

THESIS ON INFORMATICS AND SYSTEM ENGINEERING C95

**Hyperspectral Data Processing and
Interpretation in Remote Sensing Based on
Laser-Induced Fluorescence Method**

INNOKENTI SOBOLEV

TUT
PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
Department of Computer Control

This dissertation was accepted for the defence of the degree of Doctor of Philosophy in Informatics and System Engineering on July 04, 2014

Supervisors: Ph.D., Associate Professor Eduard Petlenkov, Department of Computer Control, Tallinn University of Technology

Dr., CEO at LDI Innovation Sergey Babichenko, Estonia

Opponents: Dr., Professor Rainer Reuter, Institut für Physik Meeresphysik, Carl von Ossietzky Universität, Oldenburg, Germany

Dr., Giovanna Cecchi, National Research Council of Italy (CNR), National Institute of Optics (INO), Florence, Italy

Defence of the thesis: August 26, 2014

Declaration:

Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree.

/Innokenti Sobolev/



Copyright: Innokenti Sobolev, 2014

ISSN 1406-4731

ISBN 978-9949-23-668-8 (publication)

ISBN 978-9949-23-669-5 (PDF)

INFORMAATIKA JA SÜSTEEMITEHNIKA C95

**Hüperspektraalse andmete analüüs ja
tõlgendamine kaugseires *laser-induced-
fluorescence*-meetodi põhjal**

INNOKENTI SOBOLEV

Contents

List of publications constituting the thesis	9
List of Abbreviations	11
Introduction	13
Problem statement	14
State of the art	15
Outline of the thesis	18
1 Systematization of HLIF data for OPD	21
1.1 Author's contribution	21
1.2 The HLIF spectrum	22
1.2.1 The HLIF spectra mixture	22
1.2.2 Library of spectral end-members	23
1.2.3 Natural fluorescence background	24
1.3 Specifics of water and land applications	27
1.4 HLIF models	28
1.4.1 Discriminant measure for HLIF data	29
1.5 Examples	31
1.5.1 Natural water HLIF linear unmixing	31
1.5.2 Natural water HLIF deformation detection	33
1.6 Conclusions	35
2 Exploration of HLIF signal	37
2.1 Author's contribution	38
2.2 Theory of Wavelet Transform	38
2.2.1 Wavelets	39
2.2.2 Continuous wavelet spectrum	41
2.2.3 Multiresolution Discrete Wavelet Transform	43
2.2.4 Lifting	47
2.3 HLIF signal-noise model	52
2.3.1 HLIF LIDAR noise sources	52
2.3.2 Poisson-Gaussian noise model	52

2.3.3	Noise parameters estimation of measured HLIF LIDAR data	59
2.4	Conclusions	59
3	HLIF data analysis	61
3.1	Author's contribution	62
3.2	Scale characteristics of HLIF spectra	63
3.2.1	Selection of optimal wavelet basis	67
3.2.2	Multiscale HLIF spectra comparison	70
3.2.3	Numerical example	71
3.3	Wavelet based feature extraction	75
3.3.1	HLIF feature extraction principle	75
3.3.2	Feature extraction examples	79
3.3.3	HLIF feature denoising	82
3.3.4	Feature denoising examples	85
3.4	Conclusions	88
4	HLIF diagnostics platform and OPD applications	89
4.1	Author's contribution	89
4.2	Experimental datasets	90
4.2.1	LIDAR Underwater Diagnostics in OHMSETT Facility	90
4.2.2	Gobles Fields LIDAR Operational Capabilities Validation	91
4.2.3	Vegetation oil pollution laboratory research	92
4.2.4	Oil Spill Kalamazoo River Area LIDAR Survey	92
4.3	Feature denoising in water signal spectral deformation detection - simulation experiment	93
4.4	Underwater oil pollution detection and identification	95
4.5	Land surface oil pollution identification	100
4.6	Land/water mixed diagnostics	106
4.7	Spectral Layered Analysis	113
4.8	Conclusions	117
Conclusions		119
Concluding remarks		119
Future work		120
Appendix		123
1-D lifting wavelet transform		125
Inverse 1-D lifting wavelet transform		127
Poisson-Gaussian noise model algorithm		128
Estimation of parameters		128
Feature extraction		132

ASC denoising	133
References	135
Acknowledgements	145
Kokkuvõte	147
Abstract	149
Elulookirjeldus	151
Curriculum Vitae	153
Publications	157

List of publications constituting the thesis

- S. Babichenko, V. Alekseyev, J. Lapimaa, A. Lisin, L. Poryvkina, S. Shchemelyov, I. Sobolev, and L. Vint. Airborne surveillance of water basins with hyperspectral FLS-LiDAR. In *Remote Sensing*. International Society for Optics and Photonics, 2010.
- I. Sobolev and S. Babichenko. Analysis of the performances of hyperspectral lidar for water pollution diagnostics. *EARSeL eProceedings*, 12(2):113–123, 2013a.
- I. Sobolev and S. Babichenko. Application of the wavelet transform for feature extraction in the analysis of hyperspectral laser-induced fluorescence data. *International Journal of Remote Sensing*, 34(20):7218–7235, 2013b.

Author’s Contribution to the Publications

All results of Sobolev and Babichenko (2013a,b) publications were obtained by the author of the thesis under the supervision of Sergey Babichenko.

The publication Babichenko et al. (2010) is devoted to dissemination of multiple projects technical, application and field work results. Authors of the publications are principal investigators of the projects and my contribution to the projects is in electronics and firmware design, development of control and data acquisition, spectral data processing and implementation of the necessary algorithms.

List of Abbreviations

ANN	Artificial Neural Network
ASC	Adaptive Slope Compensation
AUC	Area Under Curve
CCD	Charge-coupled device
CMOS	Complementary metal-oxide-semiconductor
CRM	Channel Relationship Method
CWT	Continuous Wavelet Transform
DOM	Dissolved Organic Matter
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
FLIR	Forward-Looking Infra-Red camera
FWD	Feature Weighted Distance
FWT	Fast Wavelet Transform
HLIF	Hyperspectral Laser Induced Fluorescence
i.i.d.	Independent and identically distributed random variables
IFO	Intermediate Fuel Oil
IR/UV	Infra Red/Ultra Violet
LCO	Light Cycle Oil
LIDAR	LIght Detection And Ranging
LIF	Laser Induced Fluorescence
LMM	Linear Mixture Model
LWT	Lifting Wavelet Transform
MAD	Median Absolute Deviation
MCP	Micro-channel plate
MDO	Marine Diesel Oil
MRA	Multiresolution analysis
MSE	Mean Squared Error
MWR	Microwave radiometer
NDVI	Normalized Difference Vegetative Index
OPD	Oil Pollution Diagnostics
p.d.f.	Probability density function
PAH	Polycyclic Aromatic Hydrocarbons
PCA	Principal Component Analysis

PMT	Photomultiplier tube
PSNR	Peak Signal to Noise Ratio
ROC	Receiver Operating Characteristic
RSS	Raman Scattering Signal
SLAR	Side-Looking Airborne Radar
SNR	Signal-to-noise ratio
SURE	Stein Unbiased Risk Estimate
SVM	Support Vector Machine
WGN	White Gaussian Noise
WT	Wavelet Transform

Introduction

The technique of laser remote sensing and its application for monitoring of land and water environments including oil pollution diagnostics (OPD) has been well explored in a course of the last decades (Hoge and Swift, 1980; Cecchi et al., 1986; Hengstermann and Reuter, 1990; Babichenko, 2008; Fingas and Brown, 2011). Among other techniques, ones based on laser induced fluorescence (LIF) method constitute a cluster of active remote sensing systems (often referenced to as *Laser Fluorosensors*) capable to deliver analytical information with high applicability (Dudeizak et al., 1991; Fingas and Brown, 2010). A fundamental advantage of active remote sensing systems is the selective excitation by monochromatic laser emission. A secondary emission of the sensed object creates an response signal, which constitutes of light scattering and fluorescence of complex molecules, like polyaromatic hydrocarbons. The selective excitation increases the applicability of technique to discriminate between the targeted and background object.

LIF technique reached its next step by inclusion of laboratory chemical spectroscopy concept of fine spectral details, which is referred to as hyperspectral. The main difference between such systems and a Laser Fluorosensor is that there are no predefined spectral ranges to detect echo-signal, and a comprehensive shape of water emission spectrum is recorded at every laser shot for consequent analysis (Babichenko, 2008). At present state the hyperspectral LIF (HLIF) sensors are promising technological solution, which brings significant additional value to effective diagnostics of events related to oil pollution (Babichenko et al., 1993; Pantani et al., 1995).

The detection and identification accuracy of oil pollution using particular sensor technology is constrained with several factors: the physical limitations of technology, unfavorable operation conditions, and the adequacy of data processing techniques. LIF sensors have been highly valued in suitability for many applications criteria setting them as the only sensor capable of working on various backgrounds including water, land, shoreline, ice and snow in night and day. In contrast, the data processing has been substantially studied mainly for multispectral approach of laser

fluorosensors. The hyperspectral data processing has been considered in the framework of passive hyperspectral imagery, presenting many state-of-the-art objects detection and identification algorithms. The topics of the HLIF technique data processing, the issues and application specifics remain unexplored in a great extent.

This thesis is focused on the HLIF data processing solutions for OPD of land and water environments addressing the improvement of sensing capabilities and meeting the criteria of the automated decision making.

The thesis considers

- establishing of HLIF data processing systematic approach for land and water environments;
- modeling and analysis of HLIF measurements using the advanced signal processing techniques;
- application of structural feature extraction/denoising for improving the identification capabilities of HLIF pattern recognition system;
- providing the HLIF automated diagnostics platform of land and water oil pollution detection and identification.

The main original contribution of this thesis is in deep, application dependent analysis of HLIF data processing in the framework of OPD technological solution. Author's contributions are discussed in more detail in the beginning of the chapters 1, 2, 3 and 4.

Problem statement

Oil exploration and transportation remain as a high risk environmental activities up to now. In particular, ninety percent of oil products are transported to Europe by sea. Therefore a large probability exists that some amount of oil products will end up in the sea. Irrespective of whether the pollution event occurs due to an accident or normal operations of ships, it still severely damages the sea environment. In addition, a heightened risk comes from the offshore oil drilling platforms. Rapidly rising demand for hydrocarbons expected to boost offshore drilling in ultra-deepwater and harsh-weather environments (GBI Research, 2011). The accidents occurred to oil tankers Prestige in 2002 and Erika in 1999 as well as the Deepwater Horizon oil spill in the Gulf of Mexico (2010) are some of the most vivid examples of possible risk and its extent. A large-scale pollution event in the sea exerts great hazards upon the environment and brings high demands to governmental agencies and commercial enterprises responsible for this domain.

In addition to potential pollution caused by accidents, some of the pollution enters the sea with dirty water created during routine operations, e.g. ballast water, the wash water from containers and tanks, and also the waste water from engine room. As a result of such pollution-causing actions, a large amount of oil enters the marine environment and it is therefore the largest source of ship-based pollution, which poses a longterm hazard to the environment of the seas and seashore areas.

In recent years the increasing demand of energy and development of new oil extraction technologies increased the production of *unconventional oil* from the onshore deposits (International Energy Agency, 2013). Unconventional sources as heavy crude oil, oil sands and shale oil are characterized by increased hazardous waste coming from production, transportation and storage. Environmental concerns with unconventional oil related to production phase are the leakages of oil and byproducts. The close proximity of water resources to oil development site and open-pit mining increases the likelihood of contamination. A byproducts hazard sedimentation of polycyclic aromatic hydrocarbons (PAH) in the Athabasca oil sands has been reported in Kurek et al. (2013).

The inland transportation of unconventional oil products is associated with high risks. The pipeline split and following oil spill can cause a long term damage to natural environment due to the low degradation and accumulation of contamination in sediment. The accumulation and post cleaning operations can take years as for Kalamazoo River oil spill of diluted bitumen (Song, 2012).

When taking into account the above-mentioned risks, the stakeholders including industries and society face the issues of detection of pollution as early as possible, localizing the pollution event and executing effective cleanup processes. In this field recent years have brought a number of different technologies, which have significantly improved the detection ability related to pollution. At the same time the existing technological solutions (SLAR, IR/UV scanner, MWR) have clear disadvantages - unable to remotely classify oil, detect oil emulsion and dissolved fractions, reveal submerged oil and operation in certain weather conditions (waves and ice) and shoreline-land environment. The main focus of the current developing is to provide a technological solution, which brings about significant additional value to effective resolving of events related to oil pollution.

State of the art

At the end of past century new laser and photonics technologies created considerable effort on development of remote sensing systems. The increased technological capabilities allowed to extend the remote sensing to

airborne applications, which remain a preferred solution for large water areas monitoring. An overview of the history of oil pollution monitoring and summary of airborne remote sensing systems can be found in several review papers (Babichenko, 2008; Jha et al., 2008; Fingas and Brown, 2014).

The theory behind and the first results of active airborne LIF LIDAR use for oil film detection were introduced in Hoge and Swift (1980), and the airborne detection of oil discharges in the sea were demonstrated in Hengstermann and Reuter (1990, 1992b); Hengstermann et al. (1992), the application of multiple sensing wavelengths in LIF LIDAR was developed in Dudelzak et al. (1991). Since that time the airborne LIF techniques were significantly enhanced and tested in numerous studies (Brown et al., 1997; Babichenko et al., 2000a; Zielinski et al., 2000). It has been proved that combined use of LIF LIDAR with other on-board sensors increases the reliability of airborne oil detection (Robbe, 2005; Zielinski et al., 2006; Lennon et al., 2006).

The first operational airborne surveillance with hyperspectral LIF LIDAR was reported in Babichenko et al. (1995). In a course of HLIF development the systems have been extensively tested in various applications on board of ships, rotary and fixed wing platforms for detection of trace level oil pollution in open seawaters and coastal areas; finding and mapping of oil spills; location and identification of submerged oil (Ohmsett, 2005); profiling of Dissolved Organic Matter (DOM) in lakes and rivers; diagnostics of agriculture and industrial run-offs, and eutrophication; monitoring of accumulated oil pollution in the port areas; general environmental assessment of natural water quality (Babichenko et al., 1995, 2000b, 2006).

The approach to LIF data processing for OPD always considered the qualitative and quantitative analysis of the fluorescence emission spectrum. The estimation of oil class and film thickness with airborne LIF were first introduced by Hoge and Swift (1980); O’Neil et al. (1980); Hengstermann and Reuter (1990). The general LIF signal pattern recognition concept first introduced in Hengstermann and Reuter (1990) was based on linear transformation of multispectral signatures into new domain by means of Karhunen–Loève transform or Principal Component Analysis (PCA) (Abdi and Williams, 2010). A classification according to predefined classes required comprehensive catalog of optical properties with extracted relevant principal component features (Hengstermann and Reuter, 1992a). Using Minimum Distance to Means classifier (Richards and Jia, 1999) the principal component feature space was partitioned into 9 classes yielding good oil identification results in tests flights (Reuter et al., 1995).

Another pattern recognition approaches were introduced in the framework of Channel Relationship Method or “differential fluorescence” (CRM) (Cecchi et al., 1986; Almhdi et al., 2007), Asymmetry (Babichenko et al., 2002), Artificial Neural Networks (ANN) and Support Vector Machines

(SVM) (Dolenko et al., 2002; Karpicz et al., 2006; Almhdi et al., 2007). First couple of methods are related to Ad hoc and deterministic methods. The CRM uses the relation of fluorescence bands similar to Normalized Difference Vegetative Index (NDVI) applied in reflectometry (Carlson and Ripley, 1997). The Assymetry has more general concept relied on the whole spectral deformation due to the presence of pollutant. In ANN and SVM the multispectral data or selected number of emission bands are used directly as inputs. The results indicate sufficient applicability of solution, however presenting some issues with overfitting and noise in real world data. From mentioned techniques the Assymetry has been used for HLIF data processing in Babichenko et al. (2004).

A statistical and decision-theoretic approaches are prevalent in multi-spectral LIF OPD applications. Directly applied supervised classification algorithms as ANN and SVM are highly dependent on a comprehensive training dataset, including a problem with number and configuration of inputs (training features), and generally represent a “black-box” in operation. The leading PCA feature extraction method introduced in Hengstermann and Reuter (1990) has been applied for state-of-the art OPD multi-sensor airborne platform in Robbe (2005).

The HLIF technology introduces all advantages of fluorosensors with additional detailed hyperspectral data. Properly extracted information contained in hyperspectral data improves the capabilities of pattern recognition as has been demonstrated in variety of hyperspectral imagery applications (Manolakis et al., 2001; Heinz and Chang, 2001; Yang et al., 2011; Cohen et al., 2011). The conventional HLIF data analysis rely on basic hyperspectral Linear Mixture Model (LMM) technique, which has been widely studied in imaging spectroscopy (Bioucas-Dias et al., 2012). The LMM assumption combined with non-negativity constraint is exploited for “mixed” fluorescence response. This approach led to applicability of HLIF technique for inner waters monitoring, where the natural water background fluorescence (DOM or Gelbstoff) can exceed pollutant in multiple times (Babichenko et al., 2010).

An extraction of relevant information from high-dimensional data in pattern recognition is referred to as feature extraction/selection. A comprehensive history background on feature extraction can be found in Saito (1994). The structural feature extraction (Pavlidis, 1977; Fu et al., 1982) describe the high-dimensional or hyperspectral data using predefined primitives like LMM end-members or wavelets. In Bruce et al. (2002) structural feature extraction based on wavelet transform demonstrate high (over 95%) classification accuracy for hyperspectral imagery. The feature extraction is critical in searching for data with known spectral signatures in the presence of unknown and variable background.

In addition to the background interference the most limiting factor

in OPD can be considered as measurement uncertainty or noise. Many conventional signal denoising techniques are build upon the estimation of signal independent and identically distributed noise model (Donoho and Johnstone, 1995; Luisier et al., 2007). In Foi et al. (2008) practical signal-dependent Poissonian-Gaussian noise model for digital imagery has been introduced. Signal denoising methods incorporating the signal-dependent noise model are relatively new (Liu et al., 2008; Hasinoff et al., 2010; Luisier et al., 2011). It has to be mentioned that there is no known denoising or modeling of noise application concerning HLIF data.

As opposed to FLIR and radar technologies, HLIF-LIDAR-based monitoring can *selectively* quantify and identify even very small amounts of oil product in difficult conditions with lowest false alarm rate. The state-of-the art development and commercialization of complementary solution for Oil Spill Detection addressing the current technology gaps (harsh environments, precision, verification) and creating the service capabilities to tend the needs of to the decision makers (data management, delivery and presentation) includes the HLIF-LIDAR technique as the *basic sensing component* (EAS project “Development of an Oil Spill Detection System based on information and laser remote sensing technology”).

Outline of the thesis

This thesis is organized as follows: Chapter 1 introduces the developing of HLIF data systematic approach. Concepts of LMM, natural fluorescence background and HLIF end-member mixtures are discussed. It presents the discriminant measure suitable for HLIF data comparison with corresponding experimental examples of OPD in natural water. An important aspect of oil pollution detection in the presence of high natural fluorescence background in water environment is evaluated. This chapter provides needed framework for subsequent developments in the rest of the thesis.

Chapter 2 begins with detailed overview of Wavelet Transform (WT). WT is used to solve many application specific tasks in computer science and it is essential tool for domain-scale analysis of HLIF data. Research presented in Chapter 2 contributed to the development of comprehensive HLIF signal-noise model. This model is based on the state-of-the art Poisson-Gaussian noise model in Foi et al. (2008) originally developed for digital images. Estimation of model parameters algorithm is upgraded for the HLIF signals. Noise parameters estimated from real-world HLIF data are used in subsequent chapters in simulation experiments.

Key parts of the HLIF data analysis are developed in Chapter 3. This includes the structural feature extraction concept applied for extraction of important information from HLIF data while removing the irrelevant

one. The robust feature extraction is inspired by the sparsity-norm optimization and uses the lifting scheme WT. Method was presented in papers Sobolev and Babichenko (2013a,b) regarding HLIF OPD and applied for standoff detection of biological aerosols in Hausmann et al. (2014). By extending the wavelet feature extraction technique a novel Adaptive Slope Compensation (ASC) denoising method for mixed Poisson-Gaussian noise has been developed. The effectiveness of feature extraction and ASC denoising is demonstrated on a number of examples.

The methods of signal processing and feature extraction, discussed in thesis, are applied for the oil spill monitoring in land/water mixed environment in 4th Chapter. The application capabilities in real-world examples such as the detection and identification of pollutant in the presence of strong fluorescence background, the recognition of pollutant in unknown mixture and monitoring of inner waters for residual organic pollution are improved. The developed pattern recognition system for HLIF OPD is presented as a conceptual scheme as final result.

Concluding remarks summarizing the results of the thesis are drawn in the “Conclusions” section with additional view of future HLIF technique development. The “Appendix” Chapter contains the implementations of thesis algorithms as Matlab code listings with comprehensive commentaries.

Chapter 1

Systematization of HLIF data for OPD

The applicability of a remote sensing technology for OPD is directly dependent on the quality of discrimination of an oil pollution from the natural environment as well as on the pollutant type classification accuracy. In general case very few objects have strong fluorescence compared to oil, especially in clean open water basins, which facilitates the LIF diagnostics. The analysis becomes more challenging when observing inner waters (coastline, estuaries, lakes, rivers, wetlands) or terrestrial targets due to the higher intensity and variability of background fluorescence spectra.

According to previous statement, the key aspects are related to the homogeneity of the underlying surface being analyzed and a quantity of spectral variations of the targets being under exploration. The effective detection and identification of organic pollution requires understanding the differences and systematization of HLIF OPD knowledge.

The systematization of HLIF data with the examples of natural water diagnostics is provided in this chapter. Important and frequently observed natural fluorescence backgrounds are shown. The needs and gaps of OPD have been specified for understanding the differences in LIDAR applications. This chapter provides the needed framework for subsequent developments in next sections.

1.1 Author's contribution

The author elaborated the basis systematic approach for HLIF LIDAR OPD diagnostics. This includes the creation, configuration and use of the end-members library concerning the specificity of LIDAR applications.

- The reference HLIF library of spectral end-members is proposed for water and land OPD applications;

- Systematization of fluorescence objects by their role in HLIF OPD is developed. As a result the general assumptions about HLIF pollution model composition are given with diagram for water and land OPD applications;
- Discriminant measure suitable for HLIF data comparison is introduced. The LMM and proposed discriminant measure are applied for natural water OPD in experiment;
- The capabilities of organic pollution detection in natural water are tested in simulation experiment and the issues of diagnostics have been revealed;
- Carrying out experimental set up, measurements/simulation, implementing the necessary algorithms, data processing in provided examples;

1.2 The HLIF spectrum

The HLIF spectrum is a continuous measurement of a real fluorescence values over a finite discrete emission spectral range. The HLIF spectrum can serve as a spectral signature of object or combination of objects which can uniquely identify its source.

Let the $\lambda \subset \mathcal{R}_+$ denote a wavelength range which is a subset of positive real numbers. Then the HLIF spectra can be given as a discrete real-valued function $f : \lambda \rightarrow I$ with finite length where I is the fluorescence intensity. The fluorescence intensity is strictly positive so $I \in \mathcal{R}_+$.

The typical range of λ is given by construction of spectral detector, excitation wavelength and application requirements. Generally the HLIF spectra are recorded in the spectral ranges 300 – 600 nm or 450 – 750 nm (Babichenko, 2008). LIDAR systems for marine applications typically employ a laser with operational wavelengths near UV or blue/green spectral range, which gives a penetration depth of a few meters in most natural waters.

Let $I \subset \mathcal{R}_+^n$ denote a n-dimensional spectrum space as a subspace of positive real numbers. In this case, the dimensionality or the number of bands of each spectrum is n . The dimensionality varies with instrumentation and typical is not less than 256. The large number of spectral bands puts HLIF data into a group of high dimensional datasets.

1.2.1 The HLIF spectra mixture

The HLIF spectra analysis of the combination of substances is possible due to the additivity of fluorescence spectra of the molecules that are not

entering chemical reaction. The conventional method of HLIF analysis is known as end-member extraction, Linear Spectral Unmixing or LMM. The end-members are the spectral signatures of single objects of interest that are predefined in a spectral library. The analysis assumes that no chemical reaction occurs in a mixture. Therefore, given the set of end-members as $m \times n$ matrix \mathbf{F} , a LMM can be expressed as

$$\mathbf{z} = \mathbf{M} + \mathbf{e}, \quad (1.1)$$

where $\mathbf{M} = \mathbf{F}\mathbf{k}$ is end-members composed model, \mathbf{z} is an $1 \times n$ measurement vector, \mathbf{k} is $1 \times m$ component intensity vector (in reflectance spectroscopy known as *abundance*) and \mathbf{e} is measurement error. Since the fluorescence intensity of each end-member denoted by \mathbf{k} is strictly positive the solution to (1.1) can be found with the nonnegative constrained least-squares optimization solved using the Lagrange multipliers (Lawson and Hanson, 1974).

1.2.2 Library of spectral end-members

As the fluorescence spectra depend on the molecular structure it serves as the spectral “fingerprint”, explicitly characterizing substance and specifying its presence among the others. Typical HLIF analysis is aimed on the identification of the exact mixture components which is possible only when the reference spectra are known. Identification is required for the quantitative analysis also, as the estimation of concentration only works with proper calibration curve created for particular substance.

Comprehensive reference library of spectral end-members is essential for every application. The number and type of end-members is also important and application specific, as for example the spectra of ground natural fluorophores (grass, leaves etc.) obviously is redundant in water diagnostic application. Generally the end-members can be divided into two groups: signals of interest and natural fluorescent background.

In environmental applications of OPD the spectra of organic pollutants enter the first group. In general the PAHs, i.e. oil products are considered as organic pollutants. The variability of oil products is high and determined by the genesis, depth and geological position of the deposit. The layout of PAH complexes uniquely determines the fluorescence signature of oil product, therefore a feasible set of oil products should be investigated to get an overview of the spectral variability of the fluorescence signatures (Hengstermann and Reuter, 1992b).

Figure 1.1a shows the HLIF spectra end-members of 18 pollutants. The end-member HLIF data are recorded with many averages and thus always referred as “ground truth” with zero measurement noise. The oil products are colored by density estimate into three common classes: light refined,

medium refined and crude. The color classes correspond with correlation grouping of HLIF spectra shown in figure 1.1b.

Generally all potential pollutants must be included in spectral library considering the local operational aspects. The inclusion of concrete oil fluorescence spectrum gives the ability to identify it even in the presence of unknown background fluorescence which is demonstrated in chapters 3 and 4. The unknown pollutant can be roughly classified only into a groups i.e. defined in figure 1.1b, without the detailed identification.

In natural environment the state of the pollutant is dynamic and is accompanied by a number of transformation processes, as for example the weathering of oil and emulsification. The transformation of aggregate state with fast chemical and biological degradation significantly influences the fluorescence response of the oil product. This means that reference library should be extended according to the specific diagnostics, like appending the end-member spectra of pollutants in emulsified state.

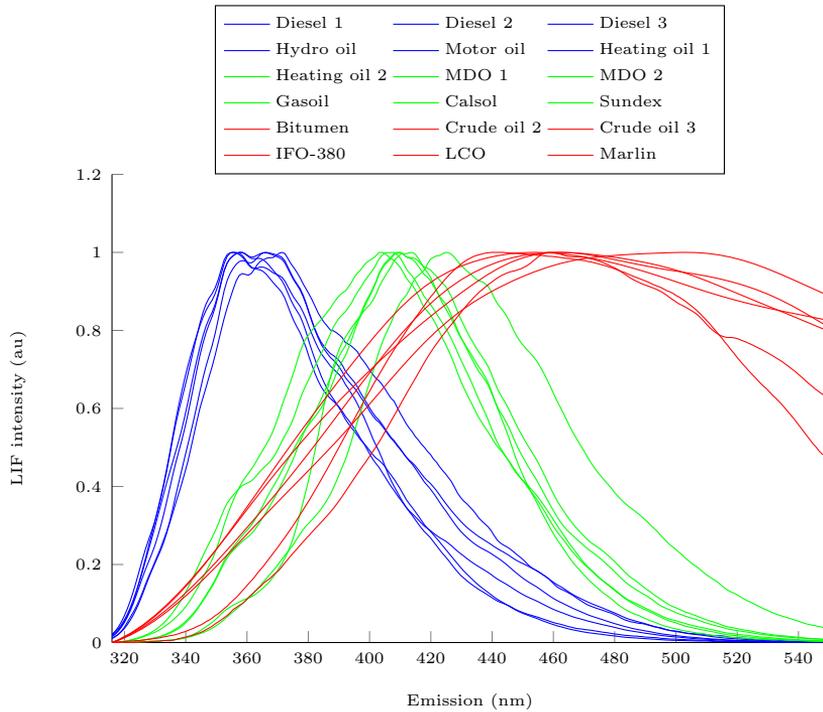
1.2.3 Natural fluorescence background

The main target of LIF LIDAR diagnostics is the discrimination of the pollutants from the natural fluorescent background. The natural fluorescent background is a definition of organic substances which naturally exist in environment. It is quite typical that fluorescence intensity of pollutant is lower than the corresponding value of background substance. The presence of background in LMM is important for diagnostics and correct estimation of pollution amount.

The group of fluorescent background substances can be entirely defined only for water diagnostic applications. The main source of fluorescence in natural water is the dissolved organic matter (*Gelbstoff*, *yellow substance*) of natural origin. The HLIF spectra of DOM under UV excitation shows wide band spectral structures with single characteristic maximum. The particular shape of DOM is uniform for single region, however it can be different in various waters depending on the different composition of humic substances (Shubina et al., 2010). Figure 1.2 shows some examples of HLIF spectra with variable organics concentration for clean natural water.

Besides the fluorescence of organic compounds the HLIF spectrum of natural water shows additional characteristic component - the Raman Scattering Signal (RSS). The Raman scattering is an inelastic scattering of photons due to the OH-stretching of the water molecules at 3400cm^{-1} from the excitation wavelength. The center position of RSS on wavelength axis in HLIF spectrum can be estimated at 345nm for 308nm excitation. The inclusion of RSS in reference library is essential for water diagnostics. It plays significant role in correct solution to LMM (1.1) at the diagnostics of the light oil products pollution, which have fluorescence maximum at near

(a)



(b)

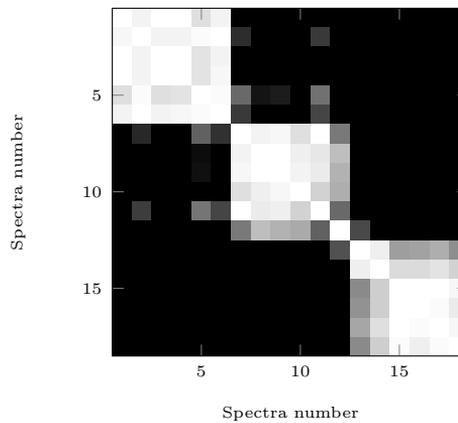


Figure 1.1: (a) - HLIF spectra of pollutants grouped by density. Light refined group (blue) include samples of diesel, hydro oil, motor oil and heating oil. Medium refined (green) are sampled catalytic gas oil, heating oil, Marine Diesel Oil (MDO) Sundex and Calsol. The crude group (red) include samples of bitumen, crude oil, IFO-380, Light Cycle Oil (LCO) and Marlin. (b) - The correlation matrix between HLIF spectra of samples, the lighter the better correlation. Three groups can be selected that corresponds to estimated density.

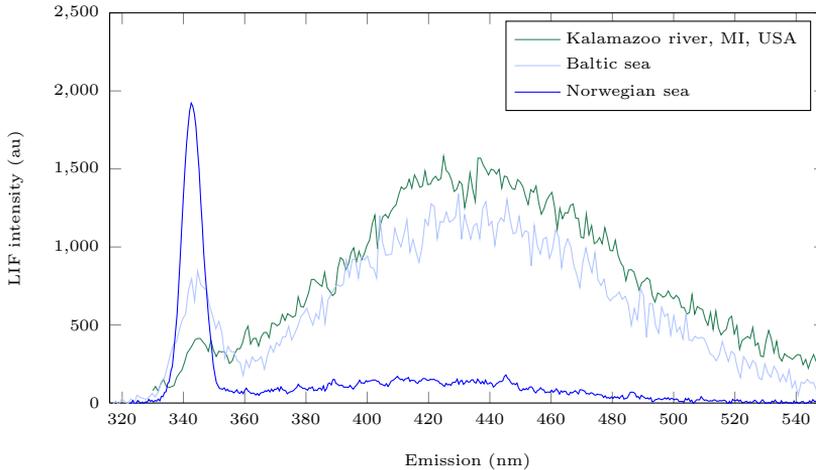


Figure 1.2: HLIF spectra of natural water. The Norwegian sea water presents ultra-low concentration of organics (estimated as $\leq 0.1\text{mg/l}$). The Baltic sea water and river are an examples of inner water basins with closed ecosystem. The value of DOM is comparable or higher than the fluorescence intensity of pollutant.

bands. The quantitative diagnostics of oil film thickness (Hengstermann and Reuter, 1990), DOM concentration and turbidity (Pantani et al., 1995) uses RSS intensity to calibrate the LIDAR spectrum.

Figure 1.2 shows the two possible conditions of water fluorescence background. The Norwegian sea example presents almost no DOM fluorescence. This means that any artificial fluorescent substance (oil products from figure 1.1a) in water will be clearly seen and modify the signal shape. The RSS intensity will be also suppressed due to the increase of water turbidity or the absorption of the surface film. Opposite case with high organics concentration require decomposition and sophisticated pattern recognition analysis in order to discriminate small additive artificial component in spectrum.

The coastal, inner waters or terrestrial applications are related with multiple and variable natural fluorescent background. Indeed, many different ground based organic objects with strong fluorescence response can be naturally encountered in such areas. In practice, it is impossible to investigate the fluorescence responses of all ground organic objects and their variations. However, common objects, like grass, can be used to create the representative library of land background end-members. Some frequently encountered organic objects with high fluorescence response are shown in figure 1.3.

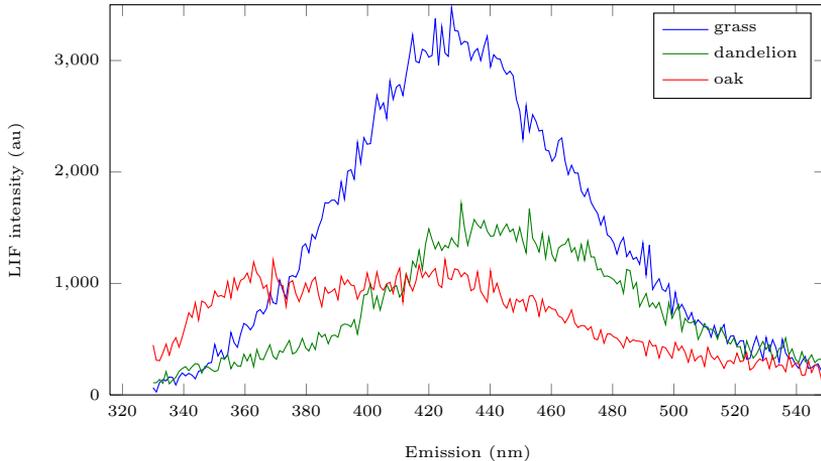


Figure 1.3: Examples of common land objects HLIF as measured by LIDAR. The most natural organic present minor to moderate fluorescence response compared to oil products. Among them the green grass has one of the strongest fluorescence.

1.3 Specifics of water and land applications

It has been mentioned that key differences in LIDAR applications are related to the homogeneity of the background and variability of targeted pollutants. The determination of application type and origin of measured sample is assumed as first step in developed HLIF OPD.

Fluorescent signal from the water body is characterized by the presence of two deterministic spectral components: RSS and DOM. Fluorescence is the result of a process that occurs in certain molecules (generally PAHs or heterocycles) thus with high probability the rest fluorescence responses, obtained from the water body are potential signals from the pollutant substances. This is a condition that a pollutant must be considered as a distortion or deformation of a water natural background.

Diagnostics of land, coastal waters, mixed areas is connected to the varying picture of the fluorescent signal. Here it is possible to claim that a quantity of fluorescence spectra and combinations is infinite. In practice some common natural land fluorophores can be measured and added into reference library, however the problem of simultaneously targeting multiple land objects creates complex combinatorial challenge. In addition, land targets present wide range of “natural” sources with typical oil product fluorescence response. This includes the road surfaces, plastic, roofs and other objects which arise from the industrial activities. Such sources will turn into hard detectable false positive alarms not connected with actual LIDAR targets.

Figure 1.4 presents the diagram of water/land application-specific diag-

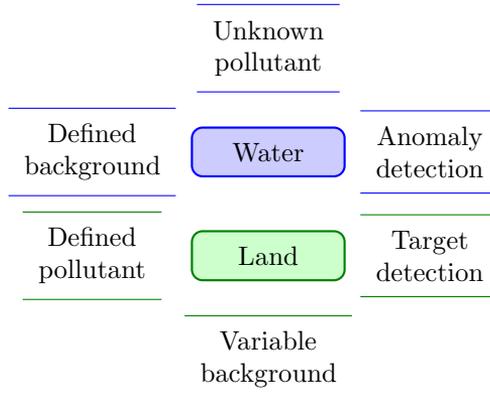


Figure 1.4: Water and land OPD application specific properties. The differences in applications are related to the variable natural background and estimation of targets.

nostics. The water diagnostics is first of all related to the routine monitoring of large territories with uniform background and unknown pollutant. The first priority of analysis is to detect spectra which differs from the natural water background - the anomaly detection. The general pollutant reference library can be used with LMM to identify the pollutant in the presence of strong background or correct the false positive answer in case of artificial object (ex. fish oil).

In land diagnostics it is assumed that the reference spectra is known, as in case of pipeline spill, integrity monitoring, exploration the source of pollution can be prior determined. Due to the variable natural background the first priority of analysis is finding the significant targets, which are closely related to the known pollutant. This analysis implies use of sophisticated pattern recognition methods, which are capable to deliver required discrimination capability of overlapped and similar spectral signatures.

1.4 HLIF models

The HLIF model is a set of expected end-members in a linear mixture, represented by matrix \mathbf{F} in LMM equation (1.1). The composition of models is related to the application specific rules defined in previous section. The general assumptions about the fluorescence mixture response are:

- *A single type of pollutant is presented in analyzed HLIF spectrum in one measurement* - This assumption follows from rather low probability of simultaneous targeting multiple pollutant objects in environmental applications.

- *For water diagnostics the end-member of DOM is known for analyzed area* - For example, the exact HLIF spectrum of DOM can be measured in preparation for airborne survey using the ground sampling.
- *The reference library of pollutants is defined* - The potential pollutants are provided before operation. Existing pollutant end-members are grouped according to spectral features (spectral shape). Grouped end-members provide coarse reference.
- *The objects in mixture do not interact, therefore the LMM equation (1.1) can be used for linear spectral decomposition* - This is general rule for oil film and water medium or oil and terrestrial organic objects interaction. However, some minor variations from linear model can be caused by *quenching* of the fluorescence (Lakowicz), this can be omitted at low optical density typical for natural waters.

Figure 1.5 shows the model composition diagram corresponding to the assumptions and application-specific tasks. According to figure 1.5 we can count the complexity of water analysis as $\mathcal{O}(p)$, where p is a number of pollutant end-members in reference library. The terrestrial pollution assumes variable fluorescent background, so the number of models grows as $p \times l$, where l is size of natural fluorescent objects library. For the real-time analysis and operational use the terrestrial pollutant models present significant overhead.

1.4.1 Discriminant measure for HLIF data

In general any pattern recognition system is based on some distance or proximity measure defined in the object or feature space. After the estimation of LMM coefficients, the models and observation can be compared in order to find the “best fit” model. The goodness-of-fit or distance between the mixture model and measurement \mathbf{z} is defined as

$$d(\mathbf{z}, \mathbf{M}) = \frac{\sum_n (\mathbf{z} - \mathbf{M})^2}{\sum_n (\mathbf{z} - E(\mathbf{z}))^2}, \quad (1.2)$$

where $E(\mathbf{z}) = \frac{1}{N} \sum_{n=1}^N z_n$ is statistical mean estimate of measured HLIF spectra.

Equation (1.2) denotes the relation between residual sum of squares and total sum of squares. It indicates the model fitness to observation in comparison with simple average. In statistics the equation (1.2) known as $1 - d$ is named - *coefficient of determination*. In general this statistics is used to estimate the *goodness of fit* for linear regression model (Colin Cameron and Windmeijer, 1997). For the regression analysis the coefficient of determination varies from 0 to 1 with 1 indicating the ideal fit. As the

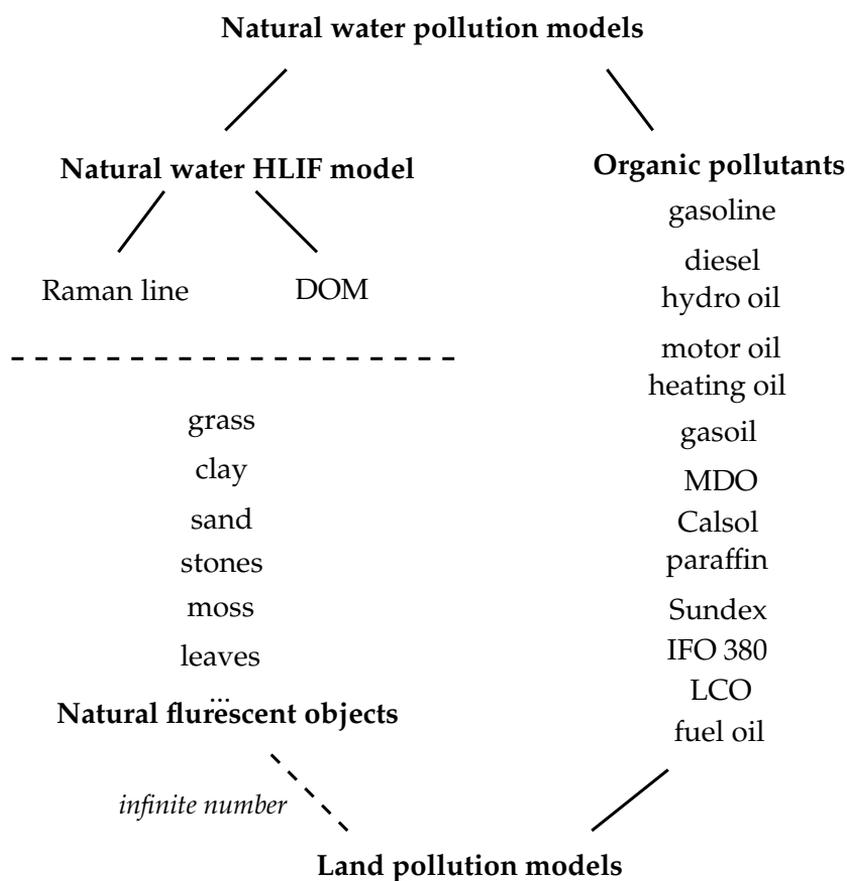


Figure 1.5: Model composition diagram for water and terrestrial applications. A reference library contains only a number of expected pollutants and some common natural fluorescence objects.

unit vector is not included in HLIF LMM, the (1.2) form varies from 0 to infinity representing the distance or proximity measure between the model and observed HLIF spectrum.

The selection of (1.2) as discriminant measure for HLIF data is not occasional. Specifically, the solution to linear mixture model (1.1) is given as minimization of the residual error, which naturally resides in (1.2). The additional reference relation to total sum of squares acts to normalize the distance which is advantageous to other common distance measures, i.e. Euclidian distance.

In general least squares regression the coefficient of determination is subject to inflation. The increasing number of end-members in model \mathbf{M} will slightly increase the value of coefficient of determination, or for (1.2) will shift it towards zero. For comparison of models with different number of end-members other statistical tests can be used, ex. Adjusted R^2 , F -test. For HLIF LMM the inflation is removed by model composition rules defined in section 1.4. In other words, the “best fit” model is found among the equal sized \mathbf{M} .

The LMM can be used to “unmix” complex signal with known natural background and pollutant additive. Thus it is useful to estimate the distance between the observation and additive residual with pollutant end-member separately as

$$d_r(\mathbf{r}, \mathbf{M}_p) = \frac{\sum_n (\mathbf{r} - \mathbf{M}_p)^2}{\sum_n (\mathbf{r} - E(\mathbf{r}))^2},$$

where $\mathbf{r} = \mathbf{z} - \mathbf{F}\mathbf{k}_b$ is the observation residual, subscripts b and p denote the end-member vectors of natural background and pollutant respectively. This approach eliminates the background end-members contribution to distance value.

1.5 Examples

1.5.1 Natural water HLIF linear unmixing

Figure 1.6 shows the spectrum of Baltic sea water with laboratory added pollutants. The equal sub-micron oil films of MDO and Crude oil were created in water tank and measured at a 15 meters distance by LIDAR. The MDO and crude oil models were measured and compared with observations.

The HLIF plot in figure 1.6a shows the deformation of natural water echo-signal in the presence of pollutant. The deformation is noticeable, however both spectra are close to clean water, due to the presence of organics structure. The RSS is suppressed due to the presence oil film.

Figures 1.6b and 1.6c show the LMM, end-member vectors and residuals for polluted samples. The residual plots indicate the similarity of MDO and

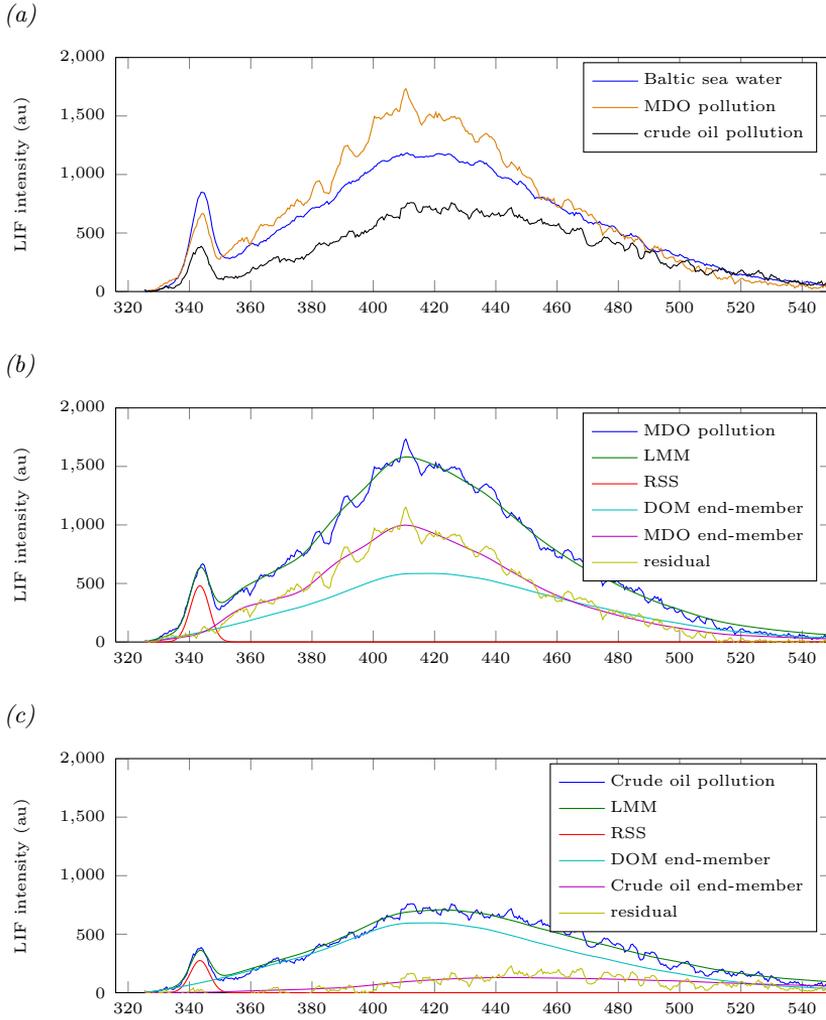


Figure 1.6: Example of clean and polluted natural water samples. Results are shown for single measurement. (a) - HLIF spectra of samples, the pollution was added to clean Baltic sea water. (b) - HLIF spectrum and LMM of MDO pollution sample. (c) - HLIF spectrum and LMM of Crude oil pollution sample.

MDO pollution	k_{RSS}	k_{DOM}	k_p	d	d_r
MDO model	479.918	585.642	997.905	0.007	0.017
Crude model	427.812	1424.937	0.000	0.034	1.116
Crude pollution					
MDO model	254.245	726.860	0.000	0.052	1.023
Crude model	275.116	596.343	129.237	0.020	0.285

Table 1.1: LMM coefficients and distances of MDO and Crude oil natural water pollution samples. k_{RSS} , k_{DOM} , k_p - are coefficients of corresponding end-members in spectrum. d and d_r are distance values for between LMM, residual and measurement.

Crude oil end-member vectors for corresponding polluted samples. One can notice the difference of fluorescence intensity between MDO and crude oil for equal film thickness.

The estimated values of model members coefficients are given in table 1.1. Following k_p values the fluorescence of MDO is 8 times higher than of Crude oil. The estimation of k_{DOM} which is related to concentration of organics is dependent on model. In experiment the DOM concentration remained constant, as water was taken from single source. One can notice that correct results give similar k_{DOM} values of 585 and 596 a.u. for MDO and Crude oil samples respectively. On the other hand, the wrong models give absolutely different estimations. This shows that correct estimation of mixture qualitative parameters requires selection of correct model.

The selection of best fit model is done using the discriminant measure. Table d and d_r values estimate correct pollutant type for both samples. The question remains of how representative the discrimination is. The relation on the residual d_r shows values at different orders for MDO (0.017) and Crude 0.285 samples. This is general problem related to high dimensionality of data space and measurement noise, which is particularly discussed in next chapters.

1.5.2 Natural water HLIF deformation detection

In previous example the pollution LMM were right off applied to contaminated samples. The clean water sample was omitted. In real case the task of water HLIF spectrum deformation detection should be considered first. It is obvious that defined spectral signature of natural water is subject to change with presence of additional fluorophore. Thus the deformation detection reveals the identity of unknown pollutant and links the analysis to the next step of pollutant identification.

Let η_d is a some positive real value, then

$$\text{clean} \equiv \begin{cases} 1 & d_{water} \leq \eta_d \\ 0 & \text{otherwise,} \end{cases}$$

denotes the binary spectral deformation classifier, where d_{water} is spectral deformation value quantified as distance between natural water model and measured HLIF spectrum (see equation (1.2)).

The selection of threshold η_d depends on two parameters: the pollutant concentration (or required detection sensitivity) and HLIF signal quality. The unknown noise component which resides in HLIF spectra will increase the distance to the model depending on the signal-to noise ratio. The latter is studied in details in section 4.3 and is not considered here.

The quantity of deformation depends on the type of pollutant or specifically the HLIF spectrum of pollutant. Generally the HLIF spectrum of DOM differs from the spectra of organic pollutants presented in figure 1.1a. However some pollutants, namely the medium refined group and crude oil products, are spectrally overlapped with DOM (see number 12. Sundex on 1.1a and LMM examples in figure 1.6). Considering the small concentrations of pollutant thus small mixture value the detection of water HLIF deformation becomes complicated.

The complexity of pollutant detection is shown in figure 1.7. The value of d_{water} is plotted against the pollutant number from figure 1.1a in simulation experiment. Here the water-pollution mixture is taken at three pollution levels corresponding to sub -,micron and optically dense oil film thickness. The relative fluorescence intensity of refined and crude oil products is selected as $\frac{10}{1}$. The film thickness is simulated using three levels of k_p in LMM and the quenching of the fluorescence is not considered.

The results indicate that oil product groups have different influence on the water deformation value, which also depends on the amount of pollutant. First group of light refined oil products creates most significant deformations in spectral shape at all pollutant concentrations (concentration is denoted as $1\mu m \sim 1mg/l$). The medium refined oil products are most overlapped with DOM spectrum. The deformation value is almost constant for all concentrations. The crude oil products have low fluorescence intensity, thus for thin films they apply most small differences into water spectrum. However, the spectrum shape difference is clearly seen for higher concentrations.

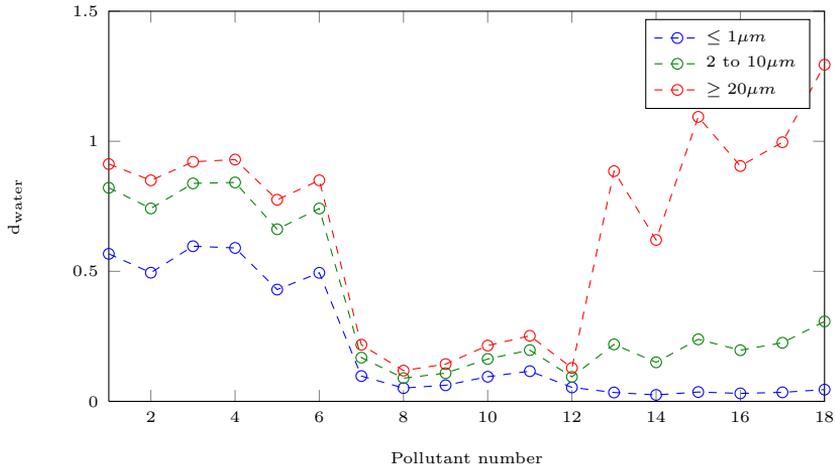


Figure 1.7: The complexity of water deformation detection from various pollutants. The lower value shows more similar spectrum to clean water. Pollutant numbers correspond to figure 1.1a. The results indicate that the binary threshold lower bound is denoted by medium refined and crude oil products. For very low pollutant concentrations $\leq 1\text{mg/l}$ the presence of crude oil products is most hard to detect.

1.6 Conclusions

The hyperspectral response from complex target, which includes multiple fluorophores can be linearly separated if the mixture components are known. This assumes that mixture components do not interact chemically and their concentrations do not influence the fluorescence spectra structure of each other, which is true for water and land OPD applications.

Under this statement the reference library and fluorescence objects systematization have been proposed by the author. The library spectra or end-members have been divided into two general groups: signals of interest and natural fluorescent backgrounds. The first group includes the oil and oil products. The oil products have been divided into three conventional classes: light refined, medium refined and crude. The division has been made by oil product estimated density, which also corresponded to general HLIF structure. The composition of signals of interest may be extended according to specific diagnostics, as an example with the reference spectra of pollutant emulsions. According to the needs of OPD the reference library also include the group of common natural fluorescence backgrounds. In water diagnostics the single natural background DOM and its variations are considered with additional water characteristic spectral component - the RSS. For land application the natural fluorescence background is highly variable, however it is also possible to include common land fluorophores, like grass, which have strong fluorescence value. However, it will be shown,

that OPD on variable background is also possible with specific HLIF feature comparison.

The general spectral response in LIDAR remote sensing comes from complex target with multiple fluorophores. The assumptions about the fluorescence mixture response have created the model composition diagram of linear mixtures for water and land OPD applications.

