicted with remotely sensed hyperspectral data. When handling a new set of data a calibration should be performed. In the simplest case this would be done by searching the geographical locations of the highest and lowest index values and performing the parameter measurements at these locations. Due to the strong correlation the scatter figure of the parameter vs the index should be close to a straight line with endpoints prescribed by the measurements.

The problem is that the properties of even the distinct basins of the same lake can be so varied that no common well performing index can be found. For example from figure 8 it can be seen that the best correlating index of the family A/B with the parameter BGA is B9 / B3 = $\mathrm{R}_{567}/\mathrm{R}_{516}$ [2](Pearson correlation 0.91). Figure 9 shows that in basin 2 the correlation of the index B9 / B3 with BGA is very weak (Pearson 0.03), whereas the index B13 / B17 = $\mathrm{R}_{610}/\mathrm{R}_{662}$ works a lot better (Pearson 0.63). In figure 14 both of these dependencies are shown as scatter plots color coded by the basins. For the basin 2 the index B9 / B3 of figure 14 (a) does not depend on BGA, but for the index B13 / B17 of figure 14 (b) the dependency is obvious.

The solution for this problem would be to find the essential properties of a water basin which determine the shape of the correlation figure concerning given parameter and index family. Then the correct index to be used in the calibration could be selected in advance based on these properties.

Another idea for utilizing the correlation figures of section 4 could be in finding variables that correlate strongly with a given parameter but mutually weakly. E. g. multilinear regression can be made more reliable using this kind of independent variables. From an index family we can always find indices with the strongest correlation with the parameter, but selection based only on the value of the correlation often gives neighboring indices which have strong mutual correlation. E. g. in figure 8 the

[2]Here e.g. B9 means band 9 and R_{567} refers to reflectance at the wavelength 567 nm.

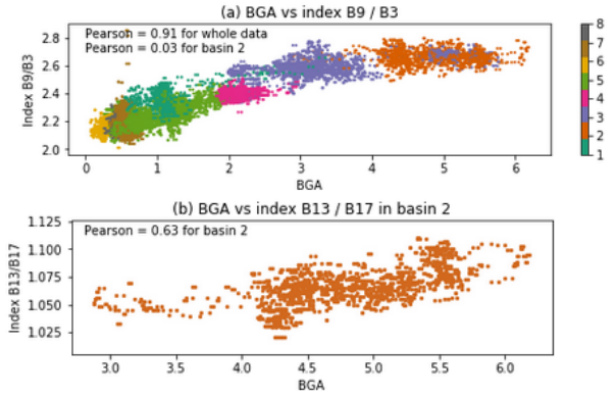

Figure 14. Scatter figures of the parameter BGA with the index B9 / B3 in the whole lake and B13 / B17 in basin 2

indices of the A/B-family which correlate best with BGA are located in the neighborhood of the optimal index B9 / B3 and by continuity are correlated also to each other. Considering the total structure of the figure instead of the correlation values, several regions of high correlation can be found. By choosing from each of these regions the index giving the local maximum of the correlation we obtain a set of indices with no built-in dependencies and from these indices the ones with low mutual correlations can be picked. Of course one can look for these non-correlating indices from several index families simultaneously.

For example the parameter BGA correlates quite well with the index B26 − B33 = $\mathrm{R}_{743} - \mathrm{R}_{838}$ from the family A-B and the index B9 / B19 = $\mathrm{R}_{567}/\mathrm{R}_{682}$ from the family A/B (correlations 0.86 and 0.78 respectively) and the mutual correlation of these indices is rather low (0.56). Figure 15 depicts the scatter plots of these indices with BGA: usually lower mutual correlation of the indices means more disparity in the corresponding scatter plots.

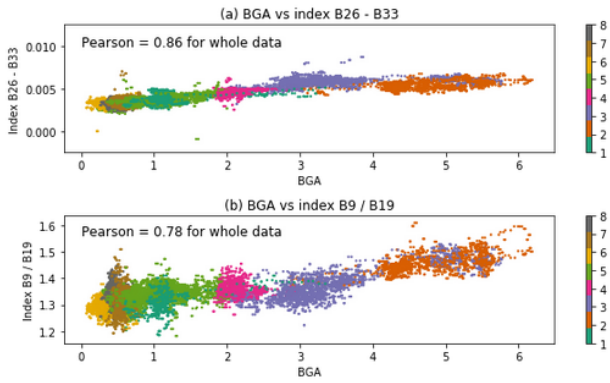

Figure 15. Scatter figures of the parameter BGA with the indices B26 − B33 and B9 / B19 (whose mutual correlation is 0.56)

The scatter plots of figure 15 have been color coded by basins in the same way as in figure 14 (a), so that the behavior of the indices in distinct basins can be investigated. It seems that among the three indices described here the index B26 − B33 seems to have the best overall correlation with BGA, even though some disjoint points lower the correlation value. Especially basins 6, 7 and 8 which are problematic for the other indices, are quite well aligned with the rest of the data for the index B26 − B33. This kind of scatter plots can be useful in examining the behavior of individual indices in the whole data and in different basins.

