UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di: Geoscienze

SCUOLA DI DOTTORATO DI RICERCA IN : Scienze Della Terra

CICLO XXVII

LASERSCANNER CYCLOSTRATIGRAPHY.
A REMOTE SENSING APPROACH FOR THE EXTRACTION OF LONG TIME SERIES FROM LARGE OUTCROPS

Direttore della Scuola: Ch.mo Prof. Massimiliano Zattin

Supervisore: Ch.mo Prof. Nereo Preto

Correlatore: Dott. Marco Franceschi

Dottorando: Dott. Luca Penasa

Luca Penasa: ***Laserscanner cyclostratigraphy,*** *A remote sensing approach for the extraction of long time series from large outcrops,* January 2015

luca.penasa@gmail.com

SUPERVISORS:
Nereo Preto
Marco Franceschi

ABSTRACT

Terrestrial Laser Scanners (TLSs) permit to capture three dimensional models of outcrops in the form of point clouds. Each point of a point cloud is the result of a sampling operation on the outcrop's surface, made trough a laser beam. This operation records the 3D coordinates of the point and the backscattered laser energy as an *intensity* value. Potentially, the intensity can be converted into a *reflectance* and used to discriminate different materials.

When series composed of limestone and marl alternations are considered, TLS intensity can be used as a *proxy* for the lithology and converted into *intensity-logs* which were demonstrated to be a *promising* source of time series for cyclostratigraphic analysis.

This thesis started from that result and had the main goal of exploiting that method to produce *long* time series, which are essential to the study of long period ($>$ 1 Myr) Milankovitch cycles in sediments. In this perspective the following themes were investigated: *a*) The effect exerted on measured intensities by shales and chert. Limestone, clay (shales) and chert make most of many deep water sedimentary successions. *b*) The identification of a simple method to normalize the intensities, to minimize the effects of distance from the outcrop and of the incidence angle of the laser beam. *c*) The creation of a software package, composed by a `C++` library and a Graphical User Interface (GUI) for simplifying the user interaction with the data that is needed for generating the time series.

Three case studies from the Central-Italy Apennines have been considered:

I) The Smirra section (*Scaglia Rossa Fm. and Scaglia Variegata Fm.*), composed of pelagic calcareous homogenites was the playgroud to compare TLS intensities to calcimetric analyses carried out on samples taken from the outcrop. Results demonstrate that TLS can be used as a proxy for $CaCO_3$ content even in series characterized by minimal lithological variations.

II) The Mulini section (*Maiolica Fm.*). TLS sensitivity to chert was investigated by comparing laboratory-measured reflectance spectra to TLS intensity. It is shown that the low-reflectance of chert can be exploited to distinguish it from limestone. A method based on a Support Vector Machine (SVM)

classifier was thus implemented to recognize chert semi-automatically on TLS point clouds.

III) The Vispi Quarry (*Maiolica Fm., Marne a Fucoidi Fm., Scaglia Bianca Fm.*). In this outcrop an almost continuous 200m-thick stratigraphic succession is exposed spanning from the upper Maiolica Fm. to the Bonarelli level, and representing ca. 20 Myrs. This outcrop was ideal to tackle the problem of retrieving long time series for cyclostratigraphic analysis. An original method and dedicated software were developed to achieve this task. With these original tools, it was possible to produce a 150m-long time series with resolution down to the centimeter, starting from ~ 30 point clouds.

The methods and algorithms introduced to cope with the long time series creation from point clouds have been implemented in a `C++` library, names `SPC`. Easy access to the data structures and methods defined in `SPC` is instead provided by a GUI, in the form of a toolbar for the CloudCompare software. The proposed toolkit is available over the internet at `https://github.com/luca-penasa`.

SOMMARIO

I Terrestrial Laser Scanners (TLSs) permettono di creare modelli tridimensionali di affioramenti, nella forma di nuvole di punti. Ogni punto di una nuvola di punti è il risultato di un'operazione di campionamento sulla superficie dell'affioramento, fatta usando un raggio laser. Questa operazione registra le coordinate 3D del punto e l'energia retroriflessa del laser, detta *intensità*. Potenzialmente, l'intensità può essere convertita in una *riflettanza* ed essere usata per discriminare matteriali differenti.

Quando si considera una serie composta da alternanze di calcari e marne, l'intensità può essere usata come *proxy* per la litologia e venir convertita in *log di intensità*, questi si sono dimostrate essere *promettenti* serie temporali per l'analisi ciclostratigrafica.

Questa tesi prende il via da questo risultato, ed ha avuto l'obiettivo principale di esplorare i metodi necessari a produrre serie temporali *lunghe*, che sono essenziali per studiare cicli Milankoviani di lungo periodo ($>$ 1 Myr) nei sedimenti. In questa prospettiva le seguenti tematiche sono state sviluppate:

a) L'effetto di argilliti e selci sull'intensità misurata. Calcare, argilla e selce formano infatti la maggior parte dei sedimenti nelle successioni di acqua profonda. *b*) L'identificazione di un metodo semplificato per la normalizzazione delle intensità, per minimizzare gli effetti della distanza dall'affioramento e dell'angolo di incidenza del raggio laser. *c*) La creazione di un pacchetto software, composto da una libreria `C++` e da una Graphical User Interface (GUI) per semplificare l'interazione dell'utente con i dati, che è necessaria per generare le serie temporali.

Tre casi studio dagli Appennini dell'Italia Centrale sono stati considerati:

I) La sezione di Smirra (*Scaglia Rossa Fm. e Scaglia Variegata Fm.*), composta da una *omogenite* calcarea pelagica, questo caso è stato usato per comparare le intensità del TLS a calcimetrie ottenute da campioni dell'affioramento. I risultati dimostrano che il TLS puó essere usato come proxy per il contenuto in $CaCO_3$, anche quando la serie è caratterizzata da variazioni litologiche minime.

II) La sezioni dei Mulini (*Maiolica Fm.*). La sensitività alla selce è stata investigata, comparando misure di riflettanza ottenute in laboratorio con

l'intensità del TLS. Si dimostra che la bassa riflettanza della selce può essere impiegata per distiguerla dal calcare. Un metodo basato su un classificatore Support Vector Machine (SVM) è stato implementato per permettere il riconoscimento semi automatico della selce sulle nuvole di punti da TLS.

III) La Cava Vispi (*Maiolica Fm., Marne a Fucoidi Fm., Scaglia Bianca Fm.*). In questo affiormanento è esposta una sezione stratigrafica continua di quasi 200m in spessore, che va dalla parte superiore della Maiolica Fm. fino al Livello Bonarelli, e rappresenta ca. 20 Myrs. Questo affioramento ha fornito il caso ideale per affrontare il problema di ottenere serie temporali lunghe per le analisi ciclostratigrafiche. Una metodologia originale e un software dedicato sono stati sviluppati per questo compito. Con questi strumenti è stato possibile produrre una serie di 150m a risoluzione centimetrica, partendo da ~ 30 nuvole di punti.

I metodi e gli algoritmi introdotti per l'estrazione di serie temporali da nuvole di punti sono stati implementati in una libreria `C++`, detta `SPC`. Un accesso facilitato alle strutture dati e ai metodi definiti in `SPC` viene invece fornito da una GUI, sottoforma di una *toolbar* per il software CloudCompare. Il toolkit proposto è disponibile in internet all'indirizzo `https://github.com/luca-penasa`.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

Contents

1.1 LASER SCANNIG DEVICES

A Terrestrial Laser Scanner (TLS) device is basically the upgraded version of a *laser rangefinder*. A laser rangefinder is a device which uses a narrow laser beam to measure the distance to a surface.

Depending on the design, a laser rangefinder would normally work by either:

1. measuring the Time Of Flight (TOF) of a laser impulse. The TOF is then converted into a distance (given the speed of light is known), or
2. by using an amplitude modulated continuous laser beam. In this case the distance is computed as difference in phase of the return signal, with respect of the emitted one.

Irrespective of the method used, the laser beam reaches the target surface and is then diffused in all directions. One of the scattered rays will eventually come back to the receiver, triggering the distance measurement.

Figure 1.1: OPTECH ILRIS 3D terrestrial laser scanner during scanning.

Coupling a laser rangefinder with a mechanical and optical directional system (made by electro-mechanical actuators, mirrors and/or prisms), which controls the direction of the laser beam, the system can be used to measure many distances in an automatic way, thanks to the controlled steering of the laser beam, followed by a range determination.

This upgrade lead to a new type of rangefinders, improved devices that can automatically collect a large number of measures. These devices are known as *profilometers* (see Figure 1.2) when they are mainly engineered to take distance measures on a line, but in the more general case the device is able to steer the laser beam on two independent axis, permitting to obtain a complete 3D object scanning, generally known as laser scanners.

When laser scanning devices are used from ground they are known as Terrestrial Laser Scanners (TLSs), in contrast to their airborne counterpart, called Airborne Laser Scanners (ALSs).

For example, considering a generic (and highly simplified) TLS, the direction of the laser beam could be expressed as a couple of angles (ϑ and φ), which together with the distance d define a point in a spherical coordinate system which can be eventually converted in cartesian coordinates:

$$x = d \sin\vartheta \cos\varphi, \quad y = d \sin\vartheta \sin\varphi, \quad z = d \cos\vartheta \tag{1.1}$$

The formulas are here diagrammatically simplified, and it would comprise more complex computations in a real-world application (i. e. terms accounting for specific design characteristics of the device would be introduced into Equation 1.1).

Each new measure made will eventually produce a triplet of values $[x, y, z]$, and thus a point in an euclidean reference frame.

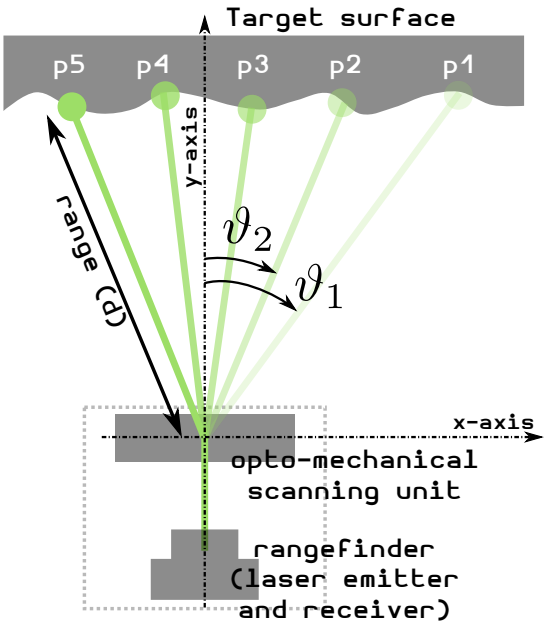


Figure 1.2: Diagrammatic representation of the working principle of a laser scanning system, a *profilometer*, in this case. The direction of the laser beam is controlled by an opto-mechanical system which steers the laser beam. This allows the system to measure the distance to the surface at incremental values of ϑ, implementing a scanning pattern. For each measure, a distance (d) together with an angle (ϑ) are recorded. These values can be converted to an euclidean reference system, centered in the device, obtaining the $[x, y]$ couples for each point (p1 to p5 in figure). This concept can be extended with an additional direction to obtain a complete 3D laser scanning system.

1.1.1 *Intensities*

The laser rangefinder incorporated in a TLS exploits the light reflected by the surface to obtain a measure of distance. The portion of light that is retro-reflected depends on the characteristics of the material of the object (namely its reflectance ρ), the angle of incidence of the laser beam on the surface, the distance from the device and some other factors (e. g. the atmospheric conditions).

Most laser scanning devices record the strength of the reflected signal as a Digital Number (DN), called *intensity* (I). This fact lead many researchers to speculate about the possibility to invert the intensity information, to obtain a *reflectance* (ρ) value. This problem corresponds to the problem of finding a transformation $T(I)$ which maps the intensity values to reflectances:

$$\rho = T(I) \tag{1.2}$$

The problem of finding a *good* $T(I)$ can be divided into two subproblems:

- determine the shape of the $T(I)$ function, namely to determine the *algebric expression* of the model. This equals to identify the constants, parameters and variables and how these relate to each other trough algebric operators;
- once the function is known, determine the values of its *parameters*, using some kind of experimental dataset for solving the problem via least squares minimization.

The solution of these problems would allow to retrieve the reflectance of the scanned object, which can potentially be used to infer information on the material scanned.

Some authors approached these issues from an experimental point of view (see e. g. Pfeifer et al., 2008; Kaasalainen et al., 2011), identifying two major factors controlling the measured intensity: the distance from the device and the angle of incidence of the laser beam with the surface. They also demonstrated that the recorded intensity values can be converted into reflectances when a *good* mapping $T(I)$ is identified (Pfeifer et al., 2008).

On the other hand, other authors were more interested in the practical use of the intensities: they were using intensities to discriminate rocks (Franceschi et al., 2011; Burton et al., 2011), to detect damaged areas on historical buildings (Armesto-González et al., 2010), to quantify moisture content in aeolian sand deposits (Nield et al., 2011), to identify biological crusts on structures (González-Jorge et al., 2012) and for studying salt marshes (Guarnieri et al., 2009).

While the experimental approach led to important advances in the understanding of the effects of distance and incidence angle, a unique and effective solution for performing a reliable *radiometric calibration* of a TLS still does not exists.

Depending on the context, the device and their needs, the authors interested in *making a practical use* of intensities designed their own correction formulas. In most of the cases, these formulas permit a *normalization* of the values, removing or at least reducing the effects of distance and in some cases of the incidence angles. This kind of correction produces a *corrected intensity* value, rather than a true reflectance measure.

In Chapter 3 of this thesis the aspect of exploiting intensities is discussed, and an original method to perform an approximate radiometric normalization is presented. This method is generic enough to be applied in most of the cases

without any special preparation and is applicable also when the original device is not accessible. It is designed to be *data adaptive* and uses a method for model calibration that is based on user-selected *keypoints*.

1.1.2 *Point clouds*

In last decades TLSs devices have become extremely popular. This popularity is probably the result of a decrease in prices and an overall maturation of the technology itself, a process driven and supported by a progressive acceptance of this technology by the community of professional surveyors (Shan and Toth, 2009).

The interest by the professionals for laser scanning devices is largely due to the fact that these provide a direct source of dense and accurate three-dimensional data, which can be used in the normal surveying practice almost instantly. In fact, TLSs produce data in the form of *point clouds*. This name well conveys what these data look like: a set of three-dimensional points scattered in space, densely representing the surface of the target.

Table 1.1 shows how the raw data of a point cloud may look like. A single point cloud can normally reach sizes of several millions of points, and a dataset is commonly composed by many point clouds.

The fact that a TLS can capture and *freeze* the geometric aspect of a scene - like a "3D photograph"- in the form of a point cloud directed the interest of many scientific communities toward these devices, that were regarded as a possible new and still unexplored source of data for their specific studies.

This interest in TLS, as a data source for highly specific tasks had, as a by product, an increasing interest in the *methodological* aspects of point cloud processing, which requires skills going from linear algebra to data mining, together with a good amount of programming practice.

The extraction of *useful* information from a point cloud is everything but an easy task. In fact, as well explained by Shan and Toth (2009), it requires highly multi-disciplinary skills:

> *[...] the full* ***exploitation of LiDAR s potentials and capabilities*** *challenges for* ***new data processing methods*** *that are fundamentally different from the ones used in traditional photogrammetry. Over the last decade, there have been many significant developments in this field, mainly resulting from* ***multidisciplinary research, including computer vision, computer graphics, electrical engineering, and photogrammetry****. Consequently, the conventional image-based photogrammetry and*

Table 1.1: An example of point cloud data. Each row represents a 3D point in an euclidean space ($\mathbb{R}^3$). For each point additional scalar values may be present as intensities or, e. g., RGB color data. While the coordinates are expressed in meters, the intensity values are normally just an *integer* Digital Number (DN) without explicit physical meaning.

X	Y	Z	intensity
6.4290	38.5237	−0.7459	837
6.4288	38.5222	−0.7416	783
6.4289	38.5229	−0.7394	883
6.4293	38.5258	−0.7401	857
6.4305	38.5358	−0.7473	920
6.4314	38.5423	−0.7508	931
6.4316	38.5438	−0.7500	905
6.4327	38.5524	−0.7557	924
6.4333	38.5574	−0.7578	927

> *vision is gradually adapting to a new subject, which is primarily concerned with point clouds data collection, calibration, registration, and* ***information extraction****.*

The problem of *information extraction* from these data is a key aspect when trying to adopt TLS as new data source for tasks which would have been accomplished in other ways. Transforming a dataset like the one of Table 1.1 into something useful is a matter of processing those numbers and transform them into a more concise representation, familiar to the user. Whether the user is trying to detect the orientation of pear leaves (Balduzzi et al., 2011), to appreciate lithologic variations in a stratigraphic sequence (Franceschi et al., 2011) or to compute forest inventory parameters (Moskal and Zheng, 2011), point clouds need to be transformed into something meaningful for the specific investigation (the orientation of the leaf, a stratigraphic log or the diameter of a log).

This problem is twofold, i. e., it is highly dependent on the possibilities the software used for processing the data offers in terms of *interaction* and *automation*:

- the user needs to *interact* with and visualize the data. Consider the example of a geologist exploring a point cloud which represents an outcrop. We shall call this digital representation of an outcrop, or any digital representation of it, a Virtual Outcrop (VO). He/she may want to measure a bedding attitude. This task is easily done in the field using a compass, but when dealing with a VO, some kind of *virtual compass* will be needed.
- the *processing* must be automated as much as possible. The dataset can become *huge* really easily and if the same operation needs to be repeated many times some kind of automation becomes important. The geologists who wants to measure hundreds of bedding attitudes will definitely appreciate some kind of automation

These are the first two obstacles which are encountered when trying to adopt point clouds as a new data source, which received special attention in this thesis. The first of these two points will be discussed in Chapter 2, where a set of algorithms for the extraction of stratigraphic information from point clouds are introduced. The implementation of these and other algorithms is also discussed and some key choices in the structure of the implementation are presented. Chapter 5 instead illustrates the second point: some specific tasks can be almost completely automated.

1.2 INTENSITY AS A ROCK-TYPE PROXY

The possibility to use the intensity measured from TLSs as a proxy of rock properties has been investigated by several authors. Pesci et al. (2008) used TLS intensity together with RGB data to recognize different lithotypes in a volcanic context, Franceschi et al. (2009) then proved that intensity from an infrared TLS could be used to discriminate marls and limestone. Burton et al. (2011) applied a similar investigation to sandstones and shales, highlighting a negative correlation between intensity and the corresponding γ-ray log, and reinforcing the concept of obtaining *stratigraphic logs* from TLS intensities.

The idea of using reflectance as fingerprints of different rock-types is a well exploited fact in remote sensing (Clark and Roush, 1984): but adopting these methods in the case of TLSs require some specific considerations.

Commercial TLSs use *monochromatic* sensors: they can provide reflectance information of the target only for the narrow band at which their lasers emits. This implies that the intensities coming from a well calibrated TLS cannot be directly compared with data from spectral libraries (e. g. the ASTER spectral

library from Baldridge et al., 2009) by means of measures of *spectral similarity* (van der Meer, 2006).

The *intensity* can potentially be converted into a *reflectance* value but in most of the cases it is preferable to normalize it obtaining a *corrected intensity*. The corrected intensity is a dimensionless scalar field which can be considered as a proxy for the reflectance.

The feasibility to exploit this proxy to extract information on the material require to take into consideration two aspects:

(this will be discussed in depth in Chapter 3)

1. the corrected intensity must be proved to be a *good* proxy of the reflectance which could be measured with dedicated devices as spectrophotometers. A correlation should be verified between the normalized intensities and the measured reflectance.

2. the reflectance itself should be thought as a proxy of some lithologic property of interest. This relationship requires to be verified and cannot be simply assumed.

This duplicity also reflects a problem of *sensitivity* of TLS intensity data. If a variation in a lithologic property is proved to affect the measured reflectivity under experimental conditions (i. e. with a spectrophotometer), that same variation could be *too small* for being appreciated by the TLS.

In Chapter 5 the results of a laboratory investigation of the reflectance spectra of chert and limestone are reported, demonstrating that these two lithologies can be distinguished at a Near InfraRed (nIR) wavelength. The corrected intensities from a scan have then be used to set up a supervised classification algorithm, proving that the corrected intensities can be used as a valid proxy for chert.

Chapter 4 illustrates the feasibility to use the corrected intensities as a proxy for the content of $CaCO_3$ on a stratigraphic sequence composed by apparently homogeneous limestone.

1.3 CYCLOSTRATIGRAPHIC APPLICATIONS

Cyclostratigraphy is the study of astronomically forced climatic cycles as recorded into sedimentary successions. *Quasi-periodic* changes in the orbital parameters induce variations in the total quantity of solar energy received by the Earth, hence contributing to the climate at a global scale.

Global climatic changes act as external *forcings* on the depositional systems that may be recorded as quasi-periodic variations in lithological, geochemical and/or palontological properties in the sediments (Hinnov and Hilgen, 2012).

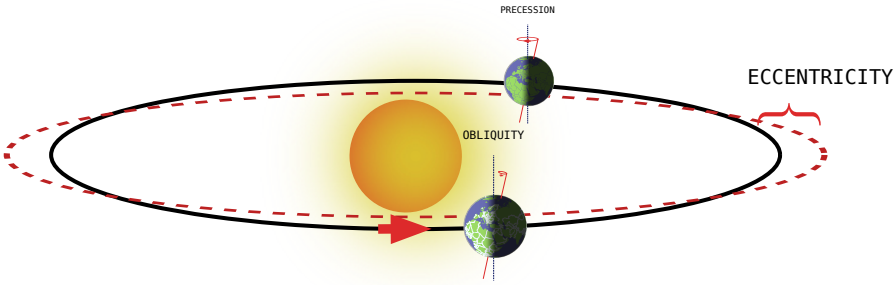


Figure 1.3: Orbital parameters in the Milankovitch band

Orbital cycles of cyclostratigraphic interest lay in the *Milankovitch band* with periods between 10 to 1000 kyr. These orbital variations change the amount and the distribution of insolation through time (House, 1995) and are *precession*, *obliquity* and *eccentricity*

.

The recognition and the study of orbitally induced forcing on climate is made through the study of *climatic time series*. A time series is a collection of scalar values, build of successive measurements of the same variable, typically repeated over a time interval. It constitutes a record of the changes of some system's property over the time.

In the case of Earth climate, these properties are *climatic proxies* which are used to infer past states of the climatic system. Henderson (2002) illustrates some proxies which can be used to reconstruct properties of the past oceans (like salinity, temperature, etc). In the context of cycostratigraphy, the proxies do not necessarily have to be representative of a specific variable of the system, proved that they can be considered sensible to the climate.

Accessing to long, densely sampled and uninterrupted time series is a fundamental aspect and it is also one of the major problems in climatic studies. Long time series can be produced by drilling the continental or oceanic sediments and submitting the retrieved cores to various measurements, but this can be made only in the perspective of expensive and large collaborative projects (see ODP).