The recognition of HLIF measurements using the proposed system of models requires the discrimination measure suitable for HLIF data. The modified fit statistic, known as coefficient of determination is introduced. The model fit or “distance” evaluation on the mixture residual, which eliminates the contribution of background end-members is proposed.

The main issue of HLIF OPD is the discrimination of pollution from natural fluorescence background and between the various pollutants. The first issue has been introduced as simulation example of natural water deformation detection. For very low trace pollution $\leq 1mg/l$ the presence of crude oil products is most hard to detect. The contribution of signal noise and discrimination of land pollution with proposed solutions will be discussed in later chapters.

The discrimination between the pollutants as demonstrated in experiment is possible using both discriminator measures. The better results have been estimated by residual distance d_r , however for crude oil sample residual the value of d_r is significantly higher than for MDO sample. This effect comes from various signal-to-noise ratio of residuals. In addition the difference between models is affected by high-dimensionality of HLIF data space. In patten recognition approach the discrimination of high-dimensional objects is preceded by a feature extraction. The HLIF feature extraction is proposed in the following chapters and provides both feature-space discrimination and denoising capabilities.

Chapter 2

Exploration of HLIF signal

Defining the structural approach to OPD in previous chapter we have examined the HLIF data as an ideal abstract term. In this chapter the HLIF data is considered and analyzed as conventional signal using the advanced signal processing methods.

By general definition of measurement, a spectrum of fluorescence intensity is always accompanied by measurement uncertainty or noise. In imaging devices equipped with CCD, CMOS and other photodetectors (PMT, photodiode, MCP) the measurement uncertainty is strongly influenced by the quantized and independent nature of light. This is specifically true for low level light detection, which is typical for fluorescence signals in active remote sensing.

A widespread approximation of noise modeling and filtration in signal and image processing uses the independent additive Gaussian white noise. With the progress in photonics technology, the main contribution to noise source in conventional digital images turned to be signal-dependent requesting the creation of suitable noise model. In near past new methods have been proposed to overcome the image signal-dependent denoising problem directly (Hasinoff et al., 2010; Luisier et al., 2011) or using signal specific transformation namely *generalized variance-stabilizing transformation* or *Anscombe Transform* (Makitalo and Foi, 2013).

For modeling the HLIF measurements, which are subject to mixed signal-dependent and Gaussian noises, the proposed Poisson-Gaussian noise model in Foi et al. (2008) with algorithm modification have been successfully applied. It has been shown that the model well corresponds to the real HLIF data and can be applied to estimate the noise parameters of HLIF LIDAR sensor. The estimated model parameters have been used to create realistic simulation data for tests and validation of OPD algorithms in chapters 3, 4 and publications (Sobolev and Babichenko, 2013a,b).

The noise modeling and related group of denoising algorithms use the

well-known mathematical instrument the Wavelet Transform (WT) (Daubechies, 1988; Daubechies et al., 1992; Daubechies, 1993; Mallat, 1989a,b). Due to the importance of the latter an overview of WT theory is given in this chapter. Various transform schemes have been developed and applied in analysis of complex non-stationary signals in many different applications (Meyer, 1993; Goswami and Chan, 2011). The three methods: Continuous wavelet transform (CWT), Fast wavelet transform (FWT) and Lifting scheme (LWT) (Sweldens, 1998) which have been applied in this work are introduced at the beginning of this chapter. The introduction is intended for readers, unfamiliar with the theory of WT.

2.1 Author’s contribution

Author’s contribution is in advancing the of state-of-the art mixed signal-dependent noise model at HLIF signals with model parameters evaluation using the real LIDAR data.

- Set up of HLIF signals specific segmentation algorithm for estimation of local expectation/standard-deviation pairs;
- Modification of global parametric model fitting algorithm for estimation of mixed signal-dependent noise model parameters. Optimization of MSE of an standard-deviation estimator with logarithmic transformation and multi-start initialization successfully finds the solution to noise model fit;
- Estimation of noise parameters of measured HLIF data;
- Implementation of segmentation, modified model fit and noise simulation algorithms within the framework of HLIF signals (see Appendix).

2.2 Theory of Wavelet Transform

Wavelet transform is well known school in the theory of time-frequency analysis, conversion of signals, images and analysis of time series. Wavelet theory historically is the kind of synthesis of many ideas in the technology, physics and mathematics. Grossmann and Morlet introduced term “wavelet” in the middle of the 80’s with the research of seismic signals (Grossmann and Morlet, 1984). At present after remarkable research in this field by such scientists as Ingrid Daubechies, Mallat, Farge, Chui and many others, theory of WT was embedded practically onto all fields of applied engineering and scientific research.

WT is used to solve many application specific tasks like: the analysis of natural time meteorological series (Lau and Weng, 1995), the analysis

of electroencephalographic data (Sun et al., 2000), the analysis of DNA sequence (Ameodo et al., 1995) and fingerprint image compression (Bradley et al., 1993). Today WT is applied in wide range of science and technology areas like radio engineering, communications, electronics, nuclear physics, seismoacoustics, meteorology, biology and economics.

2.2.1 Wavelets

It can be stated that the theory of WT is an generalization of well-known Fourier transform (FT). In the Hilbert space (a generalization of Euclidean space) an arbitrary function $f \in L$, defined on the closed limited subset $L^2[t_1, t_2]$ and satisfying the condition of finite norm

$$\|f(t)\|^2 < \infty, t \in [t_1, t_2],$$

where $\|\cdot\|^2$ denotes the norm can be expanded to the basis of some functions $\varphi(t)$ as

$$f(t) = \sum_{n=0}^{\infty} C_n \varphi_n(t).$$

The coefficients C_n are defined as

$$C_n = \frac{1}{\|\varphi_n\|^2} \langle f(t), \varphi(t) \rangle, \quad (2.1)$$

where $\|\varphi_n\|^2$ is the energy of basis function and $\langle f(t), \varphi(t) \rangle$ denotes the convolution or inner product.

The set of basis functions $\varphi(t)$ have important properties in signal analysis if they are satisfying two conditions on the interval $L^2[t_1, t_2]$

$$\langle f(t), \varphi(t) \rangle = \|\varphi_n\|^2 \delta, \quad (2.2)$$

where δ is the Kronecker's delta function and

$$\|\varphi_n(t)\|^2 = 1. \quad (2.3)$$

Equations (2.2) and (2.3) are called orthogonal and normal conditions respectively and resulting entire set of $\varphi(t)$ functions is called *orthonormal*.

The most known orthonormal basis is given by trigonometric sine and cosine functions as Fourier basis. However, many irregular functions $f(t)$ with the finite energy $\|f(t)\|^2 < \infty$ cannot be effectively represented using this basis. The more suitable basis functions must be compactly supported (or have finite effective support) in addition to properties of orthogonality and orthonormality.

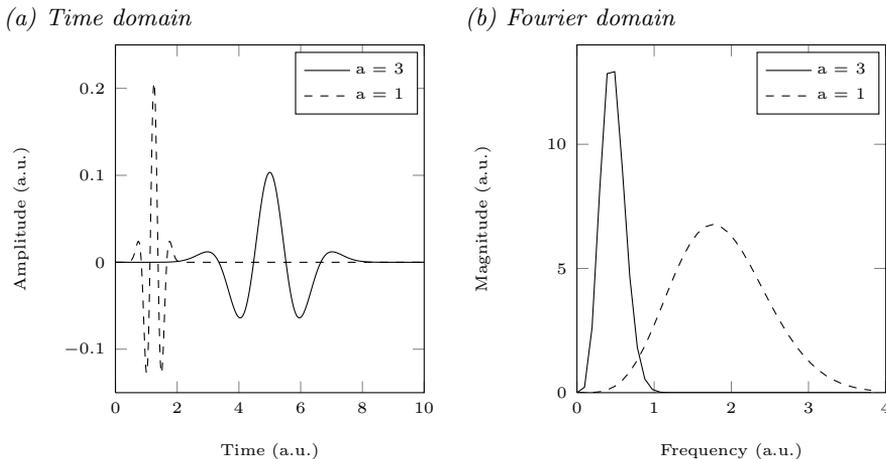


Figure 2.1: An example of mother wavelet from Gaussian wavelet family. The scale is varied between 1 and 3.

In order to satisfy compact supported constraint the systems of $\psi(t)$ functions called *wavelets* have been proposed (Daubechies et al., 1992). The basis of $\psi(t)$ functions can be represented in the form

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad (2.4)$$

where parameter a scales and b translates the wavelet function in signal domain t . The original function $\psi(t)$ is called the “mother wavelet”.

In figure 2.1 an example of mother wavelet with its translated and scaled copy from Gaussian wavelet family is given. The Fourier transform (FT) on figure 2.1b shows the corresponding frequency characteristics of selected wavelets. The wide wavelet on 2.1a in time domain corresponds to more compact and low frequency Fourier spectrum than its narrow copy. Besides that, both spectra are well localized in the frequency domain and represent band-pass filters.

Using (2.4) and (2.1) the expression for the integral Continuous wavelet transform is

$$W_{\psi}f(a, b) = \langle f(t), \psi_{ab}(t) \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt,$$

where $W_{\psi}f(a, b)$ are the decomposition coefficients of the function $f(t)$ into a wavelet basis, analogous to the Fourier series. The inverse transform is obtained as

$$f(t) = \frac{1}{C_{\psi}} \iint_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) \psi_{ab}(t) \frac{da db}{a^2}, \quad (2.5)$$

where C_ψ is the normalization coefficient (analogous to $\sqrt{2\pi}$ in the FT). In the general case expression for C_ψ , on the basis of (2.5) written in the form

$$C_\psi = \int_{-\infty}^{\infty} |\hat{\psi}(w)|^2 |w|^{-1} dw < \infty \quad (2.6)$$

is called the *admissibility condition*, where $\hat{\psi}$ is the FT of the wavelet function. For the orthonormal wavelets $C_\psi = 1$. While orthogonality and compactness of wavelets are desired properties the equation (2.6) is the main constraint for the selection of wavelet function. The wavelets which satisfy only the admissibility condition (Gaussian, Morlet wavelet families) are called *crude* wavelets. In latter case the analysis is not orthogonal and the inverse transform is not guaranteed.

The admissibility condition of the normalizing parameter (2.6) sufficiently limits the selection of wavelet functions. In particular, the $\hat{\psi}(0)$ should be 0 and ψ has to oscillate. The admissibility condition adds useful property of WT - suppression of polynomial parts of the function. From (2.6) the FT of wavelet must be equal to zero with $w = 0$. Respectively, at least zero moment of wavelet function must be equal to zero

$$\int_{-\infty}^{\infty} \psi(t) dt = 0.$$

For many applications dealt with signal analysis and denoising the suppression of polynomials of high order is useful. In general wavelet function $\psi(t)$ is having $n + 1$ zeros moments if

$$\int_{-\infty}^{\infty} t^k \psi(t) dt = 0, \quad k = 0, \dots, n. \quad (2.7)$$

The value of vanishing moments k depends on the wavelet family and mother function order.

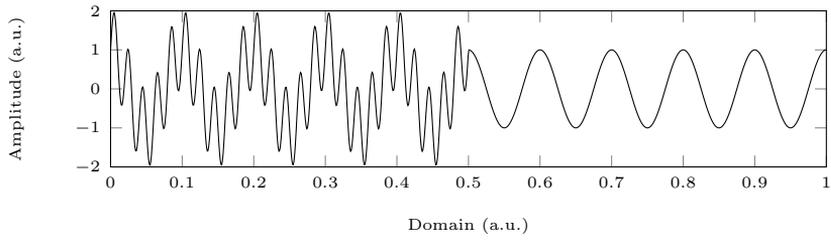
2.2.2 Continuous wavelet spectrum

Wavelet spectrum $W_\psi(a, b)$ is the two-dimensional set of the wavelet expansion coefficients. The visual analysis of $W_\psi(a, b)$ is very useful in applications as it provides the most detailed time-frequency or domain-scale information about the signal of interest.

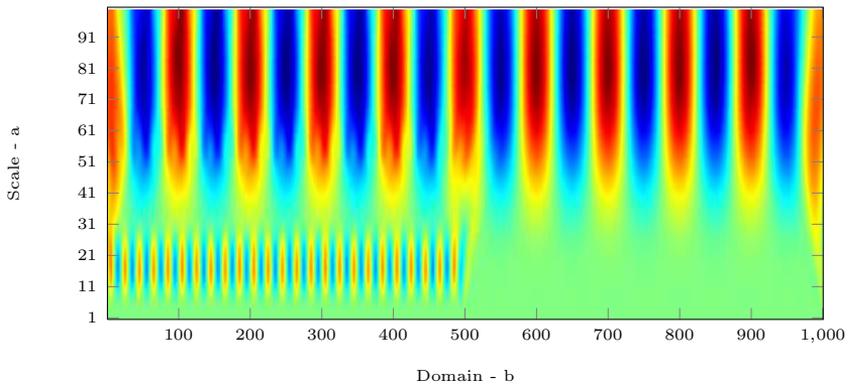
The coefficients W_ψ^{ab} are visualized as colored or gray-scaled images. The translation b and scale a are plotted along the horizontal and vertical axes respectively.

An example of variable harmonic signal is given in figure 2.2. The high frequency component is separated from overall signal and is clearly visible on small scales (ranges $11 \div 21$) in the left part of wavelet spectrum. The

(a)



(b)



(c)

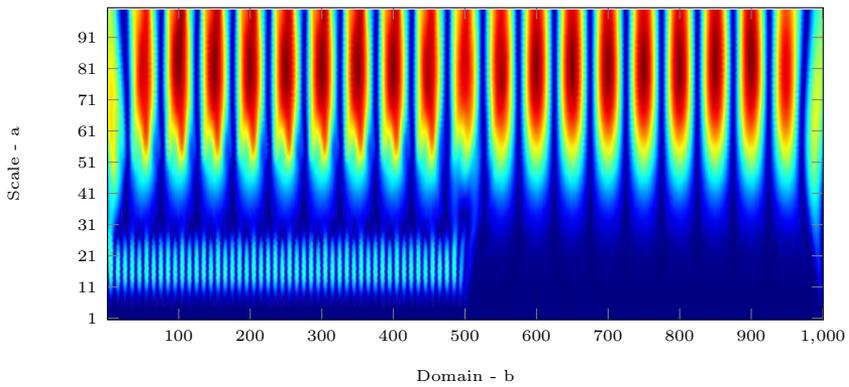


Figure 2.2: An example of continuous wavelet spectrum. (a) - signal in time domain with two harmonics, (b) - a wavelet spectrum W_ψ^{ab} , (c) - an absolute value wavelet spectrum.

low-frequency harmonics are visible over the entire signal on large scales > 60 .

In order to express the differences in wavelet spectrum an additional view 2.2c is created using the absolute values of W_{ψ}^{ab} . It is useful in preliminary signal domain-scale analysis as it amplifies the edges between the significant and hides minor wavelet coefficients.

2.2.3 Multiresolution Discrete Wavelet Transform

The compact support of the wavelet function provides the most complete and accurate domain-scale representation of non-stationary signals. However, the continuous basis presented by CWT in previous chapter is over-complete. The more optimal representation of signal in wavelet basis can be done using the Discrete Wavelet Transform (DWT).

The obvious choice for DWT is to use the discrete values of translation b and scale a coefficients. In usual for a the so called *dyadic* grid is applied, i.e let $a = 2^j$ and $\frac{b}{a} = k$ where j and k are integers.

The dyadic grid allow to use a fast algorithm, an analog to FFT, to calculate the WT for discrete sampled data. The core of fast DWT is called the Multiresolution analysis (MRA).

The MRA is defined an a sequence of nested subspaces $V_j \subset L_2(R), j \in Z$ where $\dots \subset V_{-1} \subset V_0 \subset V_1 \subset \dots \subset V_j \subset V_{j+1} \subset \dots \subset L_2(R)$ with next properties:

- *Self similarity in scale* - $v(t) \in V_j \Leftrightarrow v(2t) \in V_{j+1}$,
- *Self similarity in space* - $v(t) \in V_j \Leftrightarrow v(t + 1) \in V_j$,
- *Completeness* - $\bigcup V_j$ is dense in $L_2(R)$ and $\bigcap V_j = \{0\}$,
- *Scaling function* - there exists such scaling function $\varphi(t) \in V_0$ that a collection of integer translations $\{\varphi(t - k) \mid k \in Z\}$ form a orthonormal basis in V_0 .

As in CWT the zero moment of scaling function must equal to zero. From the properties of MRA the scaling function $\varphi_0(t)$ creates an orthonormal basis in all nested subspaces of V_0 by scaling transform:

$$\varphi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k).$$

In general any arbitrary signal $f(t) \in L^2(R)$ can be decomposed into subspaces V_j , creating a set of multiscaled orthogonal functions $v_j \in V_j$. The reconstruction of $f(t)$ is done simply by combining all v_j . Additionally the approximation of signal $f(t)$ can be achieved by removing some of the finer subspaces of V_j from reconstruction.

In practice the scaling function $\varphi_j(t)$ at level j can be viewed as a linear combination of translated and more compact $\varphi_{j+1}(t)$ using the rescaling coefficients h_k . The linear relationship between the two nested bases is called the *refinement equation*:

$$\varphi(t) = \sum_k h_k \varphi(2t - k). \quad (2.8)$$

The solution of (2.8) gives the concrete scaling function. The coefficients h_k are calculated using the orthogonality constraint as

$$\langle \varphi(t), \varphi(2t - k) \rangle = \delta_k, \quad (2.9)$$

where δ_k is Kronecker delta function. Following from the MRA properties and equation (2.8) the transfer between V_{j+1} and V_j spaces is a normed signal decimation by two. The normed decimation can be viewed as a low-pass filtration using the h_k with the cutoff frequency equal to the half Nyquist frequency of v_{j+1} . As the half of frequencies are removed from the signal it can be decimated by two into v_j .

The high-frequency information removed from v_{j+1} must be stored for further signal reconstruction. The subspaces W_j which must be orthogonal to V_j provide the additional details for signal reconstruction as

$$V_{j+1} = V_j \oplus W_j.$$

The details subspaces W_j have their own basis function that is called *wavelet*. The wavelet and scaling functions are tightly bound. If we consider that operator h_k from (2.8) is a low-pass filter then there should exist such operator g_k that

$$\psi(t) = \sum_k g_k \varphi(2t - k). \quad (2.10)$$

The (2.10) is the two-scale relation between the scaling function and wavelet. The filter coefficients $g(k)$ are obtained from the $h(k)$ using the well-known method of quadrature mirror filters

$$g(k) = (-1)^k h(2N + 1 - k), \quad k = 1, 2, \dots, 2N, \quad (2.11)$$

where the N is the length of h .

Using the pair of h and g filters the signal decomposition can be done as dyadic decimated convolution scheme also known as *two-channel subband coding*

$$cA_{j+1}(k) = \sum_n h_n cA_j(2k + n), \quad (2.12)$$

$$cD_{j+1}(k) = \sum_n g_n cA_j(2k + n). \quad (2.13)$$

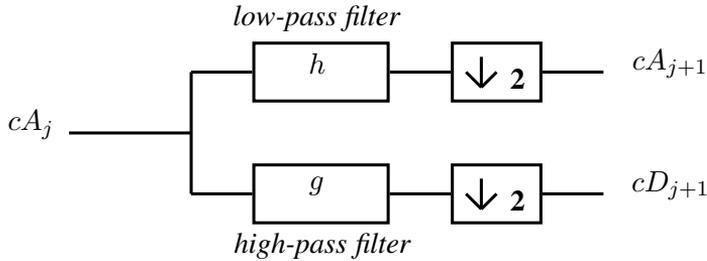


Figure 2.3: The one stage of fast DWT scheme implemented as an iterated filter bank

The scheme is implemented as iterative algorithm using the two equations (2.12) and (2.13) for computing the approximation and details subspaces (see figure 2.3). The first low-pass filter h extracts the low-half frequency information from signal $f(t)$. Another filter g extracts the upper-half frequencies. According to Nyquist theory the sampling frequency of output can be divided by two, i.e the output is decimated. This gives that output data $cA_{j+1} \oplus cD_{j+1}$ has the same number of samples as cA_j . In discrete signal processing the first $cA_0 = f(t)$.

The redundancy of CWT is removed by using the dyadic decimation so that at final stage the number of basis coefficients equals the number of samples in initial signal as in FFT. The iterative algorithm allows to process a WT without specifying the wavelet itself which is important for practical use. Specifically the use of filters h and g allows to create groups of orthogonal wavelets without defined analytical representation. An example of such group is the famous *Daubechies* wavelets.

The *Daubechies* filter coefficients h and g are calculated using the equations (2.8), (2.10) with additional wavelet and scaling functions constraints. The first equation is derived from the orthogonality constraint (2.9):

$$\sum_k h_k h_{k+2t} = \delta_t, \quad (2.14)$$

second from norm constraint:

$$\sum_k h_k = \sqrt{2}, \quad (2.15)$$

and third from wavelet smoothness using (2.7)

$$\int_t t^n \psi(t) dt = 0, \quad n = 0, \dots, N-1,$$

$$\sum_k k^n g_k = 0, \quad (2.16)$$

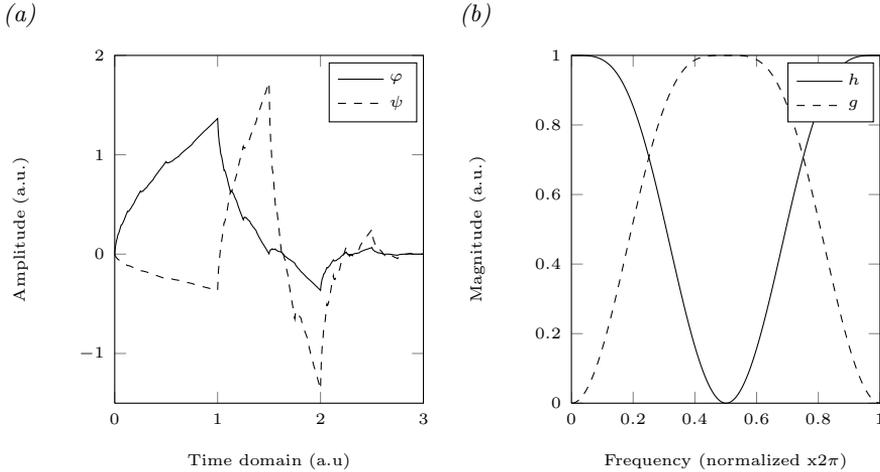


Figure 2.4: An example of wavelet and scaling function of second order from Daubechies wavelet family. (a) - reconstructed scaling and wavelet functions from filter coefficients. (b) - the Fourier spectra of h and g filters.

we can substitute h_k into (2.16) using the (2.11)

$$\sum_k (-1)^k k^n h_k = 0, \quad n = 0, \dots, N - 1. \quad (2.17)$$

Now we can rewrite the equations (2.14), (2.15) and (2.17) in explicit form for $N = 2$ as

$$\begin{aligned} h_0 h_2 + h_1 h_3 &= 0, \\ h_0 + h_1 + h_2 + h_3 &= \sqrt{2}, \\ h_0 - h_1 + h_2 - h_3 &= 0 \quad \text{for } n = 0, \\ -h_1 + 2h_2 - 3h_3 &= 0 \quad \text{for } n = 1. \end{aligned} \quad (2.18)$$

The solution of (2.18) is

$$h_0 = 0.483, \quad h_1 = 0.837, \quad h_2 = 0.224, \quad h_3 = -0.129.$$

The wavelet filter coefficients g are calculated using the (2.11)

$$g_0 = -0.129, \quad g_1 = -0.224, \quad g_2 = 0.837, \quad g_3 = -0.483.$$

The resulting scaling and wavelet coefficients for $N = 2$ are known as second-order Daubechies wavelet. The reconstructed φ and ψ functions with filters FFT responses are shown in Figure 2.4.

The orthogonality constraint (2.9) applies a limitation on the construction of wavelets. All orthogonal wavelets, except the *Haar* wavelet, do

not have symmetric support which is essential property for many applications. To overcome this limitation a group of biorthogonal wavelets was introduced. The biorthogonal WT have a dual $\tilde{\varphi}$ scaling and $\tilde{\psi}$ wavelet functions that create following pairs

$$\begin{aligned}\langle \tilde{\varphi}, \varphi(-k) \rangle &= \delta_k, \\ \langle \tilde{\psi}, \psi(-k) \rangle &= \delta_k, \\ \langle \tilde{\varphi}, \psi(-k) \rangle &= 0, \\ \langle \tilde{\psi}, \varphi(-k) \rangle &= 0.\end{aligned}$$

The biorthogonal analysis and reconstruction can be done using φ and $\tilde{\varphi}$ respectively or vice versa. The smoothness and number of zero moments for scaling and wavelet pairs can be individually controlled by selecting the functions order. This give an ability to chose the reconstruction quality independently from the wavelet properties. The fast biorthogonal MRA has the same scheme and properties as orthogonal using the h_k and g_k filters. Only difference is that a signal reconstruction is performed using the dual wavelet filters \tilde{h}_k and \tilde{g}_k .

2.2.4 Lifting

The CWT and DWT algorithms are based on the translations and dilations of one basis function on the uniform dyadic grid. In some cases the signal becomes a non-uniform sampled. This makes the samples exist on irregular grid and thus it is not possible to apply filters for scaling and wavelet functions based on the Fourier transform. Also the discrete filter sequences are not easy to apply on closed intervals. In order to process the finite length sampled signal various signal extension methods are used or some methods of wavelets on interval. However, the artificial signal extension using the periodization or linear extrapolation introduces significant artifacts on boundaries. The existence of boundary artifacts is critical for applications like feature extraction and data segmentation.

The Lifting scheme, which is also known as *Second Generation Wavelets*, introduces new framework to construct wavelets adapted to specifics signal domains. The transform is based on interpolating subdivision, average interpolation and lifting techniques. Reader can refer to original papers for more details (Sweldens, 1995; Fernandez et al., 1996; Sweldens and Schröder, 2000).

The basic idea behind the interpolating subdivision or average interpolation lies in construction of polynomials using the subset or signal samples which can be referred to as dyadic refinement scheme. For example in interpolating subdivision the estimation of some sampled function f can be

found over the twice finer grid using the Lagrange interpolating polynomials first by constructing the polynomial p as

$$p(x_{j,k+l}) = \lambda_{j,k+l} \text{ for } -D + 1 \leq l \leq D,$$

where $\lambda_{0,k}$ are the original samples of signal. The signal coefficients on the next finer level are constructed as

$$\lambda_{j+1,2k+l} = p(x_{j+1,2k+l}).$$

The result of such refinement in infinity for $\lambda_{0,k} = \delta_k$ will be the scaling function $\varphi(x)$. The resulting function has all properties of scaling function for WT: compact support, symmetry, polynomial reproduction, smoothness and refinability.

The idea of WT is to decorrelate signal, i.e to present the information contained in the signal with fewer coefficients. In the simplest case we can reduce the number of coefficients by subsampling the even samples from the original data as

$$\lambda_{-1,k} := \lambda_{0,2k} \text{ for } k \in Z.$$

The remaining information in odd samples $\lambda_{0,2k+1}$ must be stored for future signal recovery. Storing the odd samples simply as additional subset of coefficients $\gamma_{-1,k} := \lambda_{0,2k+1}$ which is known as *Lazy wavelet* is meaningless. The Lazy wavelet will produce small coefficients only in case when original odd coefficients are small.

One of possible solutions of compressing information included in $\lambda_{0,2k+1}$ samples is the interpolating scheme introduced earlier. The interpolation subdivision creates a prediction operator P as

$$\gamma_{-1,k} := P(\lambda_{-1,k}).$$

The prediction of odd samples uses the correlation present in the original even samples and thus it should decorrelate signal in wavelet coefficients. In addition the operator P should be invertible in order to reconstruct original odd samples. In practice the prediction of $\lambda_{0,2k+1}$ is not ideal, so storing the predicted values is useless. Thus, for perfect reconstruction, we need to replace the $\gamma_{-1,k}$ with error between the original $\lambda_{0,2k+1}$ and predicted $P(\lambda_{-1,k})$ values

$$\gamma_{-1,k} := \lambda_{0,2k+1} - P(\lambda_{-1,k}). \quad (2.19)$$

The equation (2.19) is called *the prediction step* or *dual lifting step*, it creates the wavelet coefficient $\gamma_{-1,k}$ to encode the error between the real data and interpolation model. If the model is close to the signal the wavelet coefficients $\gamma_{-1,k}$ will be small. In terms of frequency content the

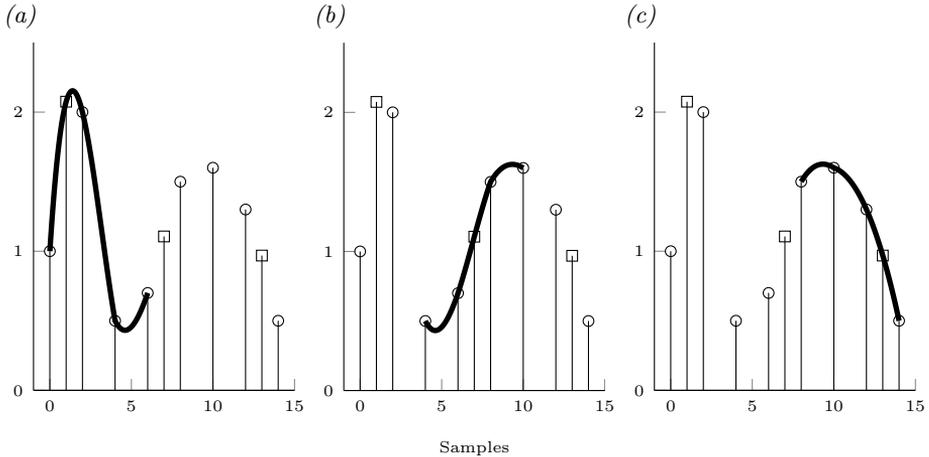


Figure 2.5: Interpolation of function near the boundary using cubic polynomial. Interpolated odd samples number 1, 7 and 13 marked with squares. (a) - interpolation affected by the left boundary, value at sample 1 calculated using one polynomial coefficient on the left and three on the right; (b) - interpolation not affected by the boundary, value at 7 calculated with two coefficients on the left and two coefficients on the right; (c) - interpolation affected by the right boundary, the flipped version of (a).

wavelet will capture the high frequencies present in signal. The two subsets $\lambda_{-1,k}, \gamma_{-1,k}$ likely to create more compact representation of original data.

The prediction operator P uses the degree of polynomial to denote the order of interpolation scheme N . For example first degree polynomial creates the piecewise linear interpolation with $N = 2$. The cubic polynomial with $N = 4$ creates the cubic interpolation. For scaling function it means that it reproduces polynomials up to degree of $N - 1$, thus the value of N denotes the smoothness of the scaling function.

The signal boundary problem introduced earlier can be easily omitted with the interpolation framework without manipulations with original signal. We can construct the polynomial of degree $N - 1$ using at least N samples. In case of signal on interval the construction of polynomial is done using the existing samples on the left or right sides of the boundary. The example of this is shown in figure 2.5.

The introduced interpolation scheme creates the wavelet and scaling functions which are adapted to interval. The wavelet extracts the high frequency component of signal and store it in $\gamma_{j,k}$. The low frequencies are stored in $\lambda_{j,k}$ as coarser signal approximations. One problem with wavelets in interpolating subdivision scheme is that they are built directly from the scaling function. Suppose we have two signal approximations A_{j+1} and A_j . The difference between two successive approximations is created first by

subsampling the even samples $\lambda_{j+1,2k}$, applying the interpolation operator on them $A_j(\lambda_{j+1,2k})$, cascading the predicted values back one level and finding the difference between A_{j+1} and A_j at predicted samples.

$$\gamma_{j,k} = \lambda_{j+1,2k+1} - A_j(\lambda_{j+1,2k}).$$

The $A_j(\lambda_{j+1,2k})$ in infinitum gives a scaling function $\varphi(x)$ so the wavelet coefficients represent the failure of signal at samples $\lambda_{j+1,2k+1}$ to follow the polynomial and as a result the wavelet function is simply the refined scaling function $\psi(x) = \varphi(2x - 1)$. We now recall to wavelet property of vanishing moments. If the wavelet is created as the difference between two successive approximations then

$$\int \psi(x) dx = \int A_{j+1}f(x) - A_jf(x) dx = 0,$$

which gives us

$$\int A_{j+1}f(x) dx = \int A_jf(x) dx. \quad (2.20)$$

We see that storing the even samples for approximations as proposed earlier will not satisfy the (2.20) and will create wavelet without even one vanishing moment as $\varphi(2x - 1)$ scaling function does not have it. In addition the coarser approximations will introduce aliasing. In order to satisfy the (2.20) we can reuse the wavelet coefficients $\gamma_{j,k}$ to update the approximation $\lambda_{j,k}$ on the same scale j with the help of additional *primal lifting* step and operator U as

$$\lambda_{j,k} := \lambda_{j,k} + U(\gamma_{j,k}).$$

The lifting step can be expressed as taking an old wavelet and build a better new by using a scaling functions of the same level j

$$\psi_{j,k} = \varphi_{j+1,2k+1} - A_{j,k}\varphi_{j,k} - B_{j,k}\varphi_{j,k+1}.$$

The coefficients $A_{j,k}$ and $B_{j,k}$ are found using the necessary condition of vanishing moments such that

$$\int_k x^{\tilde{N}-1} \psi_j(x) dx = 0.$$

As in the case of dual lifting step the primal lifting can use more neighborhood coefficients $\gamma_{j,k}$ to increase the smoothness of wavelet.

The process of lifting transform is typically described in terms of three stages of lifting:

- *Split* - The Lazy wavelet is constructed as the initial step. Other choice is to use the Haar wavelet and subtract the odd coefficients. The latter will create the primal lifting.

- *Predict* - This step establishes the number of vanishing moments of dual wavelet N and the smoothness of primal scaling function. It can be viewed as dual lifting step for Lazy wavelet. The original signal polynomial parts of order less than N will be suppressed, thus giving zero wavelet coefficients.
- *Update* - This step establishes the vanishing moments of primal wavelet \tilde{N} , or smoothness of dual scaling function. The dual scaling function preserves low frequency information in approximation coefficients λ_j .

The transform stages can be iterated to generate the 1-D fast lifting algorithm where the inverse is done using the same predict and update operators in reverse order.

The advantage of lifting scheme is that it naturally adapts to the signal domain. Actually predict and update operators can be specifically selected for the each application and they can be even be non-linear. The only constraint remains, that is needed for perfect reconstruction, is the wavelet vanishing moments property. The effectiveness of LWT has been proven in edge detection, pattern recognition and feature extraction among other examples. Customized implementations of lifting and its inverse transforms are given in Appendix section of this work and are used in other algorithms.

2.3 HLIF signal-noise model

This paragraph presents the research behind the HLIF noise modeling and estimation of noise model parameters. The HLIF signal-noise analysis uses WT, which has showed to be powerful tool for noise processing in various applications. The noise model and parameters will be used in next sections for data simulation.

2.3.1 HLIF LIDAR noise sources

The LIDAR hyperspectral linear detector based on gated image intensifier camera is subject to signal-dependent or photon noise. This main noise is generated by MCP due to the high gain up to $10^4 - 10^6$ which brings LIDAR almost to photon-counting mode of operation which is common for PMT devices (Matsuura et al., 1985). Additional noise comes from linear image sensor which is subject to photon noise, dark noise and read noise (Janesick, 2001; Xu et al., 2004). The latter two noise sources have normal Gaussian distribution. The LIDAR system generally operates under photon-noise limited conditions where photon noise exceeds both read noise and dark noise of image sensor and acquisition electronics.

The general strategy for noise reduction in LIDAR is provided as multiple accumulation of signal followed by averaging. The approach requires hitting with multiple laser pulses into single target which under many situations is not possible. The better approach is to reduce the noise in HLIF data using signal denoising technique. Due to the specifics of hyperspectral linear detector the conventional denoising algorithms, based on normal *white* noise, demonstrate poor performance, thus the adequate HLIF noise model is subject of interest.

2.3.2 Poisson-Gaussian noise model

In signal processing a widespread modeling of signal noise $\xi(x)$ is to add the linearly additive part into observation model $z(x)$ of original signal $y(x)$

$$z(x) = y(x) + \sigma\xi(x), \quad (2.21)$$

where $x \in R$ is the pixel position in domain X and $\xi(x) \in R$ is zero-mean independent random noise with standard deviation equal to σ . The assumption of noise as multivariate random i.i.d. vector created many denoising methods, like famous “Stein Unbiased Risk Estimate” (SURE) in Donoho and Johnstone (1995). Briefly, their method is based on so called “shrinkage” of wavelet expansion coefficients by *soft thresholding*. The adaptive threshold is estimated by calculating expectation of SURE, which is an unbiased estimator of MSE risk function. The expectation

minimum of it can be found without knowledge of exact y , however only assuming that noise p.d.f. is normal.

In order to incorporate the signal-dependent photon-noise in observation model let's change the (2.21) to the generic model of the form

$$z(x) = y(x) + \sigma(y(x))\xi(x), \quad (2.22)$$

known as *Poissonian-Gaussian noise model* (Foi et al., 2008). In this model a noise $\sigma(y(x))\xi(x)$ is selected as two independent components

$$\sigma(y(x))\xi(x) = \eta_p(y(x)) + \eta_g(x),$$

where η_p is *Poissonian* and η_g is *Gaussian* components with $\mathcal{P}(\chi y(x))$ and $\mathcal{N}(0, b)$ distributions. The elementary properties of Poisson and Gaussian distributions give the following standard deviation function of noise in z as

$$\sigma(y(x)) = \sqrt{ay(x) + b}, \quad (2.23)$$

where $a = \chi^{-1}$ is a signal dependent noise constant and b is a white noise variance. The error distribution of the samples $z(x)$ can be approximated with heteroscedastic Gaussian model by sum of two independent distributions. The Poisson distribution can be usually approximated by normal as

$$\mathcal{P}(\lambda) \approx \mathcal{N}(\lambda, \lambda),$$

which gives the sum of two normal distributions as

$$\begin{aligned} \mathcal{P}(ay(x)) &\approx ay(x) + \mathcal{N}(0, ay(x)), \\ \sigma(y(x))\xi(x) &\sim \mathcal{N}(0, ay(x)) + \mathcal{N}(0, b) = \mathcal{N}(0, ay(x) + b). \end{aligned} \quad (2.24)$$

Equation (2.24) gives the form of generic noise model with unknown parameters a and b . In Foi et al. (2008) the model parameters a and b are estimated by maximum-log-likelihood problem optimization

$$\left(\hat{a}, \hat{b}\right) = \arg \max_{a,b} \ln[L(a, b)] = \arg \min_{a,b} -\ln[L(a, b)],$$

where the likelihood function is obtained as

$$L(a, b) = \prod_{i=1}^N \int_0^1 \wp((\hat{y}_i, \hat{\sigma}_i)|y_i = y) dy,$$

where $i = 1, \dots, N$ are the uniform segments of spectrum.

Variables \hat{y}_i and $\hat{\sigma}_i$ are estimated from each uniform segment i as sample mean and sample standard-deviation from detail and approximation coefficients respectively. Using the Daubechies wavelets (see figure 2.4) it is

clear that the norm constraint of scaling function (2.15) will rescale the approximation coefficients. As we want to estimate the true amplitude value of uniform segment from approximation coefficients directly we normalize the scaling kernel as $\sum h = 1$. This will ensure that resulting coefficients and \hat{y}_i are unbiased. The detail coefficients, hence the $\hat{\sigma}_i$ estimator, are not biased as the Daubechies wavelet kernel has $\|g\| = 1$ naturally.

It must be noted that practical implementation of segmentation algorithm uses the Lifting scheme in order to remove the signal boundary artifacts. This greatly enhances the accuracy of $\hat{\sigma}_i$, which in turn improves the estimation of noise parameters (see Appendix section).

Instead of using the conditional probability density $\wp((\hat{y}_i, \hat{\sigma}_i)|y_i = y)$ and evaluating integral a direct minimization of Sum of Squared Difference is applied as

$$\mathcal{R}(\sigma_{fit}(y), \hat{\sigma}_i) = \sum_{i=1}^N (\sigma_{fit}(y) - \hat{\sigma}_i)^2, \quad (2.25)$$

where $\sigma_{fit}(y) = \sqrt{\max(0, \hat{a}y + \hat{b})}$ is an estimate of $\sigma(y(x))$.

The function in (2.25) has significant drawback in finding the global minimum. The distribution of $\hat{\sigma}_i$ estimate according to signal-dependent noise model can be taken as normal and in particular $\hat{\sigma}_i \sim \mathcal{N}(\sigma(y_i), \sigma^2(y_i)d_i)$. This follows that the residual \mathcal{R} is heteroscedastic. The heteroscedasticity in fitting the noise model (2.23) will degrade the fit in lower values of y_i which in turn will make estimation of $\sqrt{\hat{b}}$ unreliable. This influence on the error residuals can be efficiently compensated by applying simple non-linear transformation on $\hat{\sigma}_i$ and $\sigma_{fit}(y)$ as

$$\mathcal{R}(\sigma_{fit}(y), \hat{\sigma}_i) = \sum_{i=1}^N \left(\ln \frac{\sigma_{fit}(y)}{\hat{\sigma}_i} \right)^2. \quad (2.26)$$

The $\arg \min_{a,b} \mathcal{R}(\sigma_{fit}(y), \hat{\sigma}_i)$ is found numerically using the derivative free simplex search method (Lagarias et al., 1998). In order to guarantee that solution to (2.26) can achieve the global minimum the *multi-start* approach is used (Boender and Kan, 1987) starting with different initial guess of

$$(\hat{a}_0, \hat{b}_0) = \{a_n^* \in \mathcal{N}(\alpha, \alpha^2), b_n^* \in \mathcal{N}(\beta, \beta^2) | n = 1, \dots, N\},$$

where α and β are empirically selected values which are usually found in HLIF data and find the approximation of the global minimum as

$$(\hat{a}, \hat{b}) = \min_{a,b} \left(\arg \min \mathcal{R}_n | a_n^*, b_n^* \right). \quad (2.27)$$

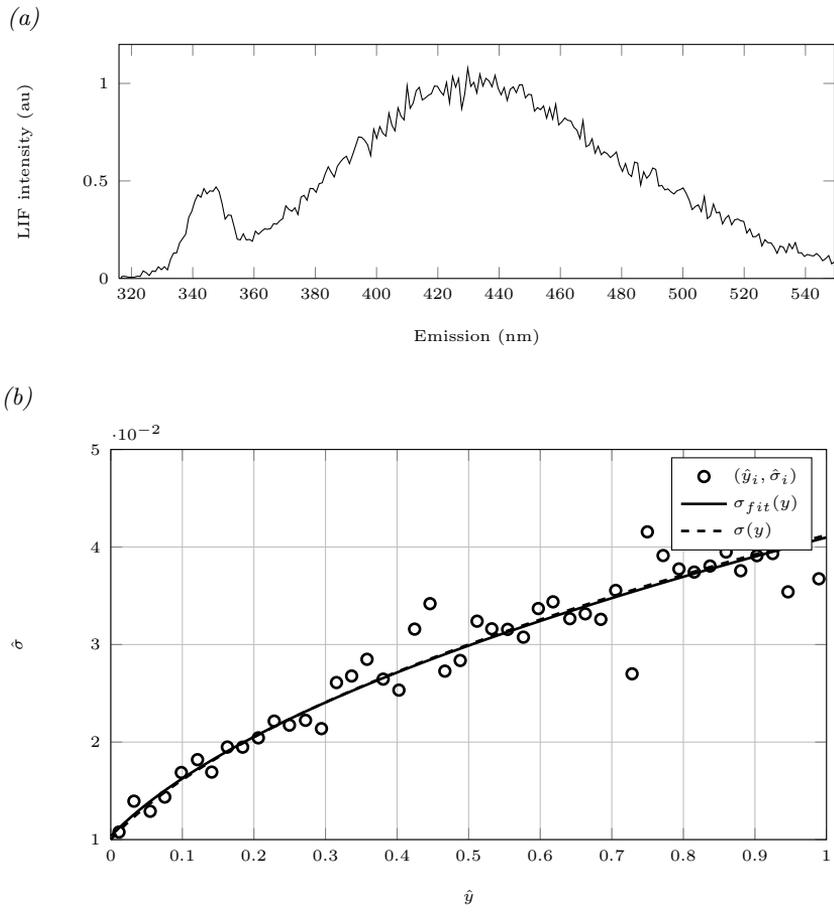


Figure 2.6: (a) - Water spectrum with simulated signal dependent noise. The noise parameters are $a = 0.04^2$ and $b = 0.01^2$. (b) - The dots represent values of signal expectation and variance using uniform segments from 100 simulations. The estimated noise parameters are $\hat{a} = 0.0395^2$ and $\hat{b} = 0.0104^2$.

Figure 2.6 shows the result of estimation of \hat{y}_i and $\hat{\sigma}_i$ pairs from noisy HLIF spectra of natural water. The noise in example was simulated using the Poisson-Gaussian noise model with parameters $a = 0.04^2$ and $b = 0.01^2$.

The solution (2.27) assumes the continuous probability density of observation model $z(x)$ from (2.22). However, generally the image sensor has limited data range which results in nonlinear clipping of observations to maximum and minimum values of sensor acquisition system. This will create a significant distortion between model and estimators $\hat{\sigma}$ and \hat{y} near the lowest and highest borders of data range. According to original paper (Foi et al., 2008) if data is given in the normalized range $[0, 1]$ the clipped observations \tilde{z} can be denoted as

$$\tilde{z}(x) = \max(0, \min(z(x), 1))$$

and the noise model is

$$\tilde{z}(x) = \tilde{y}(x) + \tilde{\sigma}(\tilde{y}(x))\tilde{\xi}(x).$$

The direct and inverse transformations that convert y and σ to \tilde{y} and $\tilde{\sigma}$ are based on the corrected p.d.f of the normally distributed random variable $\nu \sim \mathcal{N}(\mu, 1)$. For ν clipped from below $\tilde{\nu} = \max(0, \nu)$ the expectation $\mu = E\{\tilde{\nu}\}$ and standard deviation $\text{std}\{\tilde{\nu}\} = \sqrt{\text{var}\{\tilde{\nu}\}} \sim 1$ are shown in figure 2.7 as provided by numerical simulation.

The exact analytic solution also exist and can be found in Johnson et al. (1994) as

$$\begin{aligned} E\{\nu\} &= \Phi(\mu)\mu + \phi(\mu), \\ \text{var}\{\nu\} &= \Phi(\mu) + \phi(\mu)\mu - \phi^2(\mu) + \Phi(\mu)\mu(\mu - \Phi(\mu)\mu - 2\phi(\mu)), \end{aligned}$$

where ϕ and Φ are the probability density and cumulative distribution functions of the normal $\mathcal{N}(0, 1)$ distribution.

Using the above calculations and assuming distribution of $z(x)$ noise from (2.24) as heteroscedastic normal one can write clipping transformations as

$$\begin{aligned} z &\sim \mathcal{N}(y, \sigma^2(y)) = \frac{z}{\sigma(y)} \sim \mathcal{N}\left(\frac{y}{\sigma(y)}, 1\right), \\ \mu &= \frac{y}{\sigma(y)} \Rightarrow \tilde{z} = \sigma(y)\tilde{\nu}, \\ \tilde{y}(x) &= E\{\tilde{z}(x)\} = \sigma(y)E\{\tilde{\nu}\} \left(\frac{y}{\sigma(y)}\right), \\ \tilde{\sigma}(\tilde{y}) &= \text{std}\{\tilde{z}\} = \sigma(y)\text{std}\{\tilde{\nu}\} \left(\frac{y}{\sigma(y)}\right). \end{aligned}$$

The clipping from above can be defined as the inversion of clipping from below by subtracting $y, \tilde{y}, z, \tilde{z}$ from 1. The combination of clippings can

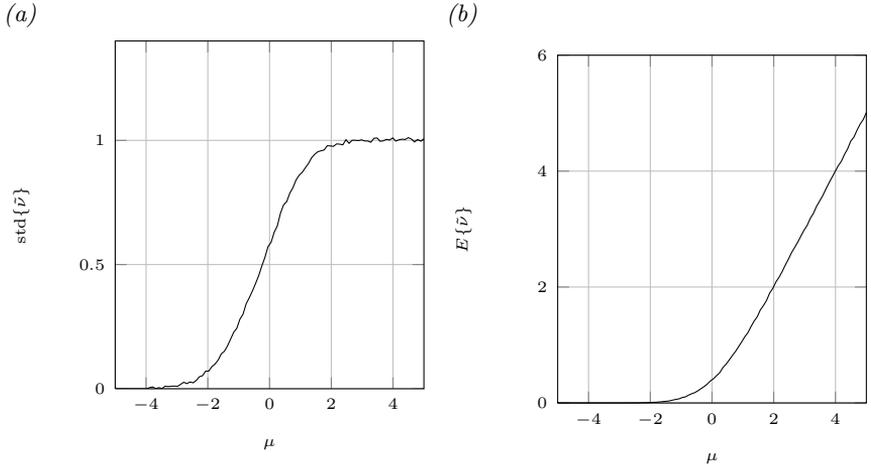


Figure 2.7: Standard deviation (a) and the expectation (b) of the clipped from below \tilde{v} data as functions of μ . Numerical simulation. The length of \tilde{v} data is 10000 samples for each μ point.

be formulated simply by summing the equations. The final equations for clipping from Foi et al. (2008) are

$$\begin{aligned}\tilde{y} &= \sigma(y)E\{\tilde{v}\} \left(\frac{y}{\sigma(y)} \right) - y + 1 - \sigma(y)E\{\tilde{v}\} \left(\frac{1-y}{\sigma(y)} \right), \\ \tilde{\sigma}(\tilde{y}) &= \sigma(y)\text{std}\{\tilde{v}\} \left(\frac{y}{\sigma(y)} \right) \text{std}\{\tilde{v}\} \left(\frac{1-y}{\sigma(y)} \right).\end{aligned}\quad (2.28)$$

The customized solution in (2.27) with the Sum of Squared Difference minimization requires to calculate clipped version only for σ_{fit} . The equation (2.26) becomes

$$\tilde{\mathcal{R}}(\tilde{\sigma}_{fit}(y), \hat{\sigma}_i) = \sum_{i=1}^N \left(\ln \frac{\tilde{\sigma}_{fit}(y)}{\hat{\sigma}_i} \right)^2$$

where $\tilde{\sigma}_{fit}(y)$ is clipped version from (2.28). The optimization results displaying clipped transformations and performance of clipped fit are shown in figure 2.8 for simple test function $f(x) = x$, $x \in [0, 1]$ clipped from below and above. With simple linear function the clipping is clearly seen near the range boundaries of y .

The Matlab implementation of HLIF spectrum noise parameters estimation using the Poisson-Gaussian noise modeling with modified algorithms, including initial HLIF segmentation is given in Appendix section. The implementation uses Lifting scheme, which naturally removes the signal boundary artifacts during the segmentation of expectation-standard deviation pairs.

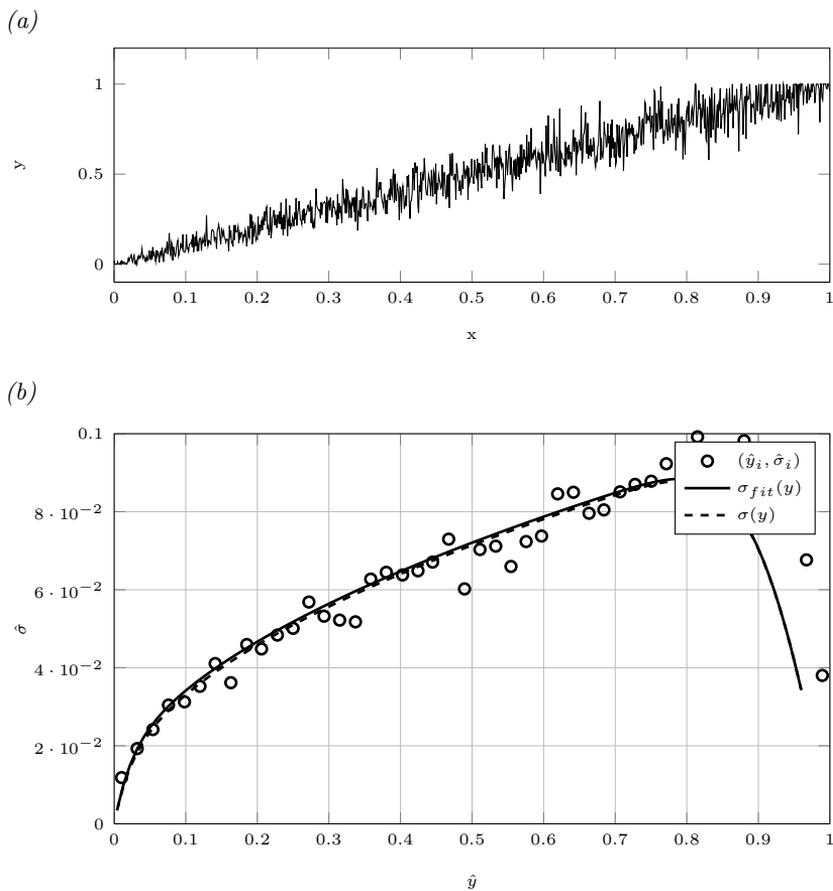


Figure 2.8: (a) - Clipped data test function with signal dependent noise. The noise parameters are $a = 0.1^2$ and $b = 0.01^2$. The clipping is seen for $y : x \rightarrow 0$ $y : x \rightarrow 1$ (b) - The dots represent values of signal expectation and variance using uniform segments from 100 simulations. The estimated parameters are $\hat{a} = 0.103^2$ and $\hat{b} = 0.011^2$.

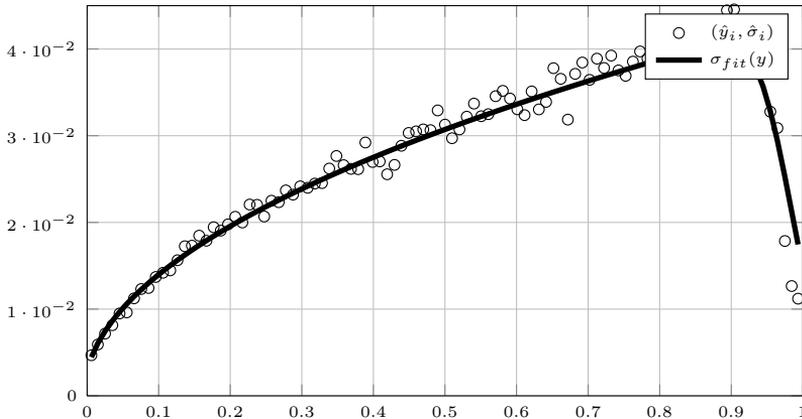


Figure 2.9: The dots represent values of signal expectation and variance using uniform segments from 500 HLIF spectra. Data derived from real world airborne HLIF LIDAR experiments. The estimated noise parameters are $\hat{a} = 0.043^2$ and $\hat{b} = 0.003^2$.