6. CONCLUSIONS

In this work the correlation behavior of hyperspectral bands and the 2-dimensional index families A/B and A-B with different water quality parameters was investigated using data from the lake Hiidenvesi. It was noted that the type of the basin seems to affect strongly on both the overall strength of the correlations and the set of the indices which correlate best with the given parameter. Typically it is not easy to find any single index which would correlate even adequately in all basins. The main reasons for the great variability of the correlation arrays among different basins are still unclear.

The parameters BGA, Turbid and CHL are absorbing substances and mutually strongly correlating. Also the parameter pH usually correlates moderately with these parameters, even if it is of a different category. The mutual correlation implies rather similar behavior for these parameters in most cases. However when dividing the data by the basins the mutual correlation of these parameters varies and so do the correlations with the index families. For the parameters BGA, Turbid, CHL and pH the shape of the spectral reflectance signature seems to depend on the value of the parameter so that larger parameter values gives sharper peak in the spectrum. Also for these parameters the overall correlation strength with an index family seems to be related to the deviation of the parameter values in a class of data.

The parameter ODO is non-absorbing and its behavior with respect to hyperspectral data departs clearly from the other parameters. E. g. the concentration of ODO does not seem to notably affect the spectral reflectance signature. In contrast the ODO concentration seems to have great impact on the correlation figures with the index families.

The correlation figure of a parameter with an index family typically contains a few areas of strong correlation. The objective is to find the essential properties of basins which determine the shape of these correlation figures. If this works out the type of a given basin could be used to choose the best index for estimating a certain parameter and quantitative predictions of the parameter values could be made based on hyperspectral data.

The methods examined in this work are simple and their fuctionality may be restricted by the non-linear behavior of the data. The same data from the lake Hiidenvesi has also been studied using convolutional neural networks (CNNs), which are neural network models that approximate the neuron model using convolutional computation. According to the preliminary results in this domain, CNNs seem to capture the non-linear behaviour of the data and estimating the parameter values in different basins of the lake using this method has been rather succesful.

ACKNOWLEDGEMENTS

This research has been co-funded by Finnish Funding Agency for Innovation Tekes (grants 2208/31/2013 and 1711/31/2016).

REFERENCES

Doraxan, D., Froidefond, J.-M., Lavender, S. and Castaing, P., 2002. Spectral signature of highly turbid waters application with spot data to quantify suspended particulate matter concentrations. *Remote Sensing of Environment* 81, pp. 149–161.

Hakala, T., 2018. Supplementary figures for the article ’Variability of remote sensing spectral indices in boreal lake basins’. Available online at https://doi.org/10.5281/zenodo.1213025.

Honkavaara, E., Saari, H., Kaivosoja, J., Pölönen, I., Hakala, T., Litkey, P., Mäkynen, J. and Pesonen, L., 2013. Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight uav spectral camera for precision agriculture. *Remote Sensing* 5(10), pp. 5006–5039.

Markelin, L., Honkavaara, E., Näsi, R., Nurminen, K. and Hakala, T., 2014. Geometric processing workflow for vertical and oblique hyperspectral frame images collected using uav. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Zurich, Switzerland, Vol. XL-3, pp. 205–210.

Näsi, R., Viljanen, N., Kaivosoja, J., Hakala, T., Pandžić, M., Markelin, L. and Honkavaara, E., 2017. Assessment of various remote sensing technologies in biomass and nitrogen content estimation using an agricultural test field. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Jyväskylä, Finland, Vol. XLII-3/W3, pp. 137–141.

Odermatt, D., Gitelson, A., Brando, V. E. and Schaepman, M., 2012. Review of constituent retrieval in optically deep and complex waters from satellite imagery. *Remote Sensing of Environment* 118, pp. 116–126.

Ranta, E., Valtonen, M. and Ikonen, E., 2015. Hiidenveden alueen yhteistarkkailun yhteenveto vuodelta 2015. Technical report, Länsi-Uudenmaan vesi ja ympäristö ry, Julkaisu 265/2016.

Saari, H.and Pölönen, I., Salo, H., Honkavaara, E., Hakala, T., Holmlund, C., Mäkynen, J., Mannila, R., Antila, T. and Akujärvi, A., 2013. Miniaturized hyperspectral imager calibration and uav flight campaigns. In: *Proceedings 8889, Sensors, Systems, and Next-Generation Satellites XVII*, Dresden, Germany, Vol. 8889, pp. 88891O–88891O–12.