Rampino et al. (2000), in his paper *Tempo of the end-Permian event: High-resolution cyclostratigraphy at the Permian-Triassic boundary*, gives an example of the data that can be retrieved from a core. In his cyclostratigraphical analysis he

used time series derived from the γ-ray log and the density estimates (which are automatically retrieved by some core-loggers) together with the isotopic record of ^{13}C obtained by mass spectrometry. The fact that the data collection of a γ-ray log is almost totally automated permitted him to obtain a time series of 1561 samples, against the 379 samples composing his carbon isotope record. Retrieving isotopic data is a time consuming and expensive operation which in general can be done on a limited number of samples.

In the seek for faster data-acquisition methods, Mix et al. (1995) demonstrated that the reflectance spectra, measured with a spectrometer (i. e. 511 data channels covering most of the visible and the nIR bands), can be calibrated to provide actual percentages of $CaCO_3$, opal, and nonbiogenic components in the sediments of the eastern tropical Pacific Ocean.

These automated methods can be applied on cores but are obviously subordinate to large limitations when working on outcrops. The usage of TLS in the nIR wavelength for fast and accurate time series retrieval was demonstrated by Franceschi et al. (2011), showing not only the feasibility of the method but also employing the data in a concrete study case. Burton et al. (2011) compared the TLS intensity to the γ-ray log of the same stratigraphic sequence, highlighting a negative correlation.

Retrieving long time series would permit to study *long* orbital periodicities ($> 2Myr$) that have been observed in recent numerical solutions of the solar system dynamic (Laskar et al., 2011) but that have been rarely observed in the geologic record.

An accurate observation of long-period cycles would permit to disentangle some important open questions on the dynamics of the solar system, with an important feedback of geology in the field of the mechanics of the solar system.

In particular, it could give the crucial information to chose among alternative numerical solutions of the solar system dynamic in deep time, prior to 50 Ma, when the chaotic behavior of the system causes different solutions to diverge.

1.4 GEOSETTING

Cyclostratigraphers are interested in time series from continuous sections, that are mostly provided by deep sea sediments which have the higher possibilities in terms of preservation of orbitally forced cycles. Most of these series are dominated by a carbonatic component in which a terrigeneous and a siliceous component of biogenic origin vary through time. Best stratigraphic sections on land are thus those of deep marine sedimentary environments.

Because of this reason, the studied sections are part of the continuous and long pelagic carbonate sequence of the Umbria-Marche basin. This sequence developed on a isolated passive continental margin, called *Adria* or *Adriatic Promontory* (Channell et al., 1979), part of the African continental platform. The Umbria-Marche sequence is part of the Central Sector of the Northern-Appenines fold-and-thrust belt. During Late Triassic and Early Jurassic, this area was occupied by a carbonate pltform that, underwent extensional tectonics. Normal faulting with block tilting produced a set of subsiding regions, each characterized by different subsidence rates, in an horst-and-graben fashion (Cresta et al., 1989). This tectonic-controlled paleogeography produced different depositional environments with carbonate platform sediments from those of inter-platform basins (Santantonio, 1994). This tectonic phase controlled the subsidence until Middle Jurassic. In regions where platform depositional rates were keeping up with subsidence, thick shallow-water sequences were deposited, as in the Lazio-Abruzzi carbonate platform. The Umbria-Marche region, in contrast, became a basin with emipelagic sedimentation (Cresta et al., 1989).

The Umbria-Marche succession is then characterized by a syndepositional extensional activity during Jurassic, that produced a lithologic differentiation between basinal and platform facies. Also the Creataceous-Paleogene formations show thickness variations due to tectonism.

From Middle Miocene, the sequence was subjected to tectonic shortening, with the development of thrusts, folds and associated contractional structures. This tectonism is associated with the development of the Apennines fold-and-thrust belt (Marchegiani et al., 1999).

A stratigraphic column showing main lithostratigraphic units is reported in Figure 1.4, original names were maintained for consistency with the literature.

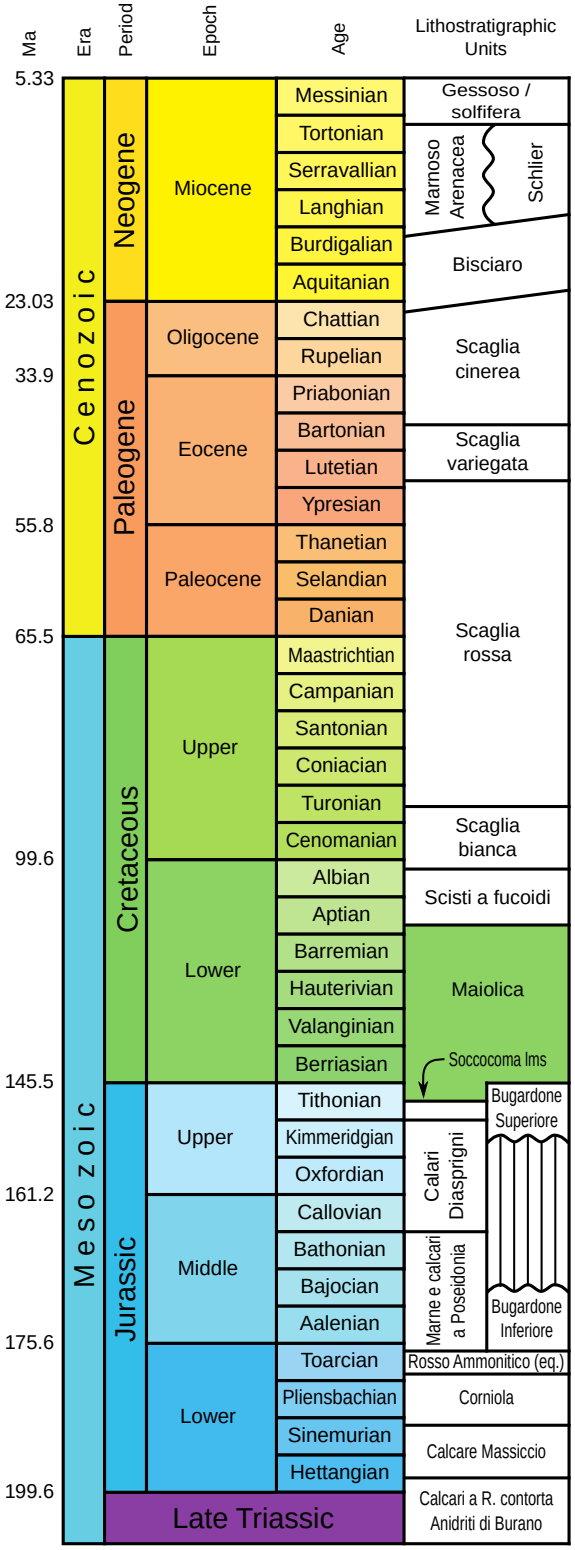


Figure 1.4: Umbria-Marche basin lithostratigraphic units. From Cresta et al. (1989)

1.5 THE QUEST FOR LONG TIME SERIES

The cyclostratigraphic application of TLS proposed by Franceschi et al. (2011) suggested the possibility to obtain cyclostratigraphic time series in a *fast* and geometrical *accurate* way. The most appealing part was probably the speed with which these time series could be reconstructed: one day of work in the field is potentially enough for building tens of meters of series. But there were still some issues to be solved and some open questions which required attention.

The nIR wavelength at which some devices worked, as the OPTECH ILRIS 3D used by Franceschi et al., was demonstrated to be sensible to the different clay content which characterizes marls and limestone alternations. It was clear that the TLS intensity of the ILRIS 3D device was correlated with hydrous minerals content (i. e. clays). But was it possible to observe also more subtle variations of this parameter? Or maybe only major lithologic transitions (as from a *limestone* to a $> 50\%$-*clay marl*) were strong enough to influence the intensity?

Furthermore another lithology is often present in the Umbria-Marche series. *Chert* is a common component in various formations, and although it is not always abundant in absolute terms, its presence has been sometimes connected with climatic conditions, making its presence a possible proxy of the orbital forcing (e. g. Ikeda et al., 2010; Erbacher et al., 1996; Muttoni and Kent, 2007; Batenburg et al., 2012).

The dimensions of the laser footprint increases with the distance, while the geometric resolution of a scan decreases. Furthermore, the intensities are normally *discretized* on a number of levels which tend to decrease with the distance (only $4-5$ levels at 50 meters). In practice, better quality scans can be obtained when scanning small portions of a wide outcrop at a time from a nearby position (Franceschi et al. suggested at a distance < 50 m). This necessity, together with the limited field of view of the ILRIS 3D ($40^\circ \times 40^\circ$), contribute in reducing the area of outcrop that can be retrieved with each scan.

Scanning large outcrops would thus require a considerable number of scans. Each one of them should be processed for normalizing the intensities and extracting the time series. Each time series requires to be stratigraphically correlated to the others, to produce an unique and long composite time series: we needed to create a dedicated software component to perform these and other operations. The processing and the management of these data required the automation and the optimization of the algorithms in use. These algorithms are described in Chapter 2.

2

SOFTWARE COMPONENTS FOR THE GEOLOGICAL-AWARE PROCESSING: SPC AND VOMBAT

Contents

This chapter presents a selection of the tools and algorithms which have been developed for point cloud processing and time series generation, and that have been extensively used throughout this thesis, especially in Chapter 6.

These algorithms have been created because available tools are not fulfilling the typical need of the sedimentary geologist.

During the development of these codes a fairly large amount of `python`[1] (e. g. Langtangen, 2008) programming was involved for the prototyping. `python` has been, in general, the language of choice for the data analysis and plotting.

When the algorithms evolved and became more complex and demanding in terms of performance, a lower level language was needed thus most of the code was migrated to `C++`. More than 250000 lines of code compose the final toolkit. All code is kept online at `https://github.com/luca-penasa` and can be freely downloaded and compiled.

We will not give a complete description of *all* the algorithms that were implemented, also because no really *high-level* math is involved in the process of extracting time series, and giving an in-depth overview of the many technical choices and *tricks* involved in writing well performing code for this task would require too much space in this thesis.

1 `www.python.org`

The C++ code which has been produced for this project is released as open-source software and we redirect the interested reader to the code itself, which has been a major effort of this thesis and represents a valuable contributions on its own.

In this chapter, we will instead review the basic mathematical background of the *time series* reconstruction problem from point clouds and then, in a second part, we will discuss how this has been approached and implemented within the two major software contributions of this thesis: the SPC C++ toolkit and the Vombat plugin for CloudCompare.

2.1 BASICS OF POINT CLOUD PROCESSING

Point clouds are a fairly recent new *data type*. A point cloud is a set of 3D points in space which provides a dense representation of objects and surfaces in a three-dimensional context. Point clouds are the basic data type that is provided by an increasing number of devices. Terrestrial Laser Scanners (TLSs), Airborne Laser Scanners (ALSs), range cameras (as the popular Microsoft Kinect), photogrammetric and structure-from-motion methods are becoming more common every day, and with them, point clouds are acquired and processed in an increasing number of contexts.

Without going into the details of the various technologies to acquire geometric information from the environment, we will assume our point clouds to be simply an *unstructured* list of 3D points as depicted into the introduction (Section 1.1.2).

Unstructured, in the case of point clouds, means that the *list* of points we are considering does not follow any specific order, thus we cannot exploit any *a-priori* knowledge about the internal ordering of this list to our advantage. This will be clearer considering an example, that will also serve to introduce some basic formalism.

Consider a point cloud P which is composed by a number n of 3D points. Each point is a triplet of scalars $\mathbf{p}_i = [x_i, y_i, z_i]$.

$$P = \{\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_n\} \tag{2.1}$$

A common problem require to find all the points of P which are contained in a given sphere with radius r and centered in $\mathbf{c} = [c_x, c_y, c_z]$. Clearly the problem consists in finding all the points which satisfy:

$$\|\mathbf{p} - \mathbf{c}\| < r \tag{2.2}$$

Although this problem may seem trivial, when dealing with large point clouds, this approach would require to compute the distance to the sphere's center for each point, and then check if this distance is lower than the radius r. This must be done for each point given we does not have any *a-priori* information that allows us to restrict the computation to a smaller set. When the number of points that must be checked is large, this solution becomes impractical, especially if this operation must be carried out many times.

In fact this operation is one of the most common *queries* that are performed on point clouds, and permits to find the *neighbors*[2] of a given point $\mathbf{p}$. Being able to access the neighbors of a point is fundamental, for example, for estimating the local normal $\mathbf{n} = [n_x, n_y, n_z]$ of the underlying surface.

This example shows how processing point clouds is inherently different from, e. g., image processing. The extraction of useful information from point clouds requires skills going from linear algebra to data mining, together with a good amount of well optimized programming practice.

By the way, various solutions to this problem exist, permitting to perform this operation in an efficient way, but we will not go more in depth than this.

2.2 FROM CLOUDS TO TIME SERIES

Franceschi et al. (2009) suggested that a way for exploiting the information contained into point clouds is to reduce the data to the more familiar representation of a time series. Time series are in fact the fundamental data used in any cyclostratigraphic analysis.

These time series should in particular represents how some *property* (quantified by a *variable*) of the rocks changed over geological *time*. These two aspects are of fundamental importance for producing reliable time series for cyclostratigraphy. Weedon (2003) identifies some guidelines for producing *good* cyclostratigraphyc time series:

1. Consistent environmental conditions. The measured variable must have been recorded in a stable environment, without major changes, e. g. of facies.

2. Unambiguity of the variable meaning. The variable should show an unambiguous relationship with one or more properties of the sedimentary system.

2 There are other ways to define the neighborhood of a point. For example the nearest N points from the given location.

3. Existence of the time-thickness relationship. For a meaningful time series analysis the stratigraphic position must be related with time.

In our case, the *properties* are represented by some kind of variable that is linked with the points of the point cloud. We will consider mainly the *intensity* provided by most TLSs but the same methods are obviously applicable to any other scalar field that may be associated with the points, either because it is inherently dependent on the method (e. g. intensity for TLS, Red-Green-Blue (RGB) for photogrammetry, etc) or because it results from some kind of data set merging, e. g. from hyperspectral cameras (see Sterzai et al., 2010; Kurz et al., 2012b,a). More insights on the relationship between the *intensity* and rock-properties are given in Chapter 4 and Chapter 5.

As regard the geological *time*, obviously we need to consider a proxy for it, that is the *stratigraphic position*.

Our first objective is thus to associate each point of the cloud with two scalar fields, one representing the rock property we want to measure and a second representing the stratigraphic position of the point. While the considered rock-property field comes directly from the normalized intensity, and hence the point cloud itself, the stratigraphic position must be estimated.

We thus seek for a function $SP(\mathbf{p})$ which, for each point $\mathbf{p}$ provides a stratigraphic position. A simple approach is to define this function as the distance from an user defined plane:

$$SP(\mathbf{p}) = \hat{\mathbf{n}} \cdot \mathbf{p} + d \tag{2.3}$$

For a given point $\mathbf{p} = [x, y, z]^T$, its stratigraphic position is given by the distance of the point from the plane defined by the unit normal $\hat{\mathbf{n}}$ and the origin-to-plane distance d. A point laying on the plane itself will have a stratigraphic position of 0, one on the same (opposite) side the normal is pointing to, will have a positive (negative) value. This simplified model represents what we will call for simplicity a Stratigraphic Reference Frame (SRF), equivalent of the folding rule used in the field for measuring stratigraphic thicknesses. Its unit normal $\hat{\mathbf{n}}$ defines a direction in space for which the stratigraphic position, and thus the time, strictly increases or possibly decreases.

The plane which defines the SRF must be aligned with the bedding planes of the stratigraphic sequence we are studying, such that all the points laying on a given layer or bed will receive an equal stratigraphic position. The normal can be promptly estimated as the best fitting plane of a subset of points which are hand-picked by the user on the same stratigraphic level.

The d parameters permits to shift the SRF in the direction of its normal, if we have a specific point $\mathbf{p}_0$ for which the stratigraphic position S_k is know we can compute a new d to match this constraint. We want to minimize the stratigraphic position computed by the SRF with respect to its expected value:

$$\hat{\mathbf{n}} \cdot \mathbf{p}_0 + d - S_k = 0 \tag{2.4}$$

and thus

$$d = S_k - \hat{\mathbf{n}} \cdot \mathbf{p}_0 \tag{2.5}$$

This simplified model allows to uniquely define a stratigraphic position for each point of the cloud. Together with the intensity, it can be used to estimate a time series via Kernel Smoothing (KS).

The statistical techniques known as KS permit the estimation of a real valued function starting from a set of unevenly spaced noisy observations (a good treatise about *kernel smoothing methods* is found in Hastie et al., 2009).

Let first define the kernel function as:

$$K_\lambda(x_0, x) = D\left(\frac{|x_0 - x|}{\lambda}\right) \tag{2.6}$$

where $x_0, x \in \mathbb{R}$, $|\cdot|$ is the absolute value, λ is a constant parameter (kernel radius or bandwidth) and $D(t)$ is a real valued function, which value is usually decreasing with increasing distance between x and x_0. Commonly used kernels $D(t)$ are reported in Table 2.1. Note that the notation $\mathbf{1}_A$ is a simple short form for the function[3]:

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \tag{2.7}$$

where A is any given set.

Given $\hat{y}(x_0)$ is a continuous function of x_0, the Nadaraya-Watson kernel-weighted average, that is the smooth estimation of the unknown $y(x_0)$ function, is defined by:

$$\hat{y}(x_0) = \frac{\sum_{i=1}^{N} K_\lambda(x_0, x_i) y_i}{\sum_{i=1}^{N} K_\lambda(x_0, x_i)} \tag{2.8}$$

3 Formally known as indicator function. Given an element x of the set X, and a subset $A \in X$ the function is 1 if $x \in A$ and is 0 if $x \notin A$.

Kernel	$D(t)$
Uniform (box)	$\frac{1}{2}\,\mathbf{1}_{\{\lvert t\rvert\leqslant 1\}}$
Epanechnikov	$\frac{3}{4}(1-t^2)\,\mathbf{1}_{\{\lvert t\rvert\leqslant 1\}}$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$
Cosine	$\frac{\pi}{4}\cos\left(\frac{\pi}{2}t\right)\mathbf{1}_{\{\lvert t\rvert\leqslant 1\}}$

Table 2.1: Commonly used kernels for Kernel Smoothing

where N is the total number of observations, and y_i is the observation at x_i. Note that this formulation actually corresponds to a weighted average, where the weights are given by the kernel function, centered in x_0 and which effect (radius) is scaled by the λ parameter.

The $\hat{y}(x_0)$ function can then be evaluated at evenly spaced x_0 values, reconstructing in this way a suitable time series to be used in spectral analysis.

A new set of x_0 positions at which the $\hat{y}(x_0)$ function is evaluated from Equation 2.8 can be computed subdividing the range of the x_0 variable at evenly spaced points, given a step Δx.

2.3 SPC AND VOMBAT

`SPC` is a `C++` library including all the algorithms needed to extract time series from point clouds. It defines a set of classes which are the basic building blocks of the processing pipeline. These *objects* can be created by the user accessing the `SPC` library from `C++` code or via a Graphical User Interface (GUI) provided by a CloudCompare's plugin (Vombat).

2.3.1 *SPC*

`SPC` contains all the object definitions, the computational algorithms and the `IO` facilities for saving and loading the data. It is highly based on the `PCL` library (Rusu and Cousins, 2011).

Here is a summary of the objects which are provided by `SPC`, which are also exposed by `Vombat`:

Reference On Disk

Point clouds are normally fairly big entities to be loaded and visualized. It then makes sense to have an object which keep track of the point cloud without actually loading it. A *Reference On Disk* object implements this concept, providing a *reference* to a point cloud which can be loaded only when really needed, e. g. when extracting the time series.

Attitude

An attitude is a measure of the orientation of a plane in 3D space, together with its position. It can be fully described in 3D by a single equation, here in normal Hessian form:

$$\hat{\mathbf{n}} \cdot \mathbf{p} - d = 0 \tag{2.9}$$

`SPC` provides a dedicated method for computing attitudes. These are used within `SPC` to represent a bedding plane attitude in a specific location of the outcrop. The user can hand-pick one or more sets of points which are expected to lay on the same plane. The parameters of the plane are then obtained by either linear or non-linear fitting of the plane in a least squares sense.

Sample

A *Sample* is a point in 3D space, which can be defined by the user and is used to represent any point that may be of interest on the outcrop, e. g. a point where an actual rock sample was collected. It is defined by a single point in $\mathbb{R}^3$: $\mathbf{p} = [p_x, p_y, p_z]$.

Stratigraphic Reference Frame (SRF)

It implements the stratigraphic positioning system explained in the last sections. It is fully defined by a single equation, which can be directly derived from an Attitude:

$$SP(\mathbf{p}) = \hat{\mathbf{n}} \cdot \mathbf{p} + d \tag{2.10}$$

In this sense any *Sample* can be projected in the stratigraphic reference system using this last equation to obtain its stratigraphic position in the considered SRF. Alternatively, the stratigraphic position of a *Sample* can be provided by the user, in such a case the stratigraphic position of the *Sample* is fixed, e. g. for defining where a SRF must have its stratigraphic zero.

Region Of Interest (ROI)

A *Region Of Interest* in `SPC` implements a simple but effective way to provide repeatable selections of points on point clouds. This object, once created, can be used to extract subsets of points from any point cloud. It is defined by a polyline in 3D with the vertexes laying on a user-defined plane. The polyline is then extruded of an user-specified offset in both the plane's sides to define a volume. The extraction is performed checking each point for being inside or outside the defined volume.

Stratigraphic Tie Constraint

A *Tie Constraint* can be defined between two or more *Samples*. When two samples, which are associated to different SRFs, are *tie constrained* their *Stratigraphic positions* are constrained to be the same, forcing the two SRFs to align in a stratigraphic sense, thus providing a composite framework.

2.3.2 *Vombat*

`Vombat`, which stays for "*Virtual Outcrop Basic Analysis Tool* ", provides an intuitive interface for handling the objects defined by `SPC`, which can thus be created and edited from within CloudCompare (Girardeau-Montaut, 2014) instead of requiring `C++` code.

This plugin allows for the definition of the objects and also to reconstruct the time series, but once the parameters and the objects have been defined the plugin can also be neglected and the data extraction process can be performed only using `SPC`.

3

SELF-NORMALIZATION OF TERRESTRIAL LIDAR INTENSITY USING SCENE-AVAILABLE TARGETS

Contents

PAPER IN PREPARATION

Luca Penasa and Dimitri Lague. Self-normalization of terrestrial lidar intensity using scene-available targets. *in preparation*, 2015

3.1 INTRODUCTION

Terrestrial Laser Scanners (TLSs) are great tools for capturing the geometry of target objects, but the geometric information does not provide any direct knowledge about the underlying material. In some cases an indirect relationship between the target material and local geometric properties of the point cloud can be drawn and exploited for classification purposes (Brodu and Lague, 2012), but in the most general case this step is not obvious and additional information is needed in order to draw conclusions on the materials under investigation.

TLSs are active devices, thus the *intensity* recorded may be considered the result of a sampling operation made under controlled conditions on the target surface. The laser beam reaches the surface, and is scattered in all directions. One of the rays will eventually return to the device more or less suppressed, depending on the reflectance of the target, the incidence angle, the range and other factors related to the atmospheric absorption and some other device-related factors. This sampling is then repeated for each three-dimensional point captured by the device.