2.3.3 Noise parameters estimation of measured HLIF LIDAR data

Using the method described in the previous subsection the HLIF noise model parameters are estimated from the real airborne LIDAR data. The estimated noise model parameters are used for subsequent data simulation in this work.

The noise parameters for the signal dependent noise model of (2.22) are estimated from experimental HLIF LIDAR data. The description of LIDAR experiments and instrumentation can be found in Alekseyev et al. (2008); Babichenko et al. (2010); Sobolev (2010). For robust estimation the selected data contains five hundred samples of targeted medium with various fluorescence intensity covering the most of detectors dynamic range. The signal expectation and variance are computed according to “Estimation of Poisson-Gaussian noise model parameters” algorithm given in Appendix section of this work and final results are present in figure 2.9.

2.4 Conclusions

The main topics of WT theory have been introduced in this section. Among the large number of WTs each suitable for different applications the CWT, fast DWT and lifting scheme have been considered here.

Each WT scheme is suitable for specific signal analysis. The CWT provides the best visual representation of signal multi-resolution structure, which will be exploited in section 3.2 for preliminary HLIF structural anal-

ysis and multiscale comparison. The lifting scheme, known as II generation wavelets, has the ability to naturally adapt to signal domain, which is desired property for feature extraction and adaptive denoising methods proposed in next chapter.

The state-of-the art mixed Poisson-Gaussian noise model has been applied for HLIF signals with some modifications of original algorithm. The modifications included changes in signal segmentation, application of lifting scheme and parametric fitting of model using the nonlinear optimization of logarithm transformed SSD of model estimator with multi-start initialization. The results indicate correct estimation of mixed Poisson-Gaussian model parameters from simulated HLIF signals and clipped test function.

Using the the Poisson-Gaussian noise model and LIDAR measurements the real LIDAR sensor noise parameters were estimated. The results showed the two orders dominance of Poissonian noise component $\hat{a} = 0.043^2$ over the Gaussian one $\hat{b} = 0.003^2$. This effect meets the extremely low light signal acquisition conditions and the physical characteristics of MCP which gain is the main noise source of current LIDAR imaging sensor. The noise model and parameters derived in this section are applied in next chapter.

Chapter 3

HLIF data analysis

The organic pollution diagnostics within the HLIF data processing scope can be viewed as a conventional pattern recognition task. A fundamental challenge in automated pattern recognition is the extraction of discriminating and characterizing information about an observation objects (Polikar, 2006). For example, in section 1.4 the extraction of the relevant information has been introduced as spectral unmixing procedure. Using the information about the mixture components and estimating the correct model we could easily exclude the natural fluorescence background from observation and extract the relevant residual of pollutant for future identification. In terms of pattern recognition the extracted residual, while discriminated from pollutant end-member, can be future viewed as object containing redundant information. Thus, in this chapter the HLIF data will be analyzed considering the pattern-structure of fluorescence emission spectrum.

As the HLIF spectrum defined in 1.2 consist of hundreds of spectral bands it can be considered as high dimensional data, and extraction of relevant information will also contribute to the dimensionality reduction. This is essential because the high dimensional data-spaces create the computational effort which is detrimental for data processing (Radovanović et al., 2010). In pattern recognition the “feature extraction” term denotes the way of extraction of important information and removal of the irrelevant part from consideration. The feature extraction can have formally two directions: the statistical and structural approach.

Most methods of the multispectral and hyperspectral remote-sensing data analysis apply the statistical or decision-theoretic feature extraction methods (Landgrebe, 2005; Chen, 2012). These methods express the linear (Principal Component Analysis, Linear Discriminant Analysis, and Projection Pursuit) and nonlinear (IsoMaps and Generative Topographic Mapping) embeddings of specific data into new, low-dimensional, efficient spaces. However, the statistical feature extraction lacks optimizing the

object specific features as these methods are designed to optimize certain decision criteria (i.e. the variance). The interpretation of extracted features and binding with physical application is also not convenient and must be explicitly determined.

The alternative structural approach describes data in terms of particular or predefined primitive components and the relationship among these components (Pavlidis, 1977). In structural feature extraction the features are strictly binded with the data-space and thus the conditional relevance can be easily evaluated. There exist number of successful examples of hyperspectral (Bruce et al., 2002; Hsu, 2007) and high dimensional feature extraction using the structural approach.

The optimality of structural analysis of HLIF data has been demonstrated in Sobolev and Babichenko (2013b), where the discriminant measure constructed in redundant but structural feature-space using the CWT significantly improved the pollutants discrimination. That approach will be discussed here as “Multiscale spectra comparison” in section 3.2.2. An alternative way is to implement structural feature extraction as signal compression using the library of biorthogonal wavelets as proposed in (Sobolev and Babichenko, 2013a). As the feature extraction is sensitive to signal domain boundary artifacts, the method applies the LWT (see section 2.2.4) for wavelet coefficients expansion. Both methods use the ”dictionary“ of wavelets as the best ”language” which can efficiently capture information in spectral-scale domains. The corresponding numerical examples demonstrate the effectiveness of wavelet feature extraction for HLIF data. The global idea has been inspired by *sparsity-norm optimization*, known as *Orthogonal Matching Pursuit* (Cai and Wang, 2011).

The wavelet feature extraction is closely related to the adaptive denoising/compression method proposed in Saito (1994). The noise can be considered as the irrelevant part of information in observation, thus the denoising is a good option to improve the discriminating capabilities of pattern recognition system. By extending the wavelet feature extraction technique a novel Adaptive Slope Compensation (ASC) denoising method for HLIF data with mixed Poisson-Gaussian noise from 2.3.2 has been proposed. The effectiveness of method will be demonstrated on a numerical examples here and in the next chapter.

3.1 Author’s contribution

The author applied the structural feature extraction using the dictionary of biorthogonal wavelets and lifting scheme for improving the quality of HLIF pattern recognition. The wavelet based approach is deeply investigated, showing that WT expansion coefficients well represent localized structure

of hyperspectral fluorescence emission spectrum. On this basis the author proposed to extract the significant meaningful fluorescence features and use them in HLIF discrimination.

- Structural characterization of HLIF data using the CWT multiscale comparison is proposed;
- A wavelet basis for optimal HLIF feature extraction is selected considering the heuristic analysis;
- The wavelet based feature extraction as sparsity-norm optimization is proposed. The OPD example showed significant improvement when using the proposed “feature weighted distance“ discriminant measure. More examples are given in next chapter;
- Novel ASC denoising method effectively processing the Poisson, Gaussian and mixed noise sources. The method performance is compared to similar AMDL and ideal loss using the numerical examples.
- Experimental set up, measurements/simulation, implementing the necessary algorithms, data processing in provided examples;

3.2 Scale characteristics of HLIF spectra

The CWT reviewed in section 2.2.1 can be efficiently exploited to analyze the scale characteristics of HLIF signals acting like a *mathematical microscope*. The continuous wavelet spectrum is the set of the coefficients of the WT series $W_\psi f(a, b)$ in time-frequency domain. In HLIF case, the coefficients of the WT series reside in the wavelength and scale rather than the time and frequency domains. The CWT has specific boundary condition problem. Here it is assumed that the HLIF spectra have compact support on selected wavelength interval $L = (\lambda_0, \lambda_{end})$, thus the zero padding signal extension is applied with $x_k \equiv 0 \forall \lambda \notin L$ over the boundary. In next sections the signal extension will be unnecessary with Lifting scheme.

The most general spectrum analyzed in LIDAR environmental applications is the fluorescence spectrum of natural water. The composition of natural water fluorescence spectrum is created by two specific light emission processes. The first one is the characteristic RSS of water molecules and the second is the DOM fluorescence (see section 1.2.2). Figure 1.2 shows the HLIF spectrum of a Baltic sea water among others. The broad spectral structure with maximum at 430 nm is the DOM fluorescence, and the thin peak at 344 nm is the RSS. The emission spectra of DOM, as of most fluorophores, is rather broad and “smooth” structure. For other fluorophores the fluorescence spectrum can have multiple local extrema points and other small-scale details.

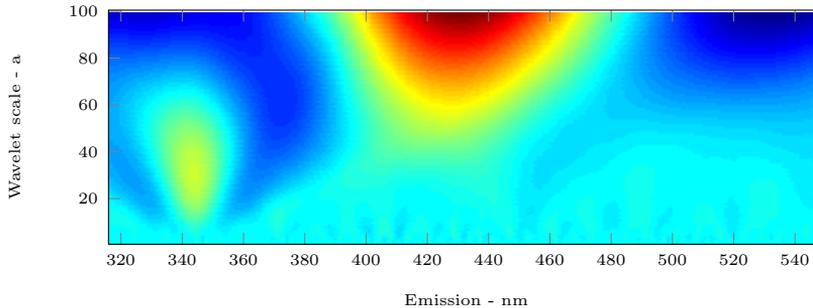
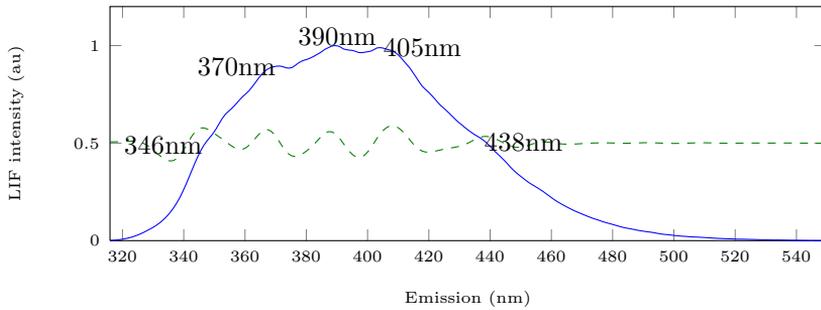


Figure 3.1: Example of natural water CWT spectrum using the reverse biorthogonal spline wavelet of order 2.4. The characteristic features of natural water are visible on small and large scales. RSS and DOM maximum are located on corresponding wavelength bands and on scales 30 and 100 respectively.

Figure 3.1 shows the CWT spectrum of Baltic sea water HLIF from figure 1.2. The wavelet used in this example is the reverse biorthogonal spline wavelet of order 2.4. Following from the CWT definition the spectrum coefficients become relatively large at scales where the signal features correlate best with the wavelet. The thin RSS is clearly seen on scales 10 to 60 and the broad DOM is separated on scales starting from 60. The spectral width of RSS in HLIF spectrum is lower-bounded by the spectral resolution of instrumentation. Thus the wavelet coefficients below the RSS scale ($a < 10$) do not correspond to spectral signal and represent the system noise. The wavelength domain localization of wavelet coefficients matches the peaks location on HLIF spectrum (344 and 430 nm).

The continuous signal features are separated by scales with CWT. For natural water example, the signal features became RSS peak and DOM, thus producing wavelet “unmixing” of complex HLIF signal. However, any smooth local features can be extracted and viewed as spectral details. Figure 3.2a displays the HLIF spectrum of Gas oil measured as optically dense film. The gas oil display complex structure which is seen as small-scale cambers at 370, 390 and 405 nm. The oscillations of wavelet coefficients on scales around 13 located from 340 to 440 nm are clearly seen on the CWT spectrum in figure 3.2b. For better visualization CWT spectrum is plotted by absolute value (for the rest of this work). The coefficients line at scale $a = 13$ is plotted separately in figure 3.2a with additional 0.5 constant shift for better visualization. These WT coefficients exactly match the locations of small-scale cambers, thus they represent the extracted spectral details of gas oil HLIF spectrum. Examination of coefficients line reveals additional spectral details on 346 and 438 nm. These are revealed locations of small-scale signal cambers that are visually hidden due to the limited spectral resolution of detector.

(a) HLIF spectrum of gas oil dense film



(b) Wavelet spectrum

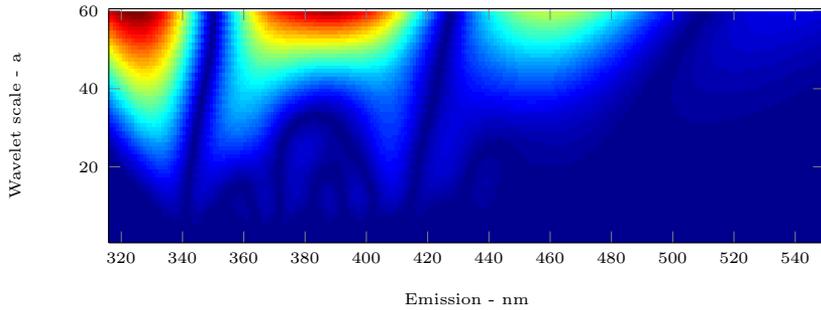
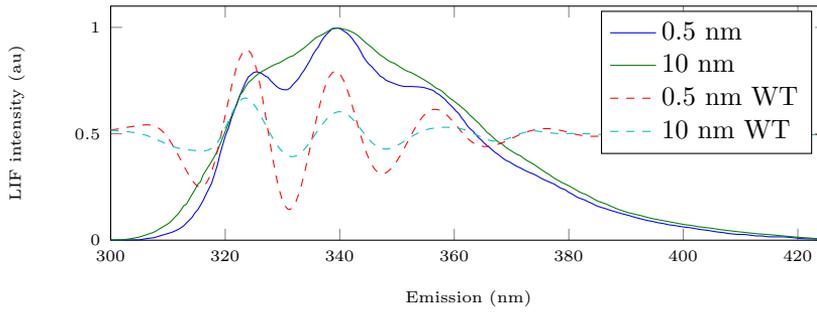


Figure 3.2: Example of gas oil CWT absolute value spectrum using the reverse biorthogonal spline wavelet of order 2.4. The CWT coefficients at scale $a = 13$ are shown on (a) as dashed line. Two additional small-scale cambers are visible at 346 and 438 nm.

(a) Emission spectra of *p*-terphenyl



(b) Wavelet spectra

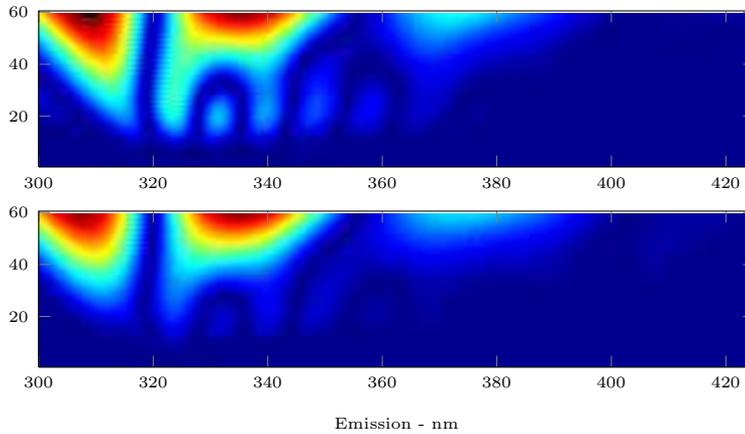


Figure 3.3: Emission spectra of *p*-terphenyl collected with spectral resolution of 0.5 and 10 nm. Spectral data taken from Lakowicz. The CWT coefficients at scale $a = 16$ are shown on (a) as dashed lines. The correlation value between the coefficients (dashed lines) is 0.96.

The last statement is significant for structural analysis. The spectral details, even though they exist, can be barely noticeable due to the broadening of fluorescence spectrum and limited detector spectral resolution. The effect of structural extraction is future demonstrated on spectra of *p-terphenyl* an aromatic hydrocarbon collected with a variable spectral resolution of 0.5 and 10 nm (see figure 3.3). The spectrum measured with high 0.5 nm resolution display vibrational structure with peaks located on 325, 340 and 355 nm. On the low resolution spectrum on 3.3a this structure is nearly lost. However, vast correlation of internal oscillations in emission is clearly seen on wavelet coefficients from scale $a = 16$ in both 0.5 and 10 nm resolutions.

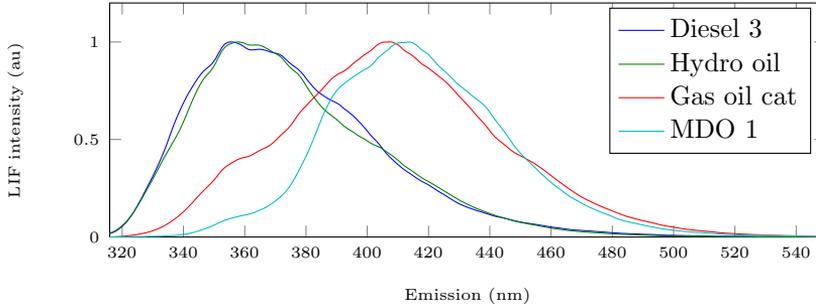
Another example of scale characteristics analysis with CWT is provided in figure 3.4. The two light and two medium refined oil products were measured as dense films. The differences between emission spectra of Diesel and Hydro oil are barely visible. The CWT spectra on 3.4b of corresponding samples introduce similar structure which is related to similar HLIF spectra. However, a small-scale detail ($a = 18$) in the range from 390 to 400 nm that present on Diesel sample is absent from the Hydro oil (figure 3.4b first and second). The Gas oil catalytic and MDO are better distinguishable by their emission spectra than previous pair of samples. The CWT spectra on 3.4b third and last image have significant differences on scales up to 40. The overall broad shape of medium refined samples resides on scales 60 and more. The location of fluorescence global maximum on scale $a = 60$ from 405 to 410 nm is visible in both samples.

The wavelet properties like compact support, regularity and vanishing moments provide very efficient basis to represent relevant information of HLIF signals. The spectral properties of importance are the related wavelength location and shape of the features. The examples demonstrate that WT preserves the spectral information and, in addition, extracts the small-scale spectral details from the HLIF data. This accurate method extends the informational components of the HLIF data so that the characteristic details of the HLIF data are preserved and are isolated for analysis.

3.2.1 Selection of optimal wavelet basis

In the previous section the analysis of HLIF data was intended to extract the small-scale details from emission spectrum while preserving the spectral properties of the latter. The wavelength localization of wavelet is dependent on the order N of its base function. Another important side of WT is the signal reconstruction from wavelet coefficients. For HLIF signals which ideally are continuous and smooth functions the reconstruction basis should be also smooth. As the result for optimal WT application we need maximal smoothness on reconstruction and maximally possible compact support on decomposition.

(a) HLIF spectra of four light refined oil products



(b) Wavelet spectra

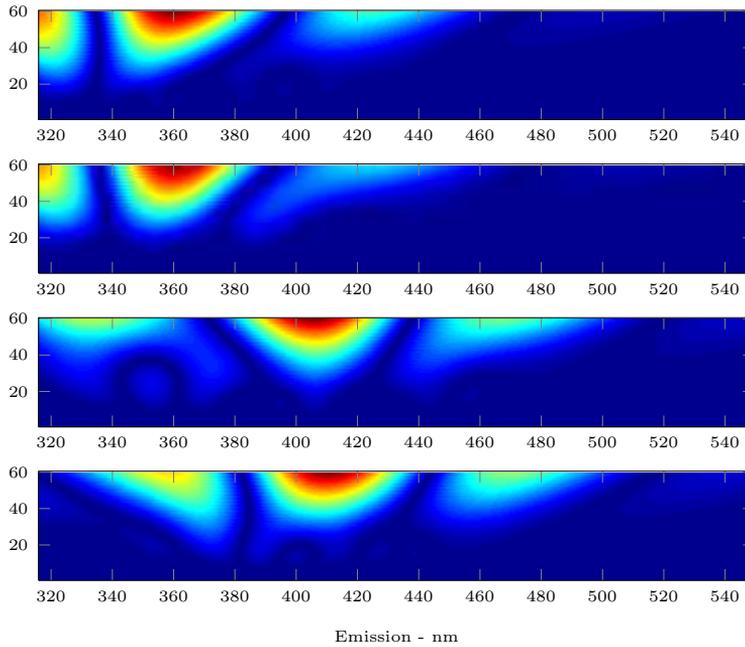


Figure 3.4: The CWT absolute value spectra of four light refined oil products. On (b) from top to bottom: Diesel, Hydro oil, Gas oil catalytic, MDO.

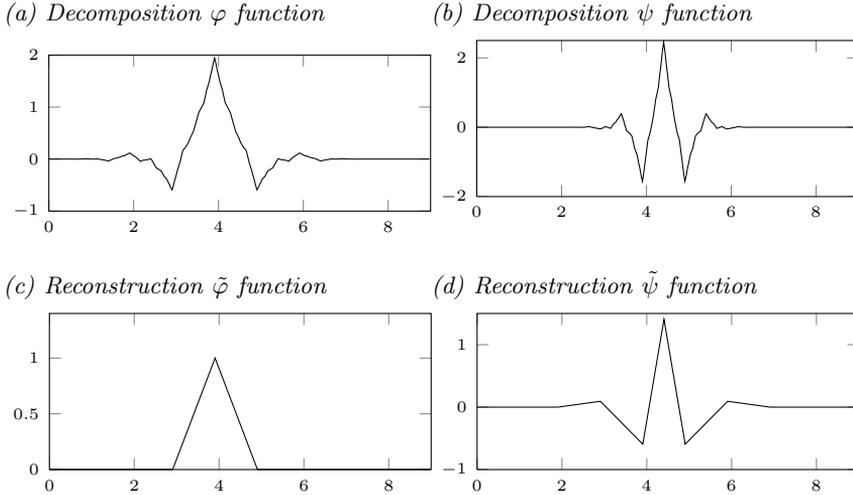


Figure 3.5: Biorthogonal wavelet analytical functions of 2.4 order. (a) - decomposition scaling function, (b) - decomposition wavelet function, (c) - reconstruction scaling function, (d) - reconstruction wavelet function

As stated in section 2.2 the core of WT is the basis function. In case of orthogonal analysis the single basis function determines the analysis vanishing moments, the wavelet support and the reconstruction smoothness. In all cases except the *Haar* wavelet the orthogonal basis functions have asymmetric support due to the phase shift in the frequency response. The latter means that the localization of wavelet coefficients in wavelength domain will be tampered with respect to HLIF small-scale features.

Considering this heuristic analysis, it is offered to use the biorthogonal wavelets for HLIF data analysis. In DWT the biorthogonal wavelets are created with the additional *symmetric* dual functions having a finite support (Cohen, 1992). In addition, the order of decomposition scaling φ and wavelet ψ and reconstruction $\tilde{\varphi}$ and $\tilde{\psi}$ pairs can be individually controlled. The support width is taken as $2N_r + 1$ and $2N_d + 1$ for reconstruction and decomposition wavelet orders respectively. Biorthogonal pair of wavelets and scaling functions used in previous section is provided in figure 3.5 in analytical form for CWT.

The wavelet functions in figure 3.5 are given for order of reconstruction filter lower than decomposition. However, as stated in section 2.2.3 the biorthogonal pairs can be used vice versa; the $\tilde{\varphi}$ and $\tilde{\psi}$ are used for decomposition and φ with ψ for reconstruction. This is required to preserve constraints on maximal smoothness on reconstruction and maximally possible compact support on decomposition. In literature this filters are commonly found as *reverse biorthogonal spline wavelets*.

The LWT interpolation naturally “creates” biorthogonal symmetric wavelets. In subsequent chapters the reverse biorthogonal filters lifting schemes of orders 1.3, 1.5, 2.4 and 2.6 are used. The coefficients of lifting wavelets are provided in Appendix examples and can be found in many research books.

3.2.2 Multiscale HLIF spectra comparison

In section 1.4.1 the quantitative discriminant measure for HLIF data is defined as the distance d between the mixture model and measurement/residual. In conventional approach the binary classifier can be constructed directly using the similarity measure by estimating the best threshold η_d .

It is proposed that the analysis of the HLIF structure – specifically, the search for differences – is not accurate if information from the complete data-space is used. As the structure of the HLIF spectrum is a set of small to medium scale local features with global baseline we want to compare spectra relying on detailed structure.

If \mathbf{z} and \mathbf{M} are observation and model (end-member) and $c_i^f = W_\psi f(a_i, b)$ denote CWT coefficients of function f then the scale distance measure can be defined as

$$sd(\mathbf{z}, \mathbf{M}) = \frac{1}{n} \sum_{i=1}^n d(c_i^z, c_i^M), \quad (3.1)$$

where d is distance measure from (1.2) and $i = (1, 2, \dots, n)$ are scale indexes. The scale distance

$$sD(\mathbf{z}, \mathbf{M}, a) = d_i(c_i^z, c_i^M), \quad (3.2)$$

as function of scale a provide detailed comparison of observation and model. In practice the full data-space distance can be weighted by the (3.1) producing the “feature weighted distance” FWD measure

$$FWD(\mathbf{z}, \mathbf{M}) = sd(\mathbf{z}, \mathbf{M}) * d(\mathbf{z}, \mathbf{M}). \quad (3.3)$$

The aim of equations (3.1) or (3.3) is to evaluate distance between model and observation on subset of scales, with stressed informational content. The selection of best comparison scales a is obvious. We select scales that maximally characterize spectrum, generally these are scales with cambers, curves, local extrema which was demonstrated for number of examples in 3.2.

In practice and numerical experiments the empirical selection of best scales for particular HLIF spectrum is straightforward and can be formulated with simple rules. The domain of function (3.2) can be divided into ranges according to scale characteristics of HLIF spectra. Thus, the details of spectra are localized on medium-scales (see figure 3.6b, range II). These are cambers which take $\frac{1}{16} \div \frac{1}{6}$ part of wavelength domain. This width value of cambers is determined by lower bound of RSS, which FWHM typically

equals $14nm$ or $\frac{1}{16}$ from full wavelength domain. The broad fluorescence baseline and system noise reside on large (ranges III and IV) and small (range I) scales respectively. These scales are preferably excluded from comparison.

It follows that best scales selection is the extraction of HLIF spectrum features. In case if spectrum is maximally flat, with no visible features, then the representative scales from range II are selected, considering the noise level. The representative scales means the scales with distance less than some small threshold value. The latter can be estimated by comparing the spectrum with its simulated noisy version.

3.2.3 Numerical example

In section 1.2.2 the number of land fluorescent objects was demonstrated. The most encountered HLIF measurement in land surveys is spectrum of green grass. The HLIF spectrum of grass with addition of Gaussian function with parameters $\sigma = 12, \mu = 485$ is shown in figure 3.6a. This can be spectrum of different object or combination of objects including the grass. It is obvious that spectra are similar and the calculated distance $d = 0.006$ is close to zero. The difference is embedded only in small scale artificial additive.

Figure 3.6b shows the function of distance from scale (3.2). The *rbio2.4* mother wavelet is used in this example. The distance maximum value 0.327 corresponds to scale 21. This shows that HLIF data significantly differs by the local features. The global baseline distance corresponding to scales in IV range equals to 0.002.

In following example, the power of multiscale HLIF spectra comparison is demonstrated on real-world land diagnostics. Unlike the vast spread in case of water oil pollution, the land oil pollution is concentrated in small areas. Sampling of land pollution area presents single entities of pollutant which are compactly located. This data was collected with airborne LIDAR system in the oil production region, where the oil has the ability to yield on surface in random places (the detailed description of application is given in 4.2.2). The single oil seep spots were randomly encountered along the flight and recorded in the observation database. The exact HLIF spectrum of oil produced in that region was measured along and used in analysis.

The conventional comparison using the distance (1.2) resulted in false positive erroneous answers like one shown in figure 3.7. Plotted finding in figure 3.7a is obviously the observed spectrum of oil, however the second finding in figure 3.7b is visually different. On the contrary the distances $d_a = 0.063$ and $d_b = 0.057$ display the opposite answer. The appearance of different HLIF spectra in figure 3.7b is the result of high dimensionality of HLIF data. Indeed, the discriminant measure in high dimensional HLIF

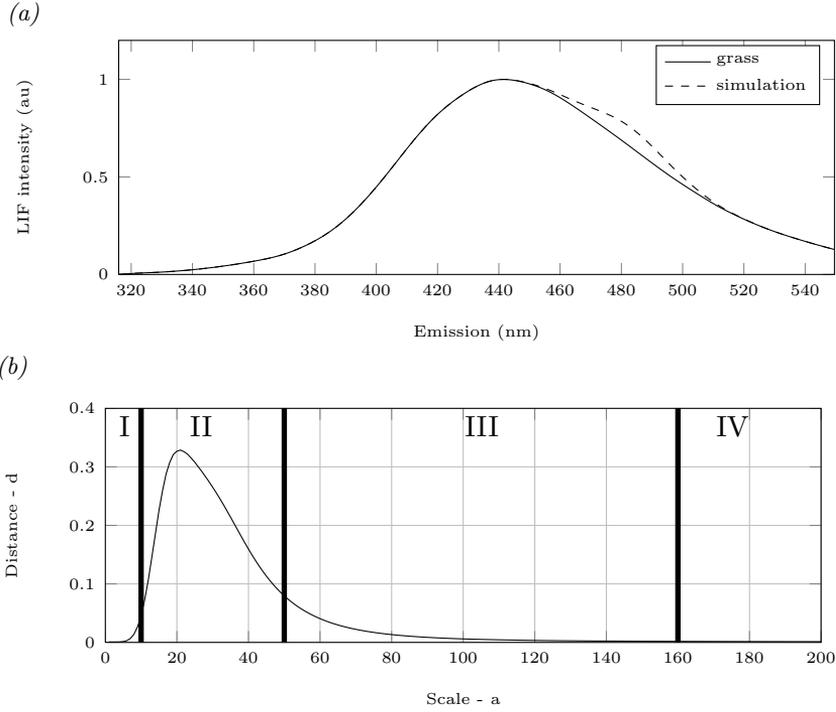


Figure 3.6: (a) - The distance-scale function plot of grass and simulated HLIF spectra. Spectra differ only locally from 460 to 500 nm range, $d = 0.006$. (b) - The plot of distance versus scale $sD(\mathbf{z}, \mathbf{M}, a)$. We extracted and compared the local feature of simulation spectrum with grass spectrum on scales in II range. The two orders increased distance $sd = 0.327, a = 21$ in that case shows that spectra are different.

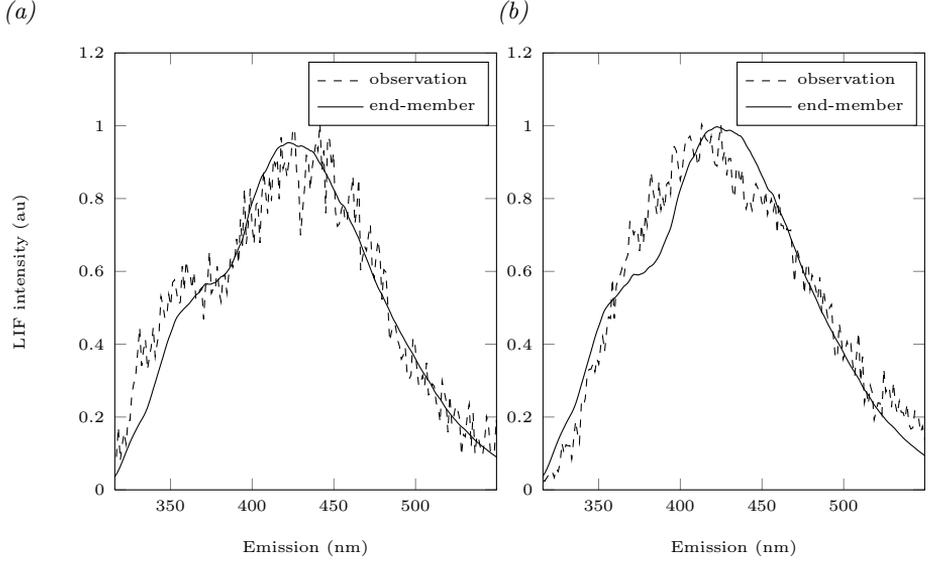


Figure 3.7: (a) - The plot of HLIF observation and end-member data for correct finding. The distance value $d_a = 0.063$. (b) - The plot of incorrect observation. The curve, visible in 350-400 nm range at end-member, is absent. However, the distance value is $d_b = 0.057$. The error is a result of high dimensionality of data, the distance in full data-space is not representative.

data-space lacks the selectivity and is not representative.

The scale structure of oil HLIF is shown on CWT spectrum in figure 3.8. This curve located between 350 and 400 nm is obviously the spectral feature which should uniquely identify the targeted oil product. The scale distance (equation (3.1)) taken on scale $a = 35$ result in $sd_a = 0.131$ and $sd_b = 5.112$ for first and second findings respectively.

The analysis of LIDAR data revealed a number of oil yields. Two values FWD and d are calculated to compare the results of FWD with conventional distance. The part of results is shown in figure 3.9, the X axis is sorted by increasing of d values and displays first most similar observations. Using the conventional distance the closest observations must point to the oil findings. The FWD results are plotted along the same sorted axis. The sample number 3 is the result demonstrated on 3.7b, which strictly lies outside of closest set for FWD. The final and correct estimation of oil locations are seven HLIF observations with indexes $l = \{1, 2, 4 - 8\}$. This FWD result matches with expert analysis.

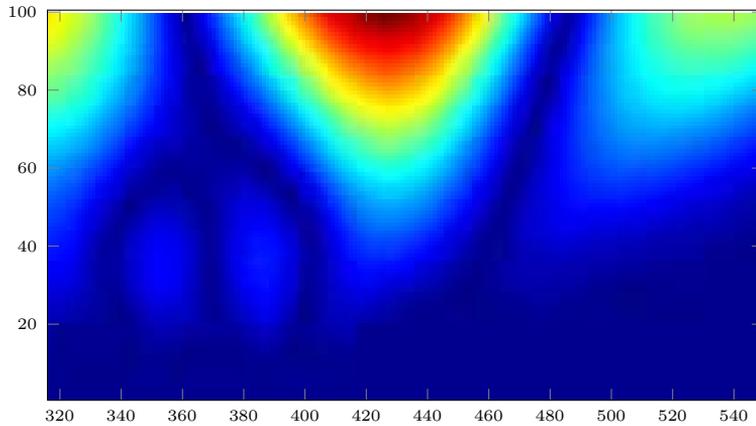


Figure 3.8: The CWT spectrum of oil product. The feature of this oil product, curve located at 350-400 nm range, is clearly seen. Location of maximum is on scale 35.

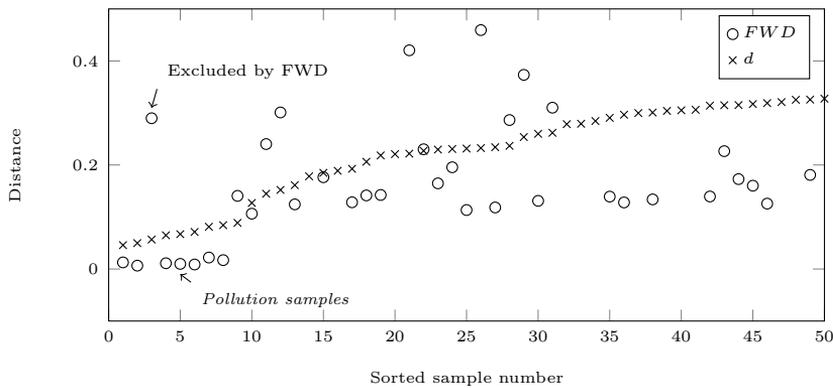


Figure 3.9: The LIDAR data analysis results in multiscale comparison. Samples are sorted according to d values in ascending order. First 10 observations which are most similar to the end-member spectrum are expected to be detected oil yield locations. The FWD analysis allocated first 8 samples excluding the number 3 (excluded sample spectrum is shown on 3.7b).

3.3 Wavelet based feature extraction

As demonstrated in previous section the CWT feature extraction serves as structural decomposition of HLIF spectrum. The redundant informational part of spectrum can be removed by exclusion of the irrelevant scales from distance calculation. The analysis is then adjusted to the particular spectral characteristics.

The relation of classification features to spectral properties, like local extreme points, is very important. For example in Polycyclic aromatic hydrocarbons (PAHs) the identical local fluorescence extrema areas in the same places yield the close related layout of poly-aromatic hydrocarbons (PAHs) and similar surrounding matrix for complexes of PAHs. In this way the structural analysis with CWT feature extraction is well explained in contrast to statistical approach, like PCA.

The CWT coefficients c_i^f calculations are computationally extensive and not suitable for real-time data analysis. The procedure can follow two classification steps. The first one uses the full data-space distance (1.2) with relaxed threshold for coarse identification, which result in many false-positive findings. The multiscale comparison is made as second step on selected dataset. Another option is to apply the feature extraction using DWT or lifting schemes, which is much faster and can be used for real-time data analysis.

3.3.1 HLIF feature extraction principle

The multiscale HLIF spectra comparison paradigm is that the local features found on some scales are binded with spectral properties of object. It is stated that HLIF data can be well represented by global and local structures in the basis of wavelets.

In multiscale comparison we have heuristically selected a number of scales with corresponding local structure information. This approach has several weaknesses. The HLIF signals cannot be reconstructed from CWT coefficients, this means that part of information is lost. Secondly, the dimensionality of coefficients on single scale is the same as in initial HLIF data-space. In addition we want to formulate an algorithm for automatic feature extraction without heuristic search for scales. As a result lets formulate the “ideal” feature extraction requirements as

- We want to have spectral features extraction algorithm without losing the full information about signal and its details.
- The signal representation in feature space must be sparse, meaning that only a small number of important features are non zero, i.e. the signal is compressed in feature space.

- The feature “importance” can be measured. For example it is denoted by absolute value of corresponding coefficient. Thus, the unimportant features can be removed automatically.

From the theory of *Orthogonal Matching Pursuit* the arbitrary signal y can be represented using the basis W by the minimization of the functional

$$\arg \min_{a \in \mathbb{R}^N} \frac{1}{2} \|y - Wa\|_2^2 + \lambda \|a\|_0, \quad (3.4)$$

where a are the unknown coefficients in basis W , $\|\cdot\|$ denotes the norm operator and $\lambda \in [0, \infty)$ is a utility variable. The first term in equation (3.4) is the “squared residual norm” or “sum of squared difference” function. This term is related to feasibility of the signal representation in basis W , i.e., the smaller the value the more information from signal is captured in a . The $\lambda \|a\|_0$ is the sparsity constraint, which is a penalization of the retained basis coefficients. It must be noted that we work with well determined linear system and W is orthogonal.

The DWT and lifting schemes introduced in sections 2.2.3 and 2.2.4 represent good orthogonal basis for HLIF type of signals. In contrast to CWT they tend to capture low-frequency sub-band information of non oscillating signals, which is known as *approximation*. The example of DWT and lifting expansion into basis W for Crude oil HLIF is shown in figure 3.10. As we deal with finite length signals it is preferred to use the lifting scheme, as it naturally removes the boundary condition problem present in DWT (compare figures 3.10b and 3.10c).

If consider the ground-truth spectrum from the spectral library the most signal “energy” is allocated between the significant local spectral properties (hills, bends, curves etc.) producing high expansion coefficients in wavelet domain. The remaining coefficients are negligibly small. Therefore, it is assumed that the first k largest signal expansion coefficients can be estimates of the best basis \hat{a} .

This estimate can be proven easily when W is orthogonal. Let $\lambda = 0$, then one can expand (3.4) as

$$\frac{1}{2} \|y - Wa\|_2^2 = \frac{1}{2} (y^T y - 2y^T Wa + W^T a^T Wa), \quad (3.5)$$

and by taking the derivative with respect to a

$$\frac{\partial}{\partial a} = -W^T (y - Wa) = -W^T y + a,$$

as $W^T W = I$ where I is identity matrix. By setting this to zero we find that minimum of (3.5) is at $\hat{a} = W^T y$.

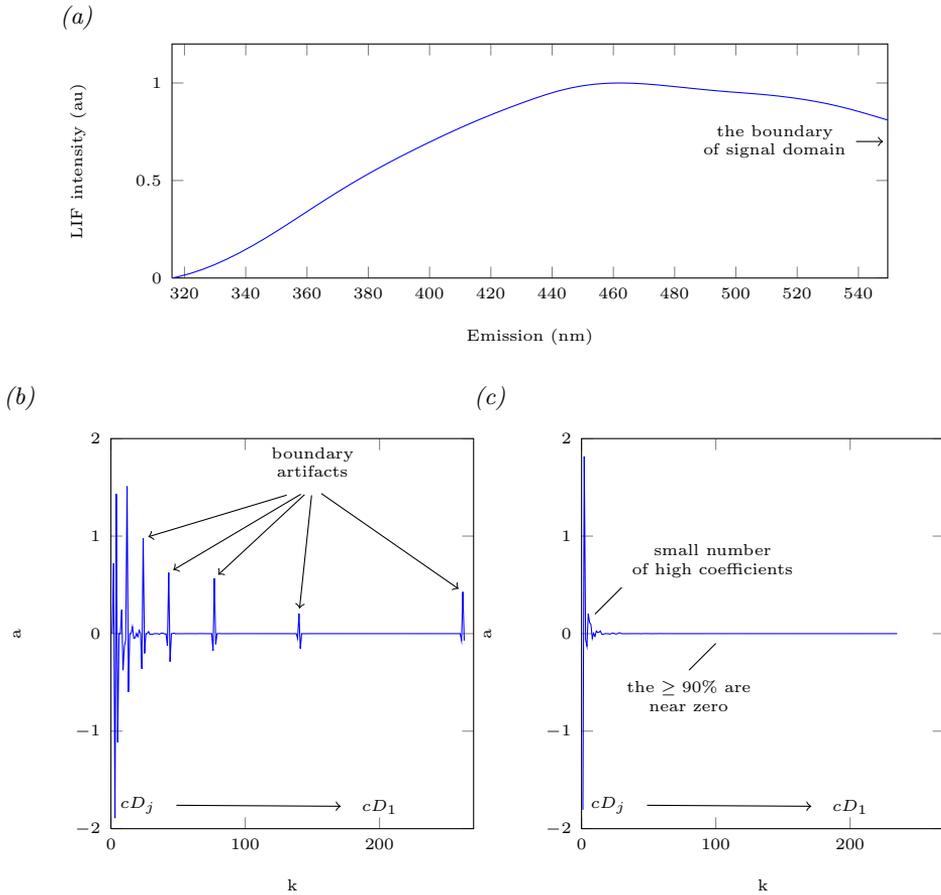


Figure 3.10: The comparison of Crude oil basis coefficients produced using the DWT and lifting schemes. (a) - The Crude oil HLIF spectrum as seen in figure 1.1a number 15. The spectrum support is beyond the detector wavelength range and the right boundary of spectrum is sufficiently high. (b) - The DWT expansion detail coefficients grouped as $\{cD_j, \dots, cD_1\}$. Due to the DWT filter support the boundary effect produced many high false coefficients on all decomposition levels. (c) - The lifting scheme coefficients. As lifting wavelet is constructed directly on signal, no boundary effect present and all coefficients have true value.

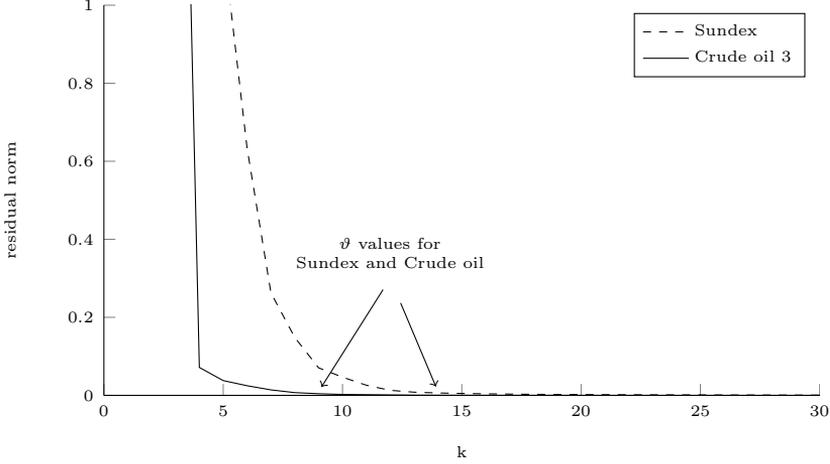


Figure 3.11: Plot of residual norm versus k for two oil HLIF spectra. The function decreases rapidly up to some value ϑ which is determined by structural complexity of spectrum.

Let k denote the number of non-zero entries in \hat{a} , then the $\|\hat{a}\|_0 = \#\{i : \hat{a}_i \neq 0\} = k$. This tells, that the number k of non-zero coefficients \hat{a} should be as small as possible. Thus, the best estimate of \hat{a} are the largest by absolute value expansion coefficients. Let Θ_k operator keeps largest k expansion coefficients of $|\hat{a}|$ then the minimum of (3.4) is found as

$$\min_k \frac{1}{2} \|(I - \Theta_k)\hat{a}\|_2^2 + \lambda k. \quad (3.6)$$

In words, at first step the complete WT representation of signal $\hat{a} = W^T y$ is created. The expansion coefficients \hat{a} sorted in descending order by absolute value, are sequentially zeroed by $(I - \Theta_k)$ operation.

Lets closer review the terms in (3.6). The first term represents the norm of retained expansion coefficients which will decrease with k (see example in figure 3.11). This value decreases fast as first highest coefficients, which contribute to the essential parts of signal, are removed from \hat{a} . Beginning with some $k > \vartheta$ the function decrease is barely noticeable, thus the retained coefficients represent insignificant signal part. The value of ϑ is determined by signal structural complexity. For example in figure 3.11 the Crude oil and Sundex can be reconstructed from 9 and 15 expansion coefficients respectively both producing equal 0.01 residual norm. It says that Sundex structure is more complex or *less compressible* than of the Crude oil.

The second term λk is the linearly increasing function with constant slope λ . The λ determines the penalty for retaining k coefficients and is related to the noise level. At this point it is assumed that $\lambda = 0$ (zero noise),

Id	Name	Sparsity, k	Compression $\frac{n}{k}$	$\#\{cA, cD_j, \dots, cD_1\}$
1	Diesel 1	17	92.83	4-3-4-4-3-0-0
2	Diesel 2	15	93.67	4-4-3-3-2-0-0
3	Diesel 3	17	92.83	4-3-3-5-3-0-0
4	Hydro oil	15	93.67	4-3-4-2-3-0-0
5	Motor oil	13	94.51	4-4-3-2-1-0-0
6	Heating oil 1	15	93.67	4-4-2-3-3-0-0
7	Heating oil 2	16	93.25	4-4-5-4-0-0-0
8	MDO 1	12	94.94	4-4-3-1-0-0-0
9	MDO 2	15	93.67	4-3-4-4-0-0-0
10	Gas oil	15	93.67	4-3-4-5-0-0-0
11	Calsol	15	93.67	4-4-4-4-0-0-0
12	Sundex	15	93.67	4-4-5-2-0-0-0
13	Crude oil 1	10	95.78	4-3-3-0-0-0-0
14	Crude oil 2	8	96.62	4-3-1-0-0-0-0
15	Crude oil 3	6	97.47	4-1-1-0-0-0-0
16	IFO-380	7	97.05	4-2-1-0-0-0-0
17	LCO	7	97.05	4-3-0-0-0-0-0
18	Marlin	7	97.05	4-3-0-0-0-0-0

Table 3.1: The results of feature extraction for reference library from figure 1.1a. The data-space is n -dimensional with $n = 237$. The sparsity column denotes the number of nonzero coefficients k . The values in compression ratio column are given as the percent of zero coefficients.

which is true for feature extraction from ground-truth end-members. For final solution the (3.6) is normalized as

$$\min_{0 < k < n/2} \frac{\|(I - \Theta_k)\hat{a}\|_2^2}{\|\hat{a}\|_2^2} \geq \tau, \quad (3.7)$$

where τ is small constant based on the maximum relative residual norm. Corresponding to previous work in (Sobolev and Babichenko, 2013a) the $\tau = 0.01$ which corresponds to relative 1% residual norm and good visual quality of reconstructed signal. The complete wavelet feature extraction algorithm implementation is given in Appendix.

3.3.2 Feature extraction examples

The feature extraction method is applied at reference spectra from figure 1.1a. The results are given in table 3.1 for parameters: $\tau = 0.01$, $n = 237$, $j = 6$ and $rbio1.5$.

The compression ratio given as percent of zero coefficients is greater than 90%. This confirms with assumptions that WT present good basis

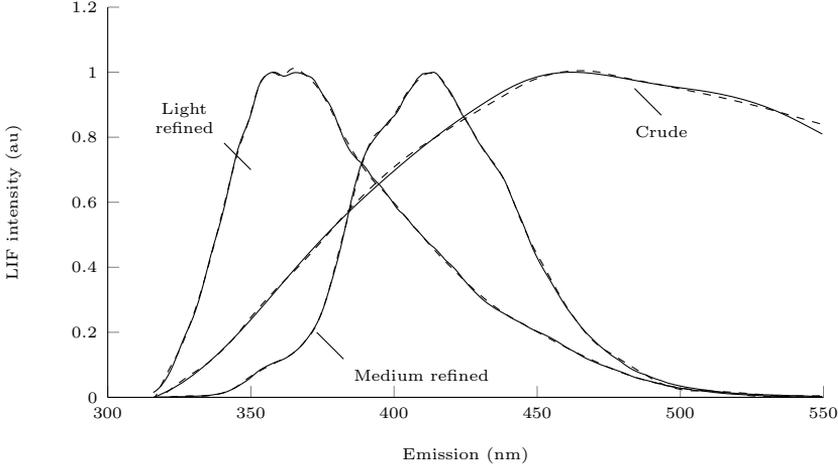


Figure 3.12: Light refined (Diesel 2), medium refined (Gas oil) and crude oil (Crude oil 3) products HLIF spectra and their reconstruction (dashed line) using the extracted features.

for HLIF type of signals. It is interesting that the sparsity of oil products HLIF correlates with type of oil product. The sparsity in table 3.1 ranges from 17 to 6 for light and crude oil products respectively. This is explained by the spectral shape complexity (see figure 3.12). The last column in table 3.1 shows the number of nonzero coefficients for each approximation and detail levels. The results for light and medium oil products show that HLIF data presents number of significant small scale features starting from scale cD_3 . The crude oil products, due to their smooth and broad shape, have large scale details on scales $cD_{5,6}$ only.

The signal information captured in small number of expansion coefficients is expected to be complete. The visual quality of reconstruction is shown in figure 3.12 for three oil products: Diesel 2, MDO 1 and Crude oil 3.

The example of Gobles oil fields in figure 3.7 is continued here using the proposed wavelet feature extraction. Previously in section 3.2.2 the CWT and heuristic analysis were applied to extract information about specific HLIF spectrum curve visible in 350 – 400 nm range. The feature extraction results with $\tau = 0.01$, $n = 237$, $j = 6$ and $rbio1.5$ for the same HLIF spectrum are

$$\begin{array}{ccc} \text{Sparsity, } k & \text{Compression } \frac{n}{k} & cA, cD_j, \dots, cD_1 \\ 14 & 94.33 & 4-3-4-3-0-0-0 \end{array} .$$

The feature curve located on scale number 35 in CWT spectrum is now encoded in nonzero detail coefficients on scales cD_{4-6} . The reconstruction of curve from selected coefficients is shown in figure 3.13.

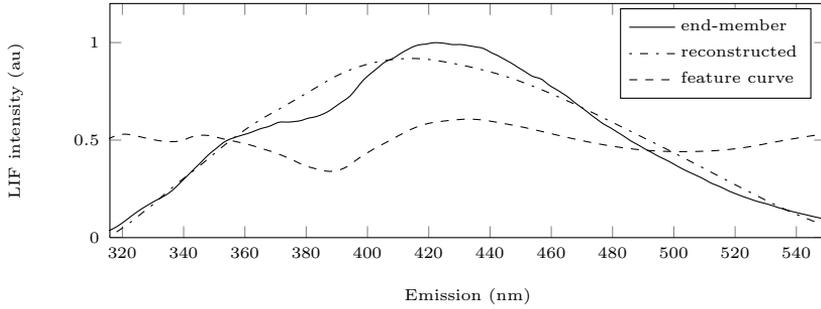


Figure 3.13: The reconstructed HLIF spectrum with removed feature curve. The curve is encoded by nonzero expansion coefficients on scales cD_{4-6} .

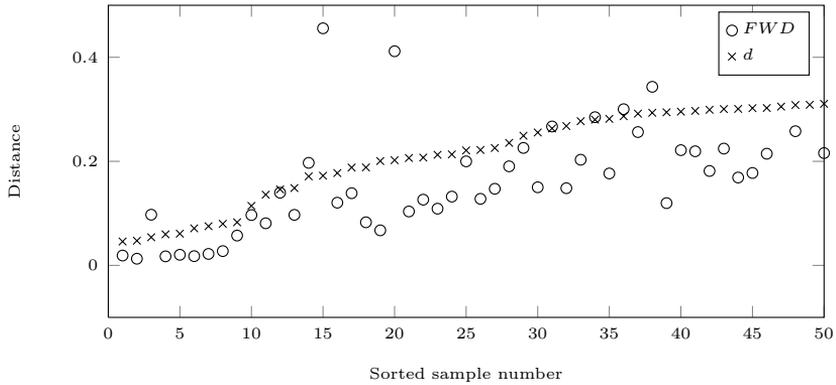


Figure 3.14: The analysis results on terrain data in feature comparison. Samples are sorted according to d values in ascending order. The detail analysis used only seven expansion coefficients for feature comparison.

The FWD is calculated using the same equation (3.3). The difference is in scaled distance computation where, instead of CWT coefficients c_i^z, c_i^M , the extracted feature coefficients for signal and model are used. New results are plotted in figure 3.14. The results show equal answers for both multiscale comparison and feature extraction methods.

3.3.3 HLIF feature denoising

The conventional approaches to adaptive denoising as mentioned in section 2.3.1 require prior estimation of noise model and parameters, like noise standard deviation. Generally the denoising methods are limited to “white” or “colored” normal noise models. Alternately, some signal processing methods, like Variance Stabilization Transform, can extend conventional algorithms to signal-dependent noise sources (Makitalo and Foi, 2013). An alternative and efficient denoising method can be applied using the sparsity-norm optimization (3.6) with noise adaptive λ . This method refers to original *adaptive Minimum Description Length* (AMDL) approach from Saito (1994) with significant improvement.

In previous section the $\lambda = 0$ as the end-member HLIF spectra represent ideal noiseless data. If y is real observation then the λ should be estimated. One constant estimator is provided in Saito (1994). The author derived the (3.6) using the informational transmission setting or MDL principle. The noise component assumed additive WGN and objective function is derived from ML estimator of normal p.d.f. with respect to σ^2 as

$$AMDL(k) = \min_{0 < k < n/2} \frac{n}{2} \log \|(I - \Theta_k)\hat{a}\|_2^2 + \frac{3}{2}k \log n. \quad (3.8)$$

The k non-zero expansion coefficients transmission penalty is denoted constant as $\frac{3}{2}k \log n$.

Lets analyze the solution of (3.8). The first term is residual norm in logarithm scale with $\frac{n}{2}$ normalization coefficient, derived from ML estimator of normal p.d.f. As both functions (logarithm and $\frac{n}{2}$) are linear they do not affect the objective final solution. The log function linearize and compact the dynamic range of objective function thus it is easier to analyze its behavior.

The second term sets $\lambda = \frac{3}{2} \log n$ which is taken as the length of information needed to transmit the k non-zero elements. It is set constant, however in hypothesis the λ must be related to signal noise, and specific value must be adapted according to noise distribution.

Let the $\varepsilon = \mathbf{a}\mathbf{y} + b$ from (2.24) is the total noise variance of measured signal \mathbf{y} . In figure 3.15a the log residual norms versus k for three variations of ε are plotted. As we move with k the residual norms starting from some $k > \vartheta$ decrease with small constant slopes β . For the signal-dependent

noise the β also depends on signal itself and is greater than for WGN. This is expected, as at each k we remove the highest expansion coefficient and for Poisson noise the distribution of magnitudes is not normal.

Following from plot 3.15a the slope β and penalty λ are related. If we assume that the breakpoint between the signal and noise components is known and β is the noise slope then the estimator

$$\hat{\lambda} = 2|\beta|, \quad (3.9)$$

is compensation for noise dependent slope in log residual norm functional. The numerical solution to (3.9) can be found using the segmented linear regression model

$$RM = \frac{n}{2} \log \|(I - \Theta_k)\hat{a}\|_2^2 = \begin{cases} \beta_1 k + o_1 & \text{for } k < B \\ \beta_2 k + o_2 & \text{for } k \geq B \end{cases}$$

where B indicates the breakpoint between two segments. The breakpoint solution is found as minimax

$$\arg \min_B \|RM - \hat{RM}\|_2^2$$

using the fast golden section search on bounded interval $\{0; n\}$. The estimated fit examples are given in figure 3.15b. The first segments correspond to the expansion coefficients of significant parts of the signal. The secondary linear segments follow the noise coefficients. The penalty for retaining the coefficients λ can be estimated from slope of $\beta_2 k + o_2$ as $\hat{\lambda} = 2|\beta_2|$. This procedure is named ‘‘Adaptive Slope Compensation’’ (ASC) denoising.

In real HLIF measurements the Poisson-Gaussian noise model cannot describe every noise component. In rare cases due to the high gain the signal is subject to events, like abrupt spikes, which produce strong artifacts. The proposed HLIF feature denoising with adaptive slope can be used to efficiently suppress such noise using simple signal-noise model assumption. The HLIF signals due to the physics are sufficiently smooth functions (have no ‘‘bumps’’, ‘‘blocks’’ or edges). Let j is the WT details index and $\hat{a} = \{cA, cD_j, \dots, cD_1\}$ then

$$\hat{a}^* = \{cA, cD_j^*, \dots, cD_1^*\} \quad \text{and} \quad cD_j^* := \frac{d}{\#\{cD_j\}^p} \quad (3.10)$$

where $d \in cD_j$, $\#\{cD_j\}$ is number of elements in cD_j and p is penalization constant. Equation (3.10) creates the penalized coefficients \hat{a}^* for operator Θ_k , which are used to sort the \hat{a} in penalized descending order. This approach applies penalty for high frequency coefficients and thus more efficiently removes ‘‘abnormal’’ high frequency noise. In experiments with

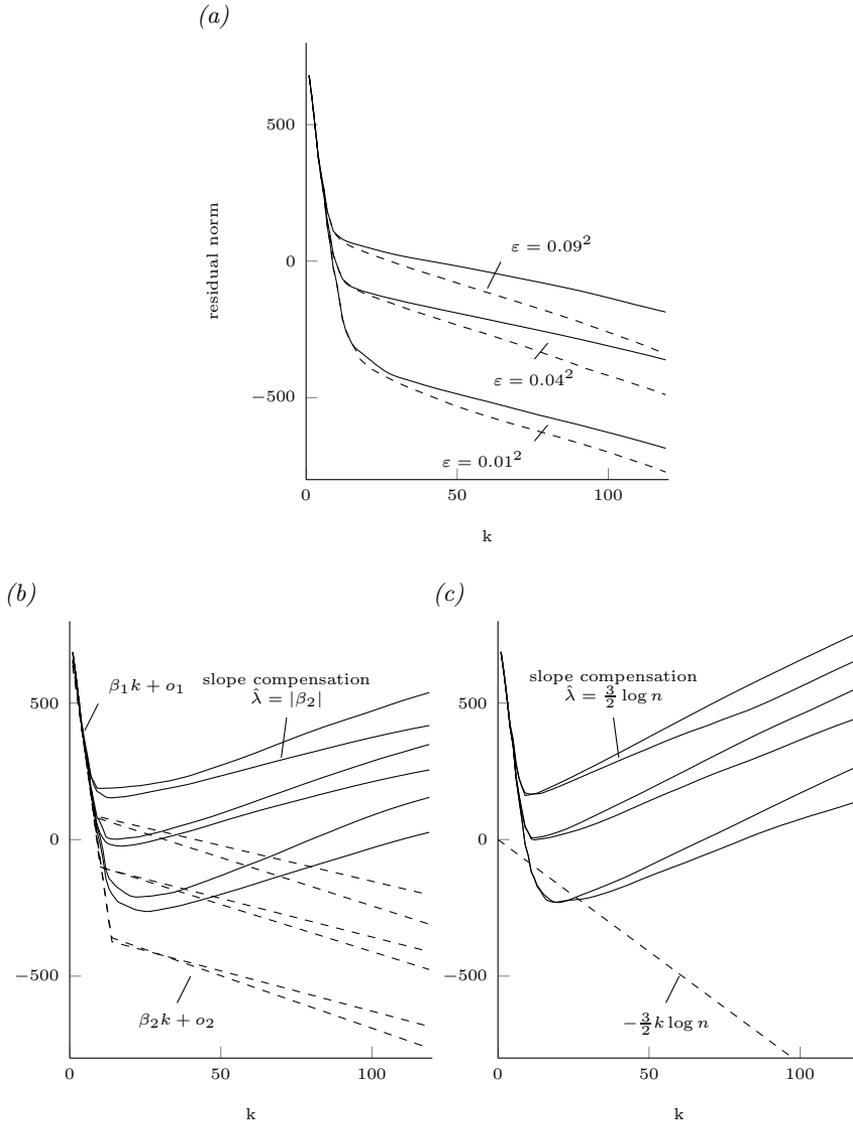


Figure 3.15: (a) - An example of residual log-norm plots for variable ϵ . The solid lines represent zero Poisson $a = 0$ noise. The dashed lines correspond to the zero Gaussian noise $b = 0$. The slope β of error norm is dependent on the type of noise distribution. (b) - Dashed lines plot the segmented linear regression model fits. The λ 's are estimated from the slopes of second linear segments β_2 . The solid lines are penalized residual norm with estimated λk function. (b) - The example of AMDL using $\frac{3}{2} k \log n$ as slope compensation. Comparing with (b), the AMDL slope value is overestimated for both Gaussian and Poisson noise.

HLIF data it was found that constant p value is ≥ 5 . This can be transformed to new operator Θ_k^j which keeps largest k expansion coefficients from specific scale j . In words, for (3.6) solution the expansion coefficients are sorted by absolute value at each scale j independently in descending order.

3.3.4 Feature denoising examples

Several application examples show the performance of feature denoising in Chapter 4. Mainly the “Feature denoising in water signal spectral deformation detection” is based on ASC denoising (see Section 4.3).

Here the simulation with signal-dependent noise model is used to test the ASC denoising performance. The ASC performance is demonstrated for various spectral responses: natural water with high and low DOM concentrations (varies the proportion of RSS and DOM), light refined, medium refined and crude oils (selected from figure 3.12). The selected spectra have various structure and represent the feasible set of spectral responses for most HLIF OPD applications. Denoising is tested for Gaussian and Poisson noise sources with $\varepsilon = 0.04^2$ and penalization of high frequency scales Θ_k^j . The summary is presented in table 3.2 for ASC and AMDL performance comparison using the *Peak Signal to Noise Ratio* (PSNR) measure.

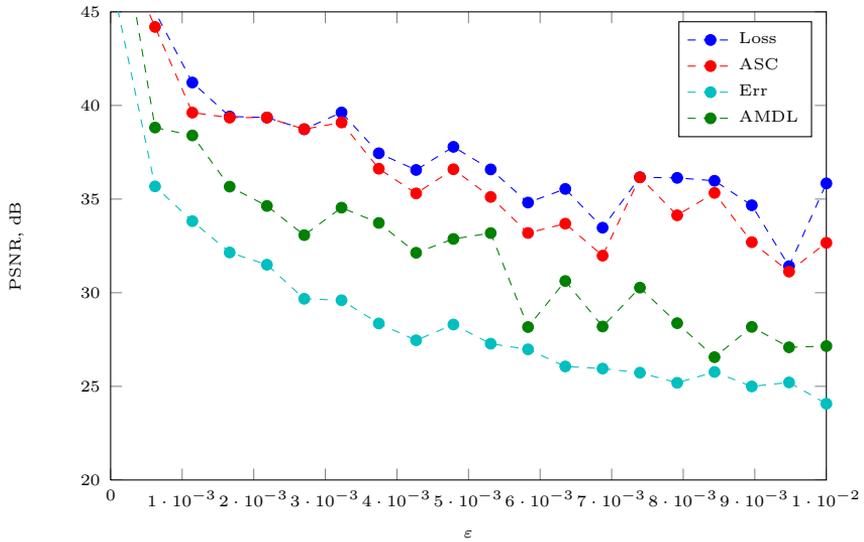
The results indicate that ASC and AMDL methods behave similar for Poisson noise denoising ($a = 0.04^2, b = 0$). This is due to the increased slope for Poisson noise residual norm, for which AMDL constant estimator fits well enough (see figure 3.15c). The Gaussian noise slope is more distinct from AMDL value, which is correctly estimated by ASC, which is seen as increasing denoising performance in table 3.2 ($a = 0, b = 0.04^2$). The penalization of high scales with operator Θ_k^j maximally improves the denoising quality for $4.79dB$ using the ASC method.

An additional performance test is created for variable ε and results are plotted in figure 3.16. The ASC method has the denoising performance close to the ideal. The ideal “Loss” minimum is calculated using the wavelet coefficients from the prior “known” HLIF signal. The AMDL method shows good performance, however the constant λ is overestimated, thus the number of retained coefficients is lower than with the correct estimate.

	high dom	low dom	light	medium	crude
$a = 0.04^2, b = 0$					
AMDL	35.87	38.84	34.48	35.9	36.26
ASC	36.6	39.14	35.16	36.25	36.63
delta	0.73	0.3	0.68	0.35	0.37
$a = 0.04^2, b = 0, \text{ high penalized}$					
AMDL	35.5	36.3	34.62	36.06	36.03
ASC	36.12	41.09	39.27	39.88	39.18
delta	0.62	4.79	4.65	3.82	3.15
$a = 0, b = 0.04^2$					
AMDL	33.65	33.81	32.53	32.07	35.82
ASC	34.57	34.44	33.53	34.02	36.25
delta	0.92	0.63	1	1.95	0.43
$a = 0, b = 0.04^2, \text{ high penalized}$					
AMDL	34.49	33.91	31.98	32.34	35.64
ASC	34.8	34.18	35.12	36.01	38.48
delta	0.31	0.27	3.14	3.67	2.84

Table 3.2: The HLIF feature denoising results for Gaussian and Poisson noise sources. The values are given as PSRN in dB. The delta corresponds to the $ASC - AMDL$. The performance improvement is affected by type of HLIF data. The most high denoising improvement is for Poisson noise HLIF spectra with high penalization operator Θ_k^j .

(a) $a = (10^{-3}, 10^{-2}), b = 0$, high penalized



(b) $a = 0, b = (10^{-3}, 10^{-2})$, high penalized

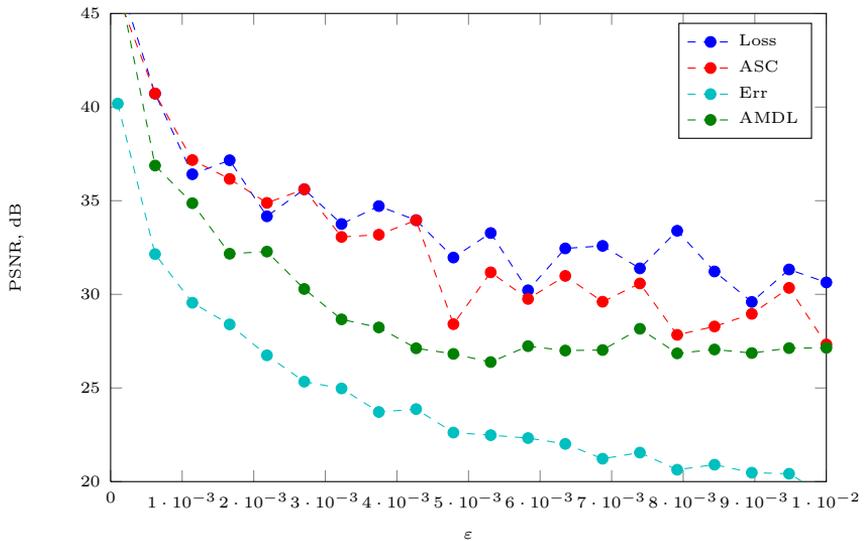


Figure 3.16: The example of denoising performance with variable noise power. Loss has the maximum performance, delivering more than 10dB PSNR increase, which is the ideal denoising if the original data is known. Performance of ASC method is close to the Loss. The AMDL method was used without high penalization, as it showed detrimental effect on method. Err shows the initial PSNR of noisy data.

3.4 Conclusions

In this chapter the structure of HLIF spectrum has been analyzed using the CWT. As demonstrated on a number of examples all spectral details have been preserved and isolated in continuous wavelet basis. This method extends the informational component in HLIF data, thus the discriminant measure can be specifically adjusted. The latter has been demonstrated on LIDAR application of land oil diagnostics. Seven correct oil findings were successfully identified, while the false positive but similar natural fluorescence samples have been excluded using the conventional discriminant measure weighted with extracted small-scale spectral information from the oil HLIF end-member. The results heavily correspond to analyst view.

The wavelet properties like compact support, regularity and vanishing moments provide very efficient basis to represent relevant information of HLIF signals. Important properties for feature extraction have been estimated as maximal smoothness of reconstruction and maximal compact support of decomposition. Based on this information the wavelets from reverse biorthogonal family were selected for subsequent application. An example of wavelet filter coefficients in lifting schemes are provided in Appendix.

The automated feature extraction based on the *Matching Pursuit* has been proposed for HLIF data. The results of the oil products end-member HLIF spectra analysis using proposed feature extraction demonstrate the dependency of the number of significant features on the oil product type. The number of significant features ranged from 17 to 7 for light refined oil and crude oil products respectively. This information corresponds to the structural complexity of fluorescence emission spectra related to those types. The numerical example of land oil diagnostics has been repeated with feature extraction providing the same results as in case of the continuous wavelet basis.

In some cases the discriminant measure can be evaluated directly in object space. This includes the detection of natural water deformation method showed in section 4.3, where the observation presents mixture of various spectral objects like RSS, DOM and oil pollutant. In such case, the proposed ASC denoising serves as improvement of discrimination quality. The numerical examples results show that denoising quality of ASC is as high as ideal loss denoising for Poisson, Gaussian and mixed noise distributions. The specific “high penalization” strategy used with ASC method improves the denoising more than $4dB$ in comparison with similar AMDL denoising method.

The feature extraction and ASC denoising methods implementations are provided in Appendix section.

Chapter 4

HLIF diagnostics platform and OPD applications

The OPD design requirements include the low false alarm rate and automatic decision-making as must meet criteria. The methods of signal processing and feature extraction, discussed in previous chapters, are designated to improve the HLIF pattern recognition system capabilities in order to fulfill the requirements of modern in-situ remote OPD designs. Application examples of HLIF structural analysis techniques for diagnostics of water and land concerning the conceptual differences are given in this chapter on the basis of real-world HLIF LIDAR missions.

The applications requirements multiplied by research and development experience in the field of HLIF OPD resulted in creation of state-of-the-art HLIF diagnostics platform. The automated diagnostics platform include the detection and identification of oil pollution in water/land mixed environments with additional quantification capabilities. Its design is optimized for combination of real-time and post-processing operation. The ultimate output of system is the georeferenced OPD map with estimated pollution and environmental indicators. By applying the proper calibration the output can be extended to qualitative estimation. The system will be integrated into commercial HLIF LIDAR platforms for OPD in marine applications (supported by EAS grant EU46064 and Barents 2020 project).

4.1 Author's contribution

Creation of base conceptual scheme of HLIF OPD platform. The HLIF OPD capabilities have been tested in a variety of real-world, laboratory and simulation experiments.

The following capabilities of the oil spill monitoring in land/water mixed environment concerned and have been improved by reprocessing of early

collected data using the developed algorithms by the author:

- The unknown pollutant detection in the presence of strong known fluorescence background by removal of acquisition noise (example “Feature denoising in water signal spectral deformation detection” - section 4.3);
- The identification of pollutant in difficult environmental conditions (example “Underwater oil pollution detection and identification” - section 4.4);
- The recognition of pollutant in the mixture with unknown fluorescence background (example “Land surface oil pollution identification” - section 4.5);
- The HLIF OPD of residual pollution in land/water mixed targets (example “Land/water mixed diagnostics” - section 4.6).

The author developed and implemented the pattern recognition system for HLIF LIDAR data processing which has been evaluated on a number of HLIF LIDAR applications. In the framework of HLIF OPD the author contributed to airborne and shipborne LIDAR experiments in LIDAR operation, data acquisition, extraction, processing and algorithms implementation. The author operated with HLIF Lidars in numerous studies in Estonia, Latvia, USA, Canada, Portugal, Spain and Norway including operational missions and R&D European projects DEOSOM (AMPERA) and Baltic Way (BONUS).

4.2 Experimental datasets

All examples are based on experimental LIDAR data obtained during the recent and past years. The major data was obtained in field works in capability demonstration projects and surveys. The examples also include the controlled laboratory research experiments and simulations with estimated instrument parameters. The field works progress was reported in several internal reports and public access articles in international remote sensing journals (Babichenko et al., 2000a; Babichenko, 2008; Babichenko et al., 2004, 2006, 2002, 1995, 2000b). The summary of the objectives and operation of the following field experiment and services is given in chronological order.

4.2.1 LIDAR Underwater Diagnostics in OHMSETT Facility

The earliest of demonstration projects in 2005 was taken to share the expertise in remote oil detection with LIDAR. In cooperation with US Coast

Guards (USCG) the airborne LIDAR analytical platform demonstrated key physical principles on which this technology is based.

The system was installed on a bridge with a near-horizontal beam configuration, with a mirror deflecting the beam down onto the target area providing approximately 15 meters of light path distance. The experiment included 29 tests of automated oil detection, characterization and submer-sion depth ranging of oil targets made with different types of light and refined oils and demonstration of system capability in presence of surface waves in various configurations, as well as during a light precipitation event. The experiment also included the heavy crude oil. Crude oil distinctive feature is that they naturally go below the water surface in hours from pollution event.

The experiments demonstrated that LIDAR system is capable to detect and identify oil beneath the surface of the water down to about 1.8 meters from a laser height of about 15 meters above the water. Very basic analytical platform was available on the time of experiment, thus the most results were developed using manual data processing. The data present an interest to evaluate and compare the quality of oil pollution detection and identification using the proposed structural analysis of HLIF data.

4.2.2 Gobles Fields LIDAR Operational Capabilities Vali-dation

The operational tests in this experiment have been conducted to demon-strate capabilities of real-time detection and automatic determination of oil type on land targets surveyed by airborne LIDAR remote sensing tech-nology. The technology was demonstrated in prospecting and oil pipeline integrity check applications.

Survey included the airborne LIDAR flights in the oil production regions of South Ontario. The database included half million of georeferenced HLIF spectral data. For prospecting analysis the spectra were filtered according to geographic location of known oil ingress spots. A oil from local oil well was measured and added into end-member library.

The experiment data is interesting for its unusual oil HLIF spectrum. The crude oil HLIF spectrum structure is highlighted by specific and strong feature curve in UV region, which is more common for refined and light oil products. The analysis results have been demonstrated and compared in previous sections 3.2.3 and 3.3.2 for “multiscale comparison” and FWD respectively.

4.2.3 Vegetation oil pollution laboratory research

The field works in land surveys required the discrimination of the oil product on the background of highly variable natural fluorescence objects. The comprehensive library of land natural fluorescent object and surfaces were obtained in the laboratory experiments in year 2010.

The dataset included various vegetation and land surfaces samples. The controlled level of pollution was added to each sample and polluted data was measured using LIDAR instrument. The distinct feature of experiments is in oil product HLIF spectrum. The light Crude oil HLIF spectrum selected for experiments represents a smooth broad structure which is common for most fluorophores. The absence of strong features makes the analysis more complicated.

4.2.4 Oil Spill Kalamazoo River Area LIDAR Survey

An airborne remote sensing survey of the heavy Crude oil (Diluted bitumen) spill along the Kalamazoo river system was conducted in 2011. The survey involved 11 hours of survey time with multiple passes over the area.

Airborne LIDAR survey was taken a year after the accident and cleaning operation. The survey task included the detection of residual oil presence on the river surface and below. At the same time the land analysis included the river bed and inundation areas.

The oil product fluorescence response was significantly influenced due to the long term hydrophysical factors. Most of heavy bitumen sank creating submerged conglomerates of oil-contaminated sediment. The organics concentration of river was very high (more than $20 \frac{mg}{l}$), it was an indication of long term oil pollution.

With respect to subsurface presence of oil, there were high levels of DOM in the water detected in the survey and verified after in the subsequent water samples. This restricted the depth to which the laser pulses could penetrate the water. Based on the level of DOM, the LIDAR was able to penetrate below the surface of the water to a depth of 20 inches. As such, it was not possible to directly survey the river beds.

The dataset present an example of mixed land/water observations, with high DOM concentrations and very variable land fluorescence backgrounds. The total number of measurements is more that 3 million, from which 200 thousands are the water targets. The final analysis showed no locations of the Crude oil film on the surface of the water along the Talmadge/Kalamazoo river system. Most of findings were located in land areas in the proximity of the river banks and in the inundation area.

4.3 Feature denoising in water signal spectral deformation detection - simulation experiment

The task of unknown pollution detection introduced in section 1.5.2 requires the analysis of spectral shape deformation due to the presence of additive component in water matrix. The definite spectral response of natural water matrix can be measured prior to operation or even updated on-line. This approach can be efficiently applied without prior knowledge of pollutant end-member in applications with high DOM content.

As mentioned in 1.5.2 the selection of deformation threshold η_d depends on the HLIF signal quality. We can separate two spectral deformations, namely the noise and presence of additional component. Figure 4.1 shows the spectra of water samples with variable signal power. In real measurements the signal power varies due to the variation of distance to target, fluctuation of laser pulse power between successive shots or changes of water medium light propagation property due to the waving.

The signals in figure 4.1 were simulated as taken from natural water with constant detector gain. The noise model is parametrized by estimated values of Poisson $\hat{a} = 0.043^2$ and Gaussian $\hat{b} = 0.003^2$ noise components (see section 2.3.3). On figure 4.1 the oil pollution is visible as increase of fluorescence response tail on the right side of spectrum. Here the pollution is given as “Crude oil 3” film from figure 1.1a. It was noticed that the fluorescence intensity of crude oils and DOM are at the same order or lower. Here the oil film thickness is at order of $1\mu m$ which corresponds to the 20% of overall signal intensity.

The deformation of water spectrum in figure 4.1 is barely noticeable for weak signal (25% signal power). This creates challenge to measure the deformation level and choosing the η_d disregarding the signal power. Figure 4.2a displays the d_{water} dependence on the signal power for simulated cases ranging from 25% to 100%. The d_{water} for polluted spectra is displayed on the same plot. It is shown that the noise disrupts the distance value at the same or greater level as deformation from pollution additive. The true value of deformation for “Crude oil 3” is plotted as dashed line (also see the corresponding figure 1.7).

The ACS feature denoising (from section 3.3.3) output is displayed in figure 4.2b. The deformation noise dependency is significantly reduced. The best threshold η_d can be estimated from plot. The denoised distance value for crude oil is close to the true one plotted as dashed line.

The minimum detectable deformation vs. pollution intensity in the presence of signal noise is shown in figure 4.3. The case of crude oil pollution is considered. Relative intensity of pollution is bounded with film thickness and by some simplification can be assumed as linearly increasing

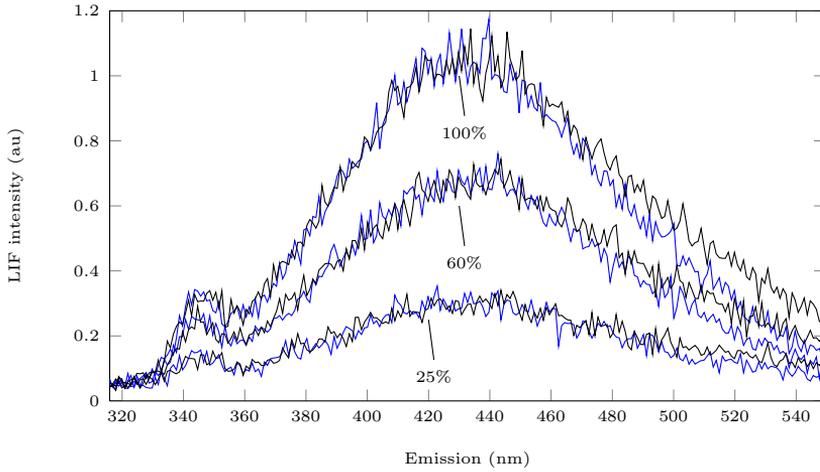


Figure 4.1: Example of natural water pollution, the simulation. The spectra show natural water matrix (blue) with high DOM content. Crude oil film pollution is black. Three levels of signal intensity ranging from 25% to full scale of detector dynamic range are shown. The level of oil pollution is 20% of overall signal intensity.

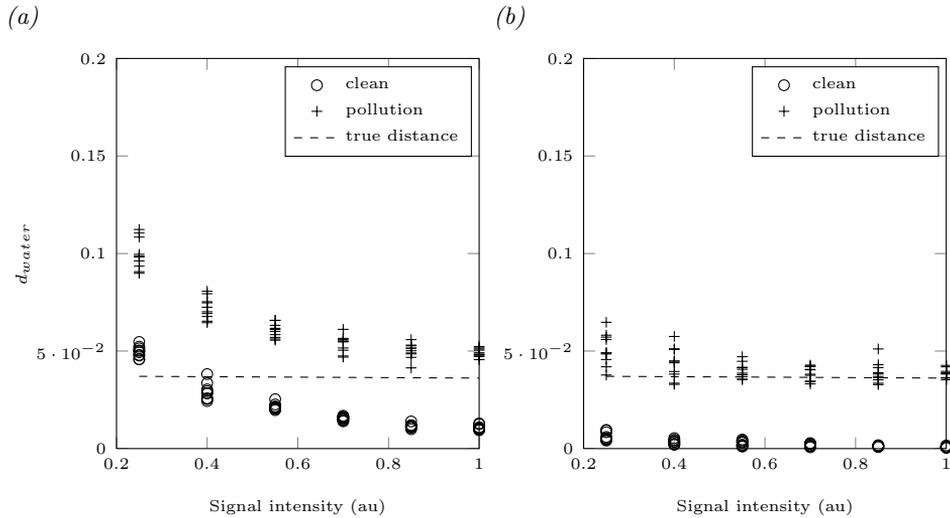


Figure 4.2: Plot of d variation with signal intensity. Ten spectra were simulated for each intensity level. The dashed horizontal line shows the distance value in case of pollution with zero noise. The pollution level correspond to $1\mu\text{m}$ oil crude film. (a) - The value of HLIF discriminant measure d is dependent on noise. (b) - Denoised signal present almost no dependency for d .

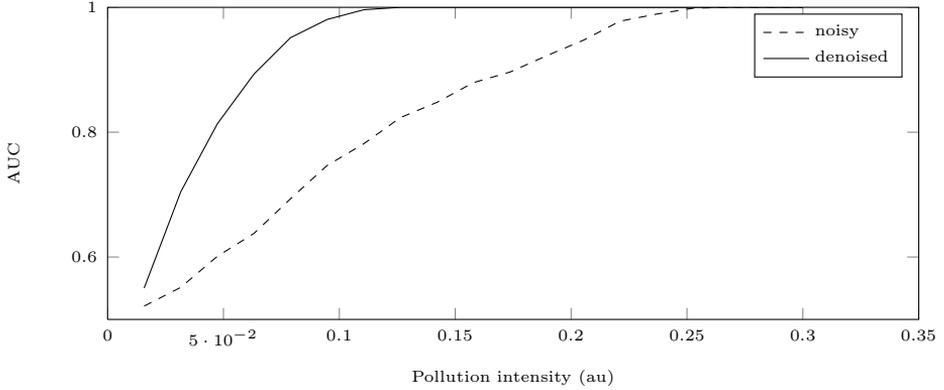


Figure 4.3: The AUC values as summary statistics for discriminating the pollution from the noise. High AUC means better discrimination (1 gives 100% discrimination). The noisy statistics are poor due to the distance measure vs. noise dependence (dashed line). The minimum value of 10% pollution from signal level can be detected with 100% confidence disregarding the total signal intensity in denoised signals (solid line).

with increasing thickness.

The figure 4.3 displays the summary statistics of true-positive versus false-positive error rates, namely Area Under Curve (AUC). The error rates were found as a function of a threshold parameter η_d . This function is known as Receiver Operating Characteristic (ROC), which illustrates the performance of binary classifier with varying threshold (Bradley, 1997). The error rate at each pollution intensity value was calculated for small, medium and high signal power as shown in figure 4.2.

According to results the detection is possible starting from relative 25% pollution intensity. Such level of deformation required to overweight the distance noise dependence, especially for low power signals (see figure 4.2a). The minimum pollutant intensity value which is 100% detectable in denoised signal equals 10% (roughly corresponds to 0.5mg/l). Comparing with the noisy results it can be stated that denoising procedure in two times increases the accuracy of crude oil pollution detection.

4.4 Underwater oil pollution detection and identification

These results are an example of the detection and identification of an oil product in difficult environmental conditions (i.e. in an environment with a high DOM content). The data were collected with a LIDAR instrument during test trials with controlled oil targets in OHMSETT experiment

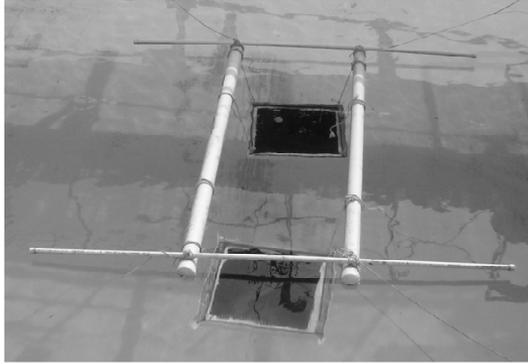


Figure 4.4: Two submerged oil targets at a depth of 1 m.

described in section 4.2.1.

Three oil products, Sundex, Calsol and Crude oil “Irene” were attached to pads and were submerged in a pool at a depth of 1 m. The targets were installed up in a line, as shown on Figure 4.4, and were anchored to the test tank. As the LIDAR instrument moved with the platform, the laser beam sequentially interacted with all the presented targets. The sequence of the targets was Sundex, Calsol, Crude oil.

The first task for the analysis system was to detect the presence of a contaminant in the water. This step was described in details in section 4.3. The spectral shape of clean water from the pool was recorded and used as a reference model for comparison with the measured spectra. As shown in figure 4.5 the denoising of HLIF data significantly increases the selectivity of spectrum deformation detection. The deformation created by the Sundex spectrum (medium refined oil product), due to its strong overlapping with DOM, is hard to detect. After the ACS denoising the analysis sensitivity is increased as seen in figure 4.5b. Other targets Calsol and Crude oils are detectable without denoising also.

The following step includes the measurement of the spectral shapes of the pollutants. The end-member HLIF spectra and their corresponding emission-scale CWT spectra are shown in figure 4.6. Both medium refined oils have strong, distinct small scale features, visible in figures 4.6b and 4.6c on scales 15 to 20. The location of spectral maximum varies for 20 nm. The small scale features are not visible on the broad spectra of Irene in figure 4.6d, some structure only arises from scale 30.

The visible details in figure 4.6 can be selected for comparison and identification of pollutant in experimental dataset. However, it is better to apply feature extraction algorithm using (3.7) and select the small number of detail coefficients which will also correspond to significant features. The algorithm parameters are $\tau = 0.01$, $n = 237$, $j = 6$ and $rbio1.5$. The

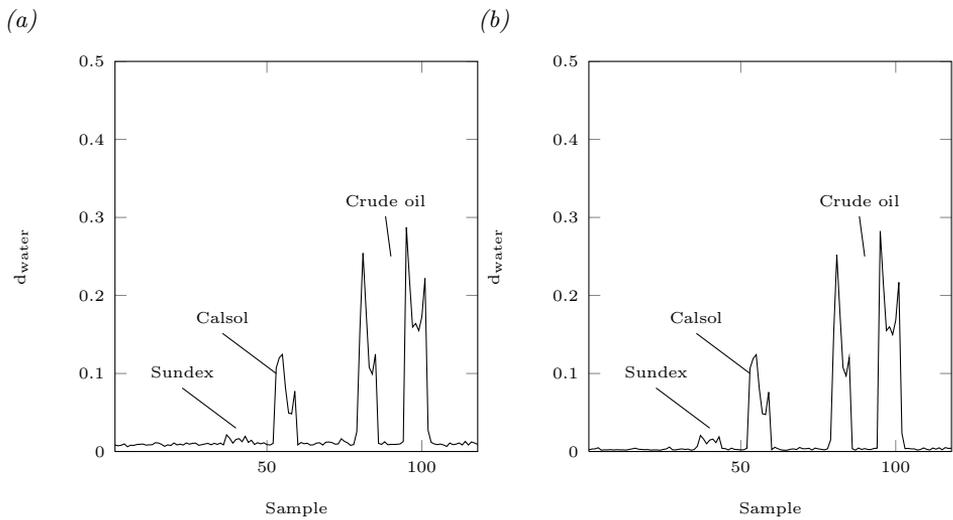
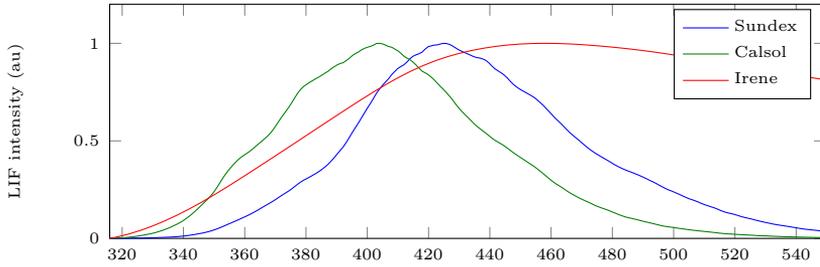
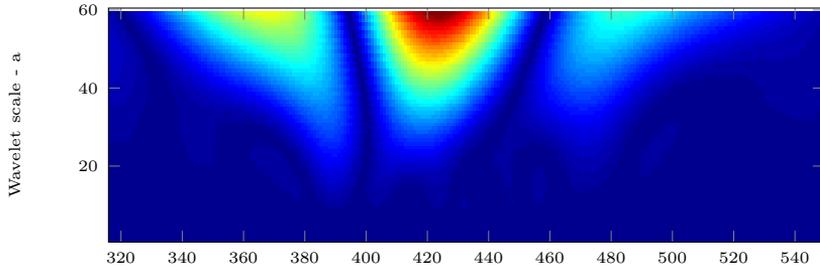


Figure 4.5: Detection of water deformation example. The target areas are marked by arrows. The presence of pollution underwater in pool as determined by distance between the HLIF of water matrix and measurement. The higher value denotes the stronger deformation of spectra. (a) - The plot indicated that Sundex pollution is not detectable due to the presence of noise additive in signal. (b) - Value of distance after the ASC denoising. The area of Sundex target is clearly seen in data.

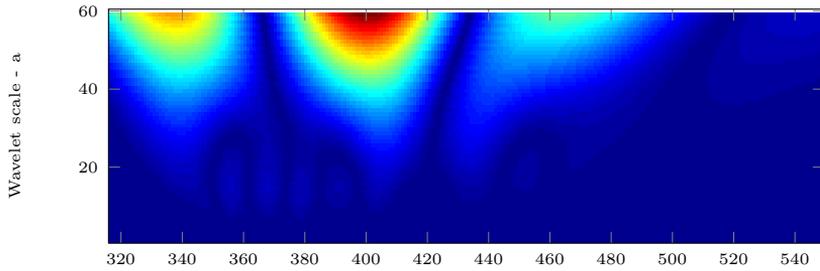
(a) HLIF spectra plot of experimental pollutants.



(b) The CWT spectra of Sundex



(c) The CWT spectra of Calsol



(d) The CWT spectra of Irene

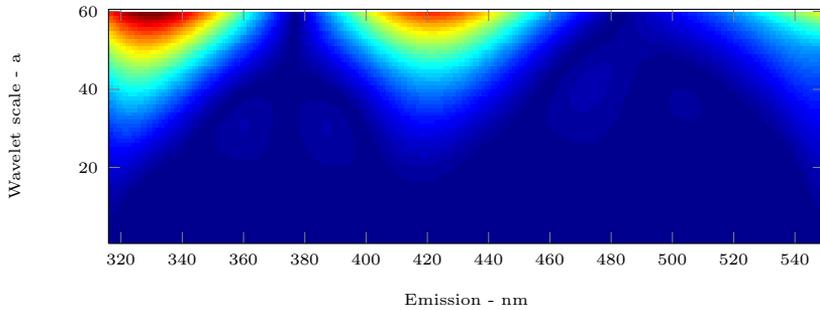


Figure 4.6: The HLIF end-member plot of Sundex, Calsol and Irene with corresponding CWT emission-scale spectra.

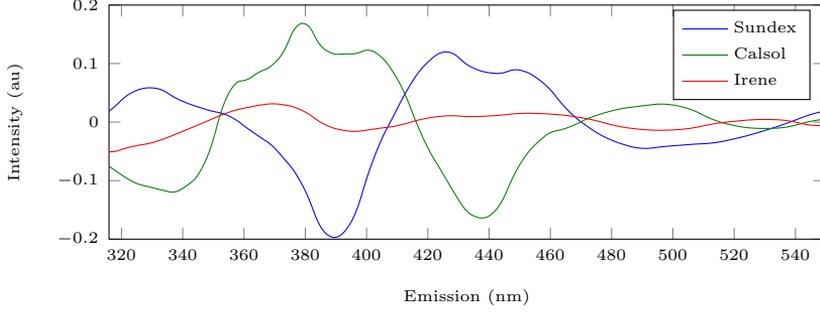


Figure 4.7: The reconstructed feature curves of Sundex, Calsol and Irene. The curves are reconstructed from detail coefficients on scales $cD_{4,5,6}$. The Irene features are extracted with $\tau = 0.005$.

non-zero coefficient number for each pollutant are

Name	Sparsity, k	Compression $\frac{n}{k}$	$\#\{cA, cD_j, \dots, cD_1\}$
Sundex	15	93.67	4—4—5—2—0—0—0
Calsol	15	93.67	4—4—4—4—0—0—0
Crude oil	7	97.05	4—3—0—0—0—0—0

The Crude oil structure is very sparse, it can be represented using only 4 approximation and 3 detail coefficients. Such small number of features for comparison can affect the performance. Thus the number of detail coefficients for analysis is increased by lowering the reconstruction error to $\tau = 0.005$. New coefficients are

Name	Sparsity, k	Compression $\frac{n}{k}$	$\#\{cA, cD_j, \dots, cD_1\}$
Crude oil	12	94.94	4—3—4—1—0—0—0

The detail coefficients on scales $cD_{4,5,6}$ are used in FWD calculations. The reconstructed details are plotted in figure 4.7.

Figure 4.8 shows the results of pollution identification. The clean water samples identified in previous step are removed from plot. The results of full signal-space distance (see figure 4.8a) show numerous classification errors specifically for crude oil samples (the results overlap with Calsol model). On figure 4.8b the type of pollutant is clearly seen for each sample. The Irene crude oil samples are distinguished from Calsol model. Some distance values of Irene samples are generally higher than overall binary classifier threshold (estimated as $5e^{-3}$). Such minor errors are caused by a high level of noise in the instrument. The distance is increased as the small features of crude oil spectrum are affected by the signal noise.

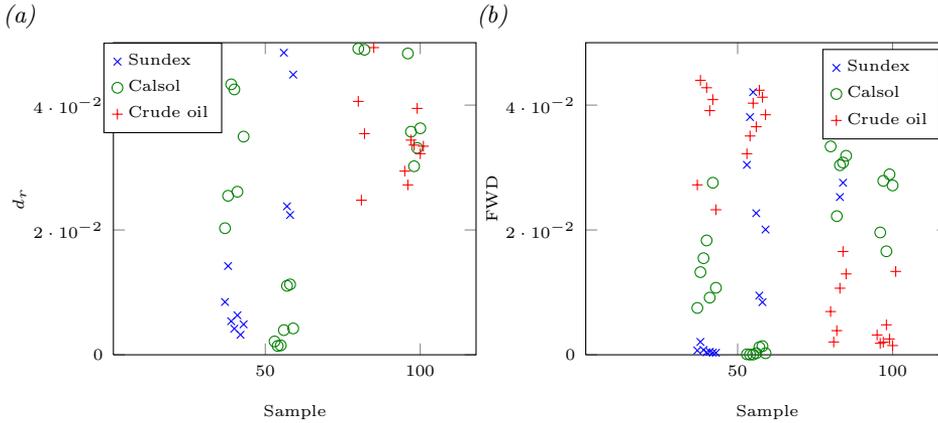


Figure 4.8: Identification of oil pollution example. The identification of pollution from underwater targets. The lower value denotes the stronger similarity of sample residual to the pollutant end-member. (a) - the conventional distance. (b) - the FWD on sample residual.

4.5 Land surface oil pollution identification

The next example is denoted to show the application of HLIF data analysis for identification of oil pollution of land surface, see description in section 4.2.3. The experiment was designed to simulate the fresh oil pollution, when oil constantly refreshes from leakage source. The research was also directed to obtain a library of land natural fluorescent objects. The library also included frequently encountered non organic surfaces like sand, stones, etc, which have some natural fluorescence response due to the covering organics layer. The dataset included porous objects, like moss and soil, which significantly affected the pollution process, as most of fresh oil instantly went underneath the surface layer.

The experimental library contained common natural fluorescence objects and underlying surfaces with additional 26 vegetation items. The vegetation data included leaves and bark samples of hardwood and coniferous trees, bushes and flowers. The complete list is given in table 4.1.

The plot 4.9a is given as normalized fluorescence response of samples. The fluorescence efficiency of polluted and clean samples is mixed, thus no general pollution detection using the intensity value is possible. The only group of sand, soil etc. surfaces located in samples from 160 to 190 has fluorescence value lower than in polluted samples. The distribution of fluorescence intensities for polluted and clean samples is shown in figure 4.9b.

Figure 4.10a plot shows selected HLIF spectra of underlying surfaces

Surfaces	Vegetation
soil	maple fir thuja juniper pine
sand	willow blueberries cherry raspberries hazel
clay	aspen apple birch rowan chestnut larch
stone	linden oak grapes nettle fern dandelion
grass moss	blueberries clover poplar currant sagebrush

Table 4.1: The experimental library of land vegetation and surfaces. Total 26 vegetation items with grass and moss as underlying surfaces.

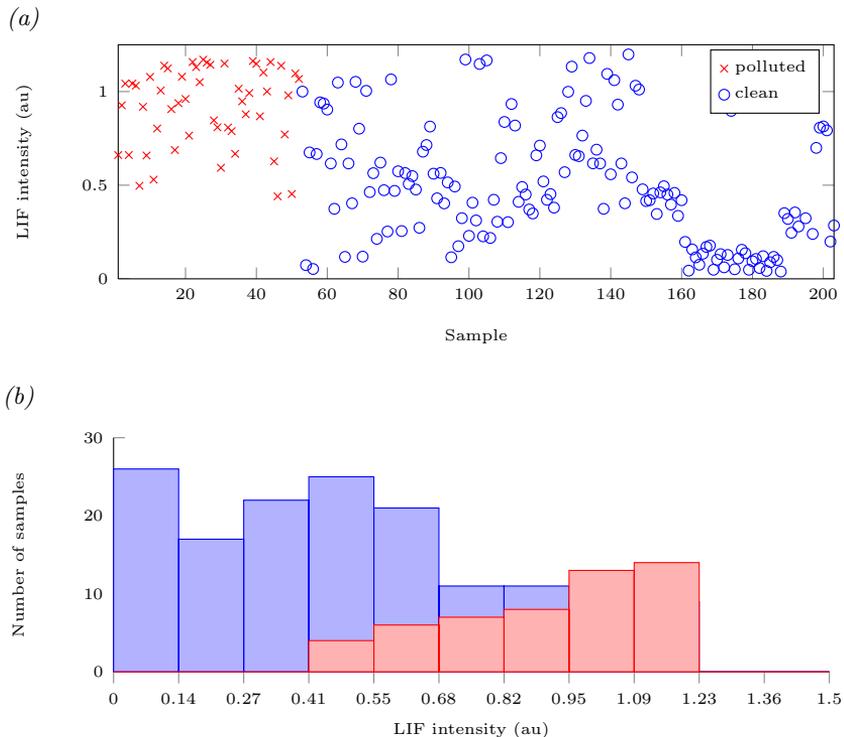


Figure 4.9: Fluorescence intensity values of land objects experimental data. (a) - The plot of normalized fluorescence values for experiment. (b) - Histogram of fluorescence values, red - polluted, blue - clean samples. The clean and polluted samples distributions overlap.

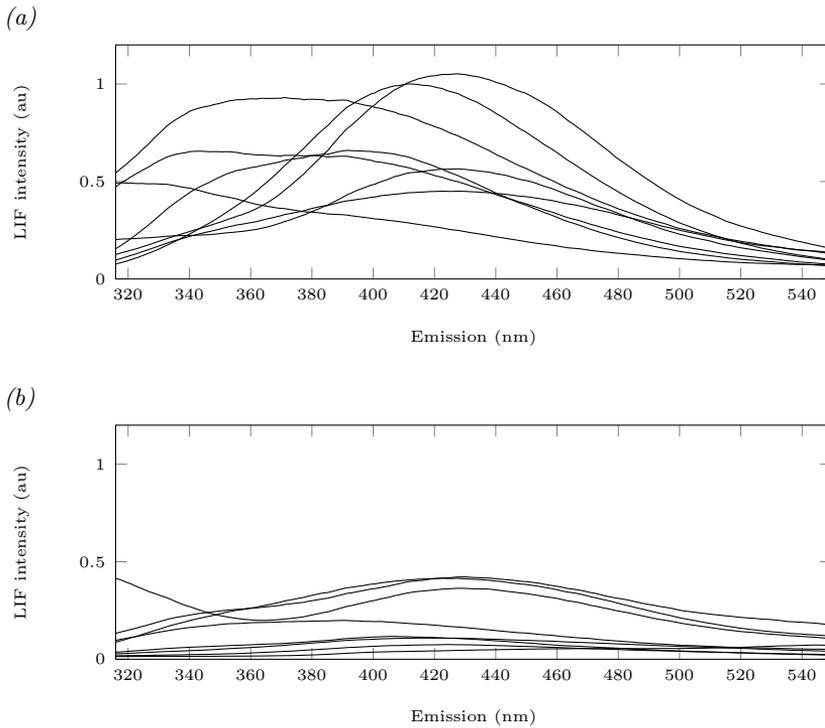


Figure 4.10: HLIF spectra examples of land objects experimental data. (a) - the spectra of fluorophores with high fluorescence intensity. (b) - the spectra of fluorophores with low fluorescence intensity. The fluorescence is compared with polluted samples.

and vegetation data. The natural fluorophores are grouped by fluorescence intensity into two groups according to histogram in figure 4.9b. It is seen that most of natural fluorophores have fluorescence comparable with polluted samples in this experiment. These includes the spectra of grass, poplar, clover and others. The low fluorescence samples include moss, soil, clay and birch. The spectra maximum are located in UV and blue range. Most of examples display a broad fluorescence spectrum with minor curvature. The visible variability of HLIF spectra shapes and intensities is high.

The fluorescence of selected natural objects was measured with and without oil pollution. The typical Crude oil was used as pollutant, as it is most general type of oil product transferred in onshore oil pipelines. The crude oil pollution was added on selected sample surface in equal amount and instantly measured. As mentioned, for some porous substances, most of oil additive eventually disappeared from target area. The final dataset contained 52 samples of polluted and 203 objects total.

The selected Crude oil HLIF spectrum with extracted and reconstructed feature curve is shown in figure 4.11. This smooth crude oil spectrum presents non standing features, thus τ is lowered to 0.005. The feature extraction parameters are $\tau = 0.005$, $n = 237$, $j = 6$ and $rbio2.6$ and the non-zero coefficients are

$$\begin{array}{ccc} \text{Sparsity, } k & \text{Compression } \frac{n}{k} & \#\{cA, cD_j, \dots, cD_1\} \\ 14 & 94.09 & 4-4-5-1-0-0-0 \end{array} .$$

In this example the HLIF analysis differ from previous experiments. Even for this experimental small library of thirty two end-members the application of LMM has low practical value (see discussion in sections 1.3 and 1.4). The applicable solution is to analyze the extracted pollutant features directly. The idea is that if pollutant spectrum is presented in linear mixture then its small scale structural components will likely to affect the resulting spectrum. This in turn will decrease distance between mixture and pollutant end-member spectrum on local scales. On the contrary, the absence of similar features will set the distance to relatively large value and exclude the false-positive sample.

Figure 4.12a shows the analysis results as plot of feature weighted distance vs sample. The first 52 samples are polluted natural objects from table 4.1. The rest of x-axis represents table 4.1 objects and their separate parts (bark, back side of leaves etc.). It is seen that generally polluted samples have FWD value close to zero. However there are some samples with distinct position on plot. The presence of such errors is due to the spongy objects that absorb oil. The experimental amount of oil was quickly absorbed, leaving only a minor or no oil on surface and consequently the pollutant HLIF signal features were extremely low in observed object.

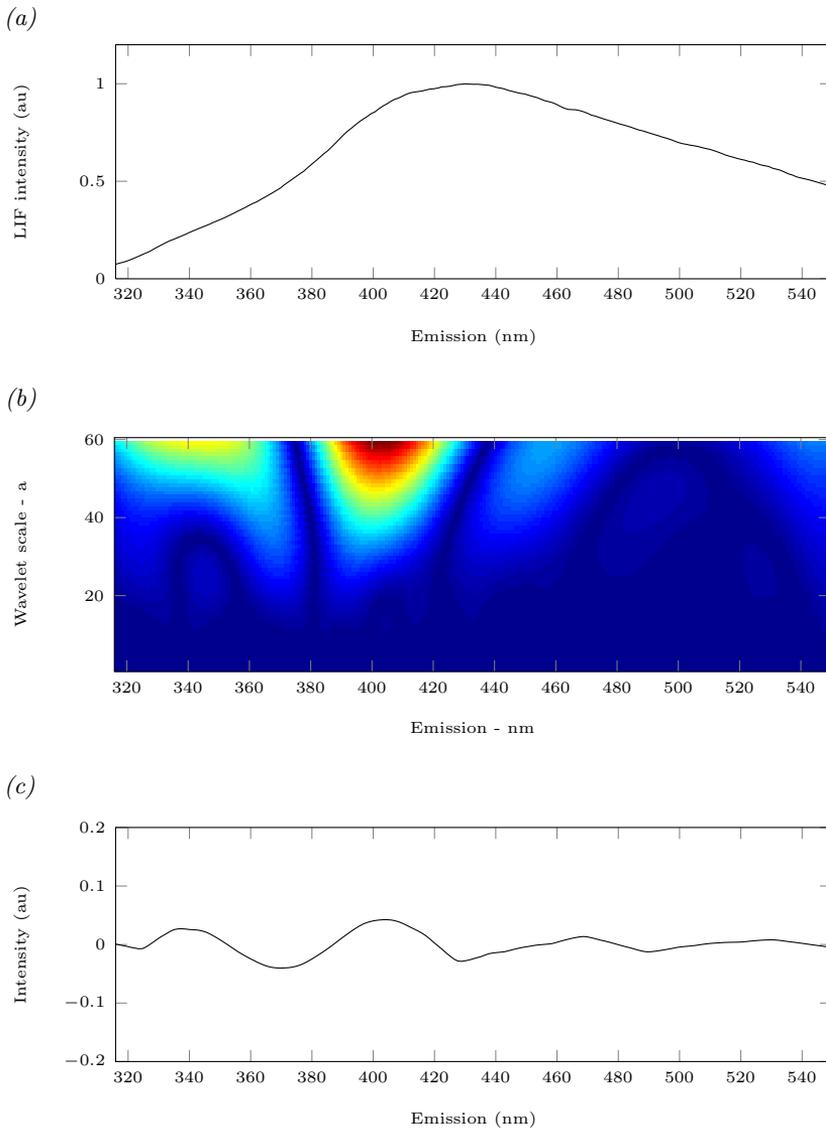


Figure 4.11: HLIF spectra of crude oil used in experiment with CWT and reconstructed feature curve. (a) - the HLIF spectra of crude oil. The spectrum structure is common, smooth and lacks of significant and unique features (ex. in comparison with oil in figure 3.13). (b)- the CWT emission-scale spectrum. (c) - the reconstructed feature curve using the scale $cD_{4,5}$ coefficients.

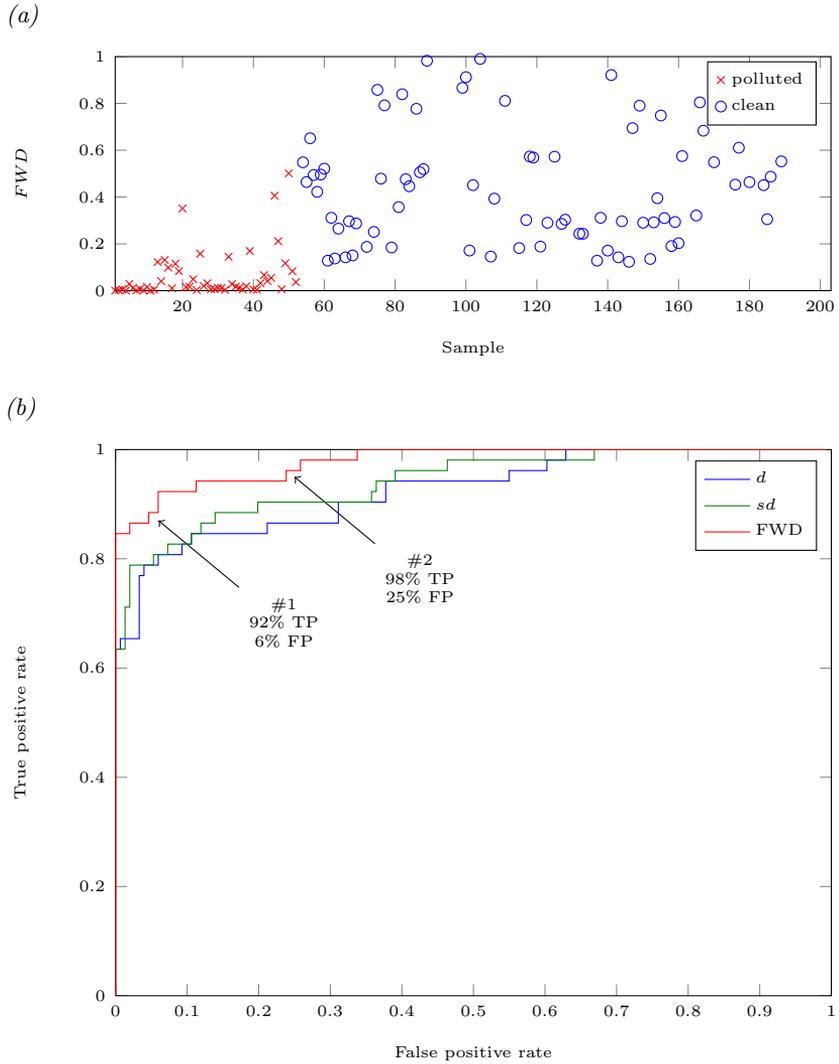


Figure 4.12: Land surface oil pollution identification results. The plot of TP vs FP rate. (a) - The plot of FWD. The lower value denotes the stronger similarity of observation to the pollutant on extracted features. (b) - The plot of true-positive vs false-positive rate or ROC characteristic for signal-space distance d , scaled distance - sd and FWD. The best results are given by FWD.