Most devices record intensity as standard practice thus most of existent points clouds coming from TLS are equipped with an *intensity* scalar field, directly associated to the points.

It is thus clear how much appealing may be the idea to invert the intensity into some kind of scalar field describing the target material (i. e. obtaining the reflectivity ρ, or some other scalar that may be considered as a proxy for the reflectivity itself). Recently, also full-waveform hyperspectral lidar based on broad-spectrum lasers emitters have been proposed (Kaasalainen et al., 2007b; Chen et al., 2010), and although they are still experimental devices their potential has been proved in some circumstances (Hakala et al., 2012; Suomalainen et al., 2011; Kaasalainen et al., 2010b).

Various authors demonstrated the viability of inverting the intensities to obtain reflectance information from commercial TLSs performing detailed experiments (e. g. Pfeifer et al., 2008; Kaasalainen et al., 2011) on various devices. The behavior of the intensity from TLSs has been particularly investigated in respect to the distance and the incidence angle (Krooks et al., 2013; Anttila et al., 2011), but also the effects of moisture (Kaasalainen et al., 2010a) and surface irregularities (Pesci and Teza, 2008) have been considered. In a recent paper, Fang et al. (2015), considered also the defocusing of the receiver's optics. The effect of incomplete overlap between the laser beam and the receiver's field of view, which may affect biaxial systems, has still to be systematically investigated in literature (Pfeifer et al., 2008; Fang et al., 2015). Airborne Laser Scanners (ALSs) systems have been subjected to a similar research activity (e. g. Ahokas et al., 2006; Höfle and Pfeifer, 2007; Kaasalainen et al., 2007a; Jutzi and Gross, 2009) and much of that work can be considered to some extent valid also for TLSs.

Additionally, time-invariance of model's parameters would require better investigations (Wang and Lu, 2009) and the manipulation of the intensities from the device itself and/or the proprietary softwares add a layer of complexity that is not easy to disentangle.

On the other hand, the viability of using the intensities as a material proxy have been demonstrated also in many practical cases, to discriminate rocks (Franceschi et al., 2011; Burton et al., 2011), to detect damaged areas on historical buildings (Armesto-González et al., 2010), to quantify moisture content in aeolian sand deposits (Nield et al., 2011), to identify biological crusts on structures (González-Jorge et al., 2012) and for studying salt marshes (Guarnieri et al., 2009). In such contexts a *normalization* of intensities have been generally adopted, rather than a complete radiometric calibration (which would provide real reflectance measures).

The normalization shall remove the effects of variables which exert an effect on the measured intensities but do not directly depend on the material itself, namely the distance and the incidence angle. Although the physics underlaying the calibration problem is mostly understood (especially after Fang et al., 2015), it is still not easy to obtain a good normalization of the intensities for practical applications.

The calibration problem requires to face with two tasks: *a*) determine the *algebric expression* of the calibration model, and *b*) determine the values of the *parameters* of the model.

Models derived from the *lidar equation* have been generally adopted, although they require to face with various device-specific optimizations (especially for near-distances) to be really effective (Kaasalainen et al., 2011; Fang et al., 2015). Another approach which has been pursued uses *black-box* models, which are not directly linked to the physics of the device, but can give better results (Pfeifer et al., 2008).

Regardless of the type of model chosen, the *parameters* of the model are estimated by means of a least squares solution, against a dataset of *observations*. These observations have been normally acquired by experimental data collection, using targets for which the reflectance is known (for a proper calibration) or can be assumed to be homogeneous and constant (for a normalization, see e. g. Fang et al., 2015, which also reports an effective scheme for placing the calibration targets). These approaches are normally fairly complex and require a certain amount of experimental setup and post-processing of the data.

We illustrate here a generalized solution for extracting a set of calibration observations from overlapping point clouds taken from different *stations*. The method is based on the user-selection of *keypoints* laying on one or more homogeneous materials. Each keypoint that is seen from more than one station defines an set of *homologous* points on different scans for which the geometric variables (distance from the sensor and incidence angle) and the observed intensity are extracted from the point clouds. The observations extracted are

accompanied by additional features which can be employed to design weighting strategies for the parametrization of the model.

Our method simplifies the extraction of calibration data and can be used to refine or fit any calibration model. We demonstrate the effectiveness of this approach using a simplified *black-box* model based on a Radial Basis Function (RBF) smoother, showing that it can be used to significantly improve the quality of intensity data in many circumstances which do not require an absolute radiometric calibration.

3.2 INTENSITY CALIBRATION AND NORMALIZATION

It is often assumed that the calibration model for lidar devices can be derived from the lidar equation, here simplified for extended, Lambertian-behaving targets (Pfeifer et al., 2008):

$$P_r = \frac{\pi P_e \rho \cos(\alpha)}{4r^2 \eta_{atm} \eta_{sys}} \tag{3.1}$$

where P_r is the received power and P_e the emitted power; ρ is the target reflectivity, α the incidence angle, r the range and η_{atm} and η_{sys} are terms accounting for atmospheric and device-related transmission.

Intensity values I provided by TLSs are normally digital numbers, and a physical meaning for them can only be inferred, so that it is common to consider them as linearly related to the received power, although this may be not strictly true, so that $I = aP_r$ with a some unknown constant.

P_e is normally not recorded by the device, so that it must be considered a constant although it has been evidenced (Wang and Lu, 2009) that in some devices it may be time-dependent. The atmospheric transmissivity term can be considered as negligible for most TLS applications while the transmissivity of the system can be considered constant. It then makes sense to group constant terms and reduce the calibration problem to:

$$I_r = k\frac{\rho \cos(\alpha)}{r^2}; \text{ with } k = \frac{\pi a P_e}{4\eta_{sys}} \tag{3.2}$$

In the perspective of finding a way to correct intensities independently from the specific device, we can further simplify the model in order to get rid of the $\cos\alpha$ and the r^2 terms. In fact, it has been shown that they cannot be considered optimal explanatory terms for all the devices (Pfeifer et al., 2008; Jutzi and Gross, 2009). We substitute them with a generic function $f(r, \alpha; \beta)$ which shape and parameters β are to be determined:

$$I = k\rho f(d, \alpha; \boldsymbol{\beta}) \tag{3.3}$$

Whether $f(d, \alpha)$ is obtained from the physical modeling of the system or a general-purpose interpolator function is used, the model calibration task reduces to finding an estimate of the parameters $\boldsymbol{\beta}$ and the constant k, so that $\rho = I/(kf(\alpha, d, \hat{\boldsymbol{\beta}}))$ can be computed for any point of the cloud under investigation. This formulation of the *calibration* would permit to obtain ρ values from the intensities. This approach requires the use of targets with known reflectance and is subordinated to some important assumptions (as the linearity between the power and the measured intensity and the stability in time of the model's parameters).

A simpler approach can be provided by a *normalization*, in place of a proper *radiometric calibration*. In this case the final goal is to produce a *correction* or *normalization* of the intensities with respect to the main factors which are unrelated to the reflectance (i. e. the distance and the incidence angle). In this case generic targets can be employed and the *normalized* intensity will provide information on the relative reflectance of the materials in respect to the targets reflectance (Fang et al., 2015).

The reflectivity of a generic target is normally unknown. To take account of this, Equation 3.3 needs to be modified. Instead of obtaining a reflectance measure ρ, a value I_{corr} is obtained: this value is proportional to ρ but also to some terms that are assumed to be constant (see Equation 3.3).

$$I = I_{corr} f(d, \alpha; \boldsymbol{\beta}) \tag{3.4}$$

This *normalized* intensity I_{corr} is then the result of the measured intensity I divided for a smooth function of the distance and the scattering angle.

$$I_{corr} = \frac{I}{f(d, \alpha; \boldsymbol{\beta})} \tag{3.5}$$

This formulation of the calibration problem leads to a *normalization* of the intensities values rather than nailing them to a reflectance scale. For many data elaborations as classification, thresholding, dataset merging and visualization, the normalization of the intensity data appears as a huge improvement over the raw intensities. In these cases an imperfect or approximate normalization has no substantial impact in terms of interpretation.

3.3 CALIBRATION DATASET FROM KEYPOINTS

The problem of finding the parameters β is normally addressed by means of least squares, minimizing the objective function:

$$\left\| I - I_{corr} \cdot f(\alpha, d) \right\|^2 \tag{3.6}$$

For solving the minimization problem of Equation 3.6 a *calibration dataset* formed by the observations of the measured intensity, under controlled incidence angles and distances, is needed. For each observation, also a I_{corr} variable (in place of the reflectance itself) should be defined (this last point will be discussed in Section 3.4.1).

It has been shown that reference targets can be used for collecting a calibration dataset (e. g. Kaasalainen et al., 2009a; Fang et al., 2015): a set of artificial targets which are homogeneous in terms of material can be placed at different distances and scattering angles combinations as in Figure 3.1.

The intensity response of the targets is then repeatedly measured for many configurations of the target and the observations of the incidence angle, the distance to the target, the measured intensity and the reflectivity (if known) are tabulated and used to parametrize the calibration model. Although this approach is clearly preferable is has some drawbacks: *a*) it requires an adequate experimental setup together with a certain amount of processing time for the extracting the information needed, contributing in making the method difficult to replicate; *b*) the targets used for calibration are often almost perfect lambertian scatterers, while natural materials often deviate from the cosine-law, as commonly observed in some natural contexts (Hartzell et al., 2014; Pesci and Teza, 2008; Franceschi et al., 2009); *c*) it is not easily applicable for scans for which the original device is not accessible (e. g. scans provided by third-parties).

We generalized a data extraction method so to become easily operable with any surface that may be considered as a calibration target. This method can be employed to simplify the calibration dataset extraction, also from many TLS datasets that already exist.

3.3.1 *Keypoint approach*

A TLS survey is normally composed by a set of point clouds, each being the result of the scanning of the target object from a different station. Overlapping scans are often present in a dataset. In this context, information for obtaining an intensity normalization may be present into the dataset in two forms:

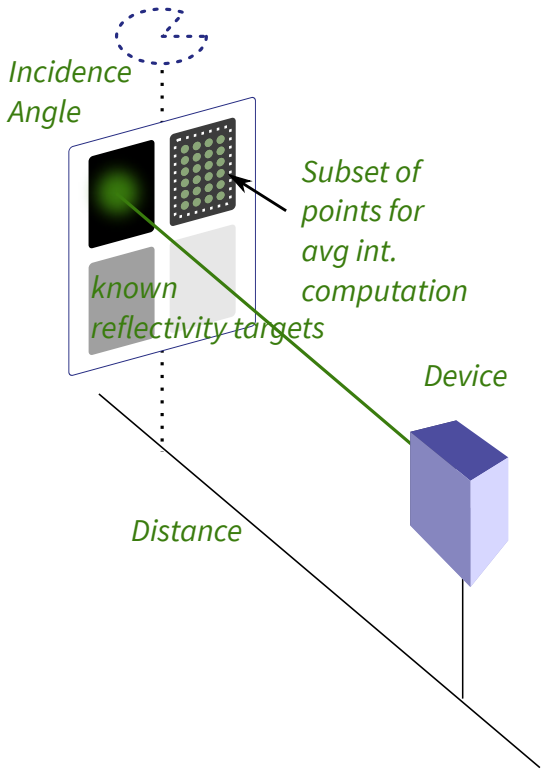


Figure 3.1: Typical experimental setup for calibration of a device based on targets with known reflectance. The target (or the scanner) can be moved in different locations and rotated to provide different readings of intensity, as the distance and the angle change. A more effective scheme is proposed by Fang et al. (2015) where many small targets are placed in the same scene following a geometric pattern.

1. as an a-priori knowledge of the materials. The user may be able to identify in the scene one or more subsets of keypoints which are expected to have the same corrected intensity after normalization

2. as overlapping scans: for the same point that was captured from two or more scan stations, the TLS records different intensities. This difference can be ascribed to the effects of different distances and/or scattering angles and should be minimized when the intensities are properly normalized.

Target surfaces are identified by the user using a set of keypoints. The problem of extracting *good* keypoints from unstructured point cloud data is a wide and well known issue for object recognition (Mian et al., 2010) and automatizing this operation would require a fairly more complex treatment. We will assume

the keypoints are defined by the user by hand-picking a set of keypoints into the scene.

Each keypoint defines a point on the object's surface that can be considered homogeneous in terms of material and locally flat. The spherical neighborhood of the keypoint is considered in order to sample the variables of interest.

For each keypoint, a unique normal of the underlying surface is computed (see Figure 3.2). A best fitting plane is computed by means of eigendecomposition of the covariance matrix (Mitra et al., 2004) . The points coming from all the available scans which sampled that location are used, see Figure 3.3. The eigenvalues deriving from the normal estimation, which can be used as a measure of *flatness* of the plane (Brodu and Lague, 2012) are also stored.

A keypoint may be seen from different scan stations, hence the same underlying surface has been scanned one or more times, from different points of view. For each keypoint, each scan will produce an additional unique observation of the calibration features: comprising the distance, the incidence angle and the measured intensity (see Figure 3.3). Table 3.1 reports the features that are estimated for each observation.

Table 3.1: The set of features estimated for each observation.

Feature	Description
Distance d	The keypoint-to-station distance
Intensity I	The measured intensity, computed as weighted average of the neighbor's intensity. Each point's intensity is weighted by its distance to the keypoint itself
Intensity Std I_{std}	The standard deviation of the intensity, computed as a weighted standard deviation of the points used for computing I
Incidence angle α	The incidence angle of the laser beam with the surface. Computing using the local normal estimated for this keypoint (see previous table)
Number of Neighbors NN	the number of neighboring points used for computing I

This keypoint-based procedure can be used to automate the process of extracting the calibration observations from a dataset, the same procedure can be employed also with artificial reference targets with know reflectance if a complete calibration is desired.

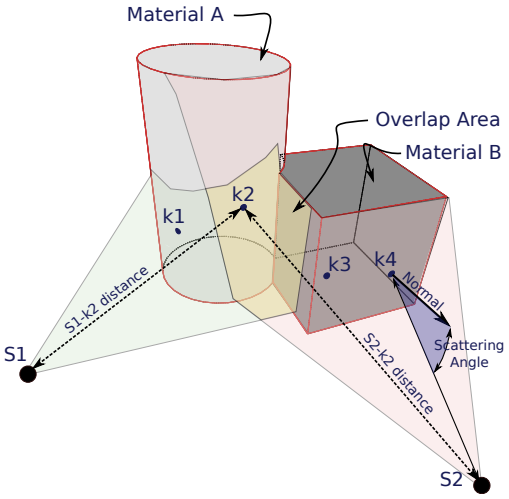


Figure 3.2: Keypoints (k1 to k4) are 3D points lying on the object surface. Keypoints are placed by the user on locations which are expected to be homogeneous in terms of material and locally flat. The keypoint k2 is seen by two different stations (S1 and S2), under different scattering angles and distances, thus providing two independent observations (one for station)

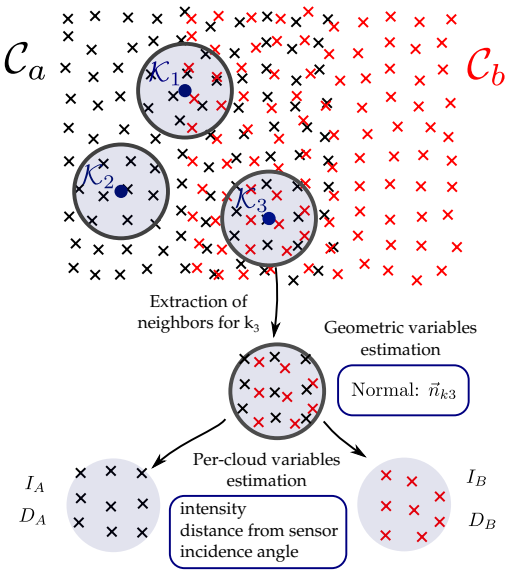


Figure 3.3: Two point clouds, $\mathcal{C}_a$ and $\mathcal{C}_b$, are the results of two scans made from different stations. For each keypoint the local normal $\vec{n}_k$ is estimated, using points from both the scans. The distance of the keypoint from the device d, the incidence angle α and the target's intensity at the keypoint depends on the device position and are estimated independently for each overlapping scan.

3.4 METHOD VALIDATION

This approach to normalization has been tested on a dataset, chosen so to represent an usage case in which normalized intensity may give important additional information over the original intensity field. The data extracted via the keypoint-based method will be presented and discussed.

Table 3.2: Device used and basic information on the dataset that will be taken into consideration

Case	Device	Context	Data Ref.
Rangitikei River	Leica Scanstation 2	Vegetation	Lague et al. (2013)

This case has been chosen so to represent an usage cases that have been proposed in literature. The first case concerns the possibility to use intensity to identify vegetation from a natural scene. In Lague et al. (2013) the approach proposed by (Brodu and Lague, 2012) was used for vegetation identification and removal. That approach is based purely on the geometric aspect of vegetation. They also proposed that a better classification result could be obtained using the normalized intensities.

In this example we will show how it is possible to identify a scattered set of keypoints which are expected to posses the same reflectance and exploit them to calibrate an adaptive model to normalize the intensity of the scene.

The dataset is composed several point clouds, each derived from a different station. The point clouds were geometrically registered. The coordinates of the device's center have been exported for each different station.

3.4.1 *A simplified normalization model*

Penasa et al. (2014) showed that in some contexts a simple correction for intensities could be obtained using a moving average method considering all the data points in the scan. Such operation implicitly requires that no correlation between the distance and the reflectivity exists and hence the *detrending* operation will reduce to a correction of the distance effect only.

Considering a specific scan for which we want to find a good normalizing function, independently from the specific device, if we can accept some uncertainty we don't need to find a normalization function able to predict with equal accuracy the whole range of scattering angles and distances that are possible

(i. e. the whole nominal range at which the device may work), but it is instead more sensible to find a function that is:

- able to accurately normalize intensities at varying angles and distances for the majority of points in the specific scan. This also implies that the scan itself can be used as a good source of data for calibrating the model: if the 95 % of the points in a scan which require normalization are at a distance between 40 to 80 meters from the station, we may be satisfied of a model able to correctly normalize the intensities in that range, and thus parametrized with observations in that range only.
- that is *reasonably well behaved* out of these bounds. Considering the remaining 5% of points, for which we are accepting an approximated correction, they will still be corrected in a reasonable way (e. g. corrected intensities not going to infinite).

Although in this case a *well-behaved* function may imply smoothness and monotonicity, we can simplify looking for a function that is unlikely to introduce unwanted artifacts (as may happen using high degree polynomials that can easily lead to bumps and periodicities).

Pfeifer et al. (2008) gave an in-depth sight on possibly eligible calibration functions $f(\alpha, d)$ for Equation 3.4. They also demonstrated the need to use highly data adaptive models in order to explain the complex intensities behavior of some devices.

Furthermore a function for which the parameters are linearly estimable (via linear least squares minimization) is highly desirable over functions requiring non-linear least squares minimizations.

Radial Basis Functions (RBFs) are a good candidate being often used for function approximation, and are especially suited for scattered and multidimensional data. This allows to use the same scheme if we want to correct for both distance and angle or just for the distance.

A RBFs interpolant is defined by a set of fixed nodes ($\mathbf{n}_i$) in the $[d, \alpha]$ space. At each node a weight (or coefficient) is associated.

For a given $\mathbf{x} = [\alpha, d]$, for which the correction factor has to be computed, the value of $f(\alpha, d)$ corresponds to the sum of all the node coefficients, weighted by a radial basis function which is proportional to the distance of $\mathbf{x}$ to the considered node $\mathbf{n}_i$. A polynomial in $\mathbf{x}$ is often added to fit the overall trend:

$$f(\mathbf{x}) = \sum_{i=1}^{N} \beta_i \, \phi(\|\mathbf{x} - \mathbf{n}_i\|) + P(\mathbf{x}), \tag{3.7}$$

If ϕ is a gaussian, points distant from the domain defined by the nodes will receive a zero weighting and $f(\mathbf{x})$ will approximate the polynomial $P(\mathbf{x})$.

In the classical RBF interpolator, a data point to be modeled corresponds to each of the nodes, with the result that for each observation (a $[d, \alpha, I]$ triplet) a new parameter β must be estimated. The method permits in fact to obtain an *exact interpolation*. In the case of intensity modeling we can use a regular grid of nodes, placed in the $[d, \alpha]$ domain, reducing the number of parameters and obtaining a smoothing effect.

In the case of a RBF smoother, $f(d, \alpha)$ function is linear in the parameters and a unique solution can be found by linear least squares (provided the system is solvable).

Considering that the set of keypoints on the same material we may expect that all of them should show the same I_{corr} value (the intensity value after correction). Being interested in a normalization of intensities, we may simply set it to 1 (or ay other arbitrary value) for all of them.

The system of equations to be solved is thus in the form:

$$\begin{cases} 1\, f(d_0, \alpha_0) = I_0 \\ 1\, f(d_1, \alpha_1) = I_1 \\ \dots \\ 1\, f(d_n, \alpha_n) = I_n \end{cases} \tag{3.8}$$

where n is the total number of observations on the considered material. Each row of the system represents an observation of a triplet $[d, \alpha, I]$, of a specific keypoint as seen from a specific station. Simplifying, after the correction all the observation must have the same intensity value (i.e 1), being on the same material.

3.4.2 *Weighted solution*

The model solution can be easily modified into a weighted last squares problem, to account for the different reliability of each observation. For each keypoint an unique local normal is computed via eigen-decomposition of the covariance matrix (Mitra et al., 2004). From the eigenvalues a *flatness index* e_r can be computed (Brodu and Lague, 2012), which represents the amount of variance

explained along the direction of the normal over the other directions (i. e. the other eigenvectors):

$$e_r = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \tag{3.9}$$

This index can be used to express how much the underlaying surface is planar. Notice that λ_0 is the lower of the eigenvalues. This implies that, for a perfectly planar surface, e_r tends to 0. On the other hand when λ_0 increases, we have $\lambda_0 \sim \lambda_1 \sim \lambda_2$ in the most extreme case, for which e_r tends to $1/3$. Designing a weighting coefficient to account for the underlying planar geometry leads to:

$$w_{er} = 1 - 3e_r \tag{3.10}$$

w_{er} is almost 1 for highly planar surfaces and tends to 0 for patches in which the points are scattered equally in all directions (e.g. vegetation, complex geometries etc.) for which the estimated normal do not completely make sense for the purposes of our investigation. Notice that the two extreme cases are more theoretical than practical, and, the case of a 0 weight will probably never happen, implying no observation is completely discarded from the solution. This point is important for certain [distance, angle] domains for which a small number of observations may be present and completely discarding them may be not a good choice.

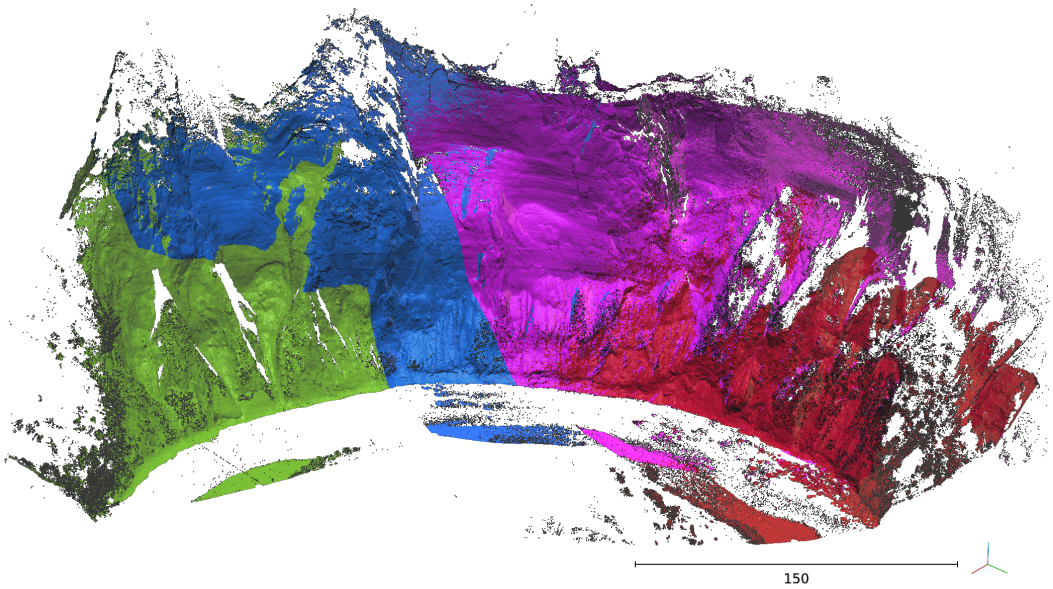
Another weight we may want to take into account is the standard deviation of the intensity values σ_{int}, when it is *high* it may suggest that the underlying patch is composed of more than just one material, and thus under-weighting this observation is desirable. A natural choice is to use the reciprocal of the inverse of the variance:

$$w_{int} = \frac{1}{\sigma_{int}^2} \tag{3.11}$$

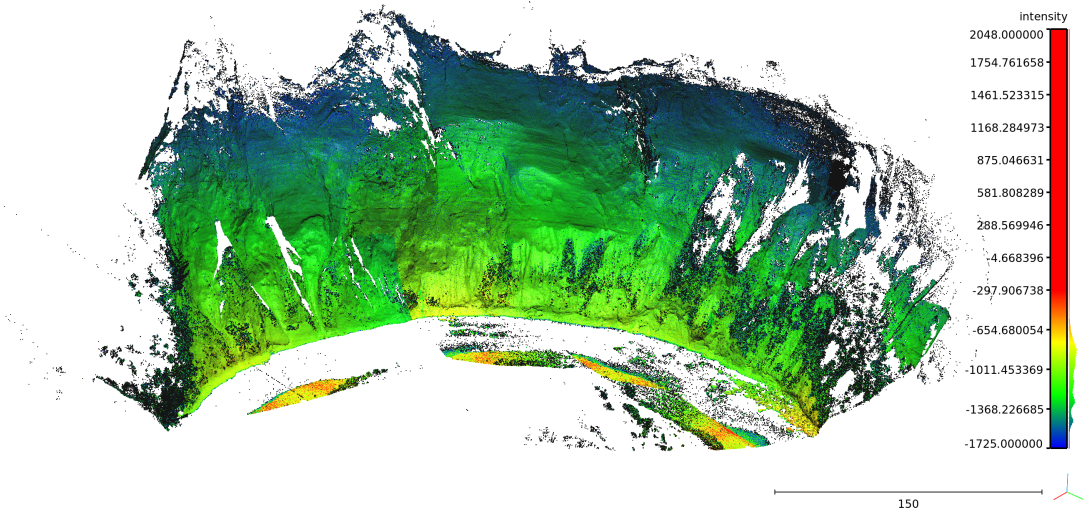
The two weighting factors can then be used to weight the observations and thus solve the calibration problem by means of weighted least squares.

3.5 RESULTS AND DISCUSSION

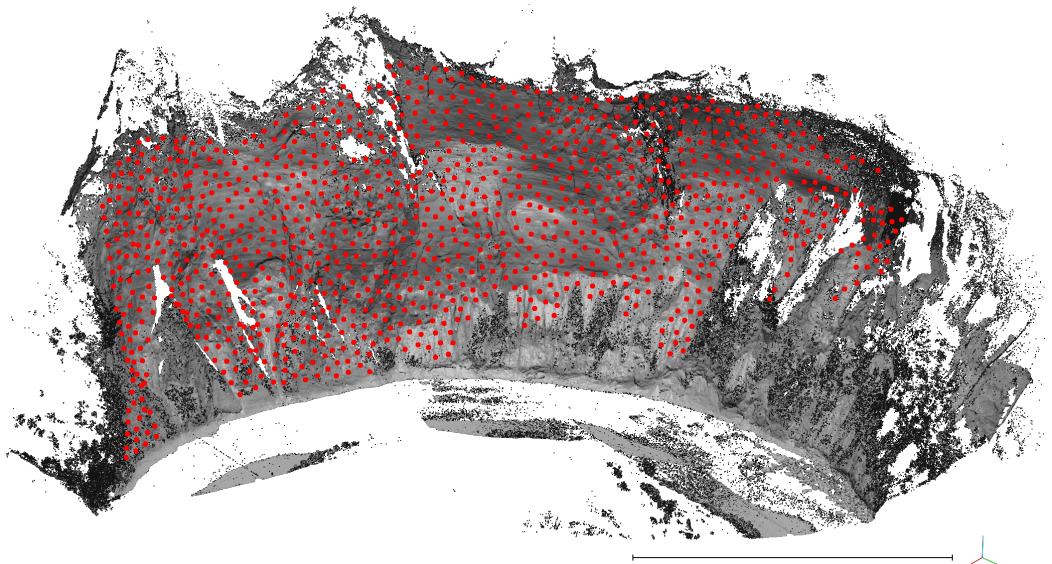
Figure 3.4 shows the Ranigitikei dataset which is composed of 4 point clouds (Figure 3.4a), each coming from a different scan operation. The scene is composed by two materials (Lague et al., 2013) that are expected to show different reflectance: vegetation and the bedrock. Minor variations of the bedrock composition are expected to have small or negligible effects on the recorded intensity due to the long ranges.



(a) The Ranigitikei dataset, composed by 4 partially overlapping clouds, each taken from a different station. Each point cloud have has a different color to visualize the overlapping areas.



(b) The intensity field of the Ranigitikei dataset, before normalization. The colorbar also shows a histogram of intensities, which is spread on a large range, mainly due to the effect of distance.



When the dataset is merged without any normalization (Figure 3.4b the effects of the distance from the device are evident. The intensities from this device (Leica Scanstation 2) are exported as a 12 bit unsigned integer (thus in the range $[-2047, 2048]$). These values have been shifted to the range $[0, 4095]$ before any further processing.

Figure 3.4c shows the keypoints that have been extracted semi-automatically from the clouds. In this case the point clouds have been merged and cleaned from most vegetation by manual editing. The resulting point cloud has then been downsampled so to end up with ~ 1000 keypoints, each separated by the nearest by ~ 2 meters.

Figure 3.5 shows the data extracted from the point clouds using the method described in Section 3.3. For each keypoint, a spherical neighborhood of 0.1 m was considered to compute local normals and average intensities. The plots at the top and at right show the relationship between the variables of interest, here highlighted by their 1st order polynomial fit.

These values have been modeled by means of RBF smoothing, the model consists of 24 nodes, 6 for the distance and 4 for the angle. The node have been evenly distributed in the $[\alpha, d]$ space. A gaussian has been used as weighting basis function (as *kernel*). The surface representing this model is shown in Figure 3.6 and highlights a clear correlation between distance and intensity. The effect of incidence angle is here less pronounced, as expected for a natural scenario.

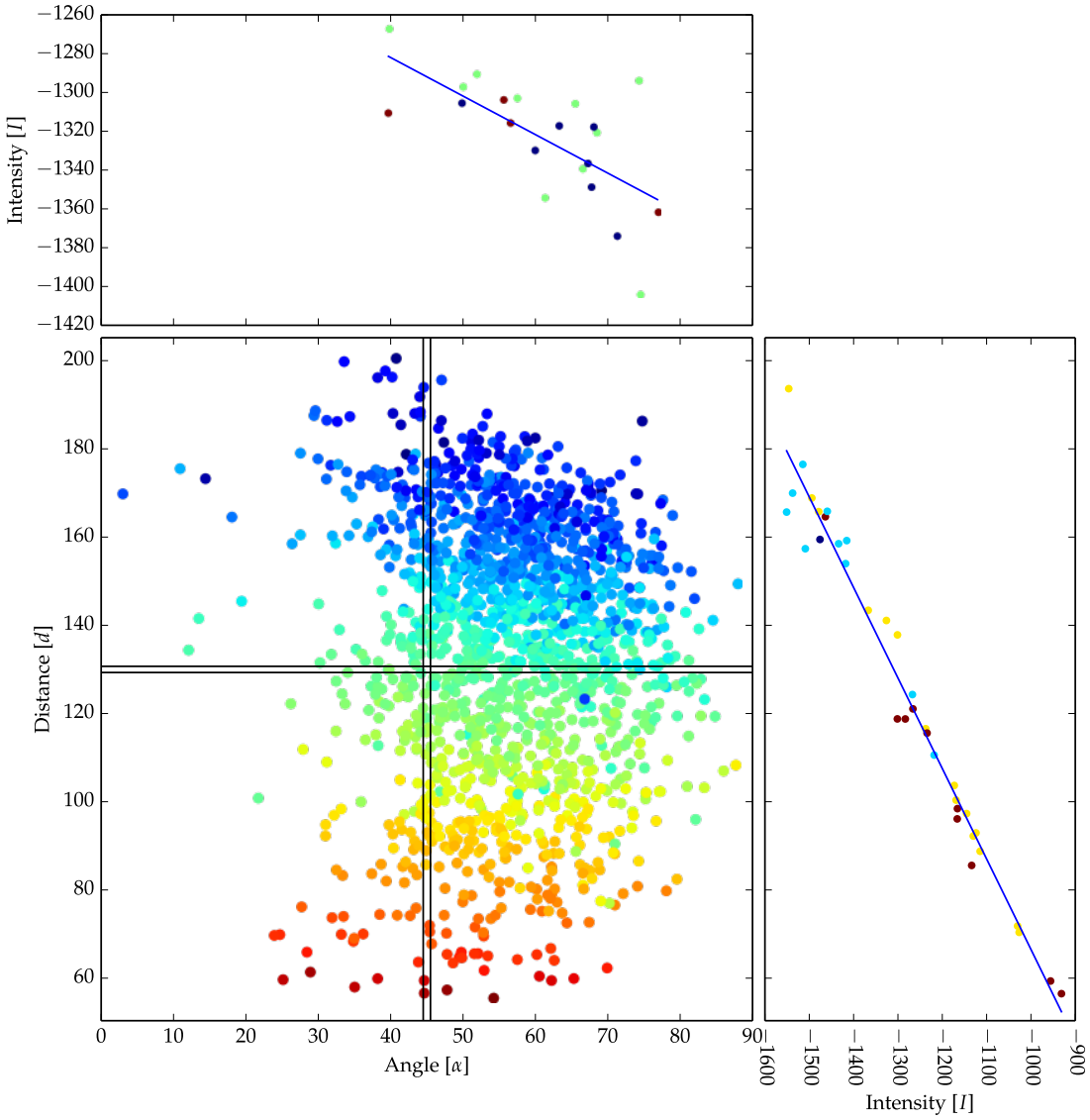


Figure 3.5: Scatterplot of the observations collected at the keypoints (Figure 3.4c). Color by measured intensity. Intensities appear to decrease both with the distance and the scattering angle. Top plot shows the relationship between intensity and angle for the fixed distance of 130 ± 0.7m (the two horizontal lines). A similar plot for distance vs. intensity (angle 45 ± 0.5 °) is given on the right. Colors for these two plots are by scan's origin.

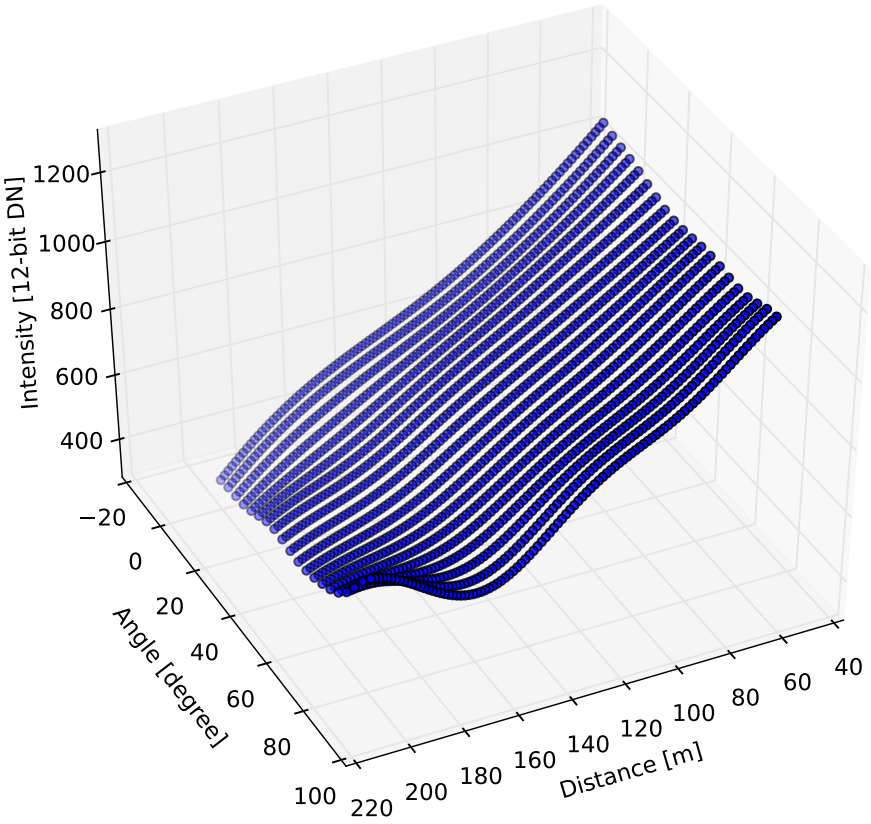


Figure 3.6: Even sampling of the resulting RBF model for the Ranigitikei dataset.

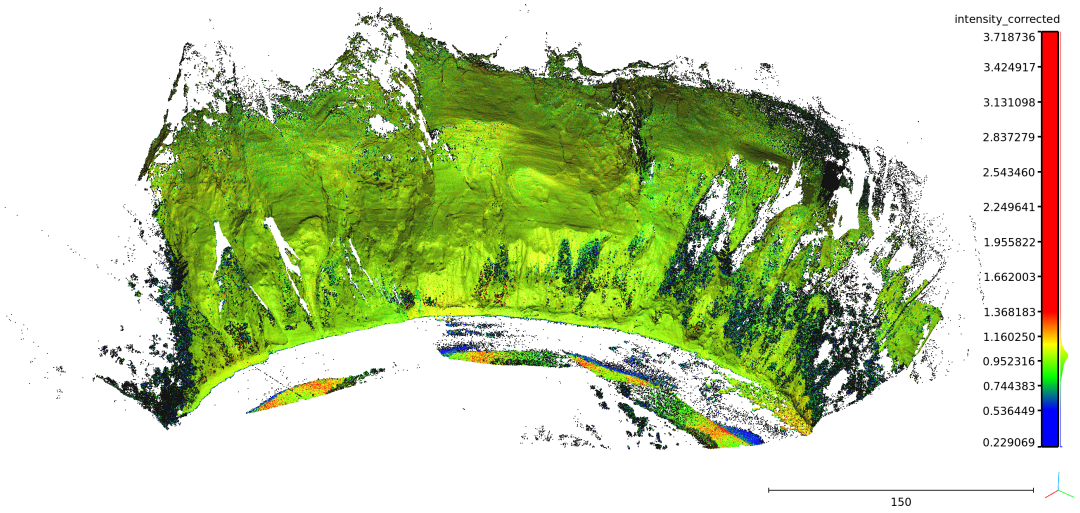


Figure 3.7: The intensity field merged after normalization. Colors by normalized intensities. The bedrock has now assumed an average value of 1 while the vegetation systematically shows lower intensity values. The overall distribution of intensities (small graph on colorbar) is now compressed and most of the effects of the distance are suppressed

4

INTENSITY RESPONSE TO CALCIUM CARBONATE. THE SMIRRA QUARRY CASE

Contents

FROM THE PAPER

Marco Franceschi, Luca Penasa, Jan Smit, Rodolfo Coccioni, Jérôme Gattacceca, Antonio Cascella, Sandro Mariani, and Alessandro Montanari. Comparison of carbonate content, facies ranking and Terrestrial Laser Scanner intensity as proxies for the cyclostratigraphic analysis of the Ypresian-Lutetian pelagic section of Smirra (Umbria-Marche Basin, Italy). *Palaeogeography, Palaeoclimatology, Palaeoecology*, submitted, 2015

4.1 INTRODUCTION

This chapter is a summary of the paper by Franceschi et al. (2015) recently submitted. In this paper we made use of the techniques introduced in Chapter 2

and 3 to investigate a continuous and well-bedded section of pelagic limestones. The outcrop exposes the R4 member of the Scaglia Rossa Fm. and the lower part of the Scaglia Variegata Fm (Figure 4.1). The section is exposed in a small abandoned quarry for a total stratigraphic length of ~ 25 meters. The section encompasses the Lutetian-Ypresian (Lu-Yp) boundary, which was defined by Molina et al. (2011) in the Gorrondatxe section in Northern Spain.

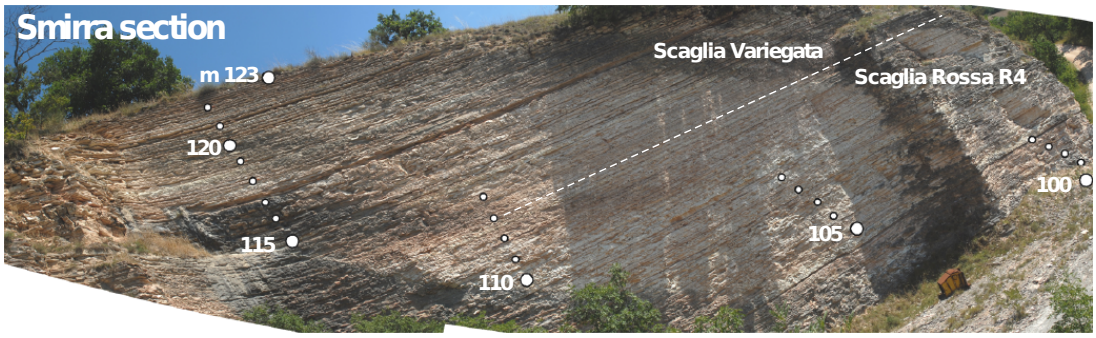


Figure 4.1: The Smira section for which $CaCO_3$ and the TLS time series were compared.

The main goal of our investigation was to locate the Lu-Yp boundary in the specific section. This problem was tackled by means of a detailed multiproxy investigation which comprised bio- and magneto-stratigraphic studies, lithological logging, carbonate content analysis and finally an intensity series obtained from an OPTECH ILRIS 3D TLS. Cyclostratigraphic analysis was then carried out on the numerical series in the tentative to place the boundary also from an astrochronological perspective.

The results of this study will not be discussed here in details but the aspects which are relevant to this thesis will be instead expanded and more carefully discussed. The application of the TLS technique to the Smirra quarry required to perform a more in depth investigation of some specific issues related to the characteristics of the outcrop and to the response of the intensities to the $CaCO_3$ content.

The intensity series which can be produced by means of TLS is an attractive way to obtain a lithologic proxy, in a fast and geometrically accurate way. A *good* proxy for a cyclostratigraphic analysis must be correlated to some other system's variable that is expected to be *sensible* to the climate.

The intensity measured with the OPTECH TLS is a Digital Number (DN) that, in the best case, can be transformed into a proxy for the reflectance of the material (see Chapter 3). Gaining a full understanding of how this remotely-

sensed variable can be related to the various properties of rocks is a key aspect for being able to consider the intensity a *good* cyclostratigraphic proxy.

Mix et al. (1995) demonstrated the use of measured reflectance for providing fast logging of cores. They were able to estimate some of the main characteristics of the sediments ($CaCO_3$ content, biogeneic silica and terrigenous component). But the spectrometer they were using permitted to measure the *reflectance* at 511 different wavelengths, covering most of Visible (VIS) ($455 - 700$ nm) and Near InfraRed (nIR) ($700 - 945$ nm) bands.

The TLSs, instead, work with just one wavelength (1541 nm for the ILRIS 3D scanner) that cannot be changed in any way, and the intensity is a proxy for the reflectance at that specific wavelength. It should therefore appear clear that the intensity value is unlikely to provide enough information to *simply* convert it into any kind of more familiar measure, for example a *$CaCO_3$ content series* or a *biogenic silica series*.

It is hence more sensible to understand in which ways the rock properties affect the intensity values in each specific case, identifying what properties are essentially captured by the intensity series.

The Smirra quarry gave the opportunity to face with this problem, allowing a direct comparison between the intensities and other observations (the $CaCO_3$ series and the lithologic log).

The R4 member of the Scaglia Rossa Formation is mainly composed of pelagic red limestones containing a variable amount of a clay. It is a well bedded series with minor interbedded shales, while the prominent bedding has been interpreted as deriving from pressure-solution stylolitization (Alvarez et al., 1985) that exagerated an originally . In the lower part also cherts are present. Toward the stratigraphic high of the Smirra section, the Scaglia Variegata Formation is richer in marls, which alternates in a rhythmic fashion with limestone.

4.2 CYCLOSTRATIGRAPHIC PROXIES

Two different *classical* proxies have been considered for the cyclostratigraphic investigation: a lithological logging, and a $CaCO_3$ wt.% series. These two datasets were provided by Sandro Montanari of the Geological Observatory of Coldigioco (Aprito, MC).

The lithological log was measured with a folding rule in the lower part of the outcrop while for the upper part the log has been measured by repelling with a rope down from the quarry's rim. The litholog was constructed considering

the three lithologies composing the series: carbonate, marlstone, and chert. Resulting lithologic series is shown in **??**.

In the same Figure the $CaCO_3$ wt. series is presented. The bulk rock samples (about 30 g each) for producing this series were collected with hammer and chisel through the lower part of the section at an average spacing of 2.5 cm for a total of 668 samples. These samples were powdered with mortar and pestle and ~ 300 mg aliquots of the $< 250\mu m$ fraction were analyzed for total calcium carbonate content ($CaCO_3$ wt.%) using a Dietrich-Frühling water calcimeter with a precision of ± 2.5 wt.%.

Notice that the rock samples which have undergone the analysis have been stratigaphically located manually in the field. Moreover, the samples were all collected in limestone beds, while shales and marly interlayers have been systematically overlooked. The masured $CaCO_3$ content is consistently greater than 85% and most of the samples recorded a value between 90 and 95 $CaCO_3$ wt.%.