The identification quality is checked with ROC characteristic curves plotted in figure 4.12b. The curves are plotted for conventional, feature and feature weighted distances. Two points marked on plot represent two possible binary classifier settings. The first point denotes the threshold configuration for optimal false-positive results with penalty on true-positive rate. The second selects the best configuration for maximizing true-positive rate. As overall the FWD value shows the classifier with best quality. The approximate improvement of classifier between conventional and feature weighted distances is 15% for equal false-positive rate.

4.6 Land/water mixed diagnostics

The final section is a application example of land/water oil pollution detection and identification in complex conditions. The airborne remote survey of large river and shore territories after the oil spill presented a sophisticated task for HLIF LIDAR system and analysis methods. The description of survey conditions and tasks is given in section 4.2.4.

Crude oil pollution which occurred a year before the survey presented a substance of mixed volatile hydrocarbon diluents with Canada's Athabasca oil sands heavy bitumen. The light fractions of hydrocarbon evaporated in a short term after a spill, leaving the bitumen in a water column. The spectrum of bitumen was measured using the airborne LIDAR system and plot is shown in figure 4.13. Analysis of spectrum structure with CWT revealed that no significant features are present on spectrum (see figure 4.13c). Yet minor details were extracted with parameters $\tau = 0.005$, $n = 256$, $j = 6$ and *rbio2.6*

Sparsity, k	Compression $\frac{n}{k}$	$\#\{cA, cD_j, \dots, cD_1\}$
16	93.75	4—4—4—4—0—0—0 .

In mixed diagnostics the on-line identification of target surface is a critical point in success oil pollutant classification. This is specifically true in diagnostics of closed waters with expected high content of natural DOM. The background surface identification can be managed using the georeferenced target location with support of topographic map meta data. However this statement is true only for post-processing analysis, as in real-time operation the information about closed water areas is typically not available. In such case the real-time analysis must lean only on incoming spectral data.

The real time simple indicator of water target is a RSS. In natural water linear mixture model (see figure 1.5) the RSS is included with spectrum of DOM. The presence of RSS on HLIF spectrum turns the algorithm towards the water analysis, which was described in detail in previous sections.

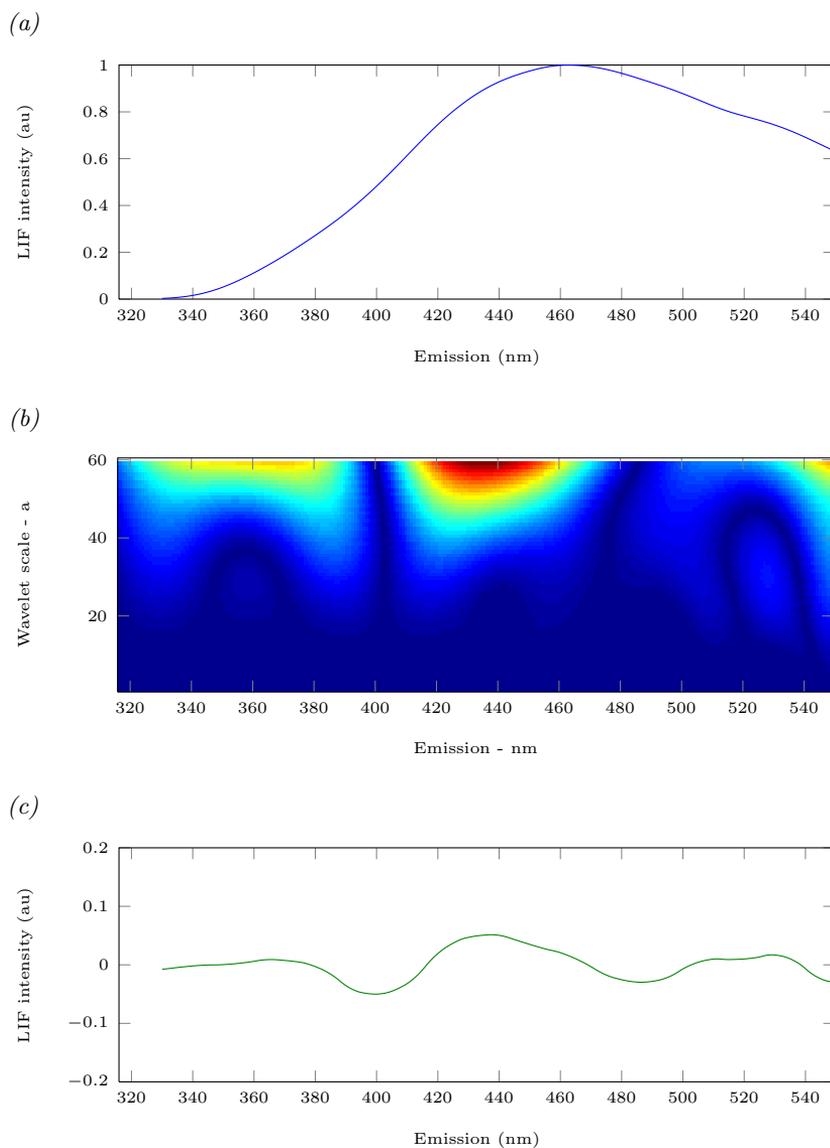


Figure 4.13: HLIF spectra of bitumen pollution from Kalamazoo experiment. The spectrum maximum is located on 460 nm. The shape is similar to previous experiment crude oil with maximum shifted towards longer wavelengths.

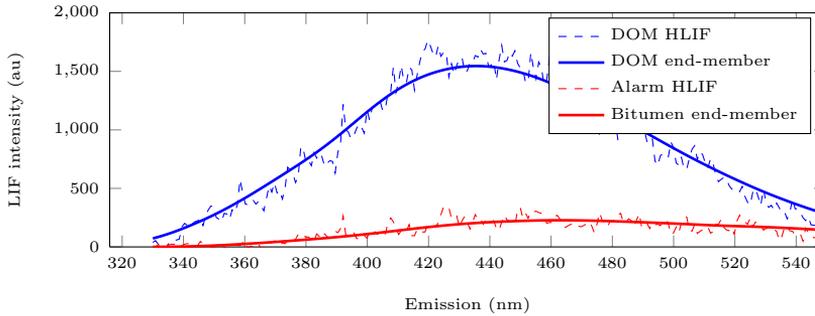


Figure 4.14: Kalamazoo ground finding HLIF spectra. The estimated oil spectra intensity is compared with DOM target from nearby location. The DOM spectrum shows no RSS, which speaks about extremely high concentration of natural organics in water.

However, due to the high concentration of natural organics, there exist cases when Raman scattering is suppressed by high water turbidity, like one showed in figure 4.14. In such case only the DOM deformation is used to determine the target surface. The latter means that DOM is subject to feature extraction as any other end-member, and FWD is used to find the samples with extremely high water turbidity. As a result, those samples are referred as a clean water.

After the target surface estimation the on-line analysis follows the land or water approaches as described previously. The post processing relies on the information of found alarms and targets with high DOM content in real time analysis. The available georeferenced meta data is used to correct the land/water surface estimation and retry the analysis for positive samples. The available meta information can be used to filter out the side findings.

Figure 4.15 shows the global result of post processing analysis summary for ground findings. The areas with detected pollution on the ground are marked with orange and red colors according to pollution signal intensity. Blue polygon is a georeferenced meta data of river estimated inundation area, which is extended from normal river bed location. There were multiple locations where oil fluorescence signals are detected. The areas with lowest distance (FWD) to oil end-member are considered as “oil detected”. Any targets with values less than one standard deviation from “oil detected” distance were classified as “oil pollution suspected” as a conservative measure to deal with this buffer area.

There were no local oil film findings on water surface, thus all HLIF spectra with RSS and high DOM deformation were classified as unknowns. Those spectra were again included in post processing analysis for land targets.

The example of land alarm is demonstrated in figure 4.14 (red line).

The fluorescence intensity of spectrum corresponds to the estimated value of bitumen on 200 meters distance. The spectral shape is discriminated using the FWD and presents high similarity to end-member.

Plot 4.16 and 4.17 show the detailed view of river inundation area at two locations selected in figure 4.15. The first section corresponds to upper river area which was the closest to the place where the pollutant entered the river. The data analysis shows residual oil pollution on land located inside the inundation area. The overall ground area remains cleared from pollution, which is confirmed by visual inspection and ground sampling. The RSS with high DOM deformation is detected on the river bank, which corresponds to mixed spectrum of land vegetation and water. No DOM deformation from oil pollutant is present in water medium.

The second area covers the larger region in downstream after the Ceresco dam. The land regions show multiple pollution locations. The number of DOM unknown deformation locations are in the proximity of the river banks and in mixed location like swamps and small ponds. Still the water surface remains clean from oil.

The application of Kalamazoo land/water mixed diagnostics combine all techniques and algorithms for HLIF data processing and interpretation. The analysis final results correspond to the estimated state of Kalamazoo region at the time of survey. The most of pollutant was removed or isolated from land surface. Large amount of heavy bitumen was developed into sediment and remained deep under the water surface. Due to the high turbidity of water the LIDAR was not able to directly detect below 20 inches, so it was not able to survey the river or lake beds.

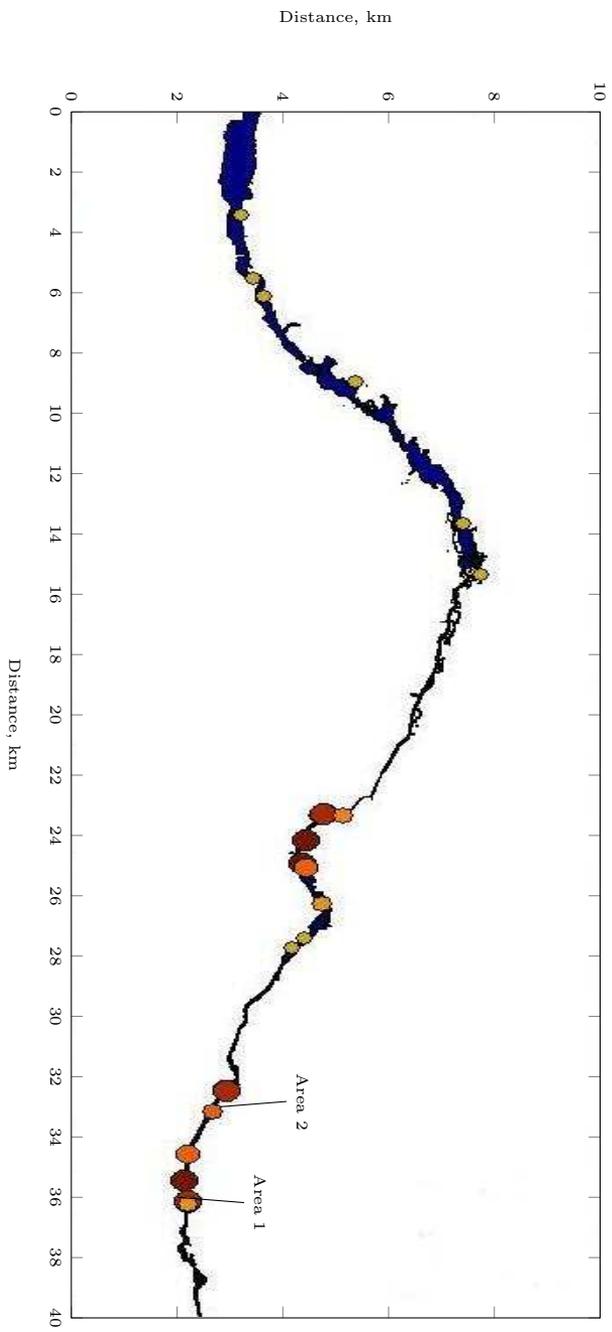


Figure 4.15: Summary of ground findings for Kalamazoo post processing. The targets with detected oil pollution on the ground are colored from orange to red according to the estimated pollution quantity (an average number of Alarm findings in area). Blue polygon is a georeferenced meta data of river estimated inundation area. Two areas are selected for detailed analysis as Area 1 and 2.

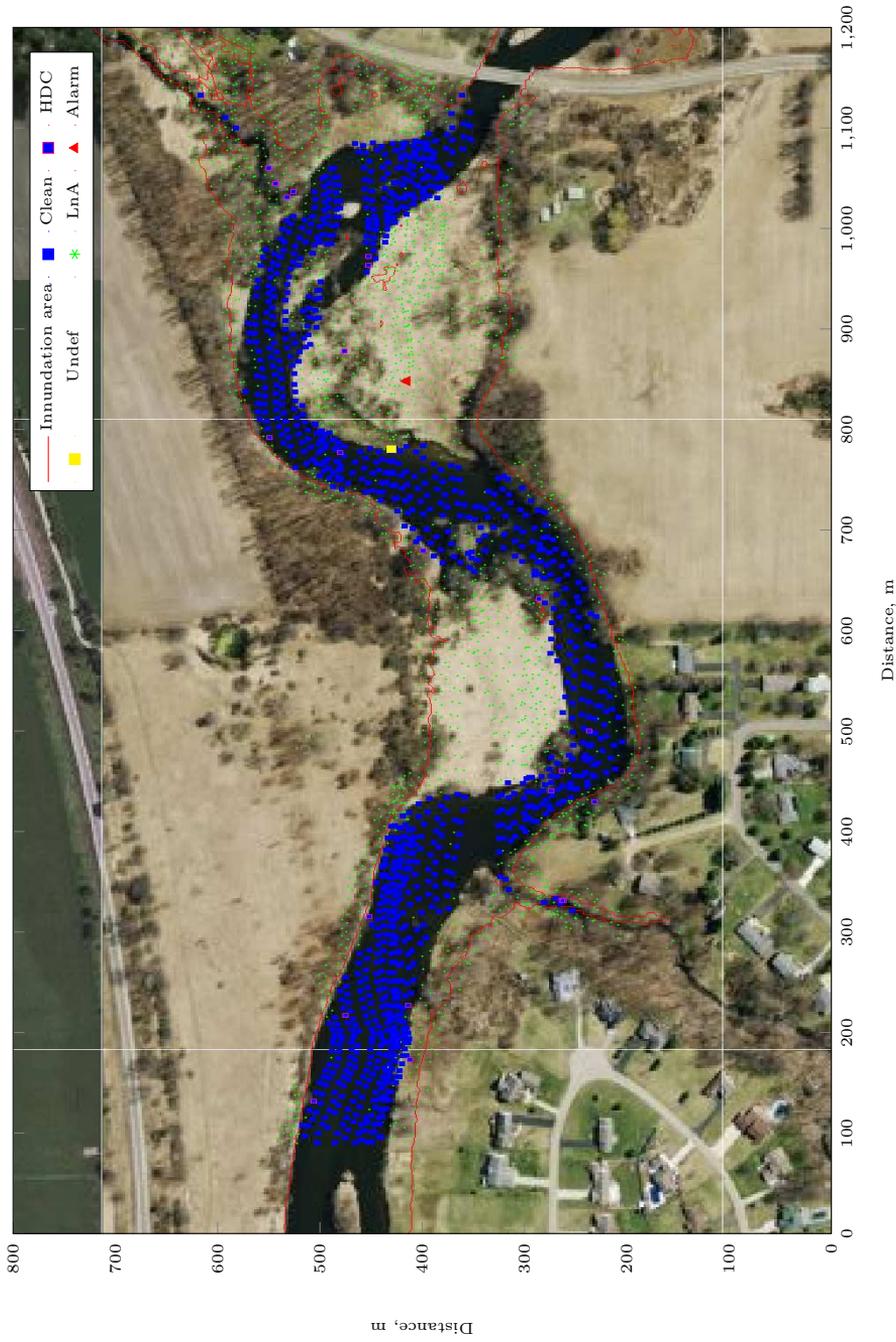


Figure 4.16: Detailed view of Kalamazoo Area 1. The map with color-scale LIDAR samples (“Clean” - clean water with RSS, “HDC” - clean water with high DOM concentration, “Undef” - unknown DOM deformation, “LnA” - land sample no oil, “Alarm” - sample with detected oil pollutant). Residual oil pollution detected inside the river inundation area is marked as red triangle. The mixed spectrum of water and vegetation is marked as yellow square.

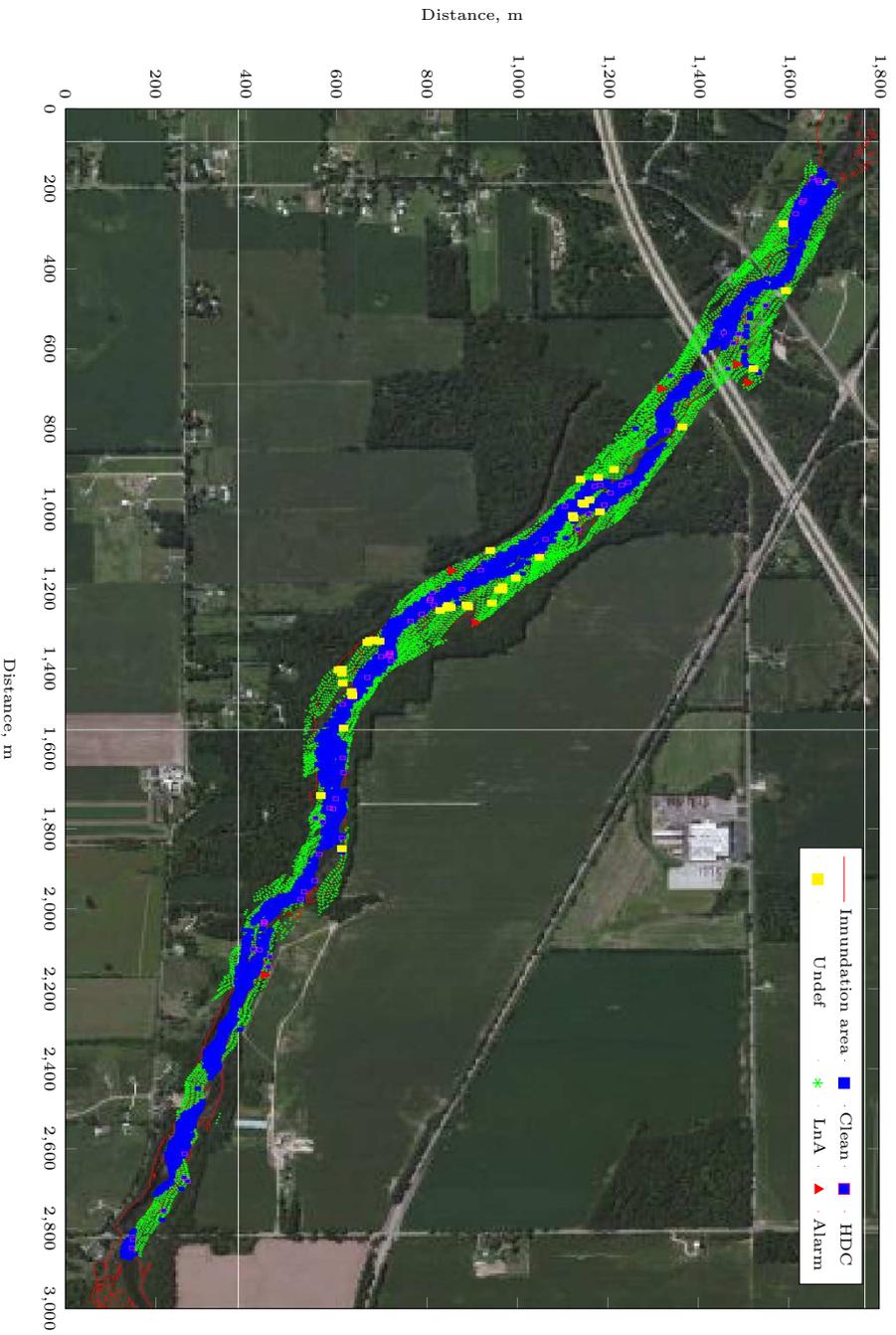


Figure 4.17: Detailed view of Kalamazoo Area 2. The map with color-scale LIDAR samples (“Clean” - clean water with RSS, “HDC” - clean water with high DOM concentration, “Undef” - unknown DOM deformation, “InA” - land sample no oil, “Alarm” - sample with detected oil pollutant). Many findings with “Undefined” DOM deformations are related to mixed spectra with RSS, DOM and land organics. Some locations of high probability land oil findings are marked as “Alarm” red triangles. The rest of river surface remains clean.

4.7 Spectral Layered Analysis

The previous application demonstrated the complete logic of pollution detection and classification. The techniques of end-member model composition, HLIF denoising and feature extraction are incorporated into the procedure. This created a framework of operational multi-tier analysis procedure which is proposed here.

The Spectral Layered Analysis (SLA) is the multi-tier system of pollution detection and identification for operational use in real-time and post processing LIDAR application. In addition to identification the analysis can estimate important parameters of water medium and can be used to classify the natural background. With proper calibration, analysis output can estimate the quantity of oil pollution. Using additional georeferenced meta-data the analysis results are used to create the thematic maps of the survey area as showed in section 4.6.

The schematic structure of analysis is given in figure 4.18. The first step of on-line analysis includes the signal preprocessing stage. The general preprocessing is used to improve the signal characteristics, like signal-to-noise (SNR) ratio. The ASC denoising is applied here. The additional corrections are device specific functions, like detector spectral response, which are used to standardize the HLIF data between the instruments. Signal filtration estimates the quality of signal and overall fluorescence intensity in order remove low and oversaturated measurements.

The second layer is the identification of underlying surface. Generally we want to discriminate between natural water and terrestrial targets. A natural water is always characterized by specific and strong fluorescence response: RSS and DOM. The presence of RSS is unambiguous indication of water medium, thus the analysis follows the water algorithm (see figure 4.19). The absence of RSS can signal about the high water turbidity. The DOM identification is used to check if the spectrum shape belongs to natural organics. The spectra which belong to DOM class without RSS are specifically marked as “High DOM Concentration” (HDC) for future processing. The rest of data moves to the land analysis procedure.

Although, RSS can be easily detected, the surface estimation error can arise due to the suppression of RSS in case of thick oil film or small water column, ex. at pool. In such case the additional information is geographic reference of measurements, which is used for supervised underlying identification and error correction in post processing stages.

The post processing of HLIF data is divided into two stages. The spectra marked as “Undefined” pollution from water analysis and all on-line alarm data are reprocessed by corresponding analysis method, with corrected geographic reference. This step is required for spectra with false-positive RSS estimation, as it removes the false-positive undefined water

pollution. Also the data with false-negative RSS estimation is rechecked with natural water LMM. The second stage is implemented for “LnA” (“Land no Alarm”) marked data. This step is optional in many cases, as the reprocessing of complete land based observations can be time consuming.

After the underlying surface identification the water surface is processed using the natural water and pollutant LMM. The FWD is calculated on measurement residual and checked against the global distance threshold. The binary classifier determines the class of pollutant. Quantitative analysis available after the classification, as the calibration data (ex. concentration vs. signal power) depends on the oil product. Some additional parameters, like concentration of organic matter, water turbidity can be estimated in water analysis (Pantani et al., 1995; Babichenko et al., 2010).

In the terrestrial areas the natural background variability present increased complexity for mixture model approach. The pollution mixture model should be managed for every background end-member not considering the background mixtures itself. Thus the LMM is created from single spectrum of pollutant and the classification distance is evaluated on important features of HLIF spectrum. Approach example and results are demonstrated in structural analysis of HLIF data in section 4.2.3.

The additional thematic map of fluorescent objects can be created if the reference library of natural terrestrial background objects is available for survey area. The benefit of additional classification is the detection of unknown or outliers, which substantially differ known natural fluorophores. This information can be used for location estimation of ground sampling in post cleaning operations.

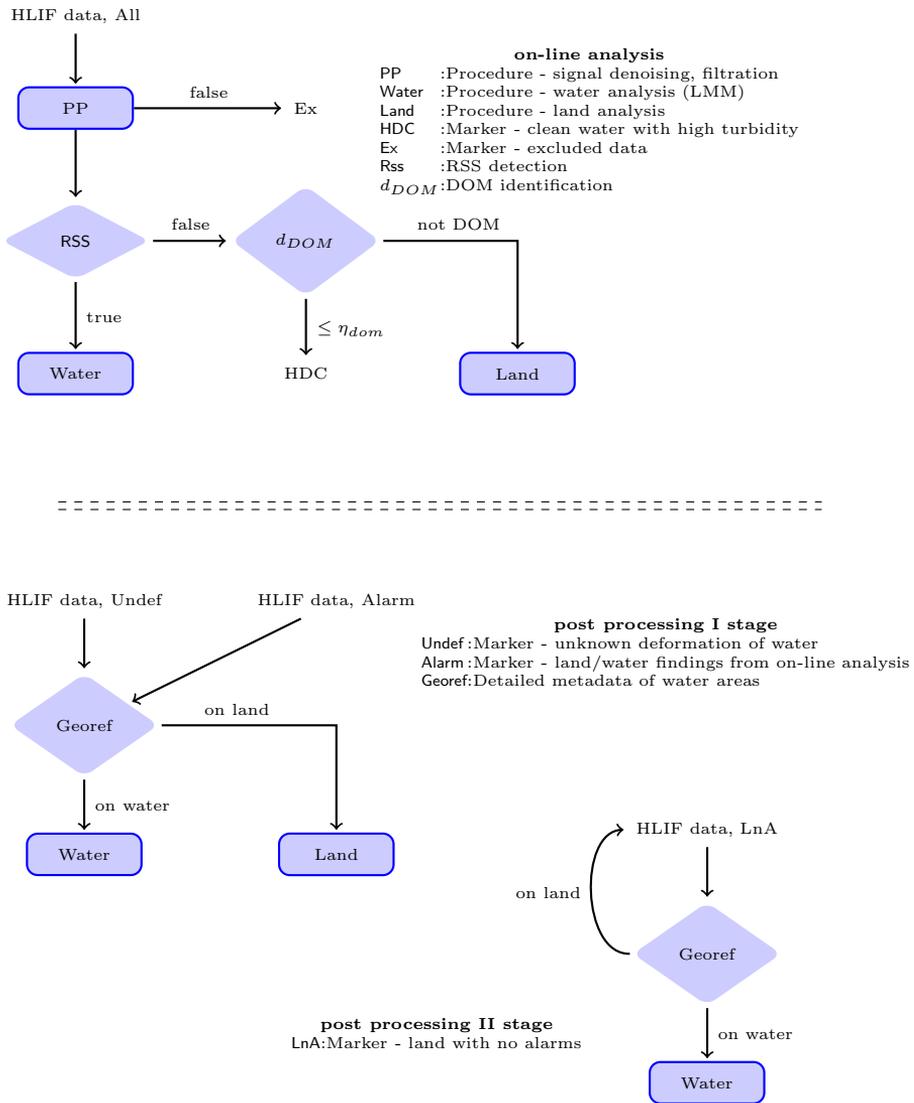


Figure 4.18: Spectral Layered Analysis diagram. The on-line analysis includes the preprocessing, RSS detection, identification of DOM and transition to land/water classification. Post processing implies additional meta data for georeferencing the measurements. Some critical observations are reanalyzed with corrected surface algorithms.

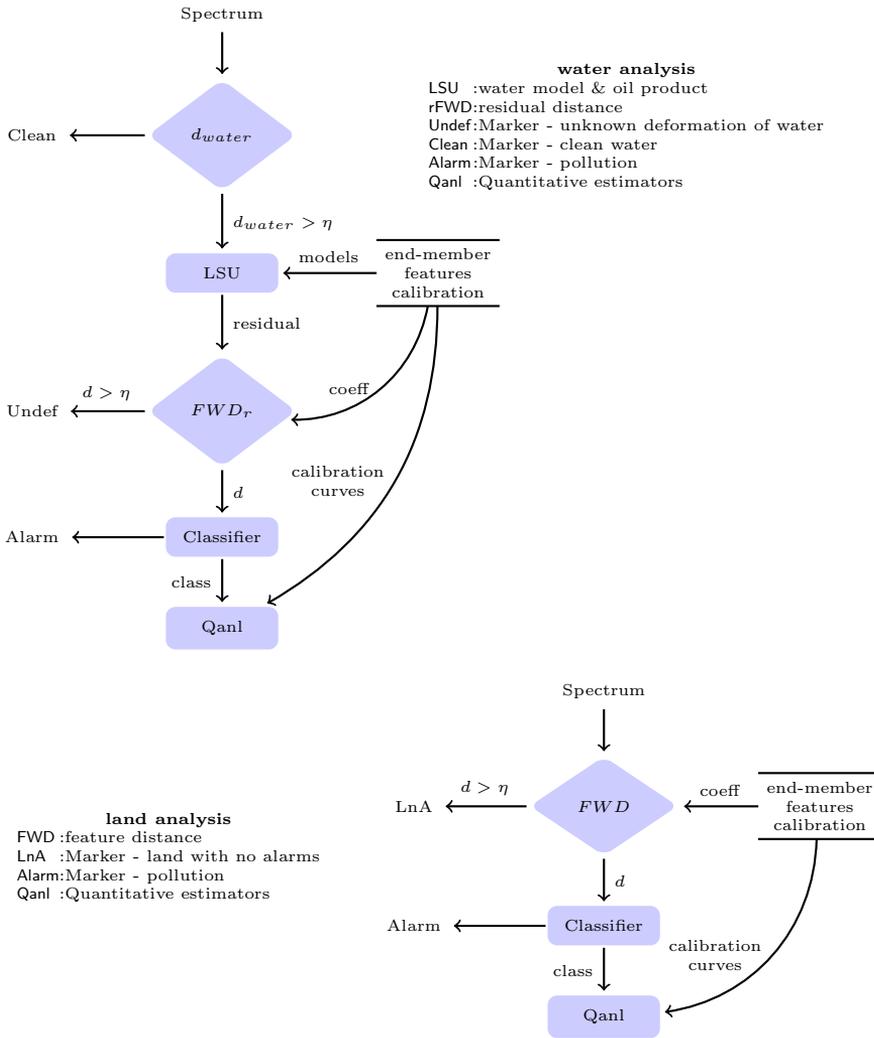


Figure 4.19: Land/water analysis diagrams. The HLIF data is processed corresponding to underlying surface. Water analysis includes the natural water LMM. Both analyses implement FWD comparison.

4.8 Conclusions

The experimental results provided in this chapter demonstrate the high reliability and feasibility of HLIF OPD method. The capabilities are further improved using the processing methods proposed in this thesis.

The developed technique of ASC denoising improves the discrimination of polluted natural water. The dependency of discrimination quality on signal power is suppressed, decreasing the minimum detectable deformation of natural water matrix from pollutant in 2.5 times as showed by experiment.

The HLIF discrimination measure weighted by extracted features (FWD) already showed the reduction of false positives for Gobles example in section 3.3.2. The results of another OPD example in difficult environmental conditions, like submerged oil targets in water with high organics, demonstrated the improvement of identification for crude oil targets.

Sophisticated experiment with land surfaces and fluorescent targets demonstrated the capabilities of HLIF OPD diagnostics on highly variable natural fluorescence background. The identification quality controlled with ROC (basic binary classifier) showed 92% of true-positive results with only 6% of false-positives. The discrimination has been performed without the linear mixture models of pollutant and background. The proposed feature extraction increased the classification performance for 15% in average.

Final example has been denoted to OPD in post-cleaning remote sensing survey of land/water areas. The results demonstrated the ability of airborne HLIF LIDAR system to detect the minor crude oil pollution on inner waters and terrain. Two local detailed georeferenced OPD maps have been generated with estimated residual pollution locations.

Considerable experimental experience allowed to refine and adjust core procedures and branches in the HLIF OPD data processing creating the state-of-the-art OPD HLIF platform, which has been introduced as a final result of the thesis. The basic structure includes the real-time and post-processing operation considering HLIF data measured on land and water. The schematic structure shown on figures 4.18 and 4.19 will be implemented in HLIF LIDAR operational software.

Conclusion

Concluding remarks

The HLIF data processing solutions for OPD are designed in this thesis. Main developments considered the improvement of pollution diagnostics in harsh operation conditions in order to meet the criteria of the automated decision making.

While the problem of pollution diagnostics with LIF technique has been well studied, a real-world measurement conditions and physical limitations of instruments (signal power, noise) have played a negative role in emergence of comprehensive operational and commercial solution of LIF-LIDAR as a *sensing component*. The conversion of LIDAR measurements, such as HLIF data, into useful information regarding the end-user requirements has been undeveloped yet, basically due to the time-consuming process of technology maturity.

The HLIF data processing begins with understanding of data structure and interpretation of data variance on the basis of scientific experience. The general assumptions about the HLIF signals, thus the fluorescence response have denoted the HLIF signal LMM approach. LMM is common for hyperspectral imagery, however in LIF the substance binded fluorescence response allows the creation of universal and calibrated end-member library and group the end-members by their role in analysis.

The LIF technique provides the direct response of targeted object molecular structure and when it is recorded with hyperspectral detector it can be analyzed as continuous spectrum. In spite of LIF advantages, the spectral broadening of fluorescence presents the challenges for discrimination between objects (for example comparing with Raman spectroscopy). This challenge has been studied across this work as discrimination of pollutant HLIF from natural fluorescence background objects (DOM or land organics) in a number of simulation and real-world experiments.

The discrimination challenge brings next statement introduced in Sobolev and Babichenko (2013a) - the relation of classification features to spectral properties, like local extreme points, is very important. First of all the discrimination between HLIF data is dependent on signal noise, which is

not related to spectral properties rather than to limitations of instruments. The ASC denoising algorithm developed in the work and tested on specific HLIF signal-dependent noise simulation demonstrated improved pollution detection capabilities in experiments with real data. The denoising for signal-dependent noise is not limited to HLIF signals, the technique will be applied for single-channel fluorosensor for detection of microalgae in water in Babichenko et al. (2014) utility model. The HLIF noise model research has been done to create realistic simulation test environments for this thesis and forthcoming work.

While the HLIF spectrum holds large amount of information the quality of discrimination is dependent on the extraction and selection of the relevant one. The selection of class discriminating information is the problem of pattern recognition and there is no general approach for all applications.

In HLIF processing the structural feature extraction/selection has been proposed. It naturally involves all relevant parts constituting the spectrum in discrimination, while the irrelevant parts are removed. The process of feature extraction is automated and can be embedded into LMM library. The developed feature extraction capabilities have been demonstrated on a number of real-world OPD field experiments and operations. A case study of land objects pollution showed the improvement at 15% comparing with conventional discrimination in data-space.

The variety of real-world, laboratory and simulation experiments have been used to test the HLIF OPD capabilities with developed processing methods at the final chapter of the work. As the result of certain development level the solution for HLIF OPD in mixed water/land environment has been created. The solution considered the real-time delivery of user information, like polluted and suspicious findings. The post-processing is designed for the refinement of the results.

The continuous detection and assessment of water/land pollution, classification and estimation the pollution level and generation of environmental warning data at earlier stage are the main assumptions behind the development of a HLIF monitoring systems. The R&D activities in this thesis made a step towards the LIDAR automated pollution diagnostics operation.

Future work

The global oil production trends create demand of technology and sensors capable working in real environments. Example is the offshore oil industry which continuously moves into the subsea areas. The subsea oil industry is the largest commercial user of Remotely Operated Vehicles (ROV), where they have become an essential tool for deepwater operations. In the primary offshore oil activity sector a long period of high oil prices and surging

deepwater activity has driven orders for offshore drilling rigs to numbers not seen for decades (GBI Research, 2011).

While being effective in many applications conventional HLIF technique still has a number of limitations, like the complex optical scheme and mild signal-to-noise ratio, that inhibit wide operational use of such instruments. The operation of sensors onboard of ROV or Unmanned Airborne Vehicles (UAV) requires the extensive miniaturization of technology. This concerns new solid state sensing hardware with low power, high sensitivity and low cost components, while leaving the advantages of hyperspectral information intact.

The promising Compressed Sensing (CS) technology which is discovered by Candes and Tao and by Donoho (Candès et al., 2006; Donoho, 2006) in 2006, in theory, gives the ability to reconstruct complex signals from very few samples. The comprehensive overview of CS advances can be found in Candès and Wakin (2008) and Lee (2011). The practical imaging devices that implement the theory have been already developed and put into marked (InView Technology Corporation, 2014).

The HLIF data presents high compression rate, which has been shown during the feature extraction in chapter 3, thus it has the potential for implementation in the framework of CS. The current status of development is the design of laboratory sensor using the conventional low-resolution discrete channel detector with static low-cost low-quality filter array. An additional theoretical developments of CS signal reconstruction considering the fluorescence spectral properties are on schedule.

Appendix

Example implementations of algorithms discussed in this thesis are provided in Appendix. The code listings are given for Matlab scripting language, however can be easily translated to any other. The codes are tested on Matlab with “Wavelet Toolbox” v4.9(R2012a). Algorithms implementations must be considered for research purposes only. The operational use require implementations in different and more computationally efficient languages.

1-D lifting wavelet transform

This algorithm is customized and corrected realization of lifting transform implementation `lwt` in “Wavelet Toolbox” v4.9(R2012a). The signal boundary adaptation error is corrected as in original `lwt` it introduces the zero padding of signal outside support. The algorithm follows the WT theory introduced in section 2.2.4.

Initialize signal y , number of scales N and lifting scheme `LS` using the Matlab function `liftwave`.

```
function [C,L] = liflwt(y,LS,N)

C = [];
ye = y;
for j=1:N
    % Splitting.
    yo = dyaddown(ye,0);
    ye = dyaddown(ye,1);

    ko = 1:length(yo);
    ke = 1:length(ye);

    % Lifting.
    for l=1:size(LS,1)-1
        koeff = LS{1,2};
        z = LS{1,3};
        if strcmp(LS{1,1},'d') % dual lifting step
            for k=1:length(koeff)
                yo = yo + koeff(k)*(interp1(ke,ye,ko+z,...
                    'linear','extrap'));
                z = z - 1;
            end
        elseif strcmp(LS{1,1},'p') % primal lifting step
            for k=1:length(koeff)
                ye = ye + koeff(k)*(interp1(ko,yo,ke+z,...
                    'linear','extrap'));
                z = z - 1;
            end
        else
            error('lifting error');
        end
    end

    % Normalization.
    ye = LS{end,1}*ye;
    yo = LS{end,2}*yo;

    C = [yo C];
    L((N+2)-j) = length(yo);
end
L(1) = length(ye);
```

```
L(end+1) = length(y);  
C = [ye C];
```

The output is coefficients C and corresponding length for each scale L .

Inverse 1-D lifting wavelet transform

This is the corrected inverse 1-D lifting transform.

```
function y = lifilwt(C,LS,N,L)

y = C(1:L(1));
idx = cumsum(L(1:end));
for j=1:N
    yo = C(idx(j)+1:idx(j+1));

    ko = 1:length(yo);
    ke = 1:length(y);

    % Normalization.
    y = y/LS{end,1};
    yo = yo/LS{end,2};

    % Lifting.
    for l=size(LS,1)-1:-1:1
        koeff = LS{l,2};
        z = LS{l,3};
        if strcmp(LS{l,1},'d') % dual lifting step
            for k=1:length(koeff)
                yo = yo - koeff(k)*(interp1(ke,y,ko+z,...
                    'linear','extrap'));
                z = z - 1;
            end
        elseif strcmp(LS{l,1},'p') % primal lifting step
            for k=1:length(koeff)
                y = y - koeff(k)*(interp1(ko,yo,ke+z,...
                    'linear','extrap'));
                z = z - 1;
            end
        else
            error('lifting error');
        end
    end

    % Merging.
    tmp = zeros(1,length(y)+length(yo));
    tmp(2:2:end) = yo;
    tmp(1:2:end) = y;

    y = tmp;
end
```

Poisson-Gaussian noise model algorithm

Algorithm to create Poisson-Gaussian noise model from (2.22). First step is to initialize Poissonian a and Gaussian b values, clipping parameters and $y(x)$ where $x \in X$ and $y : X \rightarrow [0, 1]$.

```
% y = f(x)->[0...1];  
a = 0.04^2;  
b = 0.004^2;  
clipping_above = 1;  
clipping_below = 0;
```

Generate random numbers from the Poisson distributions with mean parameters $\lambda = \frac{y}{a}$.

```
z = a*poissrnd(y/a);
```

Generate values from a normal distribution with mean 0 and standard deviation \sqrt{b} and add them to z .

```
z = z + sqrt(b)*randn(size(y));
```

Clip resulting z .

```
if clipping_below  
    z = max(0, z);  
end  
if clipping_above  
    z = min(z, 1);  
end
```

The output is $z(x) = y(x) + \mathcal{N}(0, ay(x) + b)$.

Estimation of Poisson-Gaussian noise model parameters

Algorithm estimates the Poisson-Gaussian noise model parameters using $\sigma_{fit}(y) = \sqrt{\max(0, \hat{a}y + \hat{b})}$ as curve fit to $\{\hat{y}_i, \hat{\sigma}_i\}$ segment pairs. The algorithm is divided into two parts: segmentation and curve fitting.

We consider here that dynamic range of sensor (in ex. bits per pixel, saturation level) denoted as S is known and $y(x) \geq 0 \forall x$. Initialize data $x \in X$ and $z : X \rightarrow [0, 1]$ and clipping parameters. The z can be matrix of measurements of size $m \times n$, where m is number of measurements.

```
z = z/S; % normalize z into [0...1] range  
clipping_above = 1;  
clipping_below = 0;
```

In general the HLIF signals are smooth and have no edges inside their support. Thus the estimation of local $\{\hat{y}_i, \hat{\sigma}_i\}$ pairs is based on histogram splitting of non-overlapped \hat{y} segments. The signal boundary must be

concerned, thus the noise analysis uses lifting scheme to reduce the effect of signal boundary in detail coefficients (important for signal shown on figure 2.8a). Initialize lifting scheme for 3rd order Daubechies wavelet using `liftwave('db3')` (scheme structure is given below). All wavelets in Matlab have normalized l^2 -norm to 1 and $\sum \varphi = \sqrt{2}$, thus the approximation coefficients must be normalized to $\sqrt{2}$.

```

LS = liftwave('db3');
%LS =
%'d'      [ -2.4255]      [ 0]
%'p'      [-0.0793,0.3524][ 1]
%'d'      [2.8953,-0.5614][-1]
%'p'      [ 0.0198]      [ 2]
%[2.3155]  [ 0.4319]      []

% store (A)pproximations and (D)etails of all z
A = []; D = [];
for i = 1:size(z,1)
    [C,L] = liflwt(signoise(n,:),LS,1);
    cA = C(1:L(1));
    cD = C(L(1)+1:end);
    A = [A cA/sqrt(2)];
    D = [D cD];
end

```

Compute the histogram using $k = \lceil 2n^{1/3} \rceil$ rule for number of bins k (Sturges, 1926).

```

% segment (sA)pproximations and (sD)etails
sA = {}; sD = {};

k = ceil(2*length(A)^(1/3)); % number of bins
edges = linspace(0,1,k+1);

for i = 2:1:length(edges)
    idxs = edges(i-1) <= A & A < edges(i);

    sA{end + 1} = A(idxs);
    sD{end + 1} = D(idxs);
end

```

Estimate local expectation/standard deviation pairs. The standard deviation uses robust estimator based on the median absolute deviation (MAD) with bias factor f (Mosteller and Tukey, 1977).

```

% expectation
yi = cellfun(@(in) mean(in),sA);

% standard deviation
f = cellfun(@(in) 1/(5*length(in))+norminv(3/4),sD);
sigmai = cellfun(@(in) mad(in,1),sD)./f;

% plot the results

```

```
plot(yi, sigmai, 'o');
```

In order to create robust fit of Poisson-Gaussian noise model we minimize the logarithmic risk function of (2.26) considering the nonlinear clipped versions of \tilde{y} and $\tilde{\sigma}$. The optimization uses basic multi-start approach with random initial parameters α and β for a and b respectively, which are selected empirically or by coarse least squares estimation.

```
% denote clipping transformations using anonymous functions

% clipped variance estimation with mean mu
varmu = @(mu) normcdf(mu)+normpdf(mu).*mu-normpdf(mu).^2+...
        normcdf(mu).*mu.*(mu-normcdf(mu).*mu-2*normpdf(mu));

% calculate clipped version of sigma
if clipping_below && clipping_above
    tSigma = @(y, sigma) ...
            sigma.*varmu(y./sigma).*varmu((1-y)./sigma);
elseif clipping_below
    tSigma = @(y, sigma) sigma.*varmu(y./sigma);
elseif clipping_above
    tSigma = @(y, sigma) sigma.*varmu((1-y)./sigma);
else
    tSigma = @(y, sigma) sigma;
end
```

Apply unconstrained multivariable nonlinear optimization method of (Lagarias et al., 1998) implemented in Matlab as `fminsearch` function.

```
% initialize multi-start parameters
alpha = 0.04^2;
beta = 0.004^2;
fval = Inf; % initial value of objective function
for i=1:10 % number of trials
    [abetmp, fvaltmp] = fminsearch(@(arg) ...
        optimproblem(arg, yi, sigmai, tSigma), [alpha beta] + ...
        randn(1,2).*[alpha beta].^2);

    % select global minimum
    if fvaltmp < fval
        abe = abetmp;
        fval = fvaltmp;
    end
end

% plot the results
plot(yi, sigmai, 'o', yi, tSigma(yi, sqrt(yi*abe(1)+abe(2))));
```

The `optimproblem` is the `logrisk` function implemented in separate file as

```
function out = optimproblem(ab, yi, sigmai, tSigma)
    sigma = sqrt(max(0, ab(1)*yi+ab(2)));
    sigmat = tSigma(yi, sigma); % clipping
```

```
out = sum(log(sigmai./sigmat).^2); % logrisk  
end
```

Final estimated Poisson-Gaussian model parameters are $\hat{a} = \text{abe}(1)$ and $\hat{b} = \text{abe}(2)$.

Wavelet feature extraction

Algorithm for wavelet feature extraction of HLIF signal y , which finds the sparsity-norm minimum as given in equation (3.7). Initialize lifting scheme, error constant τ and wavelet decomposition level.

```
LS = liftwave('rbiol.5');
%LS =
%'p'      [      1]          [0]
%'d'      [0.0117,-0.0859,-0.5,0.0859,-0.0117] [2]
%[0.7071] [  1.4142]          []

tau = 0.01; % 1% error
j = floor(log(length(y))/log(2));
j = j-1; % there should be at least two coefficients
```

The decomposition is created for maximum level $j - 1$ using custom implementation of lifting `liflwt` provided earlier.

```
C = liflwt(y,LS,j);
```

Apply the Θ_k operator for $0 \leq k \leq n/2$.

```
[~,IX] = sort(abs(C),'descend');

n = length(C);
for k=1:round(n/2) % number of nonzero entries
    a = C;
    a(IX((k+1):end)) = 0;
    ErrNorm(k) = abs((sum((C-a).^2)/sum(C.^2))^0.5-tau);
end
```

Find minimum of `ErrNorm` and features a .

```
[~,kmin] = min(ErrNorm);

J = setdiff(1:n,1:kmin);
a = C;
a(IX(J)) = 0; % extracted features
```

ASC denoising

This algorithm is example implementation of ASC denoising method from section 3.3.3. The solution minimizes the (3.4) with (3.6) where penalization coefficient λ is estimated as (3.9). Initialize signal y , lifting scheme LS and wavelet decomposition level j .

```
LS = liftwave('rbiol.5');
%LS =
%'p'      [          1]          [0]
%'d'      [0.0117,-0.0859,-0.5,0.0859,-0.0117] [2]
%[0.7071] [ 1.4142]          [1]

j = floor(log(length(y))/log(2));
j = j-1; % there should be at least two coefficients
```

Compute the decomposition and sort the coefficients C according to penalization rule defined in equation (3.10).

```
[C,L] = liflwt(y,LS,j);

Cpen = C; % high penalized
[~,IX] = sort(abs(Cpen(1:L(1))), 'descend');
for n=2:length(L)-1
    idx = L(n)+1:L(n+1);
    [~,tmp] = sort(abs(Cpen(idx)), 'descend');
    IX = [IX idx(tmp)];
end
```

Compute the logarithmic error norm.

```
n = length(C);
for k=1:round(n/2) % number of nonzero entries
    a = C;
    a(IX((k+1):end)) = 0;
    LogErrNorm(k) = n/2*log(sum((C-a).^2));
end
```

Using the segmented linear regression estimate best $\hat{\lambda} = 2|\beta|$. The break point between segments is estimated by fitting the linear model with unknown parameter B . Implementation is based on Matlab `fminbnd` function, which is based on golden section search and parabolic interpolation (Brent, 2013).

```
x = 1:k; % support of LogErrNorm

% segmented model
segmodel = @(B) [(x<B);x.*(x<B);(x>=B);x.*(x>=B)];
% linear regression equation
lr = @(B) LogErrNorm/segmodel(B);
% log norm fit
fit = @(B) lr(B)*segmodel(B);
```

```

% find best B which minimizes MSE
BreakPoint = round(fminbnd(@(B) sum((LogErrNorm - ...
    fit(B)).^2),0,x(end)/2));

koeff = lr(BreakPoint);

% apply the estimated slope compansation
ASC = LogErrNorm+2*abs(koeff(end))*x;

```

Reconstruct the denoised \hat{y} signal from \hat{a} coefficients.

```

[~,kmin] = min(ASC);

J = setdiff(1:n,1:kmin);
a = C;
a(IX(J)) = 0;

haty = lifilwt(a,LS,j,L);

```

References

- H. Abdi and L. J. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459, 2010.
- V. Alekseyev, S. Babichenko, and I. Sobolev. Control and signal processing system of hyperspectral FLS LiDAR. In *Electronics Conference, 2008. BEC 2008. 11th International Biennial Baltic*, pages 349–352, 2008. doi: 10.1109/BEC.2008.4657552.
- K. M. Almhdi, P. Valigi, V. Gulbinas, R. Westphal, and R. Reuter. Classification with artificial neural networks and support vector machines: application to oil fluorescence spectra. *EARSeL eProceedings*, 6(2):115–129, 2007.
- A. Ameodo, E. Bacry, P. Graves, and J. Muzyl. Characterizing long-range correlations in DNA sequences from wavelet analysis. *Physical Review Letters*, 1995.
- S. Babichenko. Laser remote sensing of the European marine environment: LIF technology and applications. In *Remote Sensing of the European Seas*, pages 189–204. Springer, 2008.
- S. Babichenko, L. Poryvkina, V. Arikese, S. Kaitala, and H. Kuosa. Remote sensing of phytoplankton using laser-induced fluorescence. *Remote sensing of environment*, 45(1):43–50, 1993.
- S. Babichenko, J. Lapimaa, L. Poryvkina, and V. Varlamov. On-line fluorescent techniques for diagnostics of water environment. In *Proc. SPIE*, volume 2503, pages 157–161, 1995. doi: 10.1117/12.221094. URL <http://dx.doi.org/10.1117/12.221094>.
- S. Babichenko, A. Dudelzak, and L. Poryvkina. Lidar application in airborne monitoring of eutrophication and chemical pollution. In *Proceedings of the VI International Conference Remote Sensing for Marine and Coastal Environments (Charleston, SC)*, volume 2, pages 377–383, 2000a.

- S. Babichenko, L. Poryvkina, and A. Dudelzak. Airborne Lidar Monitoring of Aquatic Eutrophication and Chemical Pollution. In *Remote sensing for marine and coastal environments*, volume 2, pages II–391–397, 2000b.
- S. Babichenko, A. Dudelzak, and L. Porovkina. Laser sensing technologies in studies of marine and coastal environment. In *Seventh International Conference on Remote Sensing for Marine and Coastal Environments*, page 8, 2002.
- S. Babichenko, A. Dudelzak, and L. Poryvkina. Laser remote sensing of coastal and terrestrial pollution by FLS-Lidar. *EARSeL eProceedings*, 3(1):1–7, 2004.
- S. Babichenko, A. Dudelzak, J. Lapimaa, A. Lisin, L. Poryvkina, and A. Vorobiev. Locating water pollution and shore discharges in coastal zone and inland waters with FLS lidar. *EARSeL eProceedings*, 5(1):32–41, 2006.
- S. Babichenko, V. Alekseyev, J. Lapimaa, A. Lisin, L. Poryvkina, S. Shchemelyov, I. Sobolev, and L. Vint. Airborne surveillance of water basins with hyperspectral FLS-LiDAR. In *Remote Sensing*. International Society for Optics and Photonics, 2010.
- S. Babichenko, L. Poryvkina, I. Sobolev, and L. Vint. *Sensing Device For In-Situ Non-Contact Detection Of Photosynthetic Microalgae In Water*. Number U201400015. The Estonian Patent Office, 03 2014.
- J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(2):354–379, April 2012. ISSN 1939-1404. doi: 10.1109/JSTARS.2012.2194696.
- C. G. E. Boender and A. R. Kan. Bayesian stopping rules for multistart global optimization methods. *Mathematical Programming*, 37(1):59–80, 1987.
- A. P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.
- J. N. Bradley, C. M. Brislawn, and T. Hopper. FBI wavelet/scalar quantization standard for gray-scale fingerprint image compression. In *Optical Engineering and Photonics in Aerospace Sensing*, pages 293–304. International Society for Optics and Photonics, 1993.