4.2.1 *TLS scan and Intensity preprocessing*

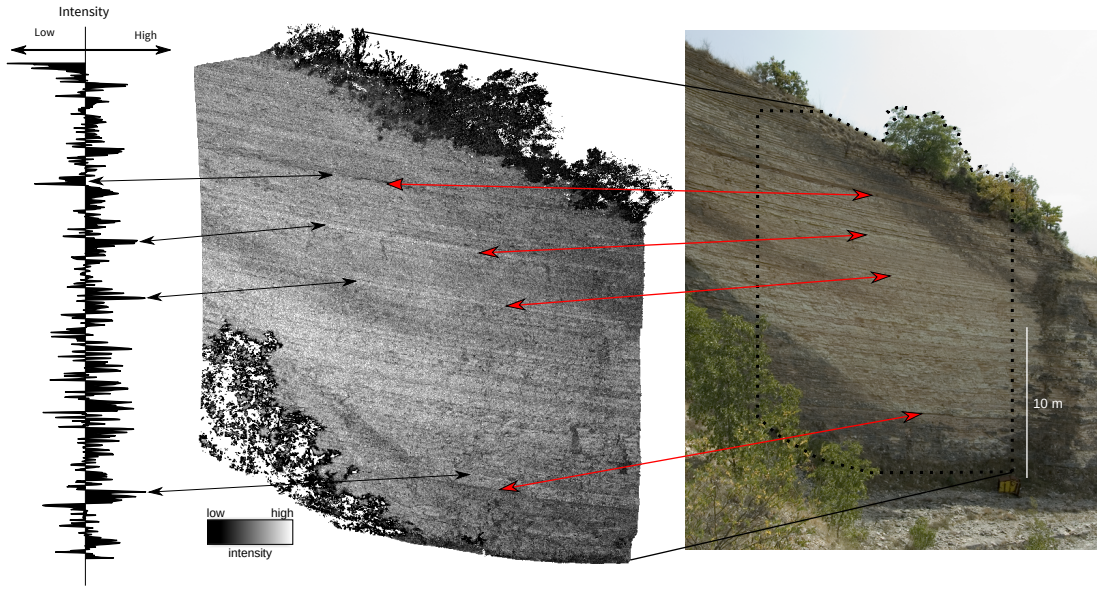


Figure 4.2: TLS scan compared to a photograph. Correlation between time series, scan and photograph highlighted by the arrows.

Single point acquisition accuracy at 50 m is ~ 0.7 cm, resolution at 50 m is ~ 1.77 cm (Lichti and Jamtsho, 2006). The Smirra section was scanned from a

it is not clear what these measures means

single station at a distance between 31 and 40 meters. At such distance, the surface is sampled with an average density of $\sim$ 7000 points per m^2, with a sampling step on the surface of $\sim$ 1.2 cm.

Before producing the intensity series, intensity values were corrected in function of the distance of the target from the acquisition position. Correction was obtained using the Kernel Smoothing approach (1.0 m bandwidth) suggested by Penasa et al. (2014) and further elaborated in Chapter 3.

For this scan, no incidence angle compensation was applied (see Chapter 3), because the scan was acquired in frontal position with respect to the outcrop, hence the $>$ 80 % of the points resulted to be viewed from an incidence angle $< 45°$ (average angle of the scan is 31°) . At these incidence angles the effect of the loss of power due to the incidence angle is small (Pesci and Teza, 2008; Hartzell et al., 2014).

The point cloud needs then to be cleaned from undesired points before generating the final intensity series, and this was obtained by removing evident vegetation patches and areas covered by debris, following the procedure suggested by Franceschi et al. (2009, 2011).

4.2.2 *Gaussian detrending*

Alteration and moisture on the outcrop result in visible intensity variations that are not related to lithological changes. When these surface alterations are uniformly present at small scale (small patches on the outcrop), their effect on the final time series is mitigated by averaging many points when the intensity series is generated (cfr. Franceschi et al., 2009, 2011).

Each sample of the final time series is in fact produced by a large number of independent observations of the intensity made at the same stratigraphic level. From these observations an unique scalar value is determined averaging them, thus providing statistical *strength* to the sample's value.

At Smirra, some large scale variations in intensity which are visibly incoherent with the stratigraphy can be observed. Large-scale variations do not display lateral persistence, hence they can be considered linked to superficial alteration and/or moisture on the rock. In the intensity series, the effect of these variations may results in low frequency trends that are not related to variations in the lithology, and thus have to be removed before time-series analysis. If not removed, it would produce low-frequency fluctuations in the final time series, impairing the signal-to-noise ratio.

These low frequency fluctuations in the intensity values can mitigated by either:

1. using a subset of the original cloud for computing the time series, as has been suggest by Franceschi et al. (2009) that is chosen to be more representative of *good* outcropping conditions (no patinas or strong weathering), and

2. by trying to suppress the *portion* the intensity field that are caused by *external* factors via further processing of the point cloud.

The first implies by meticulous cleaning of the cloud, followed by a careful selection of the points that are involved in the process of creating the time series. The points used for the creation of the Smirra time series are highlighted in Figure 4.4.

The second point be faced by either detrending the final time series or by some kind of processing of the point cloud. As described in Chapter 2, the application to point clouds of fairly basic algorithms which are commonly used, e. g., in image processing is not straightforward and requires to take into account various optimization issues.

At Smirra, we used a low-pass gaussian filtering ($\sigma = 1.0$ m) to enhance the intensity field. The result of this operation can be seen in Figure 4.3. This scalar field has been subtracted from the original intensity field, thus acting as an high-pass filter on the final scalar field.

The removal of long-wavelength trends has to be carried out carefully to avoid the loss of lithological variations that should be preserved. After this preprocessing only a subset of the original point cloud was used for producing the final intensity-series.

Considering that the final goal is to produce time series that will undergo spectral analysis, it is important to consider what effect this operation can have on the final time series.

The low-pass effect of gaussian filtering depends on its standard deviation σ. Although the implementation for point clouds its basically different by its image processing counterpart we can get an idea of the possible effects from standard signal processing theory. The cut-off frequency of the filter is:

$$f_c = \frac{1}{2\pi\sigma} \tag{4.1}$$

Thus in this case this filtering dampens all the low frequencies in the scalar field, slightly dampening periodicities of ~ 6 meter, and reaching higher attenuation towards lower frequencies.

In a cyclostratigraphic context, whether or not this filtering can be accepted will definitively depend upon the studied frequencies. The lower frequency that is expected to be detectable is the 400 kyr period Long Eccentricity (LE) that, with the expected sedimentation rate at Smirra of about < 10 m/Myr should be recorded as a < 4 m period cycle.

It is important to underline that such operation must be evaluated case by case, on the basis of the characteristics of the specific outcrop, because long-wavelength trends could be also due to lithological variations, and in that event they should be preserved. Further confirmation that low-frequency trends should be removed can be obtained from the $CaCO_3$ content time-series, that does not present any evident drift other than a possible linear trend.

4.2.3 *Time series generation*

To construct an intensity series, raw intensities were exported using the OPTECH ILRIS 3D proprietary parser as a 16 bit Digital Number (DN). The stratigraphic position of the points was defined by generating a plane in the 3D space having the average attitude of the layering and defining the distance from that plane as the stratigraphic position of each point.

Hence, a running weighted average was computed with the desired sampling step to generate a log (i. e., the intensity series) that represents the variation of intensities along the considered stratigraphic section. The intensity series was produced with a sampling step of 1.0 cm, employing a moving average with a gaussian kernel with a bandwidth of 1.0 cm. Each value of intensity series associates to each stratigraphic height an intensity that derives from the averaging of the intensities of all the points of the point cloud located at a specific stratigraphic height in the section.

As pointed out by Franceschi et al. (2011), this approach can be used effectively if the layering in the area considered for the intensity series generation displays a uniform attitude. Another assumption is that the characteristics of each layer are homogeneous at the scale of the outcrop. This is necessary to obtain a log in which a certain stratigraphic height is associated with a representative intensity value. This can be considered a reasonable assumption for most geological cases.

The final intensity series of the Smirra section is shown in Figure 4.7, where intensity values are rescaled for being comparable with the $CaCO_3$ amplitudes.

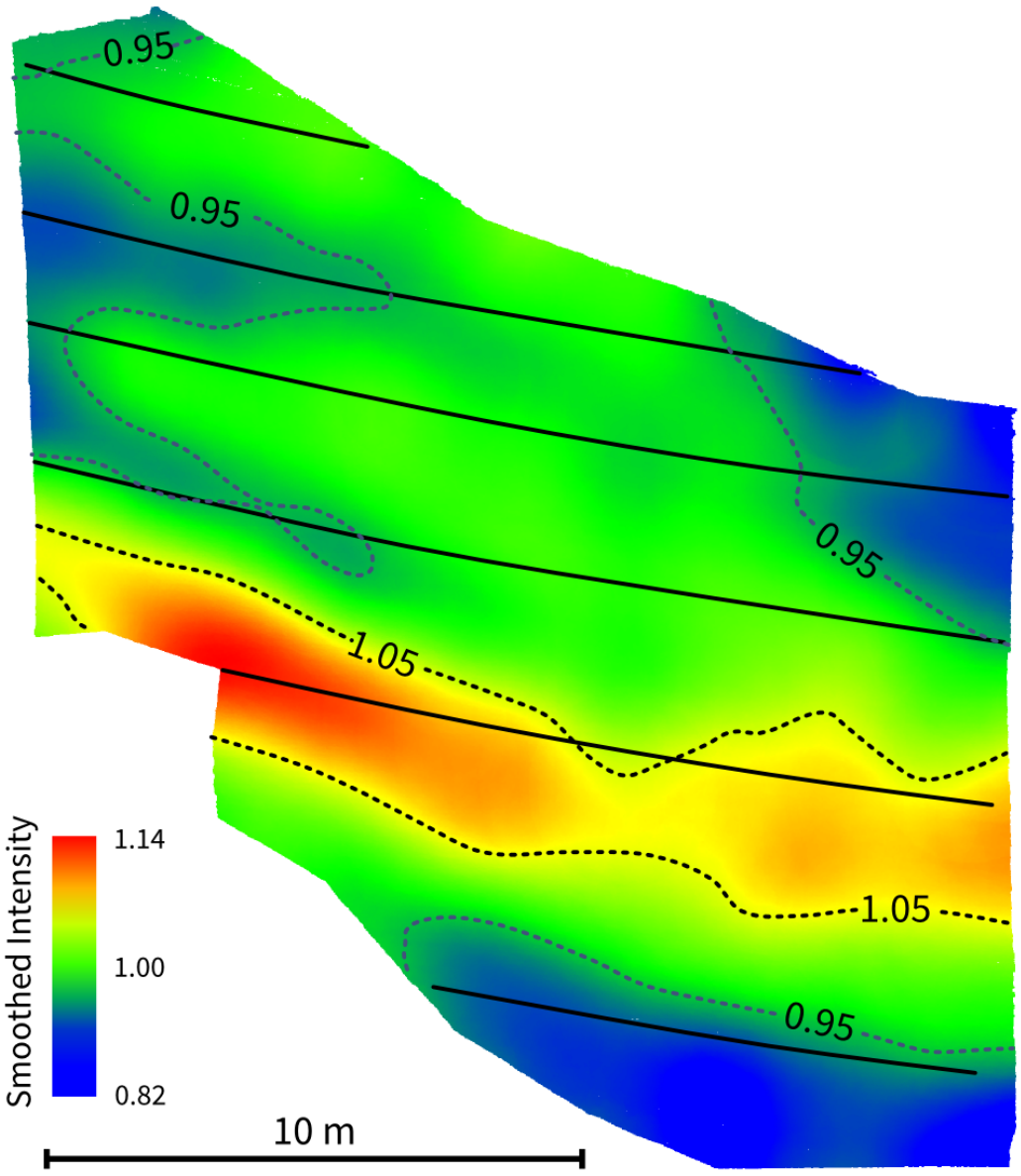


Figure 4.3: The average intensity for the point-cloud, obtained by gaussian convolution of the intensity values. These large-scale intensity variations are due to superficial alteration of the rock rather than from lithological-controlled intensity variations. Furthermore they are inconsistent with the stratification. The smoothed intensity has been subtracted from the original intensity for accounting for variations of intensities unrelated to lithology.

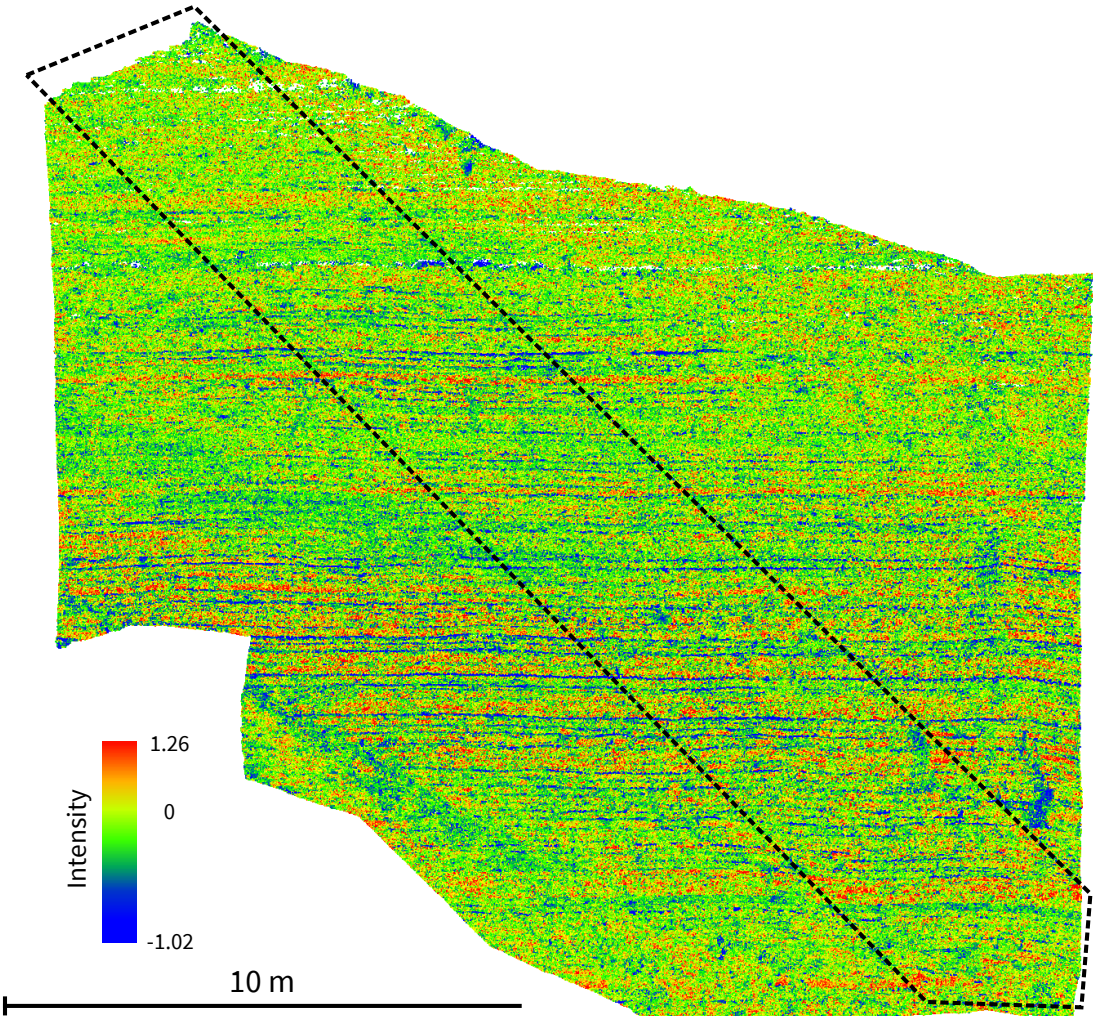


Figure 4.4: Intensity field displayed as colors. The portion used for computing the final time series is highlighted by the dashed polygon.

4.3 PROXY COMPARISON IN THE SPATIAL DOMAIN

Comparing the proxies in the spatial domain requires, in the first place, to define a common stratigraphic frame. This can be promptly done considering the stratigraphic positions reported in Figure 4.1. The same levels for which the position is known, can be recognized in the scan data, as in Figure 4.2. In this way the intensity series can be shifted in stratigraphic position to match the $CaCO_3$ series.

This allows for a *registration* of the different time series but it does have some limits when considering the time series in the spatial domain. The TLS series comes from a method which ensures a high level of geometric accuracy, while the other time series have been built by manual measuring the section with a ruler. We can expect this last technique may introduce uncertainties in terms of positioning of the samples for $CaCO_3$ measurements, which may affect also the overall length of the measured series.

The intensity series can be directly compared with the lithological log. As previously discussed, the TLS intensity shows sensitivity to both clay content and chert. Both chert and clayey layers produce local minima in the intensity series. Chert effect on the series may be of variable prominence, depending on the thickness and lateral persistence of chert layer: thinly layered or nodular chert are likely to produce shoulders on the positive peaks of limestone rather than deep local minima. Figure 4.7 shows a possible correlation of the minima recorded by the TLS intensity compared to the lithologic log.

On the other hand, samples for $CaCO_3$ analysis were collected on limestone only (85 to 91 $CaCO_3$ wt.%), thus the $CaCO_3$ wt. curve does not contain any *direct* information on shales or chert presence. Although this obvious limit the two series still present strong similarities.

Higher $CaCO_3$ contents appear to be linked with positive intensity values, while a negative peak of $CaCO_3$ is positioned near shale interlayers, which are well recorded by the intensity series. This fact suggests that within a single carbonate stratum (or pack of strata) the $CaCO_3$ is not constant and the wt.% of carbonates is lower near the shales and is higher in the middle of the stratum. Whether or not these oscillations are symmetric with respect of a stratum center may be matter of further investigations.

Interlayers that are present in the litholog but that are not clear in the intensity series and/or into the $CaCO_3$ series are probably due to psudobedding, especially abundant at the base of the section, which have been sometimes wrongly logged as shale layers. The intensity series, and the $CaCO_3$ series, are immune to this kind of disturbing elements.

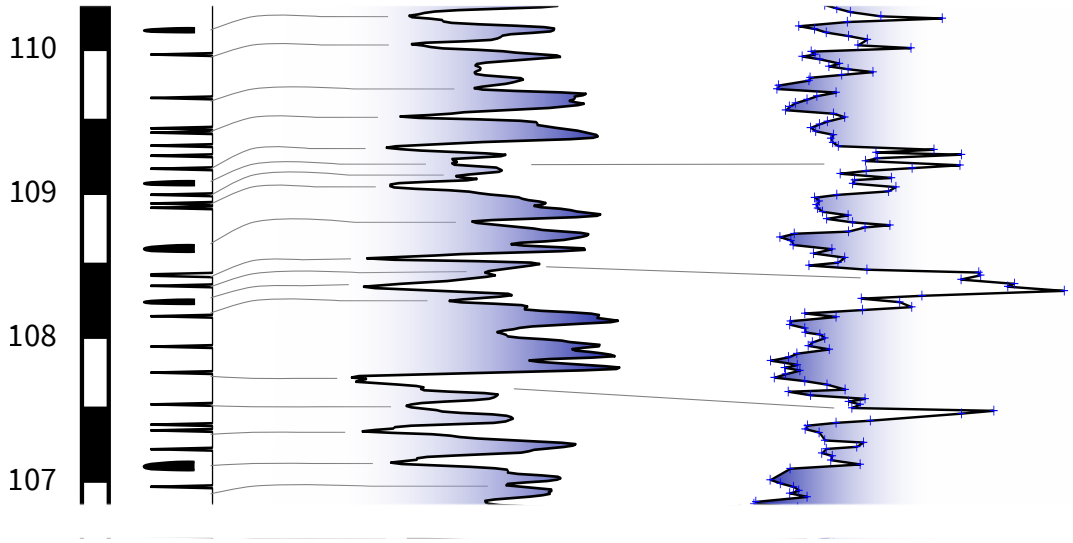


Figure 4.5: Enlargement of Figure 4.7 which shows some patterns clearly identified in both the time series. These packs of strata are of ~ 1 m in average in thickness.

These observations are summarized in the simple model of Figure 4.6, which can be used for interpreting the TLS-derived time series in the specific context of Smirra.

The observed similarities between the two proxies suggest the possibility to employ semi-automated matching for better constraining the two datasets. We give an example of this procedure performed with the software *Match* (Lisiecki and Lisiecki, 2002).

Match computes an optimal mapping between the two time series. The result of the matching is shown in Figure 4.8. Red arrows correspond to the two ties which were defined by hand before the optimization. Figure 4.9 shows the scatterplot of the intensity vs. $CaCO_3$. For producing this plot, the *matched* $CaCO_3$ time series was used. For each $CaCO_3$ sample the corresponding intensity has been sampled from the TLS series.

Pearson's r correlation coefficient (Pearson and Filon, 1898) can be computed as a simple measure of the existent correlation. The pearson coefficient is a value between $+1$ and -1, 1 means a positive correlation, 0 no correlation at all, and -1 corresponds to a negative correlation. In this case the coefficient for the dataset of Figure 4.9 is $+0.72$, with a statistically significant *p-value* ($p < 0.05$). Notice that the rescaling of the intensities for *visual* comparison, as e. g. in Figure 4.8, does not have any effect on the *Pearson's r* coefficient.

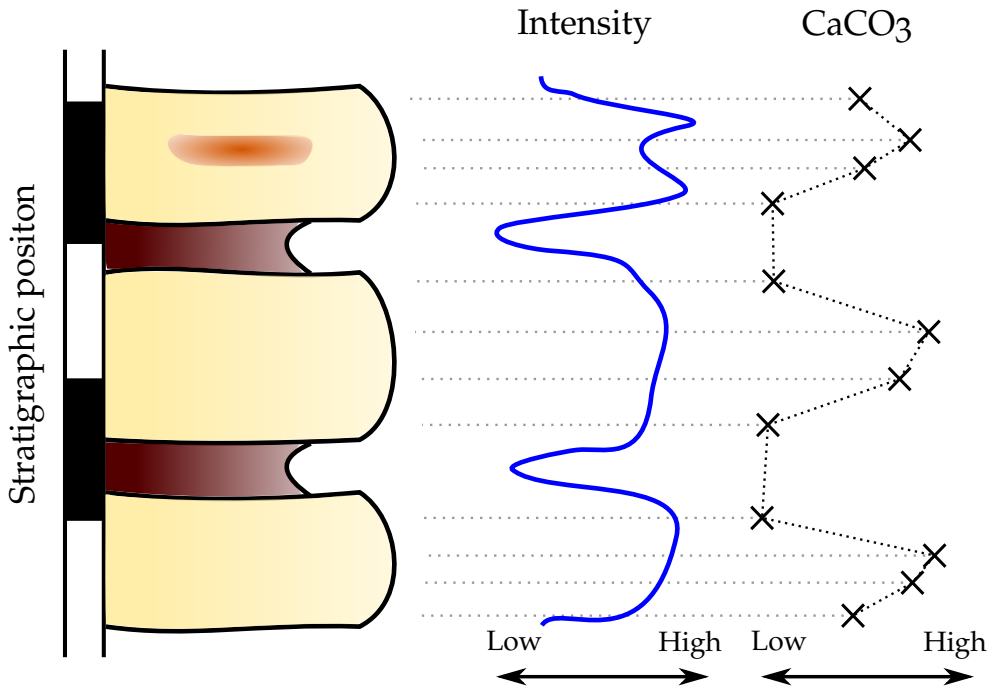


Figure 4.6: Simplified model which explains the intensity vs. $CaCO_3$ behavior at Smirra. $CaCO_3$ was collected on limestones only, while intensity is sensitive to shales and cherts. The $CaCO_3$ content of a stratum is not constant, but instead appears to be higher at the middle of the stratum while it decreases in proximity of the interlayers.