- R. P. Brent. *Algorithms for minimization without derivatives*. Courier Dover Publications, 2013.
- C. Brown, R. Nelson, M. Fingas, and J. Mullin. Airborne laser fluorosensing: overflights during lift operations of a sunken oil barge. In *Proceedings of the fourth thematic conference on remote sensing for marine and coastal environments. ERIM, Seattle*, volume 1, 1997.
- L. M. Bruce, C. H. Koger, and J. Li. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE T. Geoscience and Remote Sensing*, 40(10):2331–2338, 2002. URL <http://dx.doi.org/10.1109/TGRS.2002.804721>.
- T. T. Cai and L. Wang. Orthogonal matching pursuit for sparse signal recovery with noise. *Information Theory, IEEE Transactions on*, 57(7):4680–4688, 2011.
- E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.
- E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.
- T. N. Carlson and D. A. Ripley. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, 62(3):241–252, 1997.
- G. Cecchi, L. Pantani, P. Mazzinghi, and A. Barbaro. Fluorescence-Lidar Remote-Sensing of the Environment: Laboratory Experiments for the Characterization of Oil Spills and Vegetation. In W. Waidelich, editor, *Laser/Optoelektronik in der Technik / Laser/Optoelectronics in Engineering*, pages 652–655. Springer Berlin Heidelberg, 1986. ISBN 978-3-540-16017-5. doi: 10.1007/978-3-642-82638-2_121. URL http://dx.doi.org/10.1007/978-3-642-82638-2_121.
- C.-h. Chen. *Signal and image processing for remote sensing*. CRC press, 2012.
- A. Cohen. Biorthogonal wavelets. *Wavelets: A Tutorial in Theory and Applications*, 2:123–152, 1992.
- Y. Cohen, D. G. Blumberg, and S. R. Rotman. Sub-pixel target detection using local spatial information in hyperspectral images. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 8186 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Oct. 2011. doi: 10.1117/12.897431.

- A. Colin Cameron and F. A. Windmeijer. An R -squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77(2):329–342, 1997.
- I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7):909–996, 1988.
- I. Daubechies. Orthonormal bases of compactly supported wavelets. *SIAM Journal on Mathematical Analysis*, 24(2):499–519, 1993.
- I. Daubechies et al. *Ten lectures on wavelets*, volume 61. SIAM, 1992.
- T. A. Dolenko, V. V. Fadeev, I. V. Gerdova, S. A. Dolenko, and R. Reuter. Fluorescence diagnostics of oil pollution in coastal marine waters by use of artificial neural networks. *Applied optics*, 41(24):5155–5166, 2002.
- D. L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.
- D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal Of The American Statistical Association*, pages 1200–1224, 1995.
- A. E. Dudeizak, S. M. Babichenko, L. V. Poryvkina, and K. J. Saar. Total luminescent spectroscopy for remote laser diagnostics of natural water conditions. *Applied optics*, 30(4):453–458, 1991.
- A. E. Dudelzak, S. M. Babichenko, L. V. Poryvkina, and K. J. Saar. Total luminescent spectroscopy for remote laser diagnostics of natural water conditions. *Appl. Opt.*, 30(4):453–458, Feb 1991. doi: 10.1364/AO.30.000453. URL <http://ao.osa.org/abstract.cfm?URI=ao-30-4-453>.
- G. Fernandez, S. Periaswamy, and W. Sweldens. LIFTPACK: A software package for wavelet transforms using lifting. In *SPIE's 1996 International Symposium on Optical Science, Engineering, and Instrumentation*, pages 396–408. International Society for Optics and Photonics, 1996.
- M. Fingas and C. Brown. Oil Spill Remote Sensing: A Forensic Approach. In *Oil spill environmental forensics: fingerprinting and source identification*. Academic Press, 2010.
- M. Fingas and C. Brown. Review of oil spill remote sensing. *Marine Pollution Bulletin*, (0):–, 2014. ISSN 0025-326X. doi: 10.1016/j.marpolbul.2014.03.059. URL <http://www.sciencedirect.com/science/article/pii/S0025326X14002021>.
- M. Fingas and C. E. Brown. Oil spill remote sensing: a review. *Oil spill science and technology*, pages 111–169, 2011.

- A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data. *Image Processing, IEEE Transactions on*, 17(10):1737–1754, Oct 2008. ISSN 1057-7149. doi: 10.1109/TIP.2008.2001399.
- K. S. Fu et al. *Syntactic pattern recognition and applications*, volume 4. Prentice-Hall New York, 1982.
- GBI Research. *Offshore Drilling Industry to 2016 - Rapidly Rising Demand for Hydrocarbons Expected to Boost Offshore Drilling in Ultra-Deepwater and Harsh-weather Environments*. GBI Research, dec 2011.
- J. C. Goswami and A. K. Chan. *Fundamentals of wavelets: theory, algorithms, and applications*, volume 233. John Wiley & Sons, 2011.
- A. Grossmann and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*, 15(4):723–736, 1984.
- S. W. Hasinoff, F. Durand, and W. T. Freeman. Noise-optimal capture for high dynamic range photography. In *CVPR*, pages 553–560. IEEE, 2010. URL <http://dx.doi.org/10.1109/CVPR.2010.5540167>.
- A. Hausmann, F. Duschek, T. Fischbach, C. Pargmann, V. Alekseyev, L. Poryvkina, I. Sobolev, S. Babichenko, and J. Handke. Standoff detection: classification of biological aerosols using laser-induced fluorescence (LIF) technique. In *Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XV*. SPIE DSS, 2014.
- D. Heinz and C.-I. Chang. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 39(3):529–545, Mar 2001. ISSN 0196-2892. doi: 10.1109/36.911111.
- T. Hengstermann and R. Reuter. Lidar fluorosensing of mineral oil spills on the sea surface. *Appl. Opt.*, 29(22):3218–3227, Aug 1990. doi: 10.1364/AO.29.003218. URL <http://ao.osa.org/abstract.cfm?URI=ao-29-22-3218>.
- T. Hengstermann and R. Reuter. Optical Spectra of Fluorescent Pollutants in Aquatic Environments. *Oils*, 1, 1992a.
- T. Hengstermann and R. Reuter. Laser remote sensing of pollution of the sea: a quantitative approach. *European Association of Remote Sensing Laboratories, Advances in Remote Sensing*, 1(2):52–60, 1992b.

- T. Hengstermann, D. Loquay, R. Reuter, H. Wang, and R. Willkomm. A Laser Fluorosensor for Airborne Measured of Maritime Pollution and of Hydrographie Parameters. *EARSeL Advances in Remote Sensing*, 1(2): 85, 1992.
- F. E. Hoge and R. N. Swift. Oil film thickness measurement using airborne laser-induced water Raman backscatter. *Appl. Opt.*, 19(19):3269–3281, Oct 1980. doi: 10.1364/AO.19.003269. URL <http://ao.osa.org/abstract.cfm?URI=ao-19-19-3269>.
- P.-H. Hsu. Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS journal of photogrammetry and remote sensing*, 62(2):78–92, 2007.
- International Energy Agency. Supply shock from North American oil rippling through global markets, 2013. URL <http://www.iea.org/newsroomandevents/pressreleases/2013/may/name,38080,en.html>. Online. Accessed 01-May-2014.
- InView Technology Corporation. Compressive Sensing, 2014. URL <http://inviewcorp.com/technology/compressive-sensing>.
- J. R. Janesick. *Scientific charge-coupled devices*, volume 117. SPIE press Bellingham, Washington, 2001.
- M. N. Jha, J. Levy, and Y. Gao. Advances in remote sensing for oil spill disaster management: state-of-the-art sensors technology for oil spill surveillance. *Sensors*, 8(1):236–255, 2008.
- N. L. Johnson, S. Kotz, and N. Balakrishnan. Continuous univariate distributions. *Wiley Series in Probability and Statistics*, 1994.
- R. Karpicz, A. Dementjev, Z. Kuprionis, S. Pakalnis, R. Westphal, R. Reuter, and V. Gulbinas. Oil spill fluorosensing lidar for inclined onshore or shipboard operation. *Applied optics*, 45(25):6620–6625, 2006.
- J. Kurek, J. L. Kirk, D. C. Muir, X. Wang, M. S. Evans, and J. P. Smol. Legacy of a half century of Athabasca oil sands development recorded by lake ecosystems. *Proceedings of the National Academy of Sciences*, 110(5):1761–1766, 2013.
- J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1998.
- J. Lakowicz, editor. *Quenching of Fluorescence*. Springer US, 2006. ISBN 978-0-387-31278-1. URL http://dx.doi.org/10.1007/978-0-387-46312-4_8.

- D. A. Landgrebe. *Signal theory methods in multispectral remote sensing*, volume 29. John Wiley & Sons, 2005.
- K. Lau and H. Weng. Climate signal detection using wavelet transform: How to make a time series sing. *Bulletin of the American Meteorological Society*, 76(12):2391–2402, 1995.
- C. L. Lawson and R. J. Hanson. *Solving least squares problems*, volume 161. SIAM, 1974.
- H.-N. Lee. Introduction to Compressed Sensing. *Lecture Note, Spr*, 2011.
- M. Lennon, S. Babichenko, N. Thomas, V. Mariette, G. Mercier, and A. Lisin. Detection and mapping of oil slicks in the sea by combined use of hyperspectral imagery and laser induced fluorescence. *EARSeL eProceedings*, 5(1):120–128, 2006.
- C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman. Automatic Estimation and Removal of Noise from a Single Image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):299–314, 2008. URL <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2007.1176>.
- F. Luisier, T. Blu, and M. Unser. A New SURE Approach to Image Denoising: Interscale Orthonormal Wavelet Thresholding. *Image Processing, IEEE Transactions on*, 16(3):593–606, March 2007. ISSN 1057-7149. doi: 10.1109/TIP.2007.891064.
- F. Luisier, T. Blu, and M. Unser. Image Denoising in Mixed Poisson-Gaussian Noise. *IEEE Transactions on Image Processing*, 20(3):696–708, 2011. URL <http://dx.doi.org/10.1109/TIP.2010.2073477>.
- M. Makitalo and A. Foi. Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise. *Image Processing, IEEE Transactions on*, 22(1):91–103, 2013.
- S. G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Transactions of the American Mathematical Society*, 315(1):69–87, 1989a.
- S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693, 1989b.
- D. Manolakis, C. Siracusa, and G. Shaw. Hyperspectral subpixel target detection using the linear mixing model. *Geoscience and Remote Sensing, IEEE Transactions on*, 39(7):1392–1409, Jul 2001. ISSN 0196-2892. doi: 10.1109/36.934072.

- S. Matsuura, S. Umebayashi, C. Okuyama, and K. Oba. Characteristics of the Newly Developed MCP and Its Assembly. *Nuclear Science, IEEE Transactions on*, 32(1):350–354, Feb 1985. ISSN 0018-9499. doi: 10.1109/TNS.1985.4336854.
- Y. Meyer. Wavelets-algorithms and applications. *Wavelets-Algorithms and applications Society for Industrial and Applied Mathematics Translation.*, 142 p., 1, 1993.
- F. Mosteller and J. W. Tukey. Data analysis and regression: a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*, 1977.
- Ohmsett. Heavy Oil Detection with Laser Fluorometers. *The Ohmsett Gazette*, pages 6–7, 2005.
- R. A. O’Neil, L. Buja-Bijunas, and D. M. Rayner. Field performance of a laser fluorosensor for the detection of oil spills. *Appl. Opt.*, 19(6):863–870, Mar 1980. doi: 10.1364/AO.19.000863. URL <http://ao.osa.org/abstract.cfm?URI=ao-19-6-863>.
- L. Pantani, G. Cecchi, M. Bazzani, and V. Raimondi. Remote sensing of marine environment with the high spectral resolution fluorosensor FLIDAR3. In *Satellite Remote Sensing II*, pages 56–64. International Society for Optics and Photonics, 1995.
- T. Pavlidis. *Structural pattern recognition*. 1977.
- R. Polikar. John Wiley & Sons, Inc., 2006. ISBN 9780471740360. doi: 10.1002/9780471740360.ebs0904. URL <http://dx.doi.org/10.1002/9780471740360.ebs0904>.
- M. Radovanović, A. Nanopoulos, and M. Ivanović. Hubs in space: Popular nearest neighbors in high-dimensional data. *The Journal of Machine Learning Research*, 9999:2487–2531, 2010.
- R. Reuter, H. Wang, R. Willkomm, K. Loquay, T. Hengstermann, and A. Braun. A Laser Fluorosensor for Maritime Surveillance: Measurement of Oil Spills. *EARSeL Advances in Remote Sensing*, 3(3), 1995.
- J. A. Richards and X. Jia. *Remote sensing digital image analysis*, volume 3. Springer, 1999.
- N. Robbe. *Airborne Oil Spill Remote Sensing: Modelling, Analysis and Fusion of Multi-spectral Data*. PhD thesis, Hamburg University, 2005.
- N. Saito. *Local Feature Extraction and Its Applications Using a Library of Bases*. PhD thesis, Yale University, Dec. 1994.

- D. Shubina, E. Fedoseeva, O. Gorshkova, S. Patsaeva, V. Terekhova, M. Timofeev, and V. Yuzhakov. The “blue shift” of emission maximum and the fluorescence quantum yield as quantitative spectral characteristics of dissolved humic substances. *EARSeL eProceedings*, 9(1):13–21, 2010.
- I. Sobolev. FLS LiDAR systems and LIF spectra analysis. In *Info-ja kommunikatsioonitehnoloogia doktorikooli IV aastakonverents*. IKT, 2010.
- I. Sobolev and S. Babichenko. Analysis of the performances of hyperspectral lidar for water pollution diagnostics. *EARSeL eProceedings*, 12(2):113–123, 2013a.
- I. Sobolev and S. Babichenko. Application of the wavelet transform for feature extraction in the analysis of hyperspectral laser-induced fluorescence data. *International Journal of Remote Sensing*, 34(20):7218–7235, 2013b.
- L. Song. Kalamazoo River Spill Yields Record Fine, 2012. Lisa Song, a reporter for InsideClimate News, interviewed by Bruce Gellerman. Online. Accessed 01-May-2014.
- H. A. Sturges. The choice of a class interval. *Journal of the American Statistical Association*, 21(153):65–66, 1926.
- M. Sun, M. L. Scheuer, and R. J. ScLabassi. Decomposition of biomedical signals for enhancement of their time–frequency distributions. *Journal of the Franklin Institute*, 337(4):453–467, 2000.
- W. Sweldens. Lifting scheme: a new philosophy in biorthogonal wavelet constructions. In *SPIE’s 1995 International Symposium on Optical Science, Engineering, and Instrumentation*, pages 68–79. International Society for Optics and Photonics, 1995.
- W. Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM Journal on Mathematical Analysis*, 29(2):511–546, 1998.
- W. Sweldens and P. Schröder. Building your own wavelets at home. In *Wavelets in the Geosciences*, pages 72–107. Springer, 2000.
- X.-z. Xu, Z.-t. Li, and L.-j. Xue. Analysis and processing of CCD noise. *Infrared and Laser Engineering*, 33(4):343–346, 2004.
- Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang. Blind Spectral Unmixing Based on Sparse Nonnegative Matrix Factorization. *Image Processing, IEEE Transactions on*, 20(4):1112–1125, April 2011. ISSN 1057-7149. doi: 10.1109/TIP.2010.2081678.

- O. Zielinski, R. Andrews, J. Göbel, M. Hanslik, T. Hunsänger, and R. Reuter. Operational airborne hydrographic laser fluorosensing. In *Proceed. EARSeL-SIG-Workshop LIDAR, Dresden, Germany*, volume 8, 2000.
- O. Zielinski, T. Hengstermann, and N. Robbe. Detection of oil spills by airborne sensors. In *Marine Surface Films*, pages 255–271. Springer, 2006.

Acknowledgements

This thesis is based on my work done during the last seven years in the field of laser and spectroscopy technologies for environmental monitoring. I thank my supervisors, co-founder of *Laser Diagnostic Instruments AS, LDI Innovation* President & CEO, initiator and co-developer of key product concepts/designs Dr. Sergey Babichenko and Tallinn University of Technology Associate Professor Eduard Petlenkov for much patience and guidance. I am very grateful to my former colleagues in Laser Diagnostic Instruments AS for cooperation in field and laboratory works.

Last but not least, I thank my family for strong support, inspiration and belief in me. My sincerest gratitude goes to my wife for giving me energy and confidence in achieving my goals. This thesis would not have been possible without their support and patience.

My research activity was also supported by the Estonian Information Technology Foundation and Estonian Doctoral School in Information and Communication Technology.

Kokkuvõte

Hüperspektraalse andmete analüüs ja tõlgendamine kaugseires *laser-induced fluorescence* meetodi põhjal

Käesolev väitekiri on pühendatud HLIF (Hyperspectral Laser Induced Fluorescence) andmetööluse asetamisele süstemaatilisele alusele. Eesmärkideks on HLIF tehnoloogia analüütiliste võimaluste parandamine ja automaatse otsustamise tehnoloogilise aluse loomine vee- ja maapinna reostuse (Oil Pollution Diagnostics – OPD) detekteerimiseks ning tuvastamiseks. Uuringutes osutatakse märkimisväärsele OPD kvaliteedi paranemisele, kui rakendada sobival moel selleks spetsiaalselt kavandatud signaalitöötlemise tehnoloogiaid.

HLIF andmete esialgne interpreteerimine on uuringute ettevalmistavaks sammuks. Põhilised oletused HLIF signaalide kohta esitatakse LMM (Linear Mixture Model) kujul. Kavandatakse maa- ja veeobjektide referents andmebaasi struktuur ja süstematiseerimine, grupeerituna loomuliku taustfluorestsentsi ja huvipakkuvate signaalide järgi.

Kaugseire LIF meetodi puhul on võtmeküsimuseks saastesignaali eraldamine loomulikust taustsignaalist. Diskrimineerimise võimekust piiravad mõõtemääramatus ja signaali infosaldus. Käesolevas dissertatsioonis on põhiliselt teemaks signaali andmetöötlemisel müratasandus ja detaili väljatõmbamine. Lihtsa HLIF OPD korral osutuvadki oluliseks müratasandus ja wavelet'il põhinev struktuurse detaili väljatõmbamine.

Käesoleva dissertatsiooni tulemused on juba rakendatud ja integreeritakse edaspidi HLIF lidari ekspluatatsiooniplatvormi. Hulk HLIF OPD reaalseid, laboratoorseid ja simulatsioon andmehulki on läbinud võrdleva analüüsi pakutud töötlusalgoritmide efektiivsuse hindamiseks. Dissertatsiooni tulemusena on välja töötatud lahendus HLIF OPD jaoks kombineeritud vesi/maapind keskkonnas.

Abstract

Hyperspectral data processing and interpretation in remote sensing based on *laser-induced fluorescence* method

This thesis is devoted to establish a systematic approach for Hyperspectral Laser Induced Fluorescence (HLIF) data processing. The goals are to enhance the analytical capabilities of HLIF technique and provide the base platform for automated decision making in land and water Oil Pollution Diagnostics (OPD) and identification. It is shown in the research that proper application of specifically designed signal processing techniques can significantly improve the quality of OPD.

The interpretation of HLIF data is considered as a preliminary step of research. The general assumptions about the HLIF signals are expressed in the specific Linear Mixture Model (LMM) approach. The reference library structure and fluorescence objects systematization are proposed for land and water targets by dividing the objects into groups of natural fluorescence backgrounds and signals of interest.

The pollution discrimination from natural fluorescence backgrounds is one of the key issues of remote diagnostics with LIF method. The discrimination capabilities are always constrained by the measurement uncertainty and information content. The data analysis based on denoising and feature extraction is used as a main tool of the subject research. The developed techniques of Adaptive Slope Compensation (ASC) denoising and wavelet based structural feature extraction are approved to be essential in robust HLIF OPD.

The results of this thesis are implemented and will be further integrated in HLIF LIDAR operational platform. The number of real-world, laboratory and simulation OPD case studies are evaluated to test the HLIF OPD capabilities with proposed data processing methods. As the result of dedicated developments the solution for HLIF OPD in mixed water/land environment has been elaborated.

Elulookirjeldus

1. Isikuandmed

Ees- ja perekonnanimi Innokenti Sobolev
Sünniaeg ja -koht 23.05.1985, Tallinn, Eesti
Kodakondsus Eesti

2. Kontaktandmed

Aadress Äiatari tee 3, Maardu, 74117
Telefon +372 56488834
E-posti aadress is@ldi-innovation.eu

3. Hariduskäik

Õppeasutus	Lõpetamise aeg	Haridus
Tallinna Tõnismäe Reaalkool	2000	Kesk
Tallinna Polütehnikum	2004	Kesk-eri
Tallinna Tehnikaülikool, Mehaanikateaduskond, Mehhatroonikainstituut	2007	B. Sc.
Tallinna Tehnikaülikool, Mehaanikateaduskond, Mehhatroonikainstituut	2009	M. Sc.

4. Keelteoskus

Keel	Tase
Eesti	kesktase
Inglise	kõrgtase
Vene	emakeel

5. Teenistuskäik

Töötamise aeg	Tööandja nimetus	Ametikoht
05/2003 - 01/2004	AS Avesta Grupp	Elektrik
01/2006 - 05/2007	Maretek OÜ	Operaator
07/2007 - 06/2011	Laser Diagnostic Instruments AS	Elektroonika-insener konstruktor
06/2011 - 09/2013	Laser Diagnostic Instruments AS	Teadur, elektroonikainsener
10/2013 - ...	Tallinna Tehnikaülikool, Automaatikainstituut	Insener
10/2013 - ...	LDI Innovation	Teadur, elektroonikainsener

6. Teadustegevus

Teaduste projektide, ajakirja- ja konverentisartiklite loetelu on toodud ingliskeelse CV juures.

7. Kaitstud lõputööd

Tööstushoone ventilatsiooniautomaatika õppestend, Tallinna Polütehnikum, 2004.

Mobiilse roboti positsioneerimine ja kaardistamine kasutades SICK laser skannerit, BSc, Tallinna Tehnikaülikool, 2007.

Wavelet põhine LIF spektri analüüs, MSc, Tallinna Tehnikaülikool, 2009.

8. Teadustöö põhisuunad

Laseri ja spektroskoopia tehnoloogiad. Andmete töötlus.

Curriculum Vitae

1. Personal data

Name Innokenti Sobolev
Date and place of birth 23.05.1985, Tallinn, Estonia
Citizenship Estonian

2. Contact information

Address Äiatari tee 3, Maardu, 74117
Phone +372 56488834
E-mail is@ldi-innovation.eu

3. Education

Institution	Graduation date	Education
Tallinn Secondary Science School	2000	Secondary
Tallinn Polytechnic School	2004	Vocational Secondary
Tallinn University of Technology, Faculty of Mechanical Engineering, Department of Mechatronics	2007	B. Sc.
Tallinn University of Technology, Faculty of Mechanical Engineering, Department of Mechatronics	2009	M. Sc.

4. Language competence/skills

Language	Level
Estonian	fluent
English	fluent
Russian	native

5. Professional employment

Period	Organization	Position
05/2003 - 01/2004	AS Avesta Grupp	Electrician
01/2006 - 05/2007	Maretek OÜ	Operator
07/2007 - 06/2011	Laser Diagnostic Instruments AS	Electronics design engineer
06/2011 - 09/2013	Laser Diagnostic Instruments AS	Researcher, engineer
10/2013 - ...	Tallinn University of Technology, Department of Computer Control	Other staff
10/2013 - ...	LDI Innovation	Researcher, engineer

6. Scientific projects

DEOSOM Detection and Evaluation of Oil Spills by Optical Methods, 2008-2011.

Contribution

LIDAR electronics prototyping, data acquisition in field experiments.

Baltic Way The potential of currents for environmental management of the Baltic Sea maritime industry, 2009-2011.

Contribution

Developing LIDAR data processing algorithms for trace oil detection in sea water.

FP7 DELICAT Demonstration of LIDAR based Clear Air Turbulence (CAT) detection, 2009-2014.

Contribution

Development of CAT detection algorithms.

FP7 BONAS BOMB factory detection by Networks of Advanced Sensors. A DIAL system for explosives detection, 2011-2014.

Contribution

Development of LIDAR controller and firmware for DIAL system.

Barents 2020 or HiNODS High North Oil Detection System. Development compact shipborne HLIF LIDAR for operation in high north environment, 2013-2015.

Contribution

Developing HLIF detector and data acquisition system.

EAS grant EU46064 Development of an Oil Spill Detection System based on information and laser remote sensing technology, 2013-2015.

Contribution

Shipboard HLIF LIDAR firmware development, integrated real-time spectral- and meta-data processing, development of on-line communication with Network Operation Center.

FP7 EDEN Short range detection of biological agents, 2013-2015.

Contribution

Development of LIDAR control electronics for HLIF system.

cHELIOS Compressed HypErspectral Laser Induced fluOrescence Sensing, IKTDK interdistsiplinaarne projekt, 2013-2014.

Contribution

Exploration of the possible operational enhancements of HLIF LIDAR instruments using the combination of discrete detection channels and compressed sensing techniques.

7. Publications

List includes 2 articles indexed by the Web of Science database (1.1), 5 peer-reviewed articles in other international journals and collections (1.2, 3.1), one Utility Model (Babichenko et al., 2014) and others.

V. Alekseyev, S. Babichenko, and I. Sobolev. Control and signal processing system of hyperspectral FLS LiDAR. In *Electronics Conference, 2008. BEC 2008. 11th International Biennial Baltic*, pages 349–352, 2008. doi: 10.1109/BEC.2008.4657552.

S. Babichenko, V. Alekseyev, J. Lapimaa, A. Lisin, L. Poryvkina, S. Shchemelyov, I. Sobolev, and L. Vint. Airborne surveillance of water basins with hyperspectral FLS-LiDAR. In *Remote Sensing*. International Society for Optics and Photonics, 2010.

S. Babichenko, L. Poryvkina, M. Pärno, S. Shchemelyov, and I. Sobolev. LIF lidar as early warning tool for oil pipeline integrity monitoring. In *5th EARSeL Workshop on Remote Sensing of the Coastal Zone*. EARSeL, 2011.

S. Babichenko, L. Poryvkina, I. Sobolev, and L. Vint. *Sensing Device For In-Situ Non-Contact Detection Of Photosynthetic Microalgae In Water*. Number U201400015. The Estonian Patent Office, 03 2014.

- A. Hausmann, F. Duschek, T. Fischbach, C. Pargmann, V. Aleksejev, L. Poryvkina, I. Sobolev, S. Babichenko, and J. Handke. Standoff detection: classification of biological aerosols using laser-induced fluorescence (LIF) technique. In *Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XV*. SPIE DSS, 2014.
- V. Mitev, S. Babichenko, J. Bennes, R. Borelli, A. Dolfi-Bouteyre, L. Fiorani, L. Hespel, T. Huet, A. Palucci, M. Pistilli, A. Puiu, O. Rebane, and I. Sobolev. Mid-IR DIAL for high-resolution mapping of explosive precursors. In *SPIE Remote Sensing*. International Society for Optics and Photonics, 2013.
- L. Poryvkina, N. Tsvetkova, and I. Sobolev. Evaluation of apple juice quality using spectral fluorescence signatures. *Food chemistry*, 152: 573–577, 2014.
- I. Sobolev. FLS LiDAR systems and LIF spectra analysis. In *Info-ja kommunikatsioonitehnoloogia doktorikooli IV aastakonverents*. IKT, 2010.
- I. Sobolev. Hyperspectral LIF data analysis technique. In *IKTDK V aastakonverents*. IKT, 2011.
- I. Sobolev and S. Babichenko. Analysis of the performances of hyperspectral lidar for water pollution diagnostics. *EARSel eProceedings*, 12(2):113–123, 2013a.
- I. Sobolev and S. Babichenko. Application of the wavelet transform for feature extraction in the analysis of hyperspectral laser-induced fluorescence data. *International Journal of Remote Sensing*, 34(20):7218–7235, 2013b.

8. Defended theses

Ventilation system testbench, Tallinn Polytechnic School, 2004.

Mobile robot georeferencing, BSc, Tallinn University of Technology, 2007.

Wavelet Based Analysis of LIF Spectra, MSc, Tallinn University of Technology, 2009.

9. Main areas of scientific work/Current research topics

Laser and spectroscopy technologies. Data processing. Compressed sensing.

Publications

Publication 1

S. Babichenko, V. Alekseyev, J. Lapimaa, A. Lisin, L. Poryvkina, S. Shchemelyov, I. Sobolev, and L. Vint. Airborne surveillance of water basins with hyperspectral FLS-LiDAR. In Remote Sensing. International Society for Optics and Photonics, 2010.

Airborne surveillance of water basins with hyperspectral FLS[®]-LiDAR

S.Babichenko*, V.Alekseyev, J.Lapimaa, A.Lisin, L.Poryvkina, S.Shchemelyov, I.Sobolev, L.Vint
Laser Diagnostic Instruments AS, 113a Kadaka tee 12915 Tallinn, Estonia

ABSTRACT

The airborne FLS[®]-Lidars are based on the method of Laser Induced Fluorescence (LIF) and aimed at the analytical remote sensing of water objects. Scanning the laser beam across the flight trajectory and recording the comprehensive LIF spectrum with hyperspectral detector per every laser pulse provide detail maps of spectral properties of the water basins. A multi-tier model for integrated environmental assessment is applied for further analysis of this information to combine the benefits of “big-picture” capability of remote sensing techniques and GIS solutions with localized on-the-ground environmental data gathering.

In this concept far looking satellite and airborne systems provide the highest tier information. The airborne data acquisition with FLS[®]-Lidar is considered as the middle tier characterized by vast amount of LIF data with high spatial (less than 10 m) and spectral (less than 5 nm in UV/VIS spectral ranges) resolution. The lower tier is anchored with the geographical locations of important findings detected at the middle tier. Taken water samples are analyzed with fast-screening technology of Spectral Fluorescence Signatures (SFS) giving more analytical qualitative and quantitative results. And the base tier includes detail laboratory analysis of characteristic samples selected at the lower tier. Precisely geo-referenced LIF data of hyperspectral FLS-Lidar anchored to and calibrated by the ground SFS data allows detection of pollution incidents and mapping of environmental trends over vast water systems like coastal zone, lakes and rivers.

Keywords: lidar, airborne, fluorescence, water quality, oil pollution

1. INTRODUCTION

The environmental concern on natural water quality is growing, and the recent catastrophic oil spill in the Gulf of Mexico with its devastating effect on marine and coastal ecosystem clearly shows the need for reliable and flexible operational techniques to provide effective response to the consequences of disaster. The remote sensing and especially the satellite imaging was a major information tool in this disaster. It allowed to get the total picture of the oil spill area and to observe its development and spreading. Additionally to such large scale picture the airborne imaging systems were used to detail the distribution of oil inside the spill patches and to monitor the adjacent water areas. Nevertheless there was an obvious lack of detection capability of submerged oil appeared after spill weathering and use of dispersants. Undetectable movements of oil floating under the water surface hampered efficiency of skimming operations and made unpredictable oil pollution on shore¹. At the same time there is a technique capable of doing such detection, and it is the active remote sensing based on the method of Laser Induced Fluorescence (LIF).

The technique of laser remote sensing and its application for monitoring of water environment including detection of oil pollution has been well explored in a course of the last decades. Among other lidars, ones based on LIF method constitute a cluster of active remote sensing systems capable to deliver the analytical information on the studied water objects. The theory behind and the first results of airborne LIF Lidar use for oil film detection were introduced in 1980², and the airborne detection of oil discharges in the sea were demonstrated in 1986³, the application of multiple sensing wavelengths in LIF lidars was developed in 1991⁴. Since that time the airborne LIF techniques were significantly enhanced and tested in numerous studies^{5,6,7}. It has been proved that combined use of LIF Lidars with other on-board sensors increases the reliability of airborne oil detection^{8,9}.

The first airborne surveillance with hyperspectral LIF Lidar (FLS[®]-A) was reported in 1995¹⁰. Hyperspectral refers to a lidar operated with continuous LIF spectrum. Its distinct feature from other LIF Lidars (often referenced to as Laser Fluorosensor) is that there are no predefined spectral ranges to detect echo-signal, and a comprehensive shape of water emission spectrum is recorded at every laser shot for consequent analysis. The FLS-series is a family of hyperspectral LIF Lidars developed by Laser Diagnostic Instruments AS (LDI). This paper describes airborne FLS[®]- Lidars introduced for operational use in 2008¹¹ and their applications for pollution detection and water quality assessment.

*sergey.babichenko@ldi.ee; phone +372 656-7337; fax +372 656-7305; www.ldi.ee

2. INSTRUMENTS AND METHODS

2.1 Hyperspectral Fluorescent Lidar System - FLS

Since early beginning the concept of modular lidar was used in various models of FLS-series lidars. The basic FLS-modules are: the laser source with wavelength transformation unit, beam shaper, scanner, receiving telescope coupled with hyperspectral detector, lidar controller, data storage and processing unit, and peripheral devices. The core part of the FLS-Lidar is its expert software system capable of in-flight analysis of LIF spectra.

The main laser is rigidized excimer laser, which is used as Ultra Violet (UV) - sensing source and to pump the wavelength transformation unit. The later is typically either dye-laser or stimulated Raman scattering cell. It is used to sense the object at the most effective selective excitation or to apply multiwavelength excitation (see 2.2 below) of remote water body. In airborne applications such unit has pre-set wavelengths with fast switching from one to another. The beam shaper has double function: it is used to operate in eye-safe mode and to vary the footprint of the laser beam on the ground. The scanning system provides the scan of the laser beam and the total field of view of the telescope across the flight path. The telescope is coupled with flat-field polychromator and gated linear multichannel detector (500 channels) constituting hyper-spectral receiving system. The lidar controller with LAN capabilities serves to control the operation of lidar, data flow, and communication with peripheral and external devices. The basic configuration of FLS[®]-Lidar allows recording, processing, and storing the comprehensive LIF spectrum per every laser pulse with background signal subtraction.



Figure 1. King Air 90 aircraft (right) certified for carrying FLS-AM (Airborne Multi-wavelengths) lidar (left)

Table 1. The specification of airborne FLS-series Lidars.

FLS [®] -Lidar serial model Specification	FLS-AU	FLS-AM	FLS-A(e)	FLS-AE (in development)
Intended use	Water pollution	Water quality, eutrophication	Water pollution	Water quality, eutrophication
Sensing distance	50 - 300 m	100 – 500 m	100 – 500 m	100 – 500 m
Sampling rate per second	20	150	100	500
Beam scanning across the flight path /swath width	NO	YES / 1/5 of altitude	YES / 2/5 of altitude	YES / 2/5 of altitude
Spatial resolution (m) at: speed 200 km/h, altitude 200 m	3	3	6	2
Sensing wavelength, nm	308	308, 360, 466	308	308 (360/460)
Power consumption, kVA, 28 DC	1.2	6	1	< 6
Dimensions (L x W x H), mm	1569 x 614 x 795	1660 x 685 x 1004	1830 x 1345 x 660	1726 x 557 x 847
Full system weight, kg	160	375	180	350

The peripheral units are: GPS receiver, range-finder, and digital camera. Every recorded LIF spectrum is accompanied with GPS coordinates and metadata of flight parameters and lidar settings at the time of measurement. Range-finder information is used to calculate the position of the laser spot on the ground and to correct the intensity of LIF spectra by corresponding flight altitude. Any finding treated by the software expert system as an alarm, e.g. when the threshold in the analysis setup is exceeded, generates the capture of the digital camera frame to document the underlying area with such alarm.

The external devices may be Inertial Measurement Unit (IMU), parallel data storage, visualization displays and communication channels. The complete set of spectral, meta- and external data is stored in the database, and can be processed in retrospective mode. The software expert system with *Real-Time* and post-processing functions is used to analyze LIF spectra based on the reference LIF spectral library of individual organic compounds.

2.2 The method of Spectral Fluorescence Signatures - SFS

The technique of FLS[®]-Lidar has been continuously developed since 1990 in parallel with the method of Spectral Fluorescence Signatures (SFS). The later is based on recording induced fluorescence of an object as a matrix of fluorescence intensity in coordinates of excitation/emission wavelength $J^{SFS}(\lambda_{ex}, \lambda_{em})$ ¹². Upon proper selection of excitation and emission spectral ranges such SFS matrix of intensities manifests the combination of spectral patterns specific to the basic matrix (e.g. water - J_0^f) – and number of certain analytes J_i^f additionally to it: $J^{SFS}(\lambda_{ex}, \lambda_{em}) = J_0^f(\lambda_{ex}, \lambda_{em}) + \sum_{i=1}^N J_i^f(\lambda_{ex}, \lambda_{em})$. The position and shape of the spectral pattern is treated as a “fingerprint” characterizing a substance, and the intensity serves as a measure of its concentration. Recognition of the individual spectral patterns allows substance identification in a complex mixture, and accurate deriving its fluorescence intensity $J_i^f(\lambda_{ex}, \lambda_{em})$ provides estimation of its quantity. Such analysis does not require chemical decomposition of the sample and therefore can be done in on-line mode or remotely.

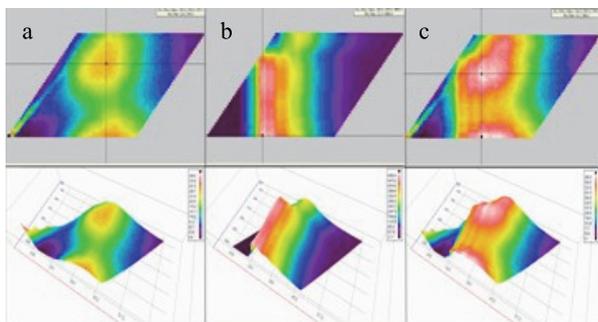


Figure 2. The SFS of seawater with DOM fluorescence (a), gasoil (b), and seawater polluted by this oil (c): 2D and 3D views.

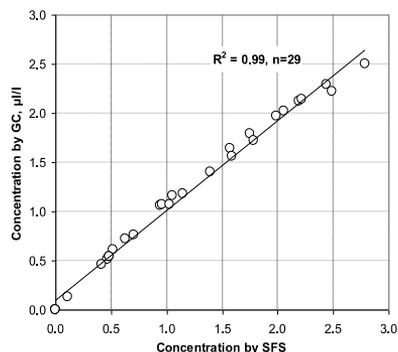


Figure 3. Correlation of SFS and GC data.

The SFS method is effectively realized with LDI analyzer Fluo-Imager^{® 13}. In a matter of minutes it measures the SFS of liquid sample, makes de-convolution of specific spectral patterns and identifies the substance based on SFS reference library. Figure 2 shows the examples of SFS for clean and polluted seawater. When the substance has been identified, the calibration curve is applied to derive its concentration in the sample. Such calibration curve is produced by parallel method, usually Gas Chromatography. Figure 3 demonstrates the correlation of SFS data by Fluo-Imager[®] and standard laboratory Gas Chromatography (GC) data for black oil.

2.3 Combined use of FLS[®]-Lidar and SFS analyzers

The FLS[®]-Lidar and SFS analyzer Fluo-Imager[®] are effectively used together. The SFS library of typical water pollutants and understanding their spectral features serve as a base for optimization of FLS[®]-Lidar operational parameters and algorithms of data processing to achieve higher sensitivity and selectivity in detection of organic compounds in a complex environment. The SFS method is also used to compile the reference library of LIF spectra and to calibrate the FLS[®]-lidar data. Indeed, the SFS library is easily convertible into LIF library used by expert system of FLS[®]-Lidar because the section of SFS along the emission axis at fixed wavelength of excitation produces the emission spectrum. At the same time the topography of SFS helps to foreknow the overlap of the LIF spectrum of the substance in question with other substances in water and adjust the expert system accordingly.

The combined use of large scale FLS data and local SFS data constitutes a valuable tool for multi-tier model of integrated environmental assessment introduced in¹⁴. In this model the highest tier includes far looking systems like satellites or airborne radars able to locate the areas of interest. The FLS[®]-Lidar is operated on the middle tier providing extensive set of geo-referenced analytical information. Real-time or post-processing analysis reveals the findings to direct the ground sampling activity. In the lower tier, the on-site analysis is carried out with SFS technique providing instant calibration and ground truthing of LIF data. And finally in the basic tier a detailed laboratory analysis of the selected samples is carried out. This model decreases substantially the number of samples through the tiers with an increase in information value of every sample at the last stage.

3. CAPABILITIES DEMONSTRATION

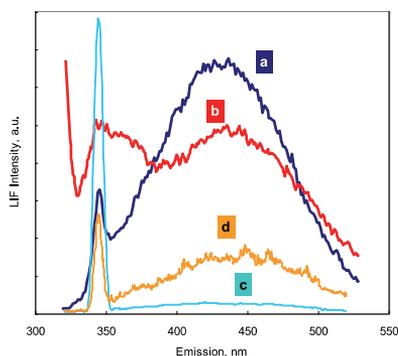


Figure 4. FLS spectra examples: (a), (b) - clean and polluted by ship motor oil Baltic Sea water (single laser shot, flight altitude 200 m); (c),(d) - clean and polluted by fuel oil Atlantic Ocean water (5 laser shots accumulated, flight altitude 500 m).

The sensitivity of airborne FLS[®]-Lidars has been proven down to *sub-ppm* range for oil products in water at the flight altitudes ranging from 100 to 500 m. Figure 4 shows typical LIF spectra of water recorded by FLS-AM (a,b) and FLS-AU (c,d) Lidars at the excitation 308 nm. The Rayleigh scattering is cut off by optical filter, and the peak at 344 nm corresponds to Raman scattering in water. The Signal to Noise Ratio (SNR) on the Raman line is higher than 20. As in lidar applications the Raman scattering signal is used as an internal spectral benchmark, it confirms the sensitivity level good enough to detect low-level water contamination.

In a course of FLS[®]-Lidars development the systems have been extensively tested in various applications on board of rotary and fixed wing platforms for detection of trace level oil pollution in open seawaters and coastal areas; finding and mapping of oil spills; location and identification of submersed oil; profiling of Dissolved Organic Matter (DOM) in lakes and rivers; diagnostics of agriculture and industrial run-offs, and eutrophication; monitoring of accumulated oil pollution in the port areas; general environmental assessment of natural water quality. Some of these applications are described in this section. Since 2008 the FLS[®]-Lidars are in operational use.

3.1 Monitoring of environmental pollution

Several research projects addressed water environment issues have been completed with shipboard FLS[®]-Lidars operating with multiwavelengths excitation in the Baltic and North Seas waters prior to the first airborne surveillance. These studies revealed the specificity of lidar operations in the waters rich by DOM, which is the case for inner seas, coastal zone, lakes and rivers. Additional specific requirements or laser remote sensing in such waters are caused by small depths of sensed water column, typically lower water transparency and higher amounts of suspended solids, and effects of bottom and shore underlying surfaces on LIF spectrum. All these factors increase the variability of LIF spectrum and complicate the detection of pollution comparing with open waters.

Figure 4 shows LIF spectrum of clean Atlantic water with expressed Raman line and almost negligible fluorescence band of DOM (see spectrum (c)). Thin oil film on the surface causes decrease of Raman line intensity and change of the spectral shape of LIF spectrum (compare (d) with (c) in Fig.4), and such changes are easy to detect. The Baltic Sea water contains noticeable and highly varied DOM (see Fig.4a) constituting variable spectral background for the fluorescence of oil in water. As a result the shape of LIF spectrum may change considerably. Fig 4b shows the spectrum of the Baltic Sea water with oil film on the surface and oil emulsion in water column: the Raman line is almost absent, the UV part of the spectrum is changed considerably due to oil emulsion in water, and the shape of the spectrum in DOM band is influenced by oil film on the surface.

The hyperspectral detection is required to analyze such spectra. The FLS expert system operates with comprehensive LIF spectrum, and is able to recognize minor variations in its shape providing reliable detection of spectral patterns even in case of overlapped spectral signals. Due to that the FLS®-Lidar is capable to detect oil pollution in water environments described above. This ability has been clearly demonstrated in numerous airborne surveillance missions with FLS®-Lidar¹⁵. An example of the results of a single surveillance mission is shown in Figure 5. The FLS-AU lidar installed on board of helicopter produced the maps of water transparency, DOM and oil pollution in the water area of one of Estonian harbors. The water transparency is expressed in arbitrary units (Raman line intensity corrected with the flight altitude), and the absolute values of DOM and oils are derived by using reference LIF library calibrated with SFS data.

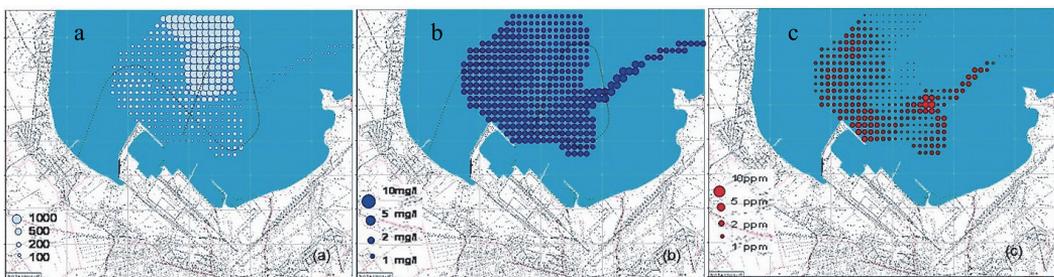


Figure 5. Surveillance of harbor area with FLS-AU lidar on board of helicopter in Nov 2004: (a) - water transparency, a.u.; (b) - DOM, mgC/l, (c) – oil pollution, ppm.

3.2 Location of submerged oil

The multiwavelength FLS®-Lidar was used in 2005 to demonstrate its capability in detection of heavy oil floating under the water surface¹⁶. The demo was carried out at the Oil and Hazardous Materials Simulated Environmental Test Tank (OHMSETT) facility in USA, famous for its testing of oil spill containment and recovery equipment. The operational performance of FLS®-Lidar was evaluated in simulated “real-life” situations at OHMSETT utilizing oils, sea water and wave generators. The lidar system was installed on a bridge above the water tank providing approximately 15 meters of light path distance to the water surface. Tests with mobile and immobile bridge and with mobile and immobile targets were executed (Fig.6).

The experiment included 29 tests of automated oil detection, characterization and submersion depth ranging of oil targets made with four types of oil, including two types of refined and two types of crude oil, and some natural and artificial targets. Samples of the two refined oils were provided prior to the tests, measured with Fluo-Imager® analyzer and added to FLS LIF library, while two crude oils were introduced as unknown targets during the tests. In the series of double-blind tests (unknown location and unknown type of oil) the response of FLS®-Lidar to these targets was tested in the presence of surface waves of various configurations, in day and night time at dry and drizzle weather conditions.



Figure 6. Installation of submerged targets with oil.

Processed results of blind tests confirmed that the FLS[®]-Lidar was able to detect submerged oil down to 3 m in the tank water comparable with clean coastal zone or lake water. The tests demonstrated the benefits of using **several excitation wavelengths**. The basic one (308 nm) was used to excite the fluorescence of submerged oil due to higher efficiency of excitation of oil fluorescence, and the second one (475 nm) served for locating and ranging the depth of floating oil due to lower attenuation in water (Fig.7). Multi-wavelength capability proved also essential for ability to detect and classify degraded and fresh oil targets made with heavy and light oils, especially with the extended reference database.

As a result, nearly-0% of false positives and very low rate of false negatives was observed even with provisional LIF library. Simultaneous presence of various oil types and unwanted surface oil pollution, combined with the task of simultaneous detection of highly fluorescent light oils and the low-fluorescing heavy oil targets challenged the dynamic range of the detector, and required on-the-spot optimizations of sensitivity and indication thresholds, to achieve correct detection. But it is not foreseen as a requirement in a typical spill assessment operation, when typically the dominating type of oil must be detected on or under the surface.

3.3 Detection of oil spills due to incidents

The capability of FLS[®]-Lidar in mapping of oil spill on water surface and estimation of its volume has been proved in the tests with controlled oil spills off the coasts of Brittany, France in 2004. In this project the down-looking FLS-AU lidar was used on board of Cessna 404 aircraft together with hyperspectral imager CASI-2³. The aircraft made a grid flight over the spilled area, and the FLS[®]-Lidar made horizontal profiling of oil film thickness along the flight path. Calibrated LIF data integrated over the flight grid produced the volumes of total spilled oil. The estimated thicknesses were quite accurate over thin oil films on water, while saturated at the critical oil film thickness (tens of microns, depending on the oil type) because of the saturation of LIF spectra intensity. The FLS[®]-Lidar has proved its capability in precise location of oil slick and mapping of its shape while estimating the minimal bound of the real volume of oil spilled, which should however be considered as reliable. It has been noticed that scanning the laser beam across the flight path would enhance the precision of FLS[®]-Lidar in oil spill detection. Such **scanning feature** has been introduced in the next models of FLS[®]-Lidar.

In winter – spring time 2006 the scanning multiwavelength FLS-AM lidar was in operation to detect spreading of oil after several pollution incidents and shipwreck in Estonian coastal waters. The measuring conditions were very specific: parts of the areas were covered by solid ice, and the navigation routs contained broken ice floating in water. The FLS-AM lidar installed on board of Estonian Boarder Guard helicopter was used for locating and mapping of real spills of oil, ship fuel and engine oils. The application of FLS[®]-Lidar has revealed specific requirements for oil detection in iced water. First, the water and ice produce non-distinguishing signals of Raman scattering. At the same time such discrimination is required because different quantification procedures should be applied for two these matrixes (see difference in LIF spectra in Fig.8). The feature of hyperspectral gated detection with FLS[®]-Lidar plays a key role in such classification procedure. When flying over the water, short gate pulse of the detector (typically bellow 500 ns) suppresses the ambient light reflected by water surface. As the reflectance of the ice is stronger by order of magnitude, the characteristic spectrum of ambient light registered by FLS[®]-Lidar between the laser pulses is noticeable, and it serves as a spectral indicator of the solid ice.

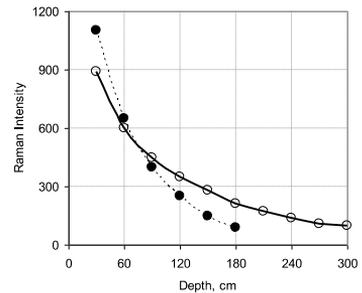


Figure 7. Attenuation of Raman line with depth at the sensing wavelength 308 nm (solid line) and 475 nm (dashed line).

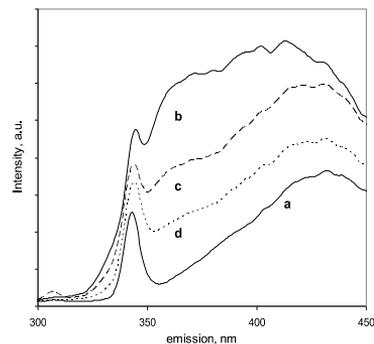


Figure 8. FLS LIF spectra of lubricant oil: (a) – on water surface; (b) – on solid ice surface; (c) – under ice layer of 2 cm; (d) – under the ice layer of 5 cm.

Such analysis procedure was used to detect spreading of the oil pollution from the open water to the coastal area and to map the location of engine oil seeps from the sunken cargo ship on the ice (Fig.9). The surveillance in the coastal area was carried out by flying along the grid ranging from open waters to the ice edge near the shore, and along the shoreline. The area was selected based on the predicted spreading of oil spill according to hydro-physical conditions in the region. The FLS-AM Lidar provided data acquisition from the 40-m corridor on the water surface with 3 m spatial resolution. In these particular cases the quantification of oil pollution was performed in retrospective mode in combination with SFS analysis. The samples of polluted water and ice were taken, and their SFS were measured to identify the pollutant. The LIF spectra were classified according to the matrix (ice or water), and then the corresponding calibration curves were applied to derive the amount of oil. Figure 9b shows the appearance of the oil pollution in the coastal area after spill in the open waters of the Gulf of Finland. The oil concentrations up to 30 mg/l were detected in the iced water along the coast line. Figure 9c shows scattered patches of engine fuel in the total amount of more than 3 tons in the area of the shipwreck.

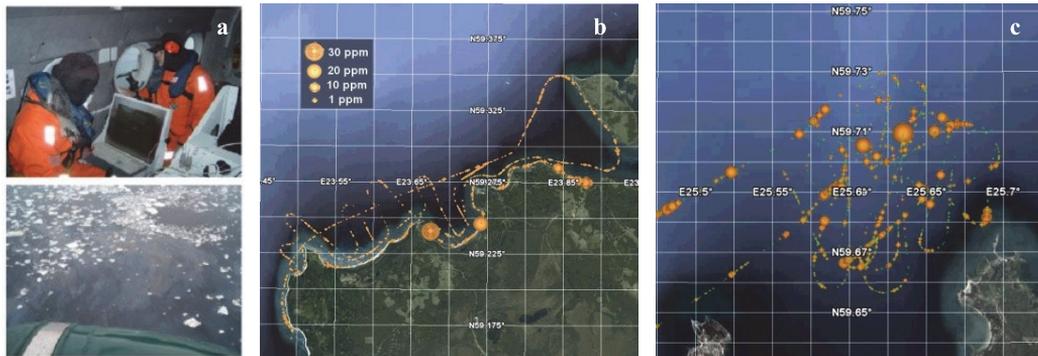


Figure 9. Shipwreck survey with FLS-AM Lidar: (a) – measuring conditions; (b) – oil pollution spreading in the coastal area; (c) – the map of oil patches on the ice in the shipwreck site, bubble size is proportional to the amount of oil.

3.4 Large scale surveillance

The assessment of natural water quality and subsequent management remains a principal environmental issue and even critical one for certain water areas, and the airborne surveillance with LIF lidar is proven to be efficient tool addressing this issue. In 2003 the multi-tier model for integrated environmental assessment has been introduced for water-quality monitoring, watershed modeling and identification of contamination source in the pilot project in Ontario, Canada. The Great Lakes and major Ontario rivers were surveyed with FLS-AU Lidar. As a result more than 800,000 LIF spectra were collected along the 2200 linear km observed with about 70 findings of deviation in water quality. The water sampling in 19 spots was done to establish reference points and to aid in the calibration of LIF data by SFS analysis with Fluo-Imager® ($r^2=0,95$). Numerous pilot projects with airborne FLS®-Lidars in 2004 – 2008 were aimed to detect chemical pollution on the navigation routes and to measure high-resolution spatial profiles of DOM, phytoplankton, and agriculture run-offs in Estonian coastal waters and lakes^{6,10}.

In 2008-09 the multi-tier concept was applied in the watershed surveillance in Atlantic Canada. Through a case study approach the project supported by Atlantic Innovation Fund (AIF) was aimed at the evaluation of the techniques which can be used by municipal clients and for environmental assessment. Two in river systems and two in lake systems in the Province of Nova Scotia were the areas of the project. The FLS-AM lidar on board of King Air aircraft was used in the project. Scanning the laser beam across the flight path was especially important for surveillance of lake and river waters, because it allowed getting the water LIF spectra and filtering off the signals from the ground, when the aircraft was not precisely following the stream canal or coast line. High sampling rate (up to 150 samples per second) of FLS-AM Lidar provided the spatial resolution of 5-7 m depending on the aircraft speed. The geo-referencing capabilities of the lidar were significantly improved using synchronized IMU accounted for the roll, pitch, and yaw of the aircraft (provided for the project by Applied Geomatics Research Group (AGRG)).