It is clear that this high level of correlation can be obtained only when the two time series have been matched. For the original time series, the *Pearson's r* coefficient is rather low, of $+0.11$, but it still shows statistically significant *p-value*, confirming a strong correlation between the two time series.

4.3.1 *Spectral analysis*

Cyclotratigraphic analysis are carried out by means of investigations in the frequency domain of the various proxies (see e. g. Ghil et al., 2001). In this section we will give some insights on the relationship that occur between the $CaCO_3$ and the intensity values. We can consider the portion of time series with overlap for comparing the $CaCO_3$ to the intensity time series. A well known method for spectrogram estimation is the Multi Taper Method (MTM) introduced by Thomson (1982) and that method will be used for investigating periodicities. An interesting way to explore what spectral features are present

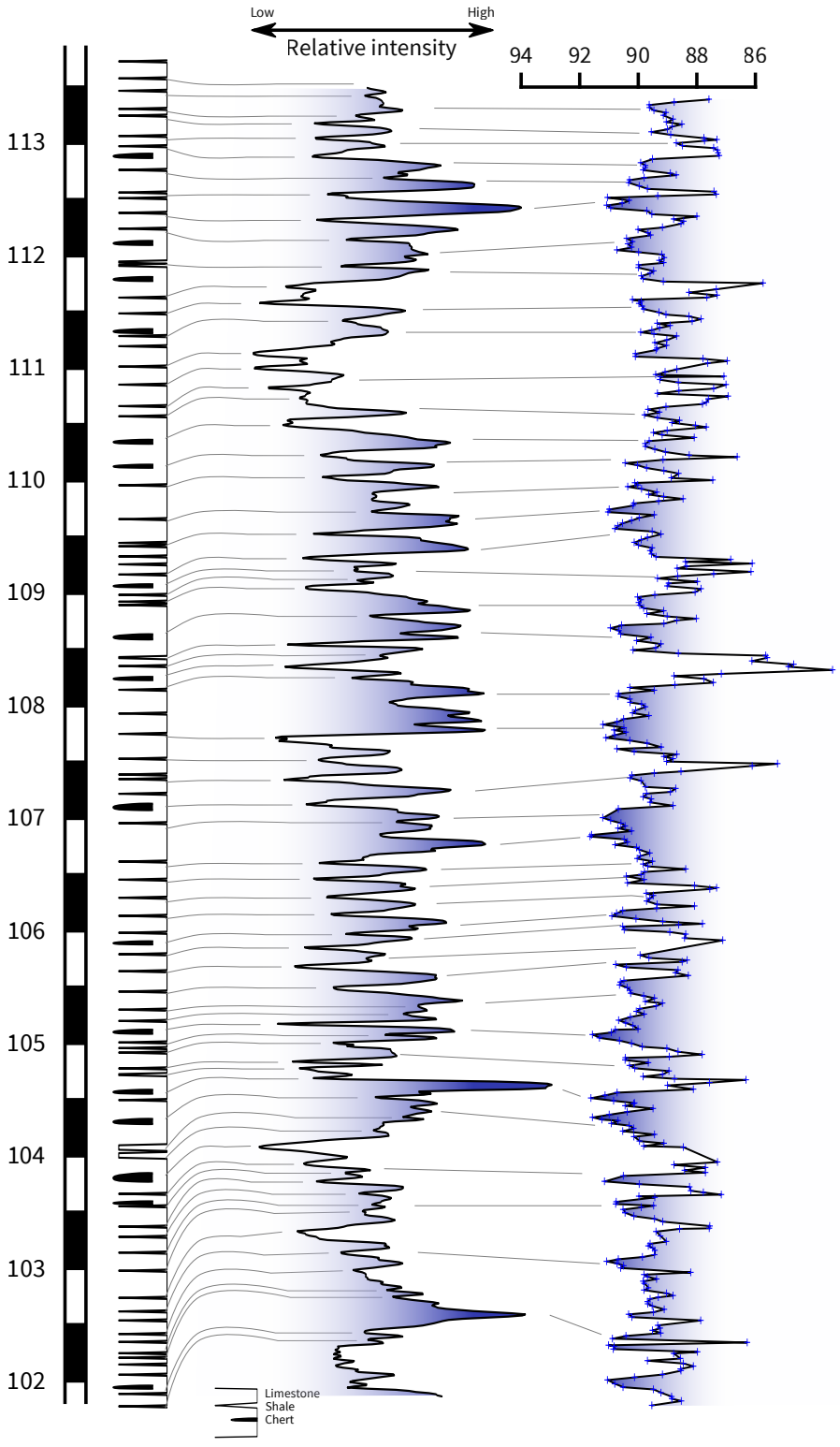


Figure 4.7: Correlating the TLS time series to the litholog and the $CaCO_3$ series requires to take account of the errors that may have been introduced by the manual logging of the lithologic log and the consequent inaccuracy in the stratigraphic positioning of the samples collected for $CaCO_3$ analysis.

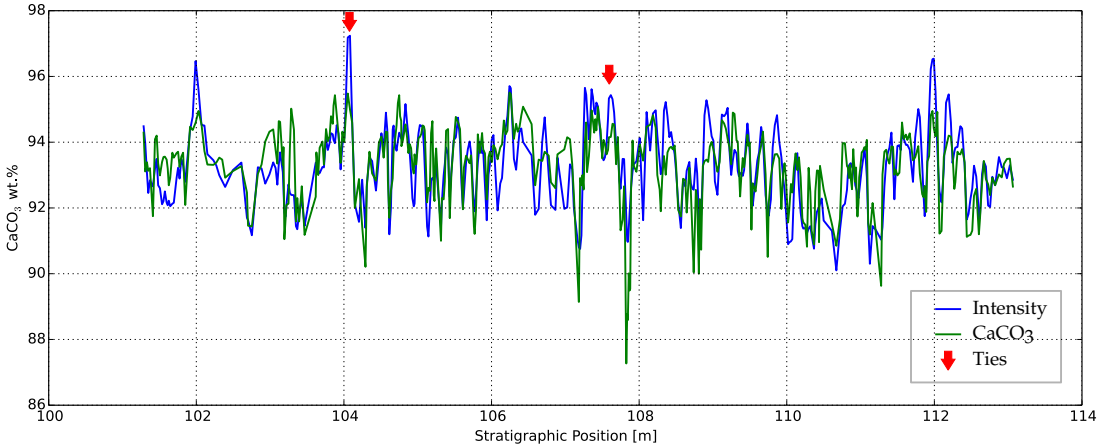


Figure 4.8: Result of the automated matching between the two time series. Ordinate scale is the $CaCO_3$ wt.%. The intensities have been rescaled both in the mean and in standard deviation to be comparable with the $CaCO_3$ time series.

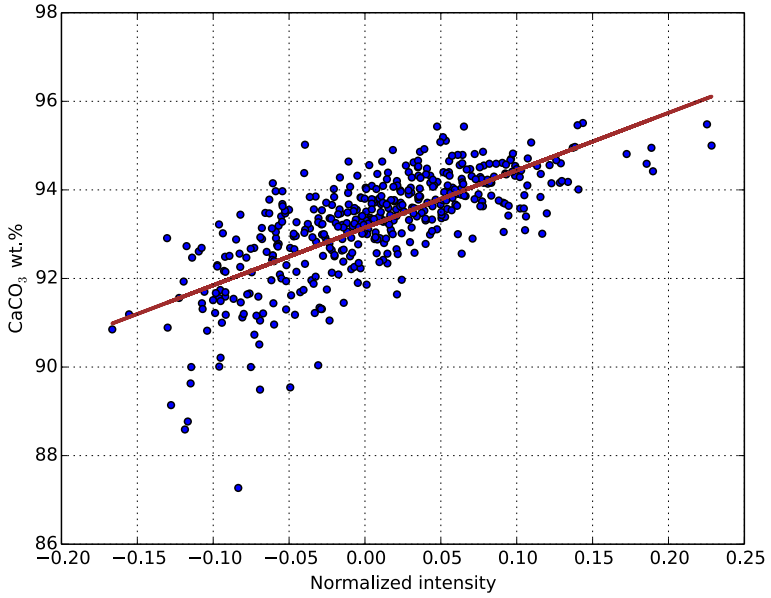


Figure 4.9: Scatterplot of the *matched* $CaCO_3$ series and the corresponding intensity values. A strong linear correlation can be observed. See text details.

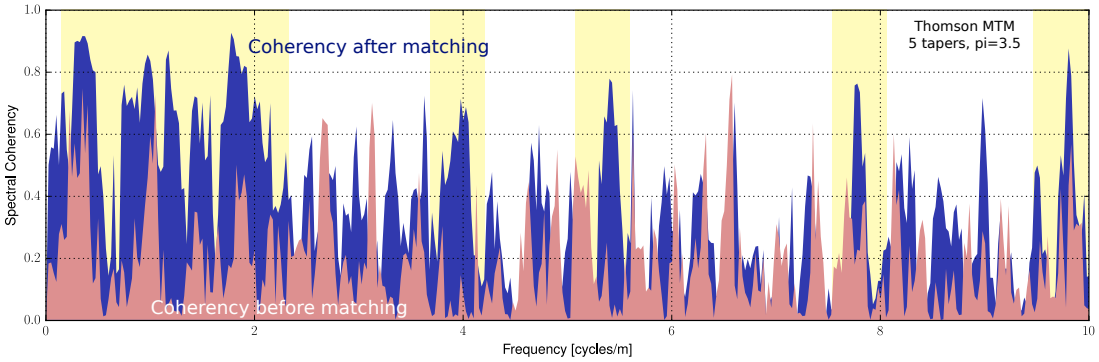


Figure 4.10: Spectral coherency between the $CaCO_3$ and intensity, before and after the matching. The matching process (which works in the spatial domain) uses the local maxima and minima to create a mapping between the two time series. This process leads to an increase of coherency of some frequency bands. In the original time series, the various components are *slightly* out-of-tune and do not show up as strong coherency values. After the matching, some specific bands are powered and their coherency increase. These bands represent cyclicities which are present in both time series.

in both the time series is the *Coherency* between the two. Figure 4.10 shows the coherency spectra of the two series before and after the automated matching.

The frequencies at which a strong increase of coherency can be observed correspond to the main spectral components of the time series which have been used by *Match* in the spatial domain to perform its matching, and should be regarded as *interesting* spectral bands that may carry most of the *common* information embedded in the time series.

The $CaCO_3$ time series underwent changes in the spectral domain after the matching. Figure 4.11 shows the two spectra computed by means of the MTM of Thomson.

The *matched* time series shows a general increase of the frequencies of the main peaks, while the spectral content is mostly preserved. The compression of the time series is of ~ 0.5m, corresponding to the $4-5\%$ of its overall length (only the part which overlaps with the intensity was used for the matching). If the intensity series is considered the best estimate of the *real* length of the section, we can argue that a $\sim 5\%$ error was introduced by the manual measurement of the section, assuming the intensity series better represents the total length. These errors exert an effect also on the spectra and can potentially influence the cyclostratigraphic interpretation if not taken into account.

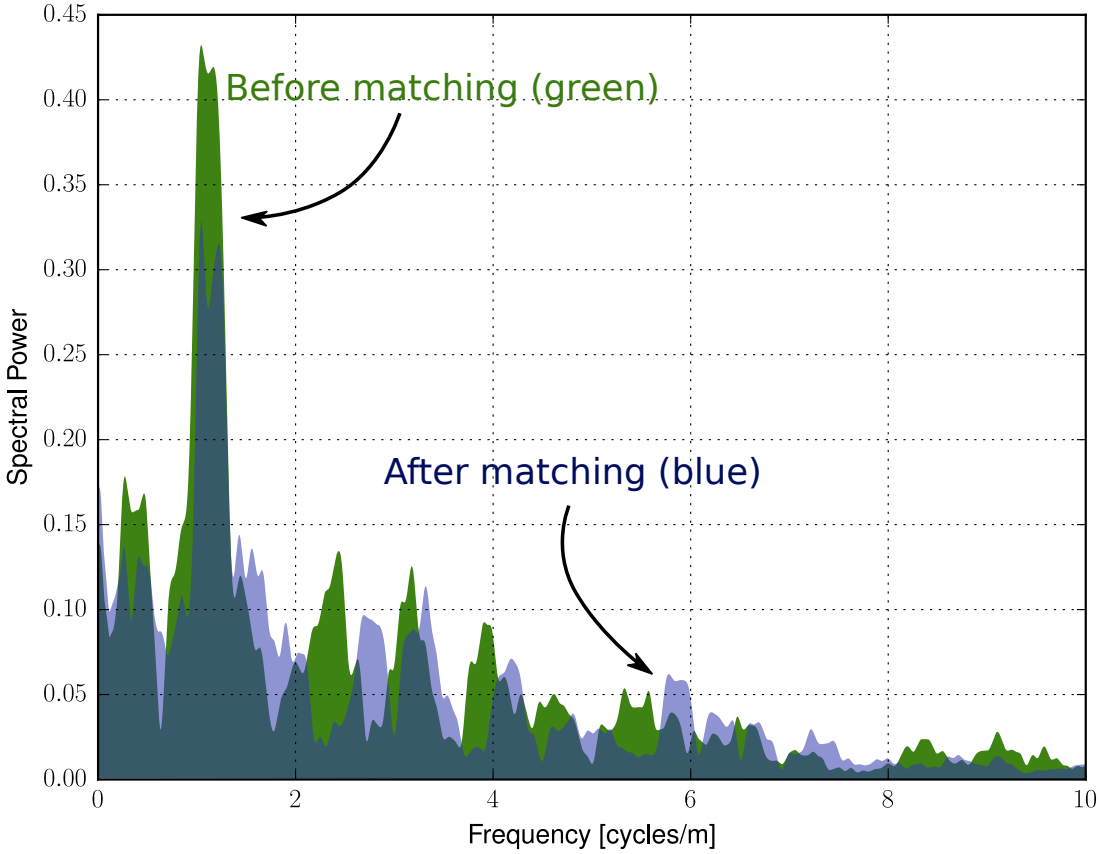


Figure 4.11: Spectra of $CaCO_3$ time series before and after matching. The main peaks are mostly preserved by the matching process, but the frequencies of these peaks have been generally shifted toward higher frequencies, compatible with a compressions of the $CaCO_3$ time series introduced by the matching. MTM with $p = 4$

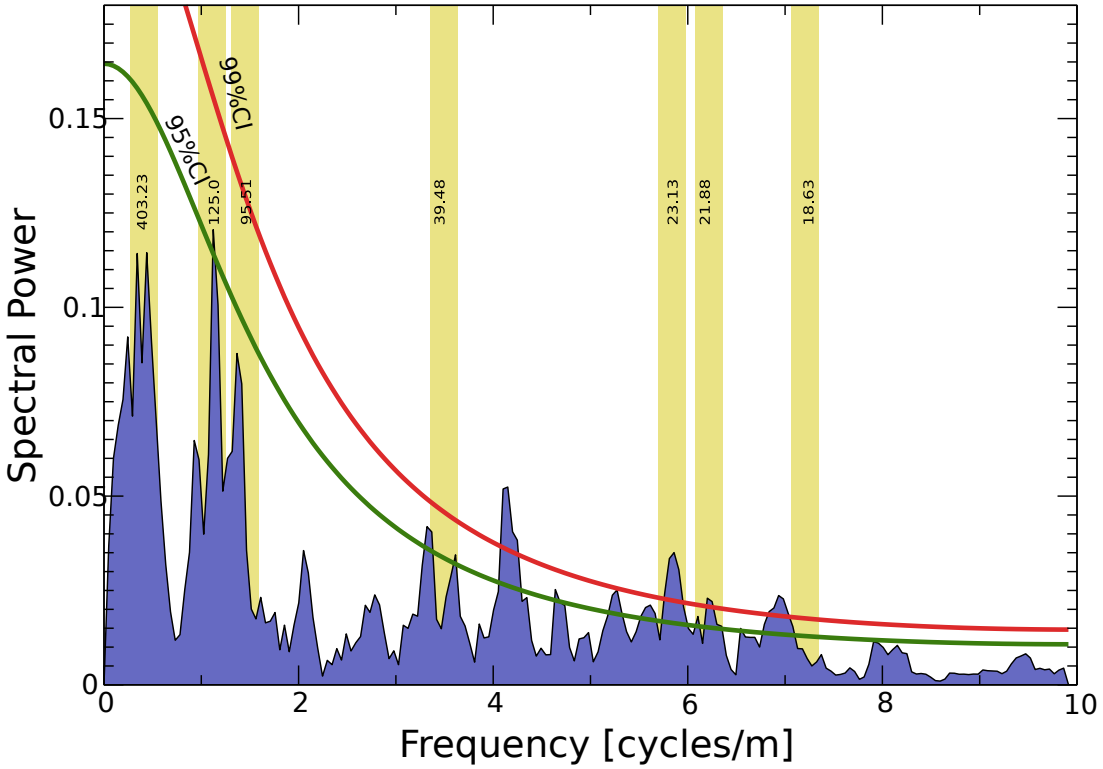


Figure 4.12: Spectrum of the intensity time series obtained from MTM, with p = 3. Overlayed the expected Milankovitch frequencies, as predicted by the Laskar10a solution for the eccentricity component and the Laskar04 for the obliquity and precession.

4.3.2 *Spectral analysis of the entire time series*

The time series collected have undergone time series analysis in their whole length, for this investigation the *unmatched* $CaCO_3$ time series was used.

Spectral analysis was conducted on the TLS intensity and the $CaCO_3$ wt.%. Thomson MTM was used for spectrum estimation (Thomson, 1982). The statistical significance of spectral peaks was evaluated against a red noise null hypothesis, estimated by means of the *robust* noise modeling proposed by Mann and Lees (1996). Results of spectral analysis are shown in Figure 4.12 where the expected frequency pattern provided by the recent numerical simulations of the solar system are also highlighted (Laskar et al., 2011, 2004). Computations were performed by the *SSA-MTM Toolkit* (Vautard et al., 1992).

$CaCO_3$ wt.% series yields a spectrum which is dominated by a prominent peak emerging well above the 95% confidence level (confidence interval (c.i.))

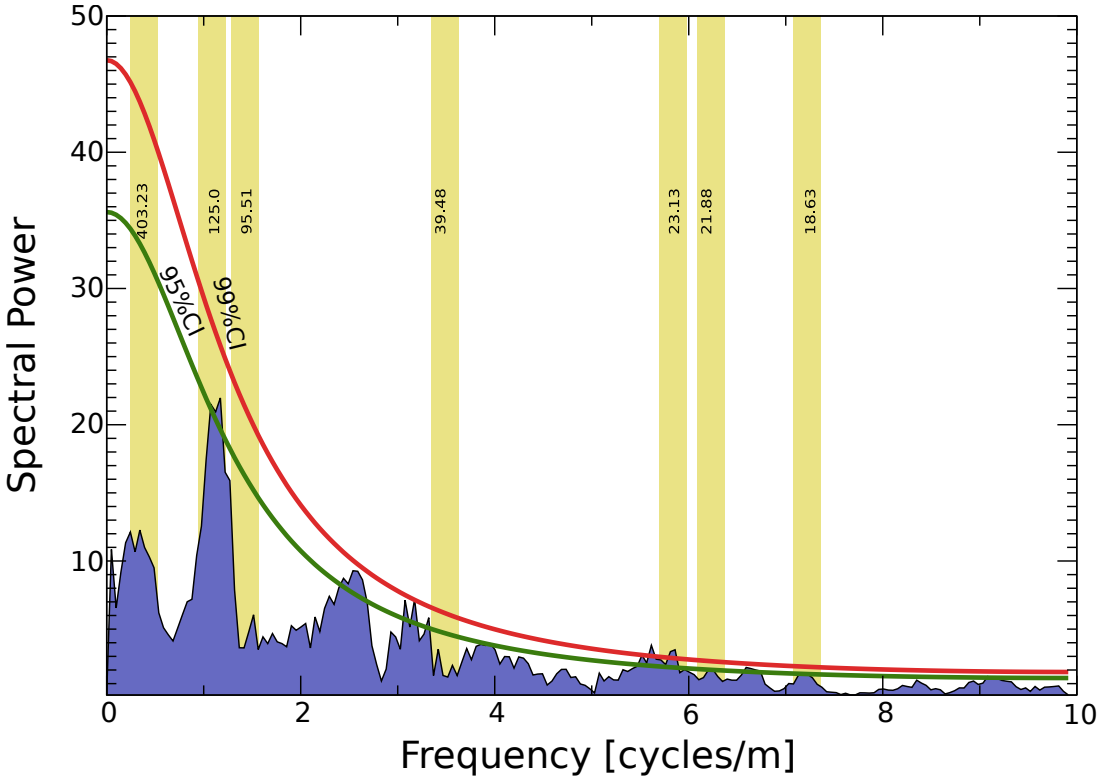


Figure 4.13: Spectrum of the $CaCO_3$ time series obtained from MTM, with $p = 3$. The original dataset of $CaCO_3$ has been resampled to produce an evenly spaced time series ($F_s = 50$ cycles/m). Overlayed the expected Milankovitch frequencies, as predicted by the Laskar10a solution for the eccentricity component and the Laskar04 for the obliquity and precession.

with a period of ~ 85 cm. The power spectrum of the TLS intensity series is also dominated by a powerful frequency signal with a period of 87 cm

An average sedimentation rate of 9 ± 1 m/Myr for the Ypresian-Luthetian interval was estimated by Coccioni et al. (2012) from the integrated biomagnetostratigraphy of the classic Contessa section at Gubbio, located a few tens of kilometers from Smirra. Hence, on the basis of this estimate, and assuming a mean sedimentation rate of ~ 9 m/Myr for the coeval Smirra section, the dominant frequency peaks with a period of 87 cm emerging in both $CaCO_3$ wt.% and TLS intensity series, would correspond to a frequency period of about 95 kyr, thus very close to the period of 95.1 kyr for the dominant short eccentricity frequency calculated by Laskar et al. (2011).

4.3.3 *TLS vs other proxies*

The application of a multi-proxy approach to the cyclostratigraphic study of the Smirra section allowed direct comparison of the results obtained by analyzing TLS intensity and the $CaCO_3$ wt.% series. This reveals that TLS imaging provided better results in terms of accuracy in the stratigraphic collocation of the measurements, and in terms of spectral resolution.

When calcareous homogenites are considered, as this succession of Smirra, outcome of hand logging can be biased by the fact that pressure solution stylolites can be easily misinterpreted as true beds or thin marly interlayers.

Correlation of TLS intensity series with the $CaCO_3$ wt.% series is high. This implies that in this calcareous homogenite, TLS intensity can be used as a reliable proxy for detecting variations in the carbonate content relative to the siliciclastic terrigenous component. Of course, TLS data cannot be used as they are, but they need to be pre-processed following the methodological pipeline described.

It is worth noticing that sampling for $CaCO_3$ wt.% analysis was conducted in limestone layers only. Shales were not sampled, hence the curve represents variations in carbonate content within carbonate layers. This can be related to cyclic increase and decrease in terrigenous content relative to primary biogenic carbonate (i.e. planktonic foraminifera and calcareous nannofossils). On the contrary, minima in the TLS series correspond to interlayers. The good correlation between intensity series and $CaCO_3$ wt.% curve, suggests that a decrease in calcium carbonate content within layers evolves in a shale interlayer.

The successful application of TLS to calcareous homogenites widens the possibilities of application of this technique, which was tested so far only in

distinct limestone-marl alternations (Franceschi et al., 2011). Moreover, TLS allowed to analyze the entire Smirra section whereas it was possible to carry out other investigations only in the limited accessible portion of the outcrop. This permitted to undertake a complete cyclostratigraphic characterization of the section.

5

INTEGRATION OF INTENSITY TEXTURES AND LOCAL GEOMETRY DESCRIPTORS FROM TERRESTRIAL LASER SCANNING TO MAP CHERT IN OUTCROPS

Contents

PUBLISHED PAPER

Luca Penasa, Marco Franceschi, Nereo Preto, Giordano Teza, and Vanessa Polito. Integration of intensity textures and local geometry descriptors from Terrestrial Laser Scanning to map chert in outcrops. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93:88–97, July 2014. ISSN 09242716. doi: 10.1016/j.

isprsjprs.2014.04.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0924271614000902

ABSTRACT

An investigation of the potential of Terrestrial Laser Scanner imaging (TLS) as a tool to map chert, an amorphous variety of silica widely diffused in sedimentary rocks, is shown here together with an original method for its automatic recognition, normally difficult to achieve in natural outcrop conditions. The presented methodology may have important consequences for geologic applications like stratigraphic studies or rock-quality assessment. Results of reflectance measurements of chert and its host rock, performed in the 400-2500 nm spectral band using a high performance UV-VIS-NIR spectrophotometer, show that chert displays low reflectance in the IR wavelengths typically operated by several commercial TLS. An infrared-light (1541 nm) TLS was used to scan a target outcrop of limestone with chert nodules and beds with the goal of mapping chert. The information deriving from distance-corrected intensity is coupled with local geometric descriptors and SPIN images, to the end of exploiting the geometrical characteristic of TLS data and improve the classification results. These features, together with the intensity values, were used to train two Support Vector Machine classifiers with the purpose of separating vegetation from rock and then limestone from chert. Principal components decomposition of the SPIN features shows that grass, limestone and chert are distinguishable with the first three principal components. The results of classification are then cross inspected in the field and with reference pictures. Results demonstrate that TLS data can be efficiently exploited to map chert when the monochromatic information of the intensity is integrated with feature descriptors and a support vector machine predictive model.

5.1 INTRODUCTION

Chert is a widespread sedimentary rock made of cryptocrystalline varieties of silica. Its origin is mainly biogenic, deriving from the settling of siliceous skeletal parts of microorganisms through the water column on the seafloor and its subsequent transformation into a stable form during early sediment burial. Chert usually occurs in the form of nodules of various size and shape, or in layers (Figure 5.1), embedded and often scattered within carbonate rocks. Silica occurs in chert in various mineralogical forms, including chalcedony, microcrystalline quartz, amorphous silica (opal-A), disordered cristobalite and tridymite (opal-CT) (Boggs, 2009; DeMaster, 2005). Chert constitutes a research topic for geologists, because its presence, absence and abundance can be related to sedimentary processes, past climatic changes and may help reconstructing the events in the environmental evolution of Earth (e.g., Ikeda et al., 2010).

Figure 5.1: Hand-sized chert samples and host rocks. Chert aspect varies from highly scattered (center) to a more hybrid banded and scattered geometry (left) to single, wide bands. Colors vary from black to yellow and brown, sometimes it is hardly distinguishable from its host rock due to similarity of colors (right).

Chert is commonly embedded in limestone. Limestone rocks are routinely used as component for concrete aggregates and the presence of chert can affect concrete quality and durability because of Alkali-Silica Reactions (ASR) (Bektas et al., 2008) or thermo-mechanical weathering (Xing et al., 2011) that may occur after concrete hardening. The amount of chert must then be monitored. This can be done directly on the aggregates (for example via image analysis, see Castro and Wigum (2012) or via petrographic methods, see Sims and Nixon (2003) but also a pre-evaluation on the quarry wall could provide useful information on the expected properties of the final product. The evaluation of chert abundance in outcrops can be, however, difficult and time consuming because chert is often

scattered and can be very similar to the host-rock in color. In this paper we illustrate a method aimed at quickly quantify and accurately map chert directly on the outcrop using terrestrial laser scanning (TLS).

The application of remote sensing techniques to discriminate various types of rocks in outcrops has been explored in several works in the last few years. Satisfactory results have been achieved using passive hyperspectral imaging sensors coupled, when necessary, with terrestrial laser scanning or photogrammetry to provide the 3D framework for the hyperspectral information (Kurz et al., 2012b, 2013; Murphy and Monteiro, 2013). Hyperspectral imaging operates in the short wave infrared (SWIR) range (1000-2000 nm), with ~ 5 nm typical frequency resolution, and therefore provides good results because the main spectral features of geological materials fall in this range. Nevertheless, the currently achievable spatial resolution is not better that some cm for a 50 m acquisition distance, limitations related to the use of passive sensors exist (e.g., necessity of radiometric calibration and dependence on conditions of illumination) and accurate registration on 3D models is necessary to facilitate quantification of material distribution and spatial relationships. This imaging technique is expected to gain increasing importance in the future because of technological and methodological developments.

TLS is an active sensor that produces an accurate, high-resolution 3D representation of an object in the form of a point cloud. Besides spatial information, each point carries an intensity value of the backscattered laser signal, that is function of the reflectance properties of the target. When chert is scanned with a TLS operating in the IR band, it displays low intensity values with respect to its host rock (typically limestone). In this paper we demonstrate that this feature, in association with the geometrical information carried by TLS data, can be exploited in order to map chert. Several papers investigated the potential of TLS for discrimination of rocks (Bellian et al., 2005; Pesci and Teza, 2008; Burton et al., 2011; Hodgetts, 2013). Franceschi et al. (2009, 2011) discuss the potential use of intensity as a proxy of rock properties. Outside the fields of geology the study of TLS intensity has been undertaken to detect damaged areas on historical buildings (Armesto-González et al., 2010), to quantify moisture content in aeolian sand deposits (Nield et al., 2011) and to identify biological crusts on structures (González-Jorge et al., 2012). The fact that commercial TLS usually work with a single-wavelength laser results in a minimal spectral information carried by the intensity signal, Hartzell et al. (2014) recently proposed to combine multiple TLS working at different wavelengths in the intent of overcoming this limitation.

Also the problem of segmenting point clouds on the base of geometric properties was faced by many authors, especially for airborne laser scanning data (Maas and Vosselman, 1999; Verma et al., 2006) and for mobile applications (Yang et al., 2013; Pu et al., 2011), for automating vehicle navigation (Lalonde et al., 2006) and for forestry applications (e.g., Koch et al., 2009; Ferraz et al., 2012; Brandtberg, 2007). Hybrid methods using geometry and colors or intensities were demonstrated to be useful in various scenarios improving the effectiveness of the classification algorithm (Höfle et al., 2009; Schoenberg et al., 2010). Although highly automated methods (both supervised and unsupervised) have reached high levels of complexity and effectiveness, they are still sparsely used to solve automated data collection tasks in geology (Brodu and Lague, 2012; Abellán et al., 2014; Ferrero et al., 2009; García-Sellés et al., 2011) where point clouds are often used for visual inspection and interpretation, due to the difficulty in extracting meaningful and coherent information from that huge amount of data.

In this work we show that the spatial characteristic of a point cloud (e.g., local geometric arrangement of points) as well as the spatial distribution of the intensity values can be used to enhance the potential of TLS as a tool for the discrimination of rocks. The proposed TLS-based approach contributes to the emerging studies on the application of active sensors as remote sensing tools through supervised classification, using both geometry and intensity descriptors in a combined way. It might as well have important practical implications, ranging from stratigraphy and cyclostratigraphy to rock quality evaluation.

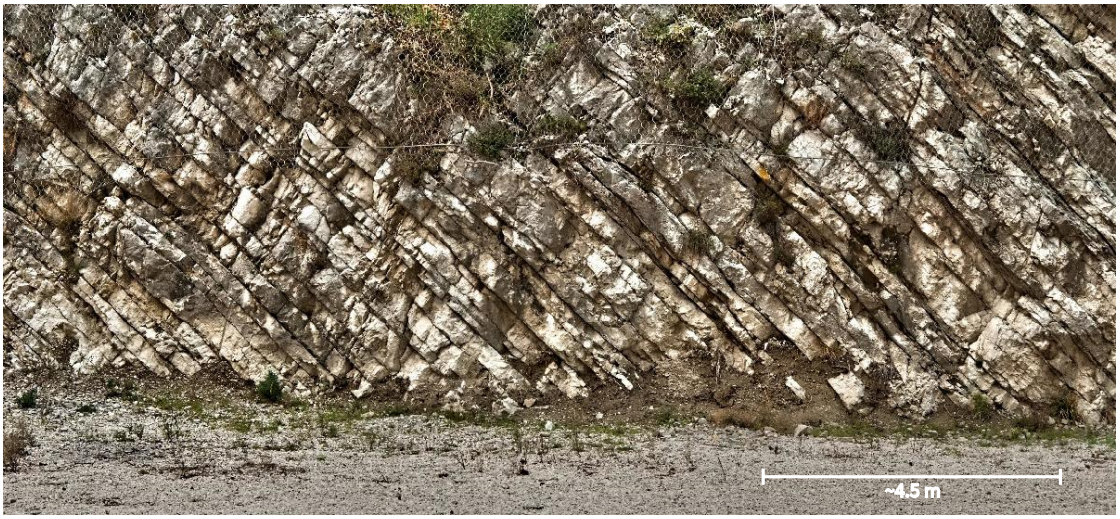


Figure 5.2: Photograph of the outcrop of Maiolica Fm nearby Gubbio taken with a standard camera. Note the noise present in the color domain. Rock color goes from pale pink to dark gray, chert is not evident.

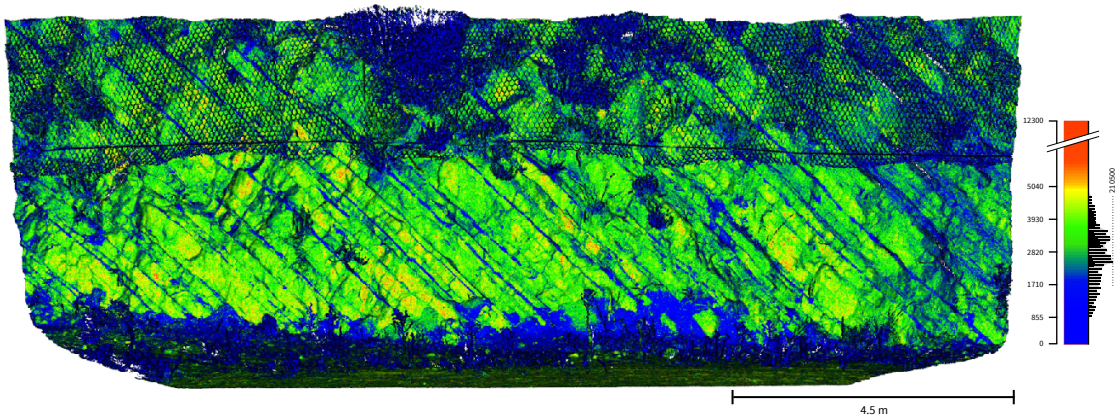


Figure 5.3: The scan, colored by intensity, and histogram of intensity distribution (right). Chert, vegetation and debris are highlighted by their low intensity values. CloudCompare software (Girardeau-Montaut, 2014) used for visualization.

5.2 METHODS

5.2.1 *Spectral and mineralogical characterization of chert*

To assess the spectral behavior of chert a set of 6 chert samples together with their corresponding host rocks (limestone) were analyzed with a spectrophotometer. All samples are made of micro-crystalline quartz, by far the most

common variety of chert. Reflectance spectra were investigated in the 400 to 2500 nm wavelength range with a spectral sampling step of 1 nm using a Cary5000 spectrophotometer. The device is manufactured by Varian and performs reflectance measurements from 200 to 2500 nm via a deuterium/halogen double light source, using an internal integrating sphere for diffuse reflectance measurements.

The samples for reflectance spectra measurement were prepared as ~ 2 x 4 cm slabs, cut with water-cooled circular saw, with no further preparations. The slabs were kept for several days at environmental humidity conditions before measurements.

5.2.2 *TLS acquisitions*

As target for testing the method an outcrop cut in cherty limestone (Figure 5.2) was selected in Central Italy, nearby Gubbio[1]. The scanned rocks (Maiolica Formation) are white micritic limestone in which chert is commonly found as centimeter to decimeter-sized nodules or layers of variable continuity. Chert color varies a lot (from white to black, red, yellow, etc...) and often resembles that of the host limestone. This in many cases makes chert indistinguishable, in the visible band of the electromagnetic spectrum, from the rocks in which it is contained.

The outcrop was scanned with an OPTECH ILRIS 3D TLS whose laser unit emits in the infrared band at 1541 nm wavelength (Larsson et al., 2007). The average distance between TLS and target ranged from 20 to 23 m. The acquisition produced a 1318 x 3277 grid of points with an average point spacing of 6.7 mm. The range image was converted from the binary format to unstructured point cloud using the OPTECH proprietary parser that removed dropout readings and outliers, slightly reducing the total number of points to 4,318,440. Intensities were exported using the option "RAW 16-bit"of the parser, as xyz ASCII format. The OPTECH ILRIS 3D works with two different gain domains and intensity exported as 16-bit results in values from 0 to 255 for the high gain while for the low gain the values are restricted to the set $\{300, 400, 500, \ldots, 25500\}$ (Larsson et al., 2007). We did not modify this arrangement given that cross calibrating low-gain to high-gain readings is not straightforward and the high-gain points are just ~ 1.0% of the total, and mainly exhibit very low intensity responses (i.e., vegetation).

1 43.360457 N, 12.579775 E

5.2.3 *Intensity preprocessing*

TLS intensities calibration is needed to use them for discrimination purposes. Although the topic has been faced by several authors (e.g., Pesci and Teza, 2008; Franceschi et al., 2009; Kaasalainen et al., 2011) a standard method for all TLS does not exist yet. Proposed calibration models try to estimate how intensities vary with distances and scattering angles, although other variables may be considered. Methods often make use of tarps with known lambertian reflectance to estimate the correction functions for a certain device (Kaasalainen et al., 2011). These calibration schemes are in fact rather complex and they cannot account for unpredictable fluctuations of the intensity values often observed for the specific instrument (Wang and Lu, 2009).

In this work we did not perform a correction of the values in function of variables other than distance (e.g., incidence angle, atmospheric corrections, etc. . .). Pesci and Teza (2008) and Hartzell et al. (2014) showed that the scattering angle effect can be considered negligible when the surface roughness is comparable with the size of the laser spot. Other studies (Krooks et al., 2013; Kaasalainen et al., 2009b) showed that the incidence angle becomes a significant variable only when it exceeds 20°for most materials. Since the outcrop was scanned from a frontal position we considered the effect of the angle of incidence as not relevant, however in the general case a more comprehensive and robust calibration is desirable (e.g., derived from Kaasalainen et al., 2011; Höfle and Pfeifer, 2007).

Our simplified calibration scheme is adaptive to the data, free of a physical model (i.e., no radar equation involved) and is based on estimating a correcting intensity-distance function $\widehat{I}_h(d)$ on an appropriate reference point cloud (a subset of the original scan or an additional scan), via a Nadaraya-Watson regression estimator (Hastie et al., 2009). The estimator computes a smooth function describing intensities as function of distances via a weighted moving average, using a Gaussian kernel as weighting function:

$$\widehat{I}_h(d) = \frac{\sum_{i=1}^{n} K(\frac{|d-d_i|}{h}) I_i}{\sum_{i=1}^{n} K(\frac{|d-d_i|}{h})}, \quad \text{with} \quad K(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \tag{5.1}$$

where n is the total number of points, d_i and I_i are respectively the distance and the intensity for the ith point. The parameter h (the bandwidth) controls the smoothness of the estimator. The corrected intensity $\dot{I}_i$ is then computed as $\dot{I}_i = I_i / \widehat{I}_h(d_i)$ for each point of the cloud. Figure 5.4a illustrates the intensity-distance relationship in a typical scan of a stratigraphic sequence and the intensities after

correction. This correction helps stabilizing the variance and normalizes the mean to 1.0 (Figure 5.4b). The plot also illustrates some characteristics of the OPTECH TLS. Intensity levels are discretized (~ 20 levels at 35 meters) although the possible range is $\{0, \ldots, 25500\}$. At longer distances the discretization is heavier with only 4-5 levels resolved at 50 m.

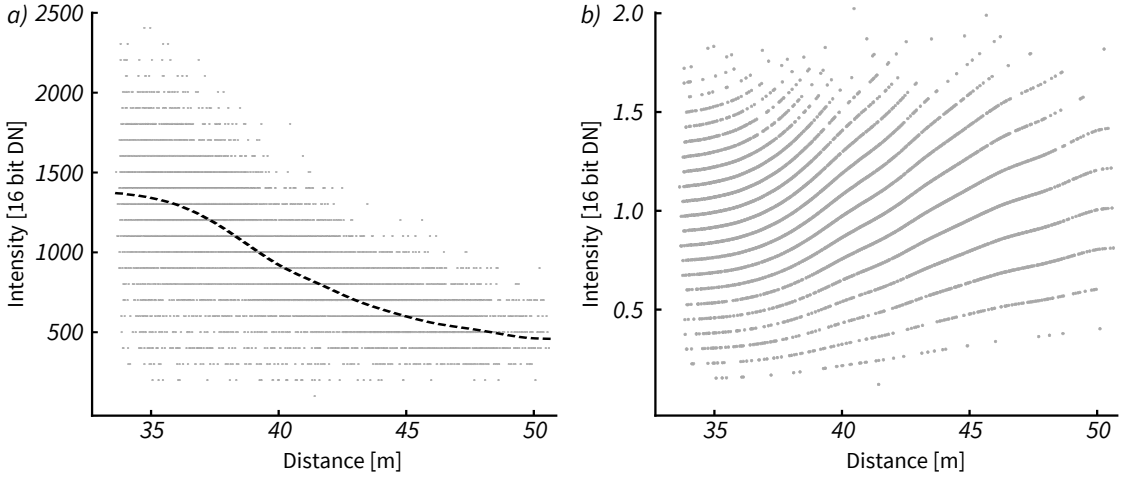


Figure 5.4: Distance-intensity scatter-plot of a typical OPTECH ILRIS 3D scan. a) Dashed line represents the trend estimated with Nadaraya-Watson estimator (bandwidth $h = 1.0$ m). b) Intensities after distance-correction. The corrected intensity is computed normalizing the "RAW 16-bit"intensities with the estimated trend.

5.2.4 *Feature descriptors: texture descriptors and local geometric dispersion indices*

The main obstacle to intensity-based classification derives from the monochromatic nature of the TLS signal. Since lots of factors influence intensity, many parts of the target that are not chert (clayey interbeds, vegetation, etc. . .) can display a lower reflectance than limestone. This means that classifying the point-cloud only on the basis of intensity can bring unsatisfying results. To overcome this issue, we took advantage of the differences in morphology that the various elements of an outcrop can display to add further variables useful for classification. Chert normally occurs in predictable shapes like nodules or bands (Figure 5.1). Moreover, due to its hardness, chert tends to resist to erosion and generate locally flat and smooth surfaces on the outcrop. Vegetation, being present in patches with scattered distribution of points, displays peculiar geometrical characteristics. In contrast, limestone is more easily fractured,

appearing angular and irregular, while clayey interlayers generate recesses. Texture and geometrical descriptors can be used to quantify these features and train a Support Vector Machine (SVM) classificator (see Section 5.2.5 of this paper and Mountrakis et al. (2011) for a review of SVM and its applications).

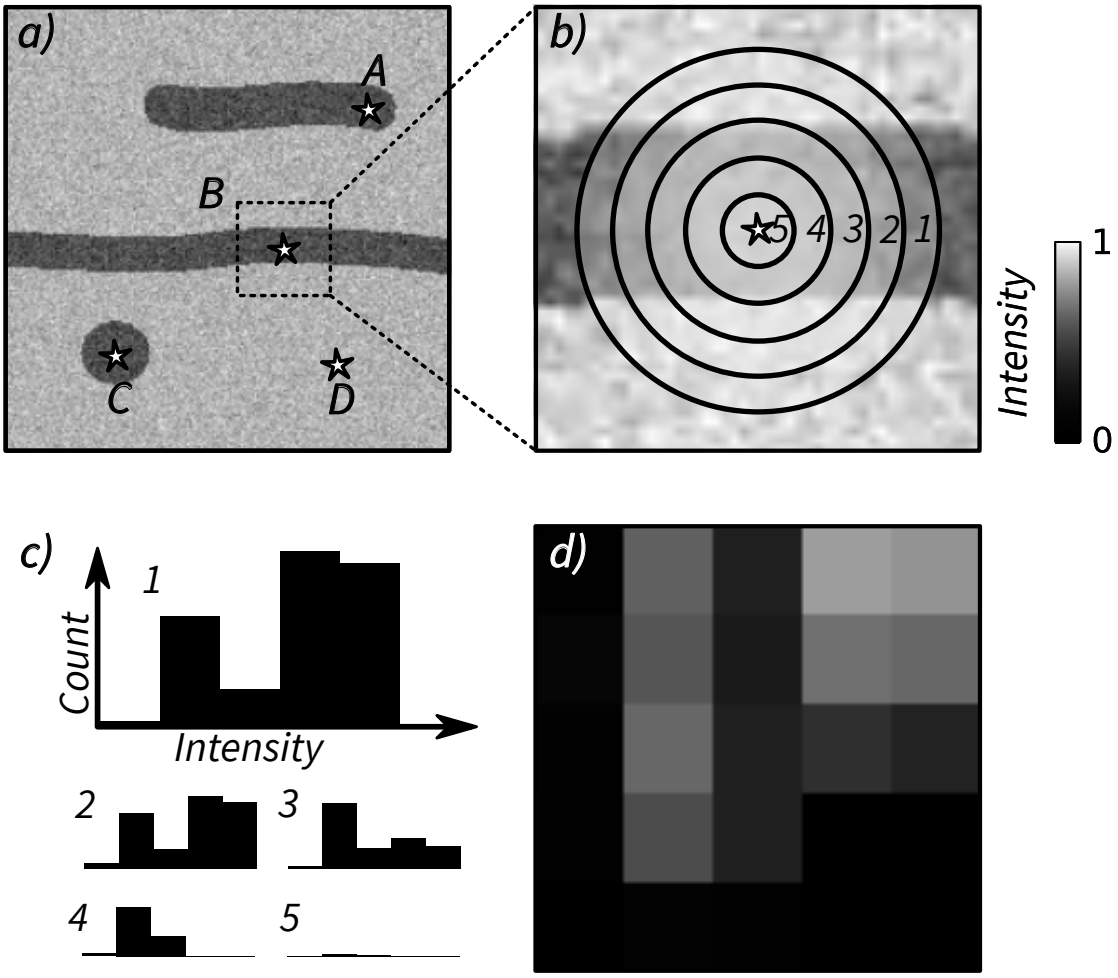


Figure 5.5: Intuitive representation of SPIN images feature description computation a) A synthetic 2D dataset representing typical shapes and aspects of chert in the intensity domain. Intensity ranges from 0 to 1 and a 0.2 sigma white noise was added. b) For each point of the cloud a set of equally spaced rings, centered on the point, are considered. Maximum radius is comparable to chert layers size. c) For each distance-ring the intensity histogram is computed and then stacked to form a matrix, 5x5 in this example, or SPIN image, in d). See Lazebnik et al. (2003) for details.