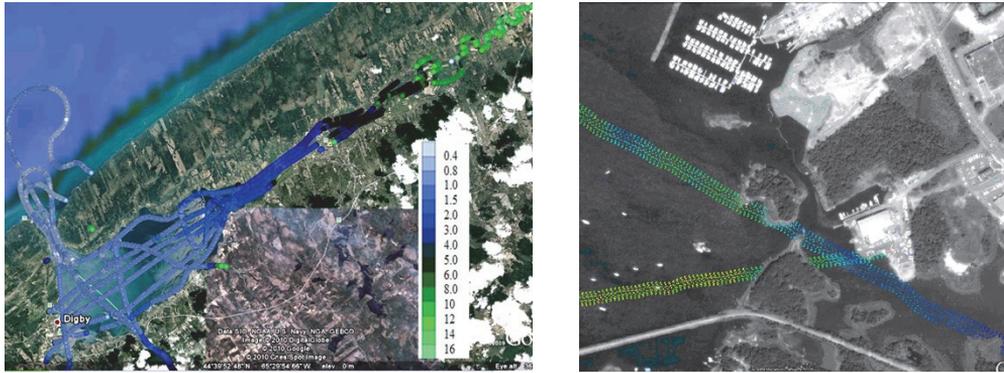


Figure 10: (left) - Color coded DOM profile in Annapolis valley, legend shows DOM in mg/l; (right) – Color coded detailed water transparency at laser footprints on the flight segment with FLS-AM: transparency increasing from blue to green.

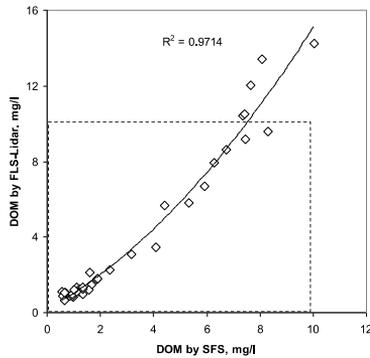


Figure 11. Correlation of FLS and SFS data of DOM.

Figure 10 shows the example of DOM profiling in the study area with zoomed flight segment showing the water transparency in every position of laser beam on the water surface inside the scanned corridor. The software expert system of FLS[®]-Lidar uses the Raman scattering line as an indicator of water, and therefore is able to discriminate the spectra of natural water from LIF spectra of adjacent lands (see detailed map in Fig.10).

The samples of water were taken in the characteristic sites to verify the correlation between the SFS and LIF data (see Figure 11). It was revealed that up to DOM concentrations of 10 mg/l the FLS and SFS data had linear correlation, while at higher DOM levels the FLS[®]-Lidar overestimated the amount of DOM and the result should be corrected with the SFS data.

4. EVOLUTION OF AIRBORNE FLS[®]-LIDARS

The FLS[®]-Lidars have proved applicability for various marine applications, and the general performances of the systems are enhanced with every new model. The upper limit of the sampling rate is increased up to 500 Hz, and the modern hyperspectral detectors records comprehensive (500 channels) LIF spectrum per every laser pulse with background subtraction providing the sensitivity for oil in water detection at the *sub-ppm* level. The scanner system is able to observe the corridor with the width of 2/5 of the flight altitude with the spatial resolution better than 3 m at the flight speed of 200 km/h and flight altitude of 200 m. Because of beam shaper it is possible to expand the laser footprint on the ground and provide full coverage of the surveyed area when it is required by application. The FLS[®]-Lidar is capable to process acquired LIF spectra in *real-time* and synchronize every reading with GPS coordinates corrected with IMU data. Such combination allows geo-referencing with the precision better than 1 m, which is important for certain applications.

Operational use of FLS[®]-Lidar determines certain requirements for airborne survey performance, and the most important ones are: to convert the acquired data into useful information and to share the information with the stakeholders, preferably

in *real-time*, for planning the measures and controlling the emergency situation. It specifies the demand for high throughput of LIF- and metadata acquisition, storage and processing, as well as the availability of shared information resources.

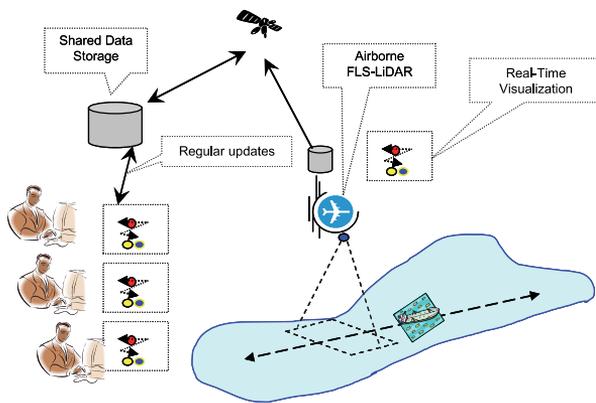


Figure 12. Communication scheme of airborne FLS®-Lidar.

Processing the hyperspectral information is a complex task, where the speed directly depends on the required quality of the result. This task is divided into preliminary data filtration and details analysis. The filtration stage implies the spoiled data exclusion from the information flow, and also the correction of useful signals for the information quality improvement. The analysis stage consists of statistical methods, algorithms of artificial intelligence (AI) and other pattern recognition and identification techniques applied for the data processing. To manage the calculations of a vast amount of data the methods of parallel processing are typically applied. However, in case of massive parallel computations the performance (Amdahl's law) and scalability limitation laws take considerable role. Based on this, the algorithms of filtration and analysis stages require optimization and high degree of parallelization.

The visualization of in-flight information and its sharing is another important issue for continuous improvement. The LIF data and metadata from GPS/IMU, range finder, digital camera, etc. while collected at different sampling rates, are integrated into information flow. The ultimate task of such integration is to define the geo-location, parameters, and the result of every single lidar measurement and to visualize it on the map of the surveyed area. To fulfill the requirements of information flow the following solutions have been implemented in the FLS software: sorting and packing the LIF and metadata according to the logics of visualization, pre-processing and generalization of data to decrease the redundancy of visual information; selection of the GIS software and communication with it in *real-time*. The Google Earth software has been selected for visualization purposes due to its interactivity and ability to handle considerable amount of data. At the same time the export function of the software allows data representation for any GIS software.

Further developments are aimed at the improvement of *real-time* reporting capabilities of FLS®-Lidars (see Fig.12). When the lidar provides in-flight *real-time* alarms according to operational set-up (e.g. oil in water), these alarms are visualized and stored into alarm file with corresponding metadata. The alarm files in Google Earth format sent periodically via communication channel to shared database are visible to ground operators. Any delay between airborne and on-ground visualization is defined by available communication channel parameters. In such communication scheme the shared FLS data can be added to any multilayer information system.

5. CONCLUSION

The FLS®-Lidars were developed as a payload for the aircraft of opportunity and can be easily installed on board of any rotary or fixed wing platform. Designed as a stand-alone system, it is ready for cooperation with other sensors on board or for integration into sensor suite because of its main controller with LAN capabilities. Such flexibility allows FLS®-Lidar using in a cost effective manner either as a single sensor on board of small aircraft to perform defined mission in particular in case of spill emergency, or as a part of multisensory survey. The multipurpose aircrafts with several sensors on board, while being more expensive in operation, are effective for airborne surveillance with broad scope of environmental targets.

The FLS®-Lidar with its operational, expert, and visualization software constitutes the airborne analytic platform capable to deliver the information to multiple stakeholders in near *real-time* mode. In many practical applications valuable

information can be obtained in combination of large-scale data with higher-resolution surveys of key locations that can be made with airborne FLS[®]-Lidar. It serves as an effective sensing tool when the survey is either scheduled for specific purposes and locations or is aimed at the periodical data collection over large areas and for long periods to enable identification of trends.

The technology of FLS[®]-Lidar reduces the cost of operational monitoring for oil contamination, is able to provide early warning for risk areas, and capable to increase the efficiency of cleanup operations. The multi-tier approach holds great promise for the optimization of environmental assessment activities by increasing the percentage of useful data (which can be constructively acted upon) and streamlining the allocation of human and technical resources.

ACKNOWLEDGEMENTS

The technical and application developments, and field works carried out by the staff of Laser Diagnostic Instruments in Estonia and Canada is greatly acknowledged.

REFERENCES

- [1] <http://www.deepwaterhorizonresponse.com>
- [2] Hoge, F.E., Swift, R.N., "Oil film thickness measurement using airborne laser induced Raman backscatter", *Appl. Opt.* 19, 3269 – 3281 (1980).
- [3] Hengstermann, T., and Reuter, R., "Lidar fluorosensing of mineral oil spills on the sea surface", *Appl. Opt.* 29, 3218-3227 (1990).
- [4] Dudelzak, A.E., Babichenko, S.M., Poryvkina, L.V., Saar, K.U., "Total luminescent spectroscopy for remote laser diagnostics of natural water conditions", *Appl. Opt.* 30, 453-458 (1991).
- [5] Brown, C.E., Nelson, R., Fingas, M.F., Mullin, J.V., "Airborne Laser Fluorosensing: Overflights During Lift Operations of a Sunken Oil Barge", *Proc of IV Thematic Conference Remote Sensing for Marine and Coastal Environments*. Orlando, FL, 1, 23-30 (1997).
- [6] Babichenko, S., Dudelzak, A., Poryvkina, L., "Lidar application in airborne monitoring of eutrophication and chemical pollution", *Proc. VI Int. Conference Remote Sensing for Marine and Coastal Environments*, Charleston, SC, 2, 391 – 397 (2000).
- [7] Zielinski, O., Andrews, R., Göbel, J., Hanslik, M., Hunsänger, T., and Reuter, R., "Operational Air-borne Hydrographic Laser Fluorosensing", *EARSeL e-Proc.* 1, 53-60 (2001).
- [8] Zielinski, O., Hengstermann, T. & Robbe, N., "Detection of oil spills by airborne sensors", In: [Marine Surface Films], Gade M, Hühnerfuss H, Korenowski GM (eds), Springer, Berlin Heidelberg New York, 255-271 (2006).
- [9] Lennon, M., Babichenko, S., Thomas, N., Mariette, V., Mercier, G., and Lisin, A., "Detection and mapping of oil slicks in the sea by combined use of hyperspectral imagery and laser induced fluorescence", *EARSeL e-Proc.* 5, 120-128 (2006).
- [10] Babichenko, S., Lapimaa, J., Poryvkina, L., Varlamov, V., "On-line fluorescent techniques for diagnostics of water environment", *Proc. SPIE* 2503, 157-161 (1995).
- [11] Babichenko, S., "Laser Remote Sensing of the European Marine Environment: LIF technology and Applications", In [Remote Sensing of the European Seas], Barale V. and Gade M. (eds), Springer, 189-204 (2008).
- [12] Poryvkina, L.V., Babichenko, S.M., Lapimaa, J., "Spectral variability of humus substance in marine ecosystems", *AMBIO*, 21(7), 465-467 (1992).
- [13] Babichenko, S., Poryvkina, L., Hoevenaer, H., and de Vos, F., "Diagnostics of organic compounds in water quality monitoring", *Environmental Technology*, 9 (1), 18-19 (1999).
- [14] Samberg A., Babichenko, S., Poryvkina L., "Improving three-tier environmental assessment model by using a 3-D scanning FLS-AM series hyperspectral lidar", *SPIE Proc.*, 5791, 318-325 (2005).
- [15] Babichenko, S., Dudelzak, A., Lapimaa, J., Lisin, A., Poryvkina, L., Vorobiev, A., "Locating water pollution and shore discharges in coastal zone and inland waters with FLS lidar", *EARSeL e-Proc.* 5 (1), 32-41 (2006).
- [16] "Heavy Oil Detection with Laser Fluorometers", *The Ohmsett Gazette*, Spring (2005).

Publication 2

I. Sobolev and S. Babichenko. Analysis of the performances of hyperspectral lidar for water pollution diagnostics. *EARSeL eProceedings*, 12(2):113–123, 2013a.

Analysis of performances of hyperspectral lidar for water pollution diagnostics

Innokenti Sobolev¹ and Sergey Babichenko²

1. Tallinn University of Technology, Department of Computer Control, Tallinn, Estonia; isobolevldi@gmail.com
2. LDI Innovation, Tallinn, Estonia; sergey.babichenko@ldi-innovation.eu

Abstract. The paper is aimed at the analysis of the performances of hyperspectral lidar for detection and classification of oil pollution in water environment in comparison with laser fluorosensor operating with discrete detection channels. We demonstrate that hyperspectral laser induced fluorescence (HLIF) spectrum holds all relevant spectral information about the target object in contrast to the discrete detection channels sensors data. In order to extract the significant features from HLIF data the multiresolutional analysis namely discrete wavelet transform (DWT) is applied. The feature extraction is automated using the sparsity-norm optimization method. The resulting features have clear spectral representation and are used in automatic object classification. The classification results and selectivity are compared with the discrete detection channel sensors data on a number of oil pollutants. The results of simulation experiments demonstrate high value of classification accuracy and the ability to sub classify similar organic compounds from the single group of objects. The comparison with discrete channels sensors data shows the significant increase in overall performance of oil pollution detection and classification.

Keywords. Remote sensing, hyperspectral, fluorescence, wavelet decomposition, feature extraction, classification

INTRODUCTION

The technique of laser remote sensing based on the method of Laser Induced Fluorescence (LIF) has been recognized in wide range of environmental applications [1][2][3]. This includes the tasks of daily monitoring in open sea areas and detailed tracking of oil spills displacement in the case of large technogeneous catastrophes. New set of arising industrial monitoring applications such as pipeline integrity monitoring, monitoring in coastal waters, waters near offshore platforms, dams, bays, and ports put the conventional active remote sensing systems into a new framework. In these conditions the lidar systems should act as a compact standalone analytical platform designed for early detection and classification of oil discharges into the environment.

The LIF lidars used for the remote sensing have mainly been coupled with discrete channel detector systems consisting of a set of photomultiplier tubes producing fixed wavelength spectra. The first instruments, namely laser fluorosensors, were first introduced as airborne lidar system for oil film detection in [1][2]. From that point, the capabilities of airborne laser fluorosensors for the detection and classification of micrometer oil films have been well demonstrated in numerous experiments. Starting from 1995 the new instruments hyperspectral LIF (HLIF) lidars were tested in airborne open sea area surveillances [4]. The distinct feature of HLIF lidar was in recording of continuous emission spectrum of target object at every laser shot. The number of detection wavelengths in HLIF lidar grew up forty times (from 10 up to 512) in comparison with the conventional fluorosensor.

The reliability of lidar system as an autonomous device is directly dependent on the quality of discrimination of oil pollution from natural environment and the pollutant type classification. The

main source of fluorescence background in natural waters is the dissolved organic matter (DOM) of natural origin. The environmental factors that impact on DOM concentration are water salinity, climate, season of the year and others [5]. Majority of experiments with laser fluorosensors were conducted in open sea areas and over coastal waters with low DOM concentration ranging from 0.1 to 3 mg/l. The absence of strong natural background produced a good identification and classification accuracy of oil pollution for those instruments [6]. A good example of oil classification with fluorosensor can be found in [7]. However the regions with high DOM content (up to 60 mg/l) like estuaries, lakes, rivers and wetlands have not been considered. The features of lidar operations in the waters rich by DOM were first studied using the HLIF shipboard lidar in the Baltic and North Seas regions. It has been demonstrated that hyperspectral detection is required to discriminate oil pollution in a variable and strong natural water background [8].

In this article we present a simulation of HLIF lidar oil pollution detection performances on various DOM concentrations and compare the results with discrete detection channels data. The detection of oil pollution in high DOM content is based on the comparison of LIF spectral shapes. Using the simulation data we provide the type classification results of selected refined and crude oil products using the HLIF spectrum features of pollutants. The application of discrete wavelet transform (DWT) is used to extract the relevant information from HLIF spectrum and significantly reduce the dimensionality of data.

SIMULATION

The simulation is based on numerous real-world experimental data sets with airborne HLIF lidar. Figure 1 left shows the selected HLIF spectra of oil products measured using the hyperspectral lidar with 308 nm excitation wavelength. The typical DOM spectrum shape is provided on Figure 1 right, the maxima of fluorescence emission is on 430 nm. The selected oil products have noticeable shape differences in comparison with DOM spectrum. For example the Calsol spectrum is shifted more into UV area with maxima around 400 nm. On the contrary the crude oils have fluorescence higher than DOM in the blue-green spectral range from 450 to 550 nm. The DOM and oil products fluorescence spectra can be discriminated using this information.

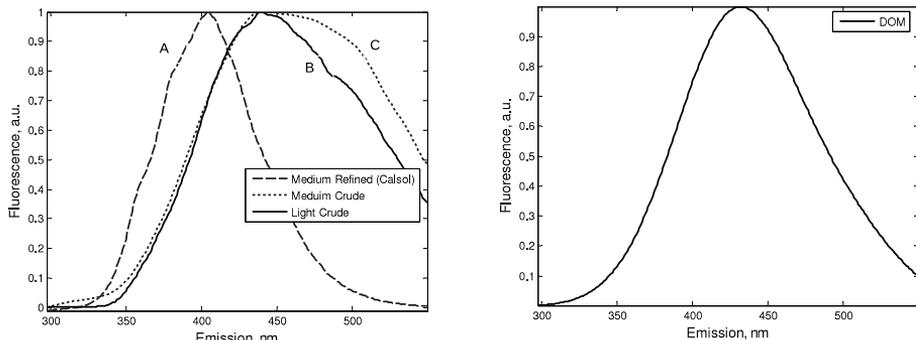


Figure 1: **(left)** The spectra of oil products selected for simulation measured as optically dense thick oil films. **(right)** The typical shape of river DOM emission spectrum at 308 excitation wavelength measured with FLS-AM lidar [8]. The Raman scattering line is removed for better visual spectra comparison.

The simulated HLIF spectrum is created as linear combination of DOM, Raman scattering line and one of three selected oil products with an additional noise vector. We explain our model selection more widely. The spectral shapes of two component mixture can be treated as a linear superposition of the individual spectral shapes of the component if no chemical reaction occurs between them, while the intensities of the fluorescence may have non-linear dependence on the concentrations of individual components due to the effect of inner filter. The non-linearities in model are not considered in performance analysis as they will effect the oil fluorescence

intensity equally in both HLIF and discrete channels data. The concentration of oil pollution is selected to be constant at 3 mg/l which is approximately proportional to 3 μm of oil film on 1 m^2 of water. The DOM concentration is varied in simulation from 1 up to 40 mg/l which corresponds to open sea and river water respectively [9]. The fluorescence intensities of DOM and oil pollution in clear water are shown in Figure 2. The value of Raman scattering is selected to be 200 units, which is typical for open sea water spectra usually recorded with referred lidar at the flight altitude of 200 metres.

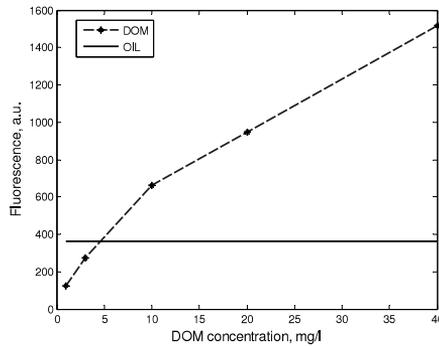


Figure 2: The calibration for DOM and fluorescence intensity of 3 ppm oil in clean water. The DOM sampled concentration values are: 1, 3, 10, 20, 40 mg/l. Values are estimated using data from [9].

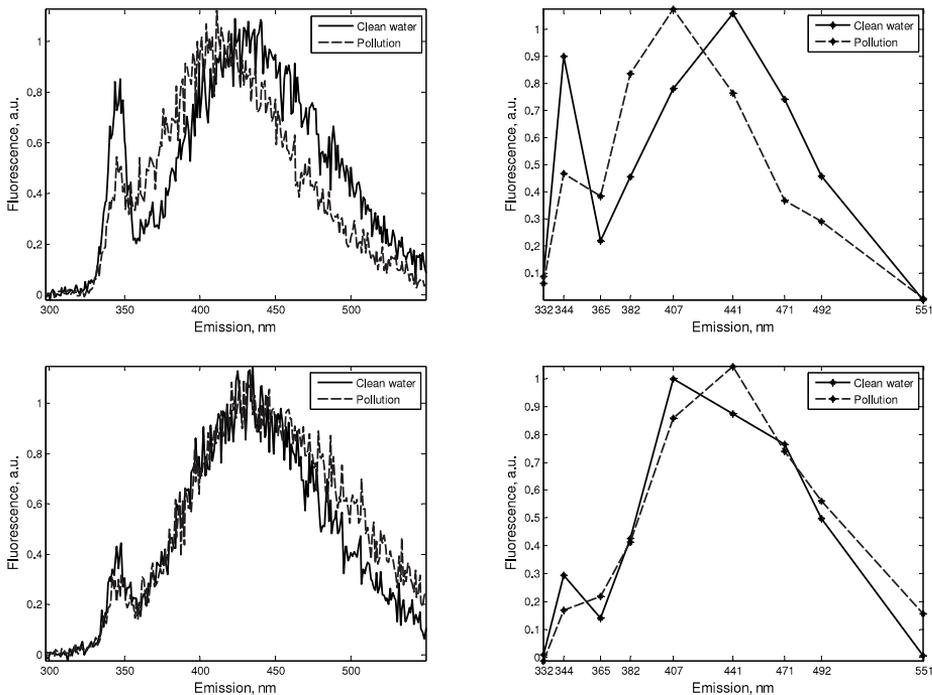


Figure 3: Simulated HLIF spectra (left) and discrete channels (right). Pollution is simulated using Calsol (top) and medium crude (bottom) oils. The DOM concentrations are 3 mg/l and 10 mg/l respectively.

The noise in FLS-AM lidar is considered to be a Shot noise where signal-to-noise ratio is equal to the square root of the fluorescence intensity. Such noise consideration is given by the lidar detector construction which is based on gated image intensifier CCD camera. The spectral intensities are expressed in au. typical for referenced HLIF lidar. Signal-to-noise (SNR) values for hyperspectral FLS-AM airborne lidar are ranging from 10 to 20 in single measurement.

The laser fluorosensor data are derived directly from HLIF spectra data. We have selected a typical configuration of discrete emission channels on PMTs and ideal interference filters with rectangular transfer function and applied this configuration to HLIF spectra. The selected discrete channels are: 332, 344, 365, 382, 407, 441, 471, 492, and 551 nm, which include the Raman scattering wavelengths. The total number of channels is 9 and the optical bandwidth is 10 nm. Modeling a discrete channel with the spectral width of 10 nm was done by integrating HLIF signal over 20 pixels (0.5 nm each) of CCD detector. It leads to 4-5 times increase of SNR value for discrete channel detection system, thus the effect of bandwidth difference has been properly taken into account. The resulting spectra of HLIF lidar and fluorosensor are shown on Figure 3.

METHODS

Pollution detection

In order to detect presence of contaminant in water with rich DOM content we need to have a model of spectral response of clean water matrix. The water matrix model is created using two vectors, one for Raman scattering signal and second for DOM fluorescence. Under the assumption that LIF spectrum is a linear combination of endmembers the Raman scattering and DOM intensity in simulated signal $S(\lambda)$ where λ is the wavelength are estimated using the traditional linear spectral unmixing (LSU) technique (regression analysis). In vector form the LSU equation can be expressed as:

$$S(\lambda) = \sum_i^n a_i \cdot F(\lambda)_i + e(\lambda), \quad (i=1, 2, \dots, m), \quad (1)$$

where F_i are endmember vectors entering the model, a_i is the intensity value of endmember i and $e(\lambda)$ is the error. Since the intensity a_i of the fluorescence is non-negative we find the solution to (1) using the non-negative least squares algorithm (NNLS). After the estimation of a_i the model of clean water is compared with the simulation spectrum using the coefficient of determination R^2 , which is the statistical measure of similarity between the model function and data. The binary classifier using the threshold $\delta \in (0, 1)$ is applied to discriminate between clean and contaminated water spectra. The quality of pollution detection with HLIF and fluorosensor data is evaluated for every DOM concentration using the receiver operating characteristic (ROC) metrics. The ROC metric shows a performance of a binary classifier as its discrimination threshold is varied [10].

Pollution identification

The next operational step is the pollutant type identification. This requires the additional sample of the pollutant to be added into the regression model. Estimated coefficients a_i contain the intensity values for all three endmember vectors (Raman scattering, DOM and pollutant). The resulting pollutant spectrum is obtained by subtracting the estimated water matrix from the complete signal. Finally the resulting spectrum is compared with known oil products from the spectral library.

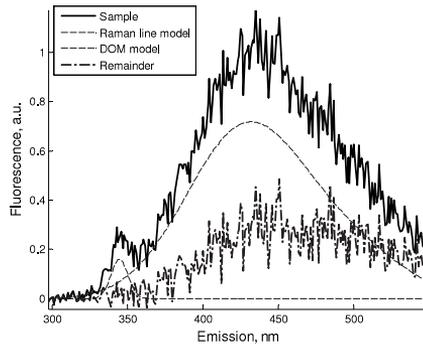


Figure 4: The example of linear extraction of pollutant from simulation sample. The pollutant is Medium crude oil, DOM concentration is 20mg/l. The complete spectrum has SNR value of 13 (set for experiment). The residual intensity is approximately 5 times lower and thus the residual SNR.

After the linear spectral unmixing the comparison of residual is complicated by the increased relative noise. Indeed the intensity of pollutant is only a part of the intensity of the complete signal and thus linearly extracted pollutant has the same deviation of unknown noise component in relation to the smaller signal value (see Figure 4). This effect is amplified with the increase of the DOM content against constant pollution concentration. Thus the high DOM value in water requires an additional analysis stage to improve the selectivity and quality of the analysis of LIF spectra.

Discrete wavelet-based feature extraction

Nowadays the wavelet transform (WT) is most widely used time-frequency analysis of signals in many engineering and scientific applications. A WT is a mathematical representation of arbitrary function using orthonormal series of wavelet coefficients. The wavelet coefficients represent a time localized frequency-scale information. The generation of coefficients can be done using three schemes: continuous, dyadic and lifting. In this article we have applied a dyadic discrete WT that produces optimal representation of signal in wavelet basis. The dyadic grid allow to use a fast algorithm, an analogue to FFT, to calculate the WT for discrete sampled data. The reader can refer to the formal description of WT in [11][12].

In the case of hyperspectral LIF data we can extract the significant part and remove the redundant and noisy parts from spectrum using the WT. The wavelet analysis has the ability to separate the small-scale and large-scale information in high dimensional data. A good example of dyadic discrete wavelet transform feature extraction method applied on a hyperspectral synthetic aperture radar data is presented in [13]. The feature extraction technique proposed in [13] is based on evaluation of the class-discriminating capability using different subsets of wavelet coefficients and determination of preferably small optimum subset of wavelet coefficients.

In our previous work we proposed a continuous wavelet feature extraction method for HLIF data processing [14]. The continuous wavelet transform (CWT) is redundant and computationally expensive method; however it at best preserves the initial spectral information and excellently extracts all data features. In CWT feature extraction the representative scales are individually selected for each pollution type depending on its spectral properties and all coefficients on selected scales participated in class comparison. The relation of classification features to spectral properties, like local extreme points, is very important. For example in Polycyclic aromatic hydrocarbons (PAHs) the identical local fluorescence extrema areas in the same places yield the close related layout of poly-aromatic hydrocarbons (PAHs) and similar surrounding matrix for complexes of PAHs. The application specific feature extraction method should count the class-discriminating capabilities of features as well as application specific

feature requirements. The latter can be used as an additional constraint of feature extraction algorithm, as for example good features in terms of the class-discriminating capabilities can be affected by noise.

In this paper we propose the DWT version that uses the main criteria for feature extraction – the extracted and selected features are bound to significant spectral properties of object. The provided feature extraction algorithm exploits the idea of sparsity-norm optimization or basis-pursuit method [15]. The signal s can be represented using the basis of many functions Φ by the minimization of the functional:

$$\|s - \tilde{s}\|_{l_2} + \lambda \cdot \|\alpha\|_{l_1}, \quad \tilde{s} = \Phi\alpha \quad (2)$$

where α is the basis coefficients and $\|s - \tilde{s}\|_{l_2}$ is a reconstruction error. The l_1 norm for α is selected to preserve the sparsity [15]. In our approach the minimization of (2) must be additionally constrained by the spectral properties of object. Instead we use a simplified application specific algorithm to automatically extract most significant and noise-immune features from HLIF spectrum.

We know that the model spectrum from the spectral catalogue is noiseless therefore the most “energy” is allocated between the significant local spectral properties (hills, bends, curves etc.) thus producing the highest corresponding detail coefficients in wavelet decomposition. The optimal and sparse representation of model is obtained by the following way. At first step the complete DWT representation of model is created on the maximum level of decomposition M . Next, the spectrum is reconstructed using only the approximation coefficients $A_m, m=M$. On the following steps the detail coefficients from all decomposition levels $D_m, m=1,2,\dots,M$, sorted in descending order by their intensity values, are sequentially added into the reconstruction and results are compared with the initial model. After that the minimum of the sparsity-norm functional is found by comparing the amount of non-zero detail coefficients with the reconstruction quality (2). In experiment we apply an error threshold of 10% for the reconstruction quality. As the highest detail coefficients are added first the applied error threshold is reached with the minimal subset of most significant spectral properties. The feature extraction method is applied to the model of each pollutant and extracted coefficients are stored in database. The pollutant is identified by computing the probability of a spectrum i to belong to the class j using the extracted feature values of a model and analysed spectrum:

$$p_{ij} = 1 - \frac{d_{ij}}{d_{\max}} \quad (3)$$

where d_{ij} is a Euclidean distance between the objects i and j , d_{\max} denotes the distance to the most distant class for i . For discrete channels data the similarity (3) is evaluated directly using the fluorescence values.

RESULTS

Pollution detection

The numbers of sets of LIF spectra of clean and polluted water were simulated for each type of pollutant using the technique and parameters described in the previous paragraphs. The simulated discrete channel data from laser fluorosensor is shown in Figure 5 left column. In case of DOM concentrations less than 3 mg/l, which corresponds mainly to open sea areas, the ROC plot has a point in the upper left corner (coordinate (0,1)), representing perfect classification (no false positives and no false negatives). In high DOM content (10, 20, 40 mg/l) the detection accuracy degrades, the false positive rate grows and ROC curve converges to a

diagonal line, which means more a random guess of a classifier. The results indicate expected degradation of detection quality for discrete channel data in waters with high DOM concentration.

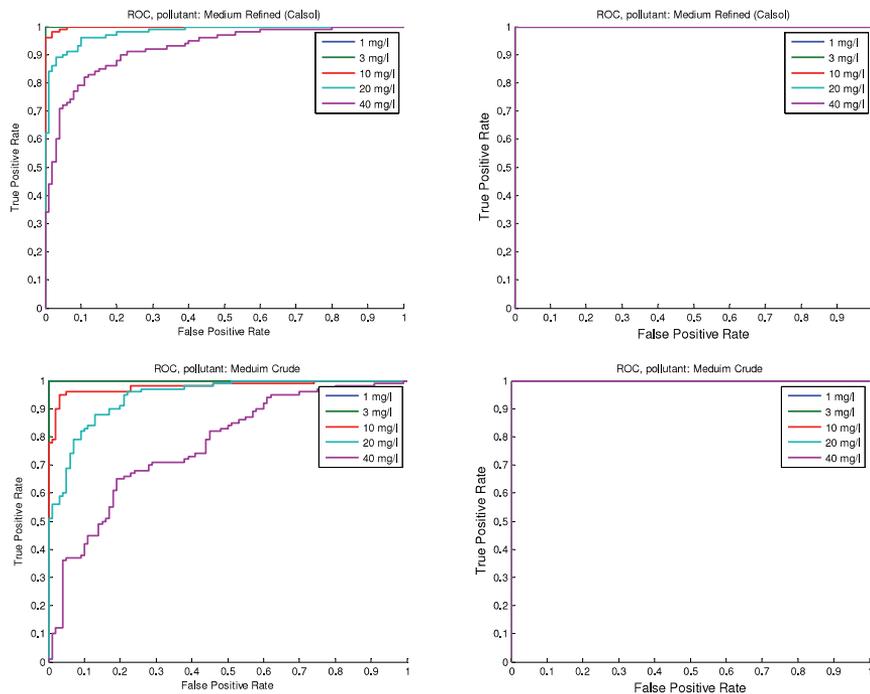
The fluorosensor simulated data pollution detection is also affected by the type of oil product. The medium refined Calsol is characterized with high fluorescence shifted to UV wavelengths and narrow spectrum comparing with DOM spectrum, which produces high deformation of the spectral shape (see Figure 3 top). The spectra of crude oils and DOM are broad and mostly overlapped and their linear superposition produces indistinguishable shape variations that are comparable with distortions from the detection noise (see Figure 3 bottom). This effect is best seen by comparing the 20 mg/l ROC curves on Figure 5 left.

The HLIF data (Figure 5 right column) is more sensible to variations of spectral shape and according to ROC curves the pollution detection results are almost perfect. The negative effect of shot noise is suppressed due to the greatly increased amount of spectral channels. The appearance of pollution adds a large-scale distortion to spectrum that affects the similarity value more than the small-scale noise distortions. The type of oil still has an effect on detection quality, which is seen for Light crude oil on Figure 5 bottom right.

Pollution identification

The pollution identification process is the following step in Lidar system autonomous operation mode. In real system the identification process relies on the previous step of rough pollution detection. However, here we ignore the non-compliant detection results for fluorosensor data because of high DOM content and assume that pollution presence is prior known.

As already stated the feature extraction takes first step in pollution identification process. The model spectra of oil pollution (see Figure 1, left) were analysed by discrete wavelet-based feature extraction algorithm and results are given in Table 1. The reconstructed versions of model spectra are plotted on Figure 6. The reconstruction error or MSE for all models is less than or equal to 10%.



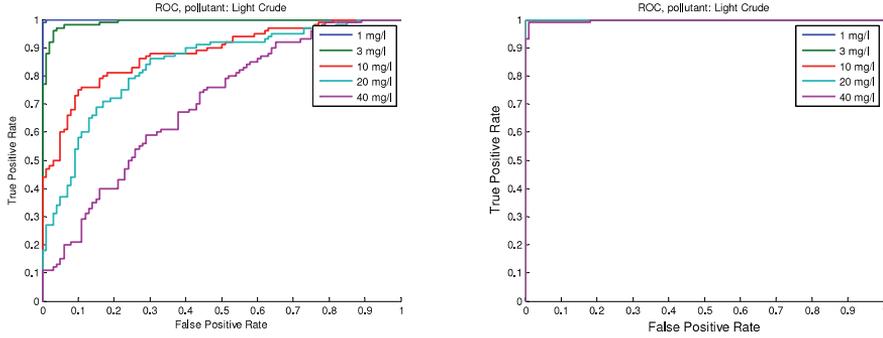


Figure 5: Pollution detection results for fluorosensor (left column) and HLIF (right column) data. The ROC curves are plotted in different colours for DOM concentrations from smallest 1 mg/l (blue) to highest 40 mg/l (violet).

The number of non-zero coefficients from Table 1 is significantly less than the number of spectral channels in original HLIF data and is close to the number of channels in conventional fluorosensor. In [13] it is pointed that approximation coefficients from DWT can be used to mimic a known spectral sensor if the spectral transfer function of the sensors is known. For example a rectangular transfer function of interference filters matches to Haar mother wavelet. In our example we used more complex and appropriate mother wavelet function from biorthogonal wavelet family and extracted the significant and sparse information from hyperspectral data. It can be referred as software imitation of a novel compressed sensing technique [16][17] as for each specific spectrum we construct a wavelet based software spectrometer.

The classification results are reported using the statistical accuracy measure (ACC) as a proportion of the true results (true positives with true negatives) to the total number of samples (see Table 2). The classification is performed on a complete set of extracted features – approximation and detail coefficients for each DOM concentration level. The ACC value is provided for three cases. The total ACC is the overall performance of samples classified into three classes: Medium Refined, Medium Crude and Light Crude. The cACC and sACC values were used to demonstrate the difference in discrimination between different groups of oil products and a sub-classification of exact oil product. The cACC is the accuracy of classification between the Refined and Crude oil groups. The sACC is the accuracy of sub classification between Medium and Light crude oils.

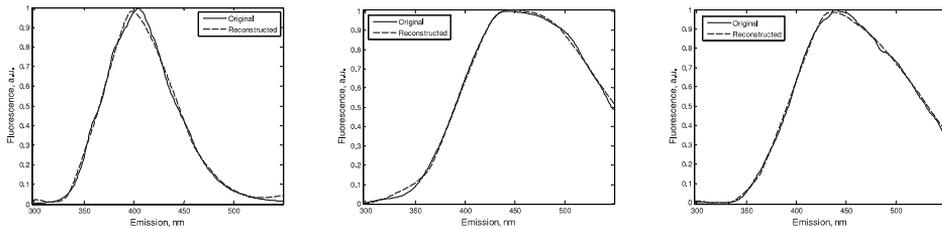


Figure 6: Medium refined (left), Medium crude (middle) and Light crude (right) oil spectra and their reconstruction using the optimal subset of extracted features.

Table 1: Feature extraction results for HLIF data. The maximum level of decomposition M is 7.

	Number of non-zero approximation coefficients (A_7)	Total number of non-zero detail coefficients (D_m)	Corresponding detail coefficients location (nm)
Medium refined	9	6	297, 332, 364, 400, 459, 491
Medium crude	9	3	297, 364, 491

Light crude	10	5	297, 332, 364, 491, 549
-------------	----	---	-------------------------

Table 2: Classification accuracy for HLIF and fluorosensor data versus DOM concentration.

DOM concentration (mg/l)	HLIF data				Fluorosensor data			
	Total (%)	ACC (%)	cACC (%)	sACC (%)	Total (%)	ACC (%)	cACC (%)	sACC (%)
1	100	100	100	100	87	100	100	80
3	98	100	100	97	80	100	100	70
10	96	100	100	94	75	99	99	62
20	90	100	100	85	72	99	99	58
40	84	100	100	80	71	97	97	57

In both methods the accuracy of classification between Medium Refined and Crude oils cACC is more than 97% for all DOM concentrations. This result was expected, as the Refined and Crude oils have noticeable differences in fluorescence responses which are detectable in presence of high noise.

On the contrary, Medium and Light crude oil spectral responses are highly overlapped which produces more difficulties in their discrimination. The increasing DOM concentration affects the SNR value in residual signal, which in turn degrades the sub-classification accuracy sACC for 20%. The classification of crude oil types is about 25% better using the HLIF data. For DOM concentrations less than 10 mg/l the sub-classification accuracy is more than 90%. This indicates that hyperspectral data holds robust information that can be well extracted with DWT even in a presence of high noise content. Comparing the total performance of the HLIF and fluorosensor data classification, it is clear that HLIF technique with DWT feature extraction method is superior.

CONCLUSION

The performance of hyperspectral lidar for the detection and classification of organic compounds in water environment was compared with discrete detection channels laser fluorosensor. The results of simulation experiments demonstrate a superior performance of HLIF technique in primary oil pollution detection and pollution identification tasks. It is important that the performances were compared in the presence of natural water fluorescence background (DOM) in various concentrations from minor 1 mg/l to high 40 mg/l. The experimental datasets were simulated using the real-world data from airborne hyperspectral lidar experiments.

The analysis of HLIF data has demonstrated perfect detection results for selected pollutants even in the presence of high DOM content. On the contrary, the detection performance with discrete channels is significantly affected by the DOM concentration producing almost random results when the ratio of oil to DOM fluorescence intensities is less than 0.5 (the case of 3 mg/l of oil at DOM concentrations above 20 mg/l for the simulations above). The experiments indicate that the fluorosensor detection performance is also affected by the spectral response of the pollutant. Thus the selected crude oil products in combination with large DOM produce insignificant variations that are undetectable because of noise distortions.

It was found that detailed spectral information can be extracted from HLIF data using the proposed method based on DWT. The feature extraction method satisfies application specific feature requirements such as sparsity and direct relation with significant local spectral properties of object. The dimension of features subset is 20 times smaller comparing with the initial dimension of hyperspectra data. The latter is essential for the classification as it eliminates the detrimental effects of high dimensionality. The pollution classification is demonstrated with three oil products selected from two distinct groups of oil: Refined and Crude. The classification

accuracy for HLIF data is 100% independently from concentration of DOM. Almost the same result at average 99% classification accuracy is received for fluorosensor data. The more complicated task of detailed sub-classification is shown on example of two Crude oil products. In that case both methods produce results that are strongly dependent on SNR value. The sub-classification accuracy decreases to 20% with increasing DOM concentration from 1 mg/l to 40 mg/l. Due to the robust feature extraction the classification of crude oil types is about 25% better using the HLIF than fluorosensor data.

In future research a bigger experimental dataset including light refined, diesel and other groups of oil products will be studied. The final result could be presented as a technical specification for hyperspectral LIF lidar analytical platform. For example, as the accuracy of detection and classification of various oil products in different environmental conditions for industrial monitoring applications. The performance of HLIF data analysis relies on the extraction of information from spectrum, thus other methods like matching pursuit could be studied for feature extraction.

ACKNOWLEDGEMENTS

This publication was supported by the Estonian Doctoral School in Information and Communication Technology and by the ESF grant Nr 8738.

REFERENCES

- [1] Hoge F E and Swift R N, 1980. Oil film thickness measurement using airborne laser induced Raman backscatter. *Appl.Opt.*, 19: 3269 – 3281 pp.
- [2] Hengstermann T and Reuter R, 1990. Lidar fluorosensing of mineral oil spills on the sea surface. *Appl. Opt.*, 29: 3218-3227 pp.
- [3] Dudelzak A E, Babichenko S M, Poryvkina L V, Saar K U, 1991. Total luminescent spectroscopy for remote laser diagnostics of natural water conditions. *Appl. Opt* 30, 453-458.
- [4] Babichenko S, Lapimaa J, Poryvkina L, Varlamov V, 1995. On-line fluorescent techniques for diagnostics of water environment. *Proc. SPIE*, 2503, 157-161 pp.
- [5] Thurman E M, 1985. *Organic Geochemistry of Natural Waters*. Springer, 9024731437, 497 pp.
- [6] Reuter R, Wang H, Willkomm R and Loquay K, 1995. A Laser Fluorosensor for Maritime Surveillance: Measurement of Oil Spills. *EARSeL Advances In Remote Sensing*, Vol. 3, No. 3 – VII.
- [7] Almhdi K M, Valigi P, Gulbinas V, Westphal R and Reuter R, 2007. Classification with Artificial Neural Networks and Support Vector Machines: application to oil fluorescence spectra. *EARSeL eProceedings*, 6(2), 115-129.
- [8] Alekseyev V, Babichenko S, Lisin A, Lapimaa J, Poryvkina L, Shchemelyov S, Sobolev I, 2010. Airborne surveillance of water basins with hyperspectral FLS-Lidar. *SPIE Remote Sensing Europe*, vol. 7825: *Remote Sensing of the Ocean, Sea Ice, and Large Water Regions*.
- [9] Babichenko S, 2001. *Spectral fluorescent signatures in diagnostics of water environment*. Tallinn : Institute of Ecology, 9985-58-208-x., 53, 60 pp.
- [10] Hanley J A & McNeil B J, 1982. *The meaning and use of the area under a receiver operating characteristic (ROC) curve* (*Radiology*) 29–36 pp.
- [11] Daubechies I, 1998. *Recent Results in Wavelet Applications*. 1998. *Proceedings of SPIE Aerosense Symposium*.

- [12] Daubechies I, 1992. Ten Lectures on Wavelets. SIAM.
- [13] Lori M B, Cliff H K and Jiang Li, 2002. Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction. IEEE Transactions On Geoscience And Remote Sensing, vol. 40, no. 10.
- [14] Sobolev I, Babichenko S, 2013. Application of the Wavelet Transform for Feature Extraction in the Analysis of Hyperspectral Laser-Induced Fluorescence Data. International Journal of Remote Sensing (in print).
- [15] Chen S S, Donoho D L and Saunders M A, 1998. Atomic decomposition by basis pursuit. SIAM J. Sci. Comput., 20: 33-61.
- [16] David L. Donoho, 2006. Compressed Sensing. IEEE Trans. Information Theory, vol. 52, no. 4, pp. 1289-1306.
- [17] Chang C C and Lee H N, 2008. On the estimation of target spectrum for filter-array based spectrometer. Optics Express, 16(2), 1056-1061.

Publication 3

I. Sobolev and S. Babichenko. Application of the wavelet transform for feature extraction in the analysis of hyperspectral laser-induced fluorescence data. *International Journal of Remote Sensing*, 34(20):7218–7235, 2013b.

Application of the Wavelet Transform for Feature Extraction in the Analysis of Hyperspectral Laser-Induced Fluorescence Data

Remote sensing approaches for environmental protection and exploration have evolved rapidly in the last decade. Among the new operational tools, hyperspectral FLS LiDAR has demonstrated high sensitivity and the ability to function in complex environments for real-time, robust oil-spill monitoring on airborne or ship-borne analytical platforms. The capabilities of such analytical platforms include real-time analysis of laser-induced fluorescence (LIF) data. Although numerous examples of the application of signal theory to the analysis of hyperspectral data analysis appear in the remote sensing literature, the conventional data analysis strategies are not well adapted to the practical issues of the application of LIF. The aim of this paper is to provide a new approach for LIF LiDAR analytical platforms that is focused on the specifics of hyperspectral LIF data. The approach is based on structural data analysis and interpretation, through which more detailed spectral matching is performed. This paper is based on a simulated experiment in which the spectra of actual sea water and well-known types of petroleum products were combined to demonstrate the wavelet-transform-based analysis of LIF data. The final part of the paper demonstrates the application of the wavelet transform to structural analysis of LIF data from field experiments for the detection and identification oil products in difficult environmental conditions.

Keywords: LiDAR, hyperspectral, feature extraction, wavelet

Introduction

At present, the need for reliable and flexible operational techniques for the detection and quantification of marine oil spills is evident. Tasks that require rapid real-time monitoring of the environment include the clean-up and recovery procedures for oil pollution. In addition to environmental protection, there are applications in the exploration and tracking of natural oil seeps. Currently, the basic technologies that have been successfully applied to both tasks are satellite imaging (Brekke and Solberg, 2005; NASA/Jet Propulsion Laboratory, 2010) and remote sensing (Babichenko et al., 2006; Lennon et al., 2006). A comprehensive overview and comparison of these technologies and sensors can be found in the review by Jha et al., 2008.

Many researchers are familiar with satellite imagery and other remote sensing technologies. Despite the broad view and high resolution offered by satellite imagery, this technology generally provides a synoptic overview of a certain area without substantial detail. For instance, satellite radar can detect an oil film on the surface of the water but not oil in the water column. Because the aggregate state of oil in water changes over time from a surface film to an emulsion in the water volume, this knowledge is important for the efficient clean-up of pollution. Different petroleum products have different toxicities and contamination lifetimes. The type of oil in a pollution area is important information for predicting the long-term impact on the environment. Therefore, the refinement of information requires not only the detection and quantification of an oil slick but also the definition of the target object class, type and weathering status. These object parameters are also critical for petroleum exploration in marine oil seeps.

Active hyperspectral remote sensing, based on the laser-induced fluorescence (LIF) method, enables refined knowledge about oil pollution to be obtained. The LIF LiDAR used for remote sensing of aqueous environments is often referred to as laser fluorosensors (Brown and Fingas, 2003). These instruments are based on detection within pre-set spectral channels

that consist of a discrete set of photomultiplier tubes coupled to optical filters. Although these systems are highly sensitive in many applications, they have rather limited selectivity when operating in complex coastal areas or areas with a large amount of dissolved organic matter (DOM). LIF LiDAR based on a linear CCD camera coupled to a polychromator that can produce continuous spectra is called hyperspectral LiDAR (Sobolev, 2010; Alekseyev et al., 2010). Consequently, the overall shape of the target emission spectrum is recorded with each laser shot, and this process allows a detailed analysis of the spectral structure to be performed.

The goal of this paper is to present the hyperspectral LIF technique to provide robust and real-time large-area characterisation of oil pollution in marine environments. Accordingly, we defined the generic spectral properties of the oils that are important for the mentioned applications of hyperspectral LIF data. Based on the results of a model experiment and a practical application, we developed a data analysis method that focuses on a non-statistical, structural approach to feature-extraction methodology and that provides high selectivity for pattern recognition in hyperspectral LIF data.

Overview of Generic Oil Properties Using Information in Hyperspectral LIF Data

A hyperspectral LIF spectrum is an emission vector with a large number of spectral bands. Compared with a discrete emission spectrum with a limited number of channels, hyperspectral LIF presents more than 500 points of smooth, detailed data. For example, the detailed shape of the fluorescence response provides the ability to identify the exact type of each petroleum product, with analyst supervision. Figure 1 shows three hyperspectral LIF spectra of oil products: sea diesel, heating oil and marine diesel oil (MDO). The data were obtained in a laboratory using a compact FLS® LiDAR with an excitation wavelength of 308 nm from an excimer laser. The oil targets were measured at close range; consequently, the resultant data contain almost no noise and represent the fluorescence models for these substances. Through simple visual comparison, an analyst can determine that the primary fluorescence intensity of sea diesel, the lightest of the oil products sampled, is in the ultraviolet (UV) range (from 350 to 370 nm) and can be easily distinguished from other samples based on its spectral shape and position. In contrast, the spectra of MDO and heating oil are localised in the same spectral range with maxima near 408 nm and only differ in minor details of the spectral shape. The detailed shape of the emission spectrum, determined by hyperspectral LIF, reveals differences between similar objects, as shown.

Aggregate State and Degradation of Oil

The state of an oil film on a water surface is dynamic and is accompanied by a number of transformation processes that result in a transition from a surface film to a mixture of aggregate states in the water column. The process of emulsification typically affects the fluorescence response of the oil product, and, in turn, this effect allows the aggregate state of the oil to be determined (Poryvkina et al., 1998). In addition, under real conditions, various hydrophysical factors of the solution and evaporation cause fast chemical and biological degradation and significantly influence the fluorescence response of the oil product in the water medium. Thus, the weathering of oil causes a change in the shape of the fluorescence spectrum.

Figure 2 shows the spectra of two oil products in the film and emulsion states at high thickness (150 µm) and concentration (5 ppm) in distilled water. The differences in the spectral shapes of the fresh film and the weathered emulsion are noticeable in different ways for the gas oil and LCO samples. The weathering of gas oil produces minor changes in the

fluorescence maxima observed between 360 and 425 nm. Compared with the fluorescence maximum of the LCO film between 330 and 390 nm, that of the LCO emulsion increases significantly. These results indicate that the transformation of the spectral shape, because of a change in the aggregate state, can be distinct (i.e., for LCO) or can be very subtle (i.e., for gas oil). The variation depends on the type of oil product and the environmental conditions. Based on the above examples, characterisation of the aggregate and the weathering state of oil products in water is also possible with the hyperspectral LIF technique.

Structural Analysis of Hyperspectral LIF Data

Numerous articles and several books on remote sensing describe the application of signal theory and machine learning for hyperspectral or high-dimensional data analysis and interpretation (Landgrebe, 2003; Chen 2006). Most algorithms for the analysis of multispectral remote sensing data begin with the decision-theoretic extraction of relevant features from the signals and the discarding of irrelevant information. Applied feature extraction methods express the linear (principal component analysis, linear discriminant analysis, and projection pursuit) and nonlinear (IsoMaps and generative topographic mapping) embeddings of specific data into new, low-dimensional, efficient spaces.

The availability of the hyperspectral LIF technique for operational use that requires real-time characterisation of oil pollution is strongly dependent on efficient data analysis. The quality of the response of the analysis system on a real-time or nearly real-time basis determines the operational fitness of LIF LiDAR. The mentioned analytical methods based on statistical feature extraction introduce excessive computational complexity (i.e., solving complex nonlinear optimisation or eigenvalue tasks) when applied to hyperspectral LIF data in real time. In addition, these methods are designed to optimise certain criteria for *given* features (i.e., the variance) rather than to capture features that are application specific.

In this paper, we focus on another structural, or synthetic, approach to hyperspectral data analysis (Pavlidis, 1977). This structural analysis describes data in terms of particular or predefined primitive components and the relationship among these components. For hyperspectral LIF data, the primitives can be efficiently represented by the spectra of the individual components and, as we have found, by some predefined basis of *good* analytical functions that express the local discontinuities in the spectral shapes.

The main task of the LIF analytical system in monitoring oil pollution is a direct search of the spectra in the sample dataset for spectra that contain a spectral pattern similar to that of a particular oil product. The search and comparison are complicated by the fact that, in most cases, the observed fluorescence spectra from actual oil-in-water targets are composite. The observed spectra typically consist of the inelastic Raman scattering line and of the spectral bands of natural DOM overlapped with the fluorescence signals of the other organic compounds. In terms of structural analysis, the primitives that describe the signal are the individual LIF spectra for the Raman scattering, DOM and oil pollution.

The first part of the simulation was designed to demonstrate the classification of composite LIF spectra with known individual components. The second part shows the novel application of predefined basis functions called *wavelets* to the analysis of hyperspectral LIF data. This algorithm is used in the latest analytical tools for LIF data processing in FLS® LiDAR. Combinations of the spectra of actual sea water and well-known types of petroleum products were used in the experiment. The ratio of the oil products to the fluorescence intensity of the DOM corresponded to an average film thickness of 1 to 1.5 microns for most oil types. To simulate coastal or closed sea water with a higher DOM content, the ratio of the intensities of the Raman scattering and DOM fluorescence was varied. In addition, a

LiDAR-specific, colored noise vector with a standard deviation of 0.1 was added to each resultant spectrum to produce more realistic models.

Decomposition of LIF Spectra on a Nonorthogonal Finite Basis of LIF

Components –Regression Analysis

Linear regression can be used to decompose the measured LIF spectra into the predefined basis of LIF components. Under the assumption that the chemical compounds in water do not interact, the observed spectral intensity $S(\lambda)$ can be represented as a linear combination of components F with unknown intensities k :

$$S(\lambda)_i = \sum_{j=1}^n F(\lambda)_{ij} k_j + e, \quad (i=1, 2, \dots, m), \quad (1)$$

where k_i is the fluorescence intensity of component i , and e is the error. The solution of

(1) defines the optimisation problem, and the values (k_1, \dots, k_n) are found by solving the

quadratic minimisation task in which the objective function is the sum of the squared

residuals.

Each model, created with LIF components F from the spectral library and the corresponding (k_1, \dots, k_n) , is compared with the measured spectra. The similarity value is calculated with the conventional coefficient of determination, R^2 , which is a statistical measure of how well the model function approximates the actual data points. This proximity measure allows qualitative assessment of the comparison value (e.g., the Chaddock scale, see Table 1) and also calibration of the sensitivity of the analysis. The sensitivity is adjusted based on the forecasted classification error rate, with an R^2 threshold of $\delta \in (0, 1)$. Thus, for very large values of δ ($\delta \geq 0.9$), many false-negative errors are observed, and, for small values (i.e., $\delta < 0.7$), false-positive errors increase.

Figure 3 shows the simulated LIF spectra of sea water polluted with diesel oil. The corresponding results of the analysis for three models are diesel combined with a water matrix (Fig. 3, left), a clean water matrix (Fig. 3, middle) and Calsol combined with a water matrix (Fig. 3, right). The components of the water matrix are the Raman line (i.e., the peak at 345 nm) and the DOM with simulated intensity coefficients of $k_{\text{dom}} = 1.00$ and $k_{\text{Raman}} = 0.13$. The intensity coefficient for the oils was $k_{\text{oil}} = 0.3$.

The similarity value, R^2 , between the first (correct) model with diesel and the simulated spectrum was “very high” (i.e., greater than 0.99), despite the small “collared” noise added to the modelled spectrum (Fig. 3, left). However, the similarity value is 0.9233 for the model spectrum with a clean water matrix. Because both values fit the definition of “very high” (Table 1), the correct model could not be explicitly selected based on the R^2 values, and qualitative assessment must therefore be used. The selection of the correct answer is further complicated upon consideration of the next model with a different oil product, Calsol. Despite the fact that the fluorescence maxima of diesel and Calsol differ by 50 nm (357 and 405 nm, respectively), in combination with the water matrix, both spectra produce shapes that are indistinguishable. In the latter case, the similarity coefficient is 0.9697, which is even closer to the result from the first model.

The regression analysis decomposition methodology provides a clear physical interpretation of the extracted features. The primitives are the LIF spectra of the components, and the solution of (1) gives the estimated fluorescence intensity of each component in the compound. The noise information, e , which is irrelevant, is successfully discarded. However, determination of the correct model based on the R^2 value is a weak point of the LIF regression analysis. The broadening of the fluorescence spectra leads to a rather strong correlation among a wide range of spectral shapes. This correlation explains the low selectivity of the comparison between the LIF model assembled from the spectral library and the measured spectral shapes. The current hyperspectral LIF analysis system is designed to meet the low selectivity of the standard comparison techniques. Accordingly, we propose an additional stage in the analysis algorithm that uses the detailed shape information of the LIF spectrum.

Wavelet Transforms

The structure of the hyperspectral LIF spectrum is a complex set of local features (extrema and curve inflection points) and additional noise on a broad Gaussian base shape. The analysis of the LIF structure—specifically, the search for differences—is not accurate if information from the global wavelength–frequency domain is used. A detailed representation of the local specific features on the corresponding frequency scales is needed. This task is the primary application of wavelet transforms (WTs).

The well-known time–frequency analysis of the signal with a Fourier transform (FT) uses basis functions with infinite energy in the time domain (in our case, the time domain is wavelength) and with marginal localisation in the frequency domain $[\cos(wt), \sin(wt)]$. This property makes FT the best analytical tool for continuous, nonstationary signals. However, the LIF spectra are irregular functions $f(t)$ in $L^2[R]$ space with finite energy and cannot be effectively represented in the FT basis. Based on this fact, it is possible to suppose that more suitable basis functions for the space must be compactly supported (or have finite effective support) and be orthogonal and orthonormal. These functions are called wavelets, $\psi_{ab}(t)$ (Daubechies, 1992; Chui, 1992). Decomposition of the signal $f(t)$ with a continuous WT is performed through convolution of the signal with the wavelet function, where the parameter a scales the wavelet function, and b shifts the wavelet function:

$$W_{\psi} f(a,b) = \langle f(t), \psi_{ab}(t) \rangle = \frac{1}{\sqrt{a}} \int f(t) \psi\left(\frac{t-b}{a}\right) dt. \quad (2)$$

In other words, the wavelet function has a mobile window that is localised near a selected moment in time b and is increased by an increase in the scale of a . The WT possesses a number of properties that enable the construction of the optimal time–frequency representation of nonstationary signals. The compact support of the wavelet function provides the most complete time–frequency information about the signal. The WT has effectively been used for signal processing and data analysis in a variety of applications, such as biomedical signal decomposition (Sun et al., 2000) and remote sensing image fusion (Gu et al., 2010). In addition to the WT, other techniques, such as detrended fluctuation analysis (Varotsos et al., 2007) and empirical mode decomposition (Huang et al., 1998), have been successfully applied to nonlinear and nonstationary signal analyses.

The continuous LIF spectra can be given by a function $f: \lambda \rightarrow I$, where λ is the emission wavelength, and I is the fluorescence intensity at each wavelength. The wavelet spectrum is the set of the coefficients of the WT series $W_{\psi} f(a,b)$. Notably, in this case, the coefficients of the WT series indicate the wavelength and frequency, or more accurately, the wavelength and scale rather than the time and frequency.

An example of the WT of the gas oil LIF spectrum is shown in Figure 4 (bottom). The local features of the LIF spectrum in the plot of the WT coefficients are localised on small scales. For example, the small peaks near 370 nm have a corresponding wavelet representation from 10 to 20. In addition, the local features are well separated from the global base spectrum, which can be found at 50 and above.

The extraction of local features is based on the spectral properties of the examined object. The LIF spectrum of gas oil is a typical example of a characteristic spectrum with small-scale spectral features, such as singularity points at emission wavelengths of 345 and 440 nm, and local maxima at 370, 390 and 405 nm. In Figure 2, another example is presented: the spectrum of the LCO IFO emulsion. This spectrum has a noticeable local feature in the emission from 330 to 390 nm. Thus, one should look for the local curves, bends and singularities of the LIF data and select the corresponding wavelet scales that are free from noise and the global baseline.

The optimal choice of wavelet function is problem dependent. For LIF, we need to accurately extract the patterns that are found on different scales of the LIF data. It is most important that the locations of the extracted wavelength-scale features do not shift during the WT. This constraint is achieved with symmetrical wavelets. We applied functions from the compactly supported, biorthogonal wavelet family; these functions are symmetrical but have finite support and controlled smoothness. Figure 6 shows two functions from the fourth-order biorthogonal wavelet family (bior4.4 in Matlab) utilised in our application.

Structural Analysis of Hyperspectral LIF Data with the WT

The wavelet decomposition of the LIF functions is exploited in a proximity evaluation. At best, the WT preserves all the spectral information and, in addition, extracts information about the spectral details from the LIF data. The entire process can be named “wavelet feature extraction”. This accurate method extends the informational components of the LIF data so that the characteristic features of the LIF functions are preserved and are isolated for future analysis.

The selectivity of the proposed method is demonstrated with a set of five oil products, whose spectral responses are shown in Figure 5 (left). The most complicated case was considered in this experiment. The water pollution models (Fig. 5, right) were created with linear combination coefficients as defined on page 8. The ratios of the fluorescence intensities of the oil products to the DOM were chosen so that the resulting LIF spectra were nearly identical (Fig. 5, right). The task of the analysis system was to detect the pollution by rejecting the clean water matrix and also to determine the correct type of contaminant in an explicit way. The R^2 threshold value was set to 0.95.

The results of the regression- and wavelet-based analyses are shown in Tables 2 and 3, respectively. The columns represent the type of simulated pollution, and the rows correspond to the test models created with the reference library. As previously asserted, the selectivity of the R^2 value calculated from the spectral shape without the capture of localised features is very low. This statement is consistent with the results shown in Table 2. According to the results in Table 2, the similarity for all the simulations and all the models are “very high”. Many erroneous answers are above the 0.99 level (e.g., MDO-Calsol, Water-Sundex). The results show that neither detection nor classification is possible using this analysis.

In the wavelet-based approach, the R^2 value was only calculated for the part of the spectrum from which information was extracted (in the scale domain); thus, irrelevant energy scales were removed from the comparison, and this removal strongly increases the selectivity of the analysis. As previously stated, the selection of the best scales for the analysis depends on the spectral properties of the examined object. The scales were individually selected for

each pollution type considering the noise level in the simulation, which impairs selectivity. Scales from 40 to 50 were selected for the given samples. The lower parts of the scale (30 and lower) with more details were affected by the noise component and were removed from the comparison. The results are shown in Table 3. A definite, correct identification exceeds a threshold value for all the oil products in the table. In all cases, the analysis detected the correct oil product and rejected the clean water matrix (last row in Table 3).

Selectivity of the Analysis

Some notes on the selectivity should be added. We define the selectivity of the analysis to be the ratio of the mean of the similarity values \bar{r} below the selected detection threshold to the threshold level δ :

$$L = 1 - \frac{\bar{r}}{\delta}, \quad \bar{r} = [R^2 | R^2 < \delta], \quad \delta \in (0, 1), \quad L \in (0, 1) \quad (3)$$

We apply the selectivity to compare the results in Tables 2 and 3, where we have small (preferably one) subset of answers above the threshold—the *winners*. For example, the analysis of the gas oil model in Table 2 with $\delta = 0.998$ has $L = 0.029$, whereas, in Table 3, with $\delta = 0.987$, this analysis has $L = 0.7348$; i.e., this value determines the confidence of the analysis for the *winner* sample of a particular test dataset. The selectivity should not be entangled with false-negative and false-positive errors. The special case in which \bar{r} is the empty set is also considered.

The selectivity of the analysis depends on the signal-to-noise ratio (SNR) of the LIF spectrum. For hyperspectral LIF data, the SNR parameter is estimated with the ratio of the signal amplitude to the standard deviation σ_{noise} of the noise. However, SNR_k is defined as

$$SNR_k = \frac{\mu_k}{\sigma_{noise}}, \quad (3)$$

where μ_k is the intensity of the selected spectral component k (e.g., an oil product) in the

composite signal. The relationship between selectivity and SNR_k in the wavelet-based

analysis is shown in Figure 7 for the gas oil model.

The simulation results show that the identification of various oil products in the water matrix with a wavelet-based feature extraction method outperforms the conventional regression analysis in all parameters. The accurate detection and extraction of spectral features that are most specific to the selected oil product are the main features of the wavelet-based approach. The proposed method can consistently identify the correct model with a wide range of noise levels; thus, the detection threshold can be defined through qualitative assessment. Table 3 shows that various oil products can have different selectivity values, which arise from the complexity and uniqueness of the hyperspectral shapes of the oil products.

Hyperspectral detection, the complete spectrum shape, the optimisation of the optical system to improve the SNR and a new analytical approach based on wavelet feature extraction were implemented with the laser remote sensing system FLS® LiDAR; these steps allowed the oil product and the properties of the product to be identified in complicated environmental conditions.

Practical Application Results

These results are an example of the detection and identification of an oil product in difficult environmental conditions (i.e., in an environment with a high DOM content). The data were collected with a shipboard FLS® LiDAR instrument during test trials with controlled oil targets. Two oil products, Sundex and Calsol (LIF spectra are shown in Figure 8 right), and an unknown type of crude oil were attached to pads and were submerged in a pool at a depth of 1 m. The targets were set up in a line, as depicted in Figure 8 (left), and were anchored to the test tank. As the LiDAR moved with the platform, the beam of the LiDAR sequentially interacted with all the presented targets. The sequence of the targets was Sundex, Calsol, crude oil 1 and crude oil 2.

The first task for the analysis system was to detect the presence of a contaminant of unknown nature in the water. The spectral shape of clean water from the pool was recorded previously and was used as a reference for comparison with the measured spectra. The R^2 values show the dissimilarity between the clean water matrix and measurements; thus, detection should be indicated by a rapid decrease in the graph of R^2 . As shown in Figure 8 (right), the spectra of the Sundex film and the sea water matrix (DOM) completely overlap. Consequently, the regression analysis fails to identify submerged Sundex (Fig. 9, left). The other targets (Calsol and crude oil) were still differentiated from the clean water matrix; however, some of the values are very close to the threshold (measurement numbers 57 and 58 for Calsol). With the proposed WT-based method, the noticeable decrease in the R^2 value for contaminated water is observed for each pollutant type (Fig. 9, right).

For the next task, the spectral shapes of the pollutants were measured and were compared with the sample data to identify the pollutant type. A high R^2 value indicates the presence of a known pollutant type. The identification results are shown in Figure 10. For the regression analysis (Fig. 10, left), this task is impossible, and numerous classification errors were demonstrated. The regression method fails to discriminate between oil types in all three cases. The wavelet method successfully classified all the oil products with a minimum of false-negative errors (Fig. 10, right). Although the wavelet method gives false-negative results for some spectra because of the high optical noise, the overall performance and the quality of the analysis are very high. Minor false-negative errors are caused by a high level of optical noise in the FLS-S device, and this noise strongly distorts some spectral shapes.

Wavelet Analysis in Field Experiments

After the new wavelet-based analysis was developed, the method was integrated into the FLS-LiDAR software for online data processing. This technique was experimentally tested in 2011 with the airborne LiDAR system FLS-Am. During an experimental survey of a region of West Virginia, minor pollution on the surface of river water was detected. The observed section of river was 200 km long and 30 m wide and was measured with a ground spatial resolution of 3 m. In total, more than 200,000 spectra were captured from the water surface. Oil film was repeatedly detected on the water surface along the river channel at 36 locations. The oil films at various locations varied in size and were less than 1 μm thick.

We have selected a continuous subset of 60 samples from one location, and we analysed this subset using regression- and wavelet-based methods. The subset samples numbered 20 to 25 were identified as water polluted with diesel oil. Figure 11 shows a continuous plot of the R^2 values calculated using the mentioned methods. The upper plots in Figure 11 represent a comparison of the river water model with the sampled spectra where the

threshold value (red line) was set to 0.95, the default value. The regression analysis results (Fig. 11, top left) demonstrate no visible change in the R^2 values for samples 20-25 compared with the rest of the results. The results from wavelet-based feature extraction (Fig. 11, top right) show a noticeable decrease in the R^2 values for the oil samples. Numerically, these results are standard deviation values for clean water of $R^2 = 0.0022$ and 0.0059 , mean values for clean water of $R^2 = 0.9774$ and 0.9900 and mean values for the oil pollution samples of $R^2 = 0.9669$ and 0.9336 for the regression and wavelet analyses, respectively. The results for the identification of the type of pollutant are shown in the bottom of Figure 11. In this case, the samples were tested with a specific diesel oil spectrum that was measured before the survey. Because of the high optical noise, the diesel component cannot be identified with regression analysis (Fig. 11 bottom left). Wavelet analysis shows that samples numbered 20 to 25 are close to the supplied diesel oil spectrum.

Conclusions

In this paper, we demonstrated the capabilities of hyperspectral LIF for the detection and characterisation of oil pollution with simulated and real-world experimental data. We used the shape of the hyperspectral fluorescence response to identify the type, aggregation and weathering state of petroleum products. In particular, we showed that, although linear regression can be applied to decompose the measured LIF spectrum into a predefined basis of LIF components, the LIF spectral shapes impose a high similarity among themselves if this method involves the comparison of spectral shapes by the R^2 value.

We have demonstrated that this similarity can be eliminated through use of the detailed shape information of LIF spectra. To extract the relevant features of the hyperspectral shapes, we applied a continuous WT. The selectivity of the classification in the simulated experiments was, on average, 20 times greater with the WT. The proposed analytical method showed high performance in real-world examples of the detection and identification of oil products with hyperspectral, airborne LIF LiDAR in difficult environmental conditions. The approach was successfully applied to the detection and identification of oil pollution from submerged targets and to the detection of minor oil pollution on the surface of river water with a high DOM content. The developed method of hyperspectral LIF data analysis will enhance the analytical capabilities of airborne or ship-borne FLS® LiDAR to provide nearly real-time, large-area characterisation of oil pollution in complex marine environments.

Acknowledgements

We gratefully acknowledge the technical and application development and field work performed by the staff of Laser Diagnostic Instruments in Estonia, Canada and the USA.

References

- Alekseyev, V.; Babichenko, S.; Lisin, A.; Lapimaa, J.; Poryvkina, L.; Shchemelyov, S.; Sobolev, I. 2010. Airborne Surveillance of Water Basins with Hyperspectral FLS-Lidar. *SPIE Remote Sensing*, 20-23 september 2010, Toulouse, France.
- Babichenko, S.; Dudelzak, A.; Lapimaa, J.; Lisin, A.; Poryvkina, L.; Vorobiev, A. 2006. Locating Water Pollution and Shore Discharges in Coastal Zone and Inland Waters with FLS Lidar. *EARSeL eProceedings*, No. 5.1, pp. 32-41.

- Brekke, C.; Solberg, A. 2005. Oil Spill Detection by Satellite Remote Sensing. *Remote Sensing of Environment*, No. 95, pp. 1–13.
- Brown, C.E.; Fingas, M.F. 2003. Review of the Development of Laser Fluorosensors for Oil Spill Application. *Marine Pollution Bulletin*. Vol. 47, pp. 477-484.
- Chen, C.H. 2006. *Signal and Image Processing for Remote Sensing*, CRC Press, ISBN 9780849350917.
- Chui, C.K. 1992. *Wavelets: a Tutorial in Theory and Applications*, Academic Press, ISBN 9780121745905.
- Daubechies, I. 1992. Ten Lectures on Wavelets. *CBMS-NSF Regional Conference Series in Applied Mathematics*, ISBN-10: 0898712742, ed.1.
- Gu, H.Y.; Li, H.T.; Yan, Q.; Han, Y.S. 2010. Object-Based Image Fusion Method based on Wavelet and PCA for Remote Sensing Imagery. In *GEOBIA 2010-Geographic Object-Based Image Analysis*. Ghent University, Ghent, Belgium, 29 June – 2 July. ISPRS Vol.No. XXXVIII-4/C7, Archives ISSN No 1682-1777.
- Huang, N. E.; Shen, Z.; Long, S. R.; Wu, M. C.; Shih, H. H.; Zheng, Q.; Yen, N.-C.; Tung, C. C.; Liu, H. H. 1998. The Empirical Mode Decomposition and The Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Royal Society of London Proceedings Series A*, vol. 454, Issue 1971, p.903.
- Jha, M.N.; Levy, J.; Gao, Y. 2008. Advances in Remote Sensing for Oil Spill Disaster Management: State-of-the-Art Sensors Technology for Oil Spill Surveillance. *Sensors*, No. 8, pp. 236-255.
- Landgrebe, D.A. 2003. *Signal Theory Methods in Multispectral Remote Sensing*, Wiley-Interscience, ISBN 9780471420286.
- Lennon, M.; Babichenko, S.; Thomas, N.; Mariette, V.; Mercier, G.; Lisin, A. 2006. Detection and Mapping of Oil Slicks in the Sea by Combined Use of Hyperspectral Imagery and Laser Induced Fluorescence. *EARSeL eProceedings*, No. 5, pp. 1-9.
- NASA/Jet Propulsion Laboratory. 2010. NASA Satellite Views Massive Gulf Oil Spill. *ScienceDaily* [online]. Available from <http://www.sciencedaily.com/releases/2010/05/100504114957.htm> [cited October 16, 2012].
- Pavlidis, T. 1977. *Structural Pattern Recognition*, Springer-Verlag.
- Poryvkina, L.; Babichenko, S.; Davydova, O. 1998. SFS Characterization of Oil Pollution in Natural Water. In *Proc. V Int. Conference Remote Sensing for Marine and Coastal Environments*, San Diego, CA, 5-7 October 1998, vol.2., pp. 520 - 524.
- Sobolev, I. 2010. FLS® LiDAR Systems and LIF Spectra Analysis. In *Info- ja kommunikatsioonitehnoloogia doktorikooli IV aastakonverents*, 26-27 november 2010, Essu mõis, Estonia. IKT doktorikool, Tallinn. pp. 129-132.
- Sun, M.; Scheuer, M.L.; ScLabassi, R.J. 2000. Decomposition of Biomedical Signals for Enhancement of Their Time-frequency Distributions. *Journal of The Franklin Institute*, Vol. 337, No. 4, pp. 453-467(15).
- Varotsos, C.; Assimakopoulos, M.N.; Efstathiou, M. 2007. Technical Note: Long-term memory effect in the atmospheric CO₂ concentration at Mauna Loa, *Atmospheric Chemistry and Physics*, No. 7, pp. 629-634.

List of Figures

Figure 1. Examples of hyperspectral LIF spectra. Sea diesel has an emission maximum at 357 nm, heating oil has a maximum at 407 nm, and marine diesel oil (MDO) has a maximum at 411 nm.

Figure 2. Hyperspectral LIF examples of the film and emulsion states: gas oil (left) in emulsified state has a maximum shifted toward the UV range with a minor change in shape; LCO IFO (right) in the emulsified state has a relative intensity increase at 370 nm with significant changes in the spectral shapes. The large amount of oil makes the medium completely opaque, which explains the absence of the Raman line.

Figure 3. Example of the LIF spectral decomposition with the regression analysis: (left) - model spectrum of an oil film of sea diesel on water, $R^2 = 0.9965$; (middle) - clean water, $R^2 = 0.9233$; (right) – model spectrum of sea water with Calsol, $R^2 = 0.9697$.

Figure 4. LIF spectrum of gas oil (top). Wavelet representation (bottom).

Figure 5. (Left) – LIF spectra of five oil products: sea diesel, gas oil, Calsol, MDO and Sundex; (Right) – modelled spectra of five oil products in water matrices. (Bottom) – scalograms of the spectra of the oil products.

Figure 6. Biorthogonal, fourth-order wavelet and scaling functions: (a) analytical scaling function, (b) analytical wavelet.

Figure 7. Relation of selectivity, L , to the SNR value for the gas oil model (wavelet-based analysis).

Figure 8. Two submerged oil targets at a depth of 1 m (left). Spectra of Calsol and Sundex oil films on the water surface with a DOM measured by a FLS-S LiDAR system (right).

Figure 9. Results of “the finding of unknowns” analysis—the evaluation of the similarity between the clean water matrix and the measured spectra. R^2 values for the classical regression analysis; the plot indicates false-negative errors for the Sundex targets (left). The WT-based method reported no errors (right). The dissimilarity threshold (red line) is set to 0.95.

Figure 10. Results for the identification of the pollution type. (left) - conventional regression analysis; (right) – wavelet-based method; (top) – detection of Sundex oil product; (middle) – detection of Calsol; (bottom) – detection of crude oil. The dissimilarity threshold (dashed line) is set to 0.95.

Figure 11. (Top) – Analytical results for the clean water matrix; (left) – classical regression analysis; (right) – wavelet-based method. The dissimilarity threshold (red line) is set to 0.95. The spectra with numbers from 20 to 25 are marked as the “pollution area”. This mark means that dissimilarities were found in comparison with the clean water matrix. (Bottom) – Analysis results for the identification of diesel oil. The wavelet-based method identified pollution in samples 20 to 25.

Figure 12. Two averaged spectra for the observed water medium with and without oil pollution. The diesel oil film caused a small variation in the spectrum shape in the UV area from 350 to 400 nm.

Fig. 1.

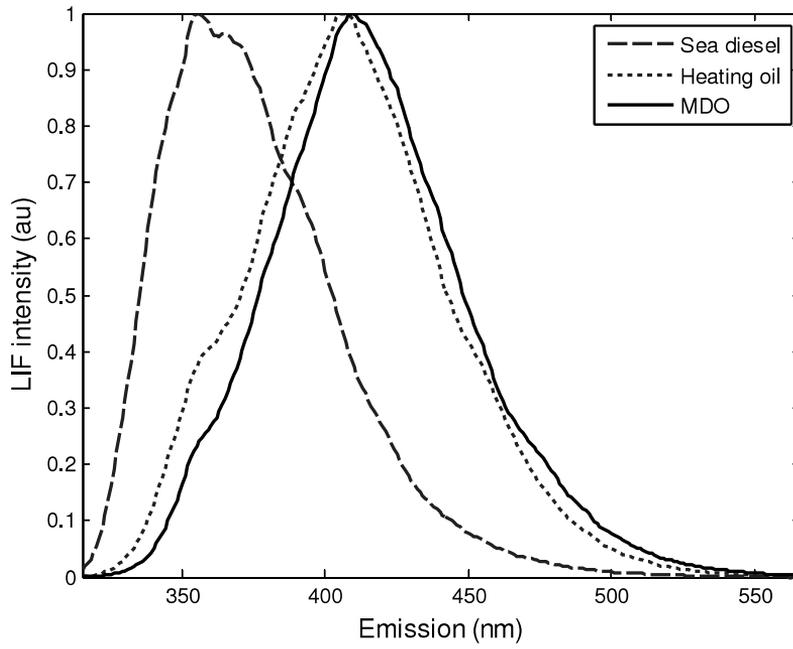


Fig. 2.

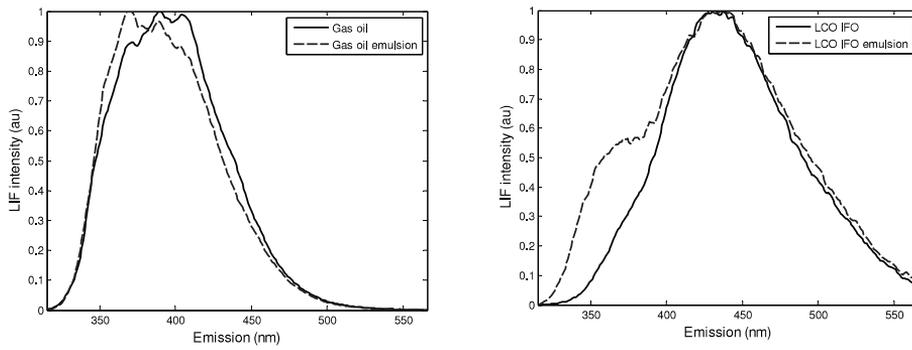


Fig. 3.

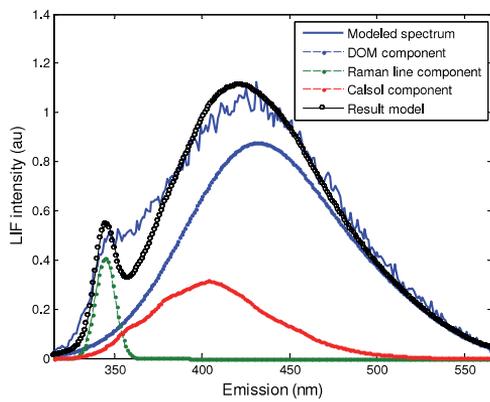
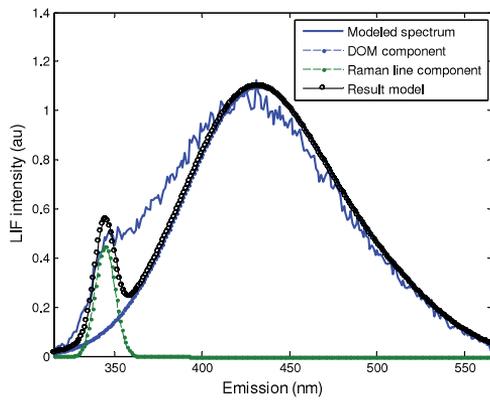
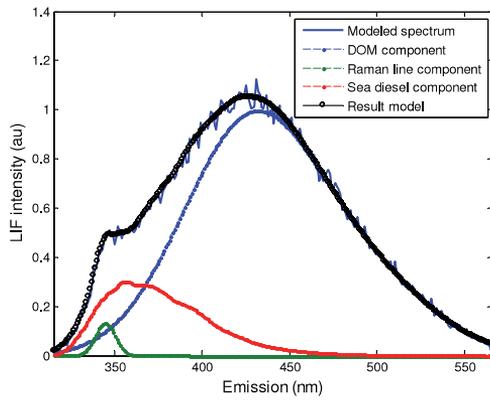


Fig. 4.

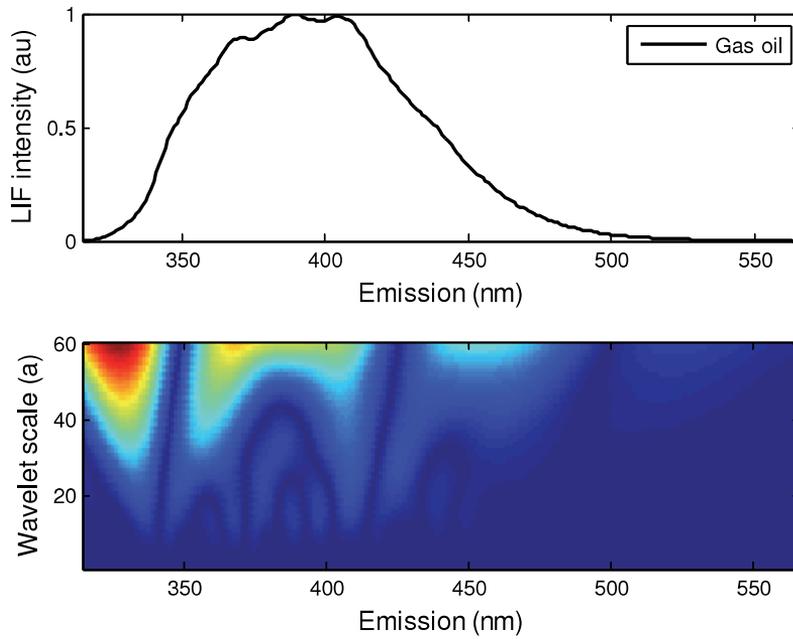
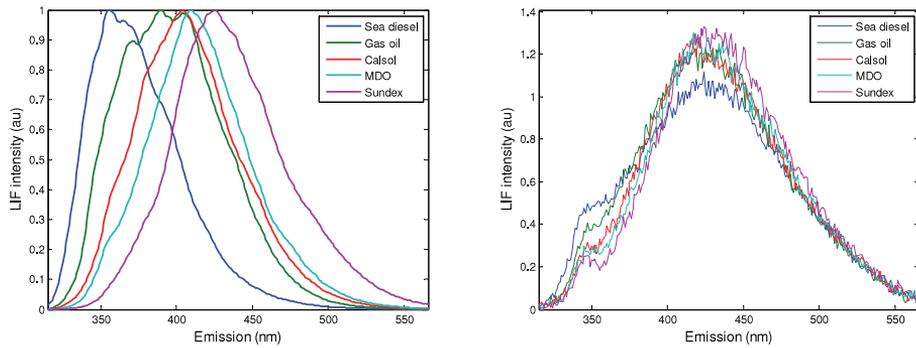


Fig. 5.



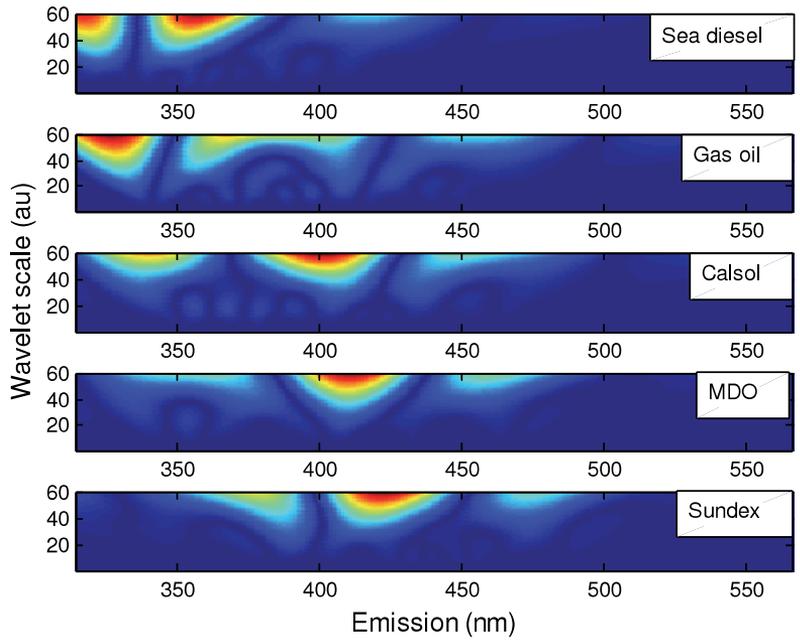


Fig. 6.

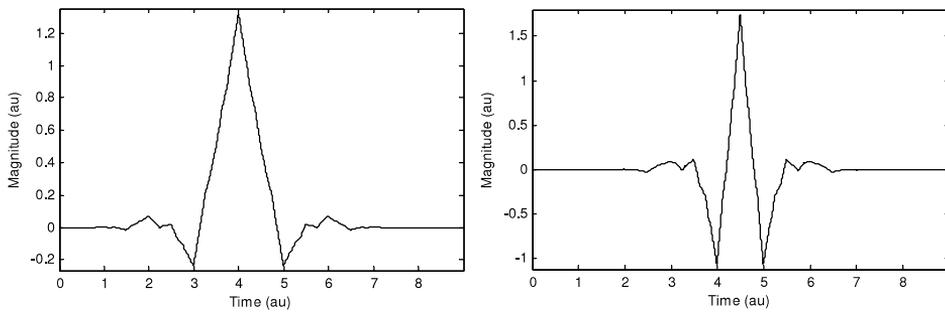


Fig. 7.

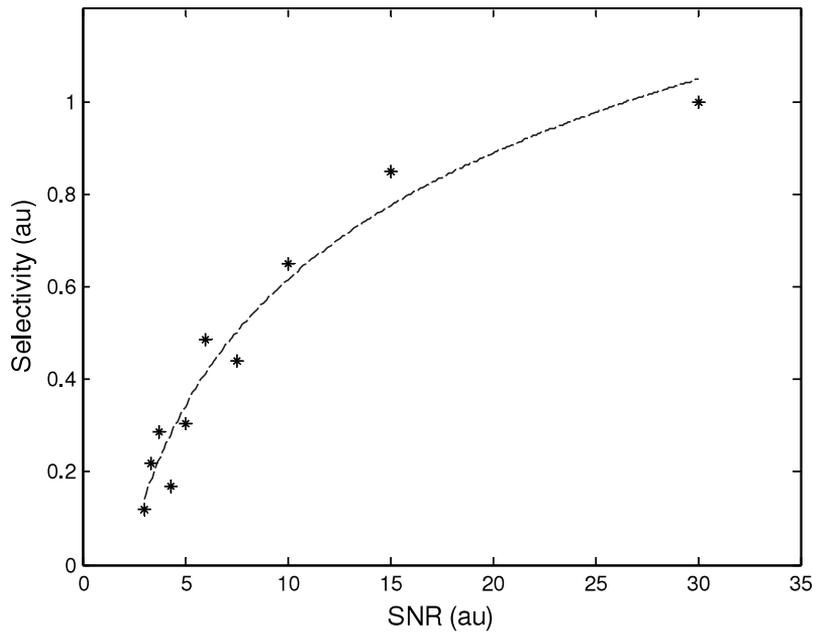


Fig. 8.

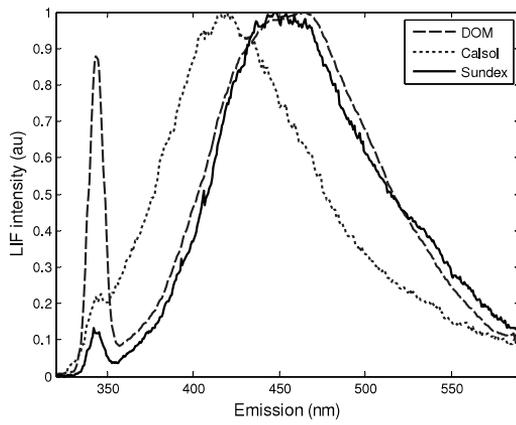
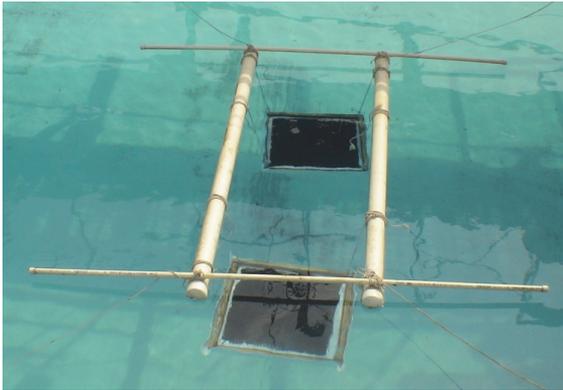


Fig. 9.

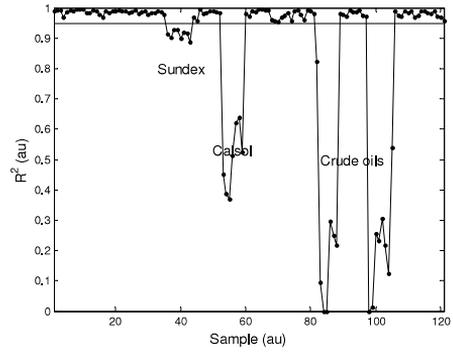
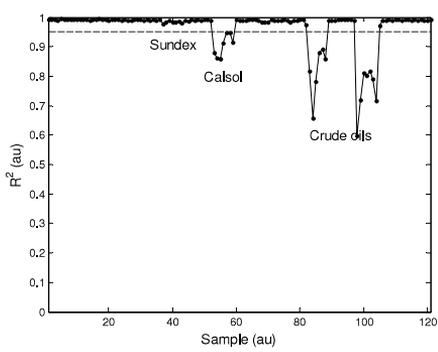


Fig. 10.

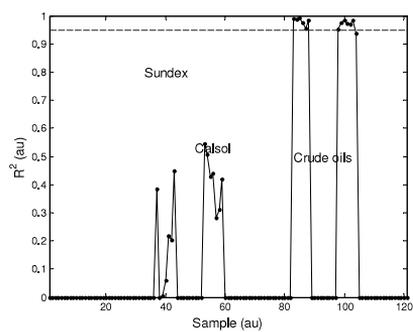
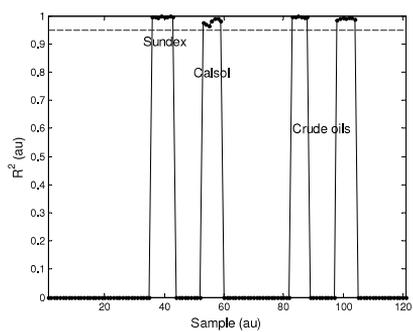
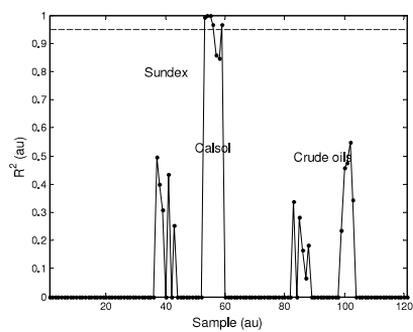
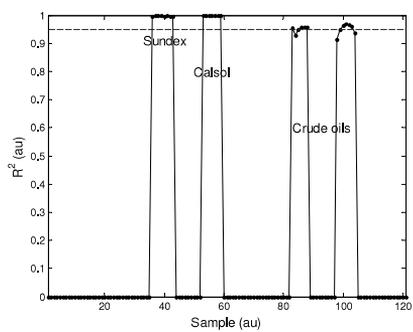
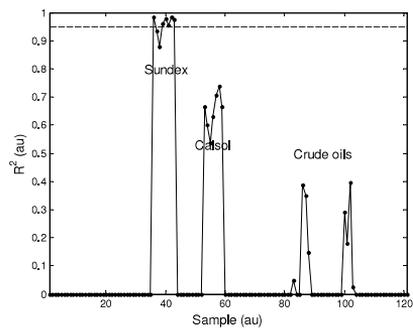
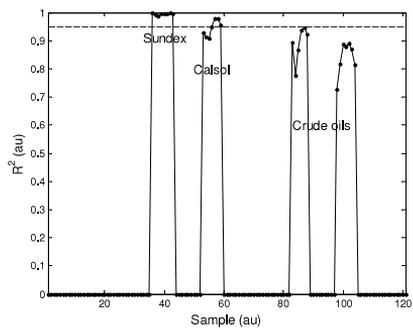


Fig. 11.

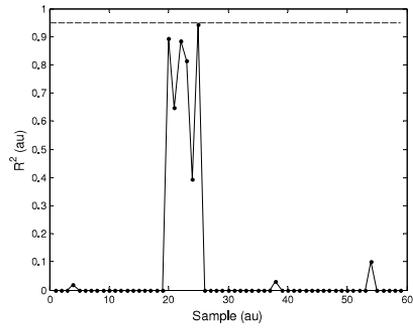
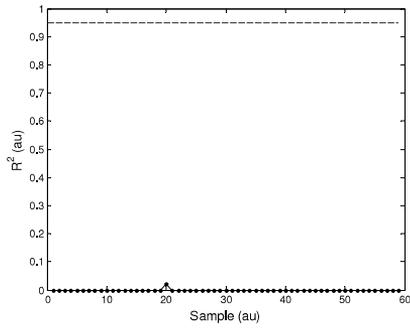
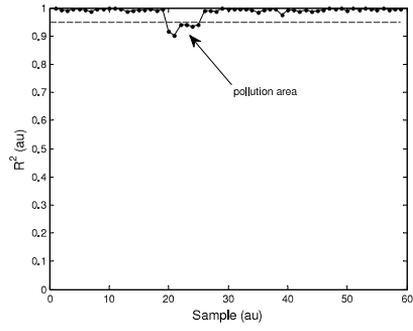
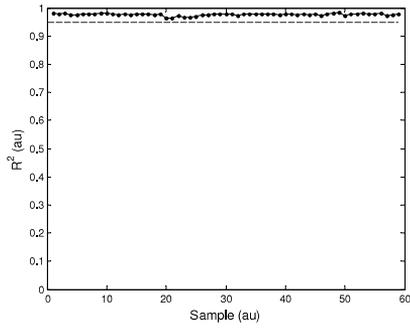


Fig. 12.

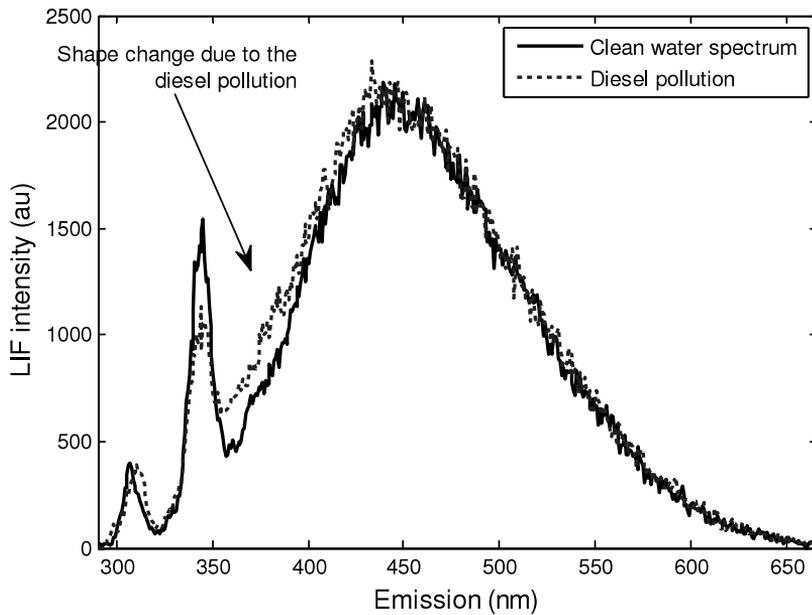


Table 1. The application of a qualitative measure of the coefficient of determination with the following table to pre-set the similarity threshold in the LIF analysis is useful. The table is also known as the Chaddock scale.

Determination coefficient	Qualitative characteristic
0.1-0.3	Weak
0.3-0.5	Moderate
0.5-0.7	Salient
0.7-0.9	High
0.9-0.99	Very high

Table 2. The R^2 values for the regression analysis results for the identification and detection of oil in the water matrix. The columns represent pollution type. The rows are the tested models from the reference library. The selectivity L value can be calculated for each pollution type via equation (3).

	Sea diesel	Gas oil	MDO	Calsol	Sundex
Sea diesel	0.997	0.986	0.982	0.982	0.994
Gas oil	0.984	0.998	0.991	0.995	0.994
MDO	0.951	0.985	0.998	0.995	0.995
Calsol	0.969	0.994	0.996	0.998	0.994
Sundex	0.919	0.942	0.984	0.968	0.998
Water	0.919	0.938	0.976	0.962	0.994

Table 3. The R^2 values for wavelet feature extraction-based analysis results. For all modelled spectra, the SNR_k ratio of the oil product component is 10. The columns represent the pollution type. The rows are the tested models from the reference library. The selectivity value L can be calculated for each pollution type via equation (3).

	Sea diesel	Gas oil	MDO	Calsol	Sundex
Sea diesel	0.982	0.649	0.000	0.000	0.000
Gas oil	0.031	0.987	0.000	0.241	0.000
MDO	0.000	0.053	0.975	0.642	0.000
Calsol	0.000	0.607	0.315	0.964	0.000
Sundex	0.000	0.000	0.552	0.000	0.978
Water	0.000	0.000	0.000	0.000	0.000

**DISSERTATIONS DEFENDED AT
TALLINN UNIVERSITY OF TECHNOLOGY ON
INFORMATICS AND SYSTEM ENGINEERING**

1. **Lea Elmik**. Informational Modelling of a Communication Office. 1992.
2. **Kalle Tammemäe**. Control Intensive Digital System Synthesis. 1997.
3. **Eerik Lossmann**. Complex Signal Classification Algorithms, Based on the Third-Order Statistical Models. 1999.
4. **Kaido Kikkas**. Using the Internet in Rehabilitation of People with Mobility Impairments – Case Studies and Views from Estonia. 1999.
5. **Nazmun Nahar**. Global Electronic Commerce Process: Business-to-Business. 1999.
6. **Jevgeni Riipulk**. Microwave Radiometry for Medical Applications. 2000.
7. **Alar Kuusik**. Compact Smart Home Systems: Design and Verification of Cost Effective Hardware Solutions. 2001.
8. **Jaan Raik**. Hierarchical Test Generation for Digital Circuits Represented by Decision Diagrams. 2001.
9. **Andri Riid**. Transparent Fuzzy Systems: Model and Control. 2002.
10. **Marina Brik**. Investigation and Development of Test Generation Methods for Control Part of Digital Systems. 2002.
11. **Raul Land**. Synchronous Approximation and Processing of Sampled Data Signals. 2002.
12. **Ants Ronk**. An Extended Block-Adaptive Fourier Analyser for Analysis and Reproduction of Periodic Components of Band-Limited Discrete-Time Signals. 2002.
13. **Toivo Paavle**. System Level Modeling of the Phase Locked Loops: Behavioral Analysis and Parameterization. 2003.
14. **Irina Astrova**. On Integration of Object-Oriented Applications with Relational Databases. 2003.
15. **Kuldar Taveter**. A Multi-Perspective Methodology for Agent-Oriented Business Modelling and Simulation. 2004.
16. **Taivo Kangilaski**. Eesti Energia käiduhaldussüsteem. 2004.
17. **Artur Jutman**. Selected Issues of Modeling, Verification and Testing of Digital Systems. 2004.
18. **Ander Tenno**. Simulation and Estimation of Electro-Chemical Processes in Maintenance-Free Batteries with Fixed Electrolyte. 2004.

19. **Oleg Korolkov**. Formation of Diffusion Welded Al Contacts to Semiconductor Silicon. 2004.
20. **Risto Vaarandi**. Tools and Techniques for Event Log Analysis. 2005.
21. **Marko Koort**. Transmitter Power Control in Wireless Communication Systems. 2005.
22. **Raul Savimaa**. Modelling Emergent Behaviour of Organizations. Time-Aware, UML and Agent Based Approach. 2005.
23. **Raido Kurel**. Investigation of Electrical Characteristics of SiC Based Complementary JBS Structures. 2005.
24. **Rainer Taniloo**. Ökonoomsete negatiivse diferentsiaaltakistusega astmete ja elementide disainimine ja optimeerimine. 2005.
25. **Pauli Lallo**. Adaptive Secure Data Transmission Method for OSI Level I. 2005.
26. **Deniss Kumlander**. Some Practical Algorithms to Solve the Maximum Clique Problem. 2005.
27. **Tarmo Veskiõja**. Stable Marriage Problem and College Admission. 2005.
28. **Elena Fomina**. Low Power Finite State Machine Synthesis. 2005.
29. **Eero Ivask**. Digital Test in WEB-Based Environment 2006.
30. **Виктор Войтович**. Разработка технологий выращивания из жидкой фазы эпитаксиальных структур арсенида галлия с высоковольтным р-п переходом и изготовления диодов на их основе. 2006.
31. **Tanel Alumäe**. Methods for Estonian Large Vocabulary Speech Recognition. 2006.
32. **Erki Eessaar**. Relational and Object-Relational Database Management Systems as Platforms for Managing Softwareengineering Artefacts. 2006.
33. **Rauno Gordon**. Modelling of Cardiac Dynamics and Intracardiac Bio-impedance. 2007.
34. **Madis Listak**. A Task-Oriented Design of a Biologically Inspired Underwater Robot. 2007.
35. **Elmet Orasson**. Hybrid Built-in Self-Test. Methods and Tools for Analysis and Optimization of BIST. 2007.
36. **Eduard Petlenkov**. Neural Networks Based Identification and Control of Nonlinear Systems: ANARX Model Based Approach. 2007.
37. **Toomas Kirt**. Concept Formation in Exploratory Data Analysis: Case Studies of Linguistic and Banking Data. 2007.
38. **Juhan-Peep Ernits**. Two State Space Reduction Techniques for Explicit State Model Checking. 2007.

39. **Innar Liiv**. Pattern Discovery Using Seriation and Matrix Reordering: A Unified View, Extensions and an Application to Inventory Management. 2008.
40. **Andrei Pokatilov**. Development of National Standard for Voltage Unit Based on Solid-State References. 2008.
41. **Karin Lindroos**. Mapping Social Structures by Formal Non-Linear Information Processing Methods: Case Studies of Estonian Islands Environments. 2008.
42. **Maksim Jenihhin**. Simulation-Based Hardware Verification with High-Level Decision Diagrams. 2008.
43. **Ando Saabas**. Logics for Low-Level Code and Proof-Preserving Program Transformations. 2008.
44. **Ilja Tšahhиров**. Security Protocols Analysis in the Computational Model – Dependency Flow Graphs-Based Approach. 2008.
45. **Toomas Ruuben**. Wideband Digital Beamforming in Sonar Systems. 2009.
46. **Sergei Devadze**. Fault Simulation of Digital Systems. 2009.
47. **Andrei Krivošei**. Model Based Method for Adaptive Decomposition of the Thoracic Bio-Impedance Variations into Cardiac and Respiratory Components. 2009.
48. **Vineeth Govind**. DfT-Based External Test and Diagnosis of Mesh-like Networks on Chips. 2009.
49. **Andres Kull**. Model-Based Testing of Reactive Systems. 2009.
50. **Ants Torim**. Formal Concepts in the Theory of Monotone Systems. 2009.
51. **Erika Matsak**. Discovering Logical Constructs from Estonian Children Language. 2009.
52. **Paul Annus**. Multichannel Bioimpedance Spectroscopy: Instrumentation Methods and Design Principles. 2009.
53. **Maris Tõnso**. Computer Algebra Tools for Modelling, Analysis and Synthesis for Nonlinear Control Systems. 2010.
54. **Aivo Jürgenson**. Efficient Semantics of Parallel and Serial Models of Attack Trees. 2010.
55. **Erkki Joasoon**. The Tactile Feedback Device for Multi-Touch User Interfaces. 2010.
56. **Jürgo-Sören Preden**. Enhancing Situation – Awareness Cognition and Reasoning of Ad-Hoc Network Agents. 2010.
57. **Pavel Grigorenko**. Higher-Order Attribute Semantics of Flat Languages. 2010.
58. **Anna Rannaste**. Hierarcical Test Pattern Generation and Untestability Identification Techniques for Synchronous Sequential Circuits. 2010.

59. **Sergei Strik**. Battery Charging and Full-Featured Battery Charger Integrated Circuit for Portable Applications. 2011.
60. **Rain Ottis**. A Systematic Approach to Offensive Volunteer Cyber Militia. 2011.
61. **Natalja Sleptšuk**. Investigation of the Intermediate Layer in the Metal-Silicon Carbide Contact Obtained by Diffusion Welding. 2011.
62. **Martin Jaanus**. The Interactive Learning Environment for Mobile Laboratories. 2011.
63. **Argo Kasemaa**. Analog Front End Components for Bio-Impedance Measurement: Current Source Design and Implementation. 2011.
64. **Kenneth Geers**. Strategic Cyber Security: Evaluating Nation-State Cyber Attack Mitigation Strategies. 2011.
65. **Riina Maigre**. Composition of Web Services on Large Service Models. 2011.
66. **Helena Kruus**. Optimization of Built-in Self-Test in Digital Systems. 2011.
67. **Gunnar Pihõ**. Archetypes Based Techniques for Development of Domains, Requirements and Software. 2011.
68. **Juri Gavšin**. Intrinsic Robot Safety Through Reversibility of Actions. 2011.
69. **Dmitri Mihhailov**. Hardware Implementation of Recursive Sorting Algorithms Using Tree-like Structures and HFSM Models. 2012.
70. **Anton Tšertov**. System Modeling for Processor-Centric Test Automation. 2012.
71. **Sergei Kostin**. Self-Diagnosis in Digital Systems. 2012.
72. **Mihkel Tagel**. System-Level Design of Timing-Sensitive Network-on-Chip Based Dependable Systems. 2012.
73. **Juri Belikov**. Polynomial Methods for Nonlinear Control Systems. 2012.
74. **Kristina Vassiljeva**. Restricted Connectivity Neural Networks based Identification for Control. 2012.
75. **Tarmo Robal**. Towards Adaptive Web – Analysing and Recommending Web Users` Behaviour. 2012.
76. **Anton Karputkin**. Formal Verification and Error Correction on High-Level Decision Diagrams. 2012.
77. **Vadim Kimlaychuk**. Simulations in Multi-Agent Communication System. 2012.
78. **Taavi Viilukas**. Constraints Solving Based Hierarchical Test Generation for Synchronous Sequential Circuits. 2012.

79. **Marko Kääramees.** A Symbolic Approach to Model-based Online Testing. 2012.
80. **Enar Reilent.** Whiteboard Architecture for the Multi-agent Sensor Systems. 2012.
81. **Jaan Ojarand.** Wideband Excitation Signals for Fast Impedance Spectroscopy of Biological Objects. 2012.
82. **Igor Aleksejev.** FPGA-based Embedded Virtual Instrumentation. 2013.
83. **Juri Mihhailov.** Accurate Flexible Current Measurement Method and its Realization in Power and Battery Management Integrated Circuits for Portable Applications. 2013.
84. **Tõnis Saar.** The Piezo-Electric Impedance Spectroscopy: Solutions and Applications. 2013.
85. **Ermo Täks.** An Automated Legal Content Capture and Visualisation Method. 2013.
86. **Uljana Reinsalu.** Fault Simulation and Code Coverage Analysis of RTL Designs Using High-Level Decision Diagrams. 2013.
87. **Anton Tšepurov.** Hardware Modeling for Design Verification and Debug. 2013.
88. **Ivo Mürsepp.** Robust Detectors for Cognitive Radio. 2013.
89. **Jaas Ježov.** Pressure sensitive lateral line for underwater robot. 2013.
90. **Vadim Kaparin.** Transformation of Nonlinear State Equations into Observer Form. 2013.
92. **Reeno Reeder.** Development and Optimisation of Modelling Methods and Algorithms for Terahertz Range Radiation Sources Based on Quantum Well Heterostructures. 2014.
93. **Ants Koel.** GaAs and SiC Semiconductor Materials Based Power Structures: Static and Dynamic Behavior Analysis. 2014.
94. **Jaan Übi.** Methods for Coepetition and Retention Analysis: An Application to University Management. 2014.