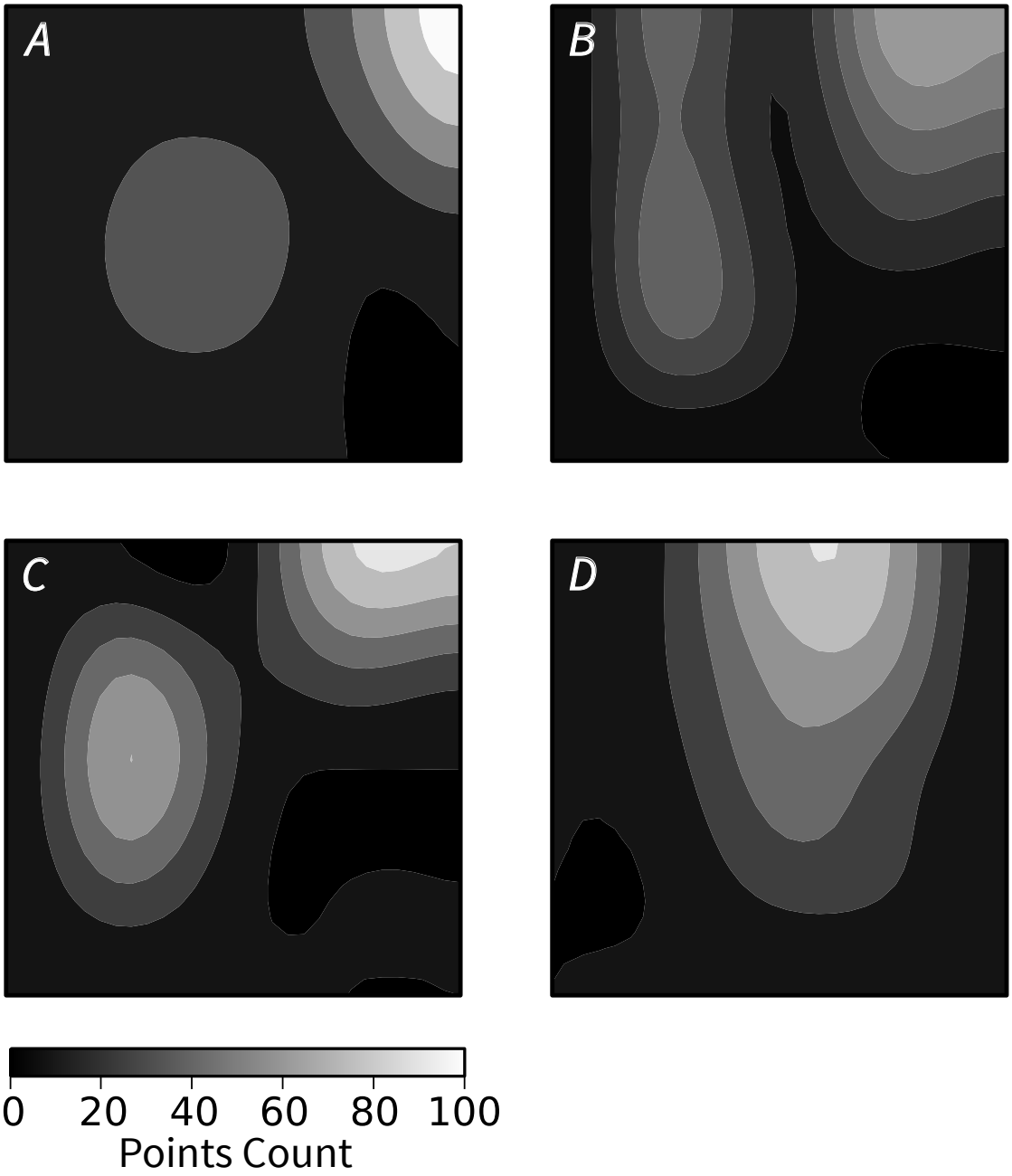


Figure 5.6: Four SPIN images for the points A, B, C and D highlighted in Figure 5.5). The SPIN images are here represented as smoothed contour lines.

In addition to raw intensity, we computed two different features for each point of the point-cloud: SPIN images feature descriptors and local geometric dispersion indices. SPIN images feature descriptors were originally developed

as affine-invariant texture representations for similar images retrieval and classification tasks (Lazebnik et al., 2003). In order to show that SPIN images can help in the task of recognizing chert we created a 2D synthetic dataset that resembles what can be typically obtained when cherty rocks are scanned (Figure 5.5a). Figure 5.5b,c and d illustrate in an intuitive way the computation of a SPIN image for a point and how in this case it is able to capture the local chert-limestone intensity/spatial relationships when the neighborhood radius has size comparable to the chert layers. Figure 5.6 shows the results of SPIN images computation for four points of the synthetic dataset of Figure 5.5. The SPIN images for the whole cloud were computed using the class `pcl::IntensitySpinEstimation` provided by the Point Cloud Library (PCL) (Rusu and Cousins, 2011).

Local geometric dispersion indices based on Principal Component Analysis are used to express how much the points are locally scattered and to capture the local dimensionality (Tang C. and Medioni, 2002; Demantké et al., 2011; Yang and Dong, 2013). For each point of the cloud, given a set of neighboring points within a fixed search radius, a local covariance matrix is estimated. The three eigenvalues of the covariance matrix λ_0, λ_1 and λ_2 can then be employed to compute two dispersion indices:

$$c_i = \frac{\lambda_i}{\lambda_0 + \lambda_1 + \lambda_2} \qquad i \in \{0, 1\} \tag{5.2}$$

Where c_i is a local dispersion index, i.e., a normalized measure of the variance of the points in the direction of the ith eigenvector. (see Brodu and Lague, 2012, for an extended discussion).

Computation was performed with a modified version of the class `pcl::NormalEstimation` provided by PCL (Rusu and Cousins, 2011) that estimates local normals via eigendecomposition of the local covariance matrix (Klasing et al., 2009). In fact the eigenvector associated with the lowest eigenvalue provides the unit normal vector, while eigenvalues themselves can be used to capture how much the points are locally dispersed along the directions of the eigenvectors.

5.2.5 *Support Vector Machine predictive model*

Support Vector Machines (SVMs) are binary classifier that are widely used for data mining (e.g., Inglada, 2007; Melgani and Bruzzone, 2004). SVMs try to find a separating hyperplane between two classes forming the training dataset of labeled vectors. For non-linearly separable classes, the input data can be

mapped to an higher space using the "kernel trick"(Schölkopf and Smola, 2001), seeking the separability in that space. We used a gaussian radial basis kernel:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2),\ \gamma > 0 \tag{5.3}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are vectors in the feature space and $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel value. The γ parameter controls the amount of influence of the single training vector on the classifier. The C parameter of the SVM classifier is a penalization term applied to misclassified observations (low C creates a smooth decision surface). A wrong selection of these parameters may easily lead to over-fitting thus we adopted the gridded inspection of the parameters space (with 5-fold cross-validation) provided by the LIBSVM software (Chang and Lin, 2011).

5.3 RESULTS

5.3.1 *Spectral characteristics of chert and its host rock*

Figure 5.7 Illustrates the reflectance curve of chert and its typical host rock (limestone), over the wavelength range 400-2500 nm, as measured with the spectrophotometer. The reflectance values were obtained from total scattered plus reflected light from different samples (chert and matching limestone), compared to the same well-known reference sample. Limestone resulted to be characterized by two water absorbance peaks at ~ 1400 and 1900 nm (Rossel and McBratney, 1998) summed with the typical CO_3^{2-} molecule spectrum (Crowley, 1986), with absorptions at 2000 nm (partially masked by the 1900 nm water absorption peak), 2300 nm, and 2500 nm. Chert by contrast does not show carbonate-related spectral features but exhibit well developed water absorbance peaks, particularly evident at the wavelength of ca. 1800 nm. For the considered samples differentiability between chert and limestone may be completely absent in the visible domain, while it is highest near 2000 nm. At the wavelength of 1541 nm of the OPTECH ILRIS 3D TLS, the separability is well within the 95% confidence interval for these samples.

5.3.2 *Classification Results*

Since SVMs in their basic formulation are binary classifiers we adopted a 2-step classification procedure. This formulation is intended to correctly extract chert vs. host-rock, excluding vegetation, enabling to compute the chert to host-rock ratio. Three different sets were prepared to train the models: Chert (CT),

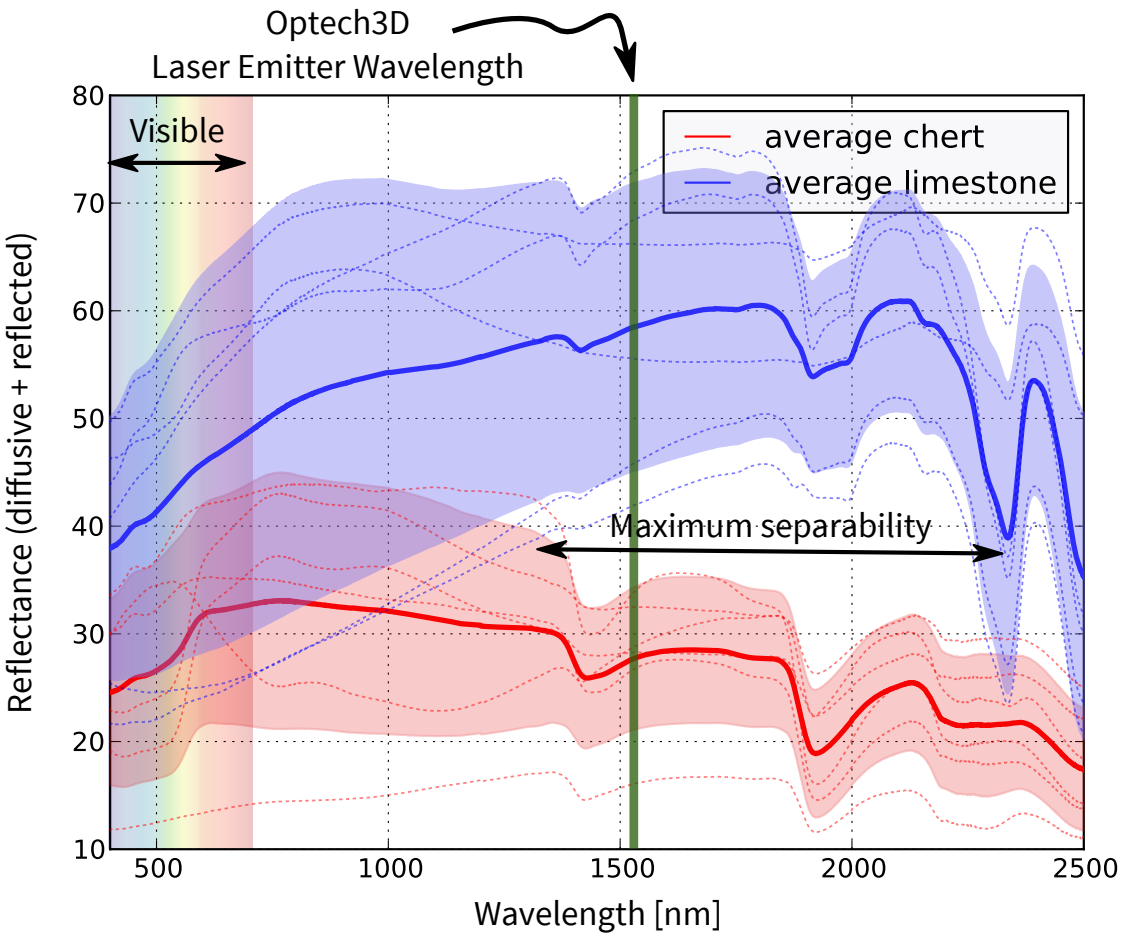


Figure 5.7: Summary of the data collected with the Cary-5000 spectrophotometer on several samples both of chert nodules and the corresponding host limestone. Three different rock formations comprising cherty limestone were sampled for this analysis. Dashed curves are the spectra of the single samples, while solid curves represent the average spectrum of chert (red) and limestone (blue). A spectral interval within the IR band shows the maximum separability for these samples. Shaded areas represent the 95% confidence intervals, estimated for each wavelength considering it as an independent measure and assuming normal distributions.

Host-Rock (HR) and Vegetation (V) while debris was not considered. Table 5.1 summarizes the number of training vectors for each labeled category. CT was carefully chosen to be representative of the most typical shapes of chert (Figure 5.5a, points A, B, and C). The training datasets were created by segmenting the original cloud by hand, using the open-source software CloudCompare

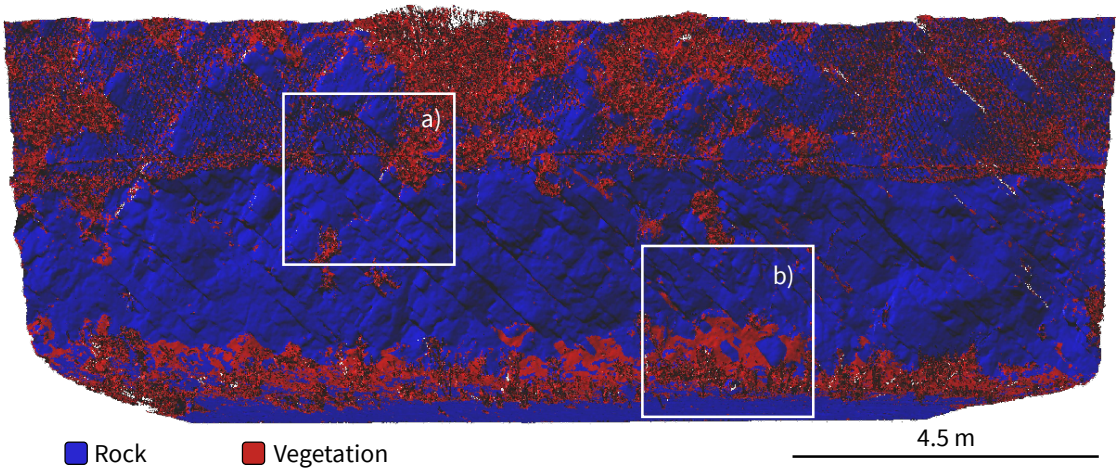


Figure 5.8: Results of the first SVM classification. Although the training data were representative of grass and rock only the classifier permitted to separate also the rock protection net in the cases in which the net was far enough from the rock surface, classifying it as vegetation. Gravel on ground has been classified as rock. Enlargements of insets a) and b) in Figure 5.10.

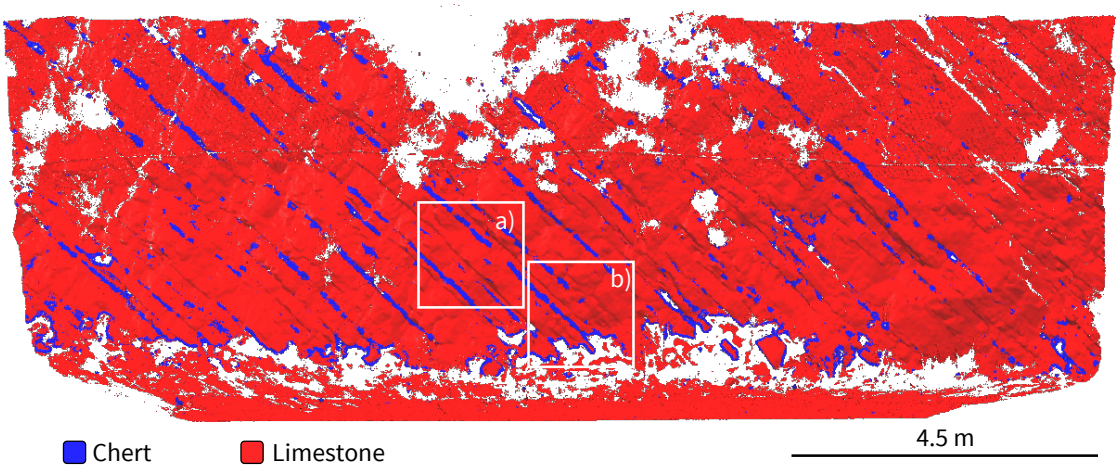


Figure 5.9: Results of the second classification step performed on the parts of the cloud that were classified as rock during the first step. Chert in blue and limestone in red. Debris (e.g., gravel) has been classified as limestone being similar in the feature domain. Note that the SVM classifier has been able to correctly classify also points laying behind the rock-protection net. Enlargements of insets a) and b) in Figure 5.11.

(Girardeau-Montaut, 2014). HR refers to the host rock. For both CT and HR, some

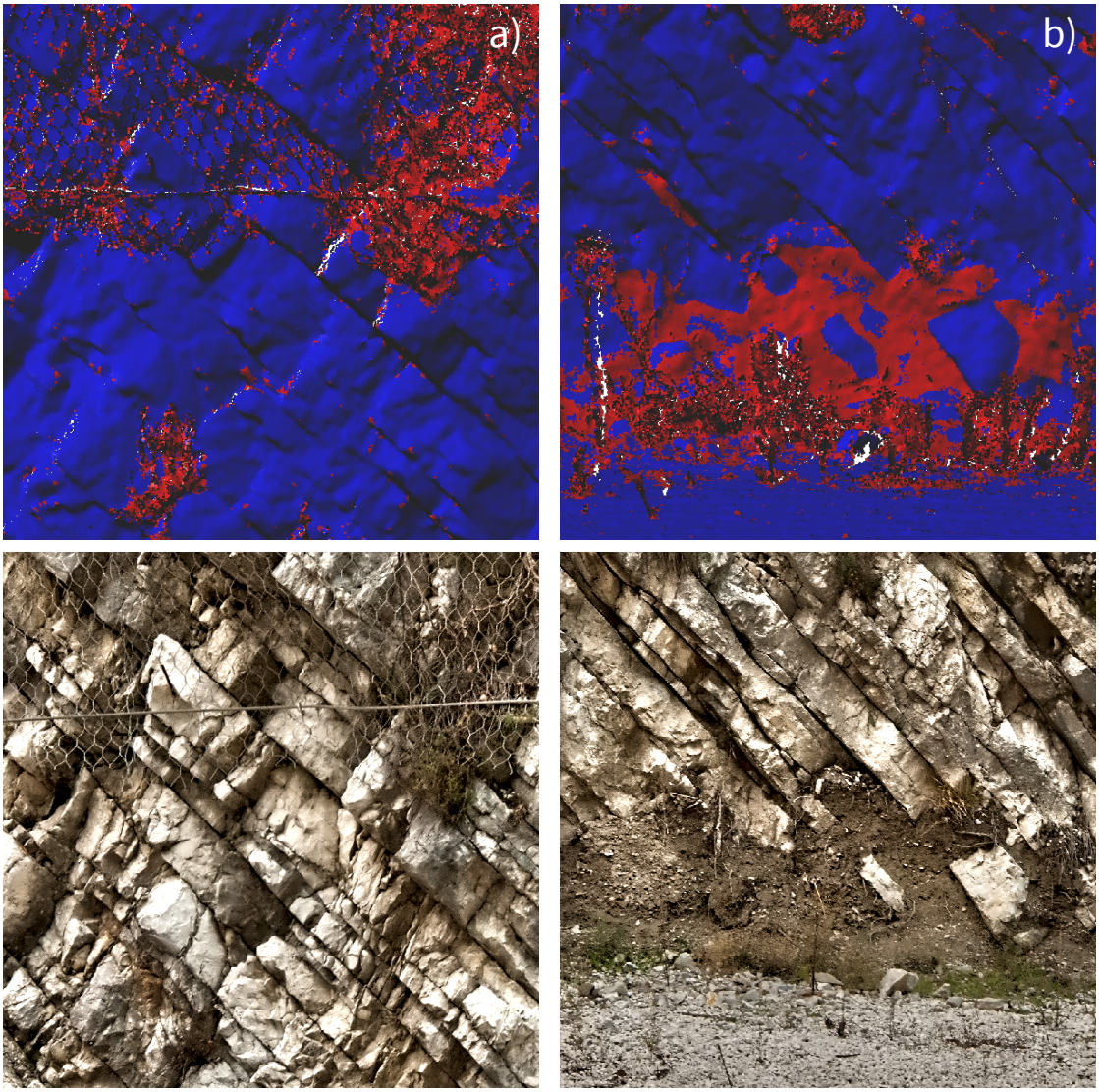


Figure 5.10: Details showing areas highlighted in Figure 5.8 a) Grass patches have been correctly segmented while the rock protection net has been partially attributed to the vegetation class due to its low intensity and the high geometrical dispersion of points that do not form a smooth surface. b) Although not specifically trained for, the SVM classifier attributed the debris to the V class.

patches covered by the iron protection-net were included to take into account its effect.

In the first classification step CT and HR were merged together to form a training set representative of all rock types (R). Parameters for the kernel were

Table 5.1: Number of training vectors for each labeled category

Label	N
Chert	5418
Host-rock	60224
Vegetation	77761

chosen via gridded inspection of the parameters space, looking for the higher 5-fold cross-validation score. The trained model was then used to classify the whole dataset, separating V from R (Figure 5.8). Results were used as input for the second classification step (Figure 5.9). Cross-validation scores, parameters of the kernel and number of points are summarized in Table 5.2.

Results have been carefully cross-inspected with the ground-truth, comparing the classification results with pictures of the outcrop in which chert was carefully individuated.

Table 5.2: Results of the two-steps classification

Step	C	γ	X-Validation	Input	Class A	Class B
I	2000	0.125	98.9%	4'318'440	3'322'520 (R)	995'920 (V)
II	32	2	98.8%	3'322'520	204'374 (CT)	3'118'146 (HR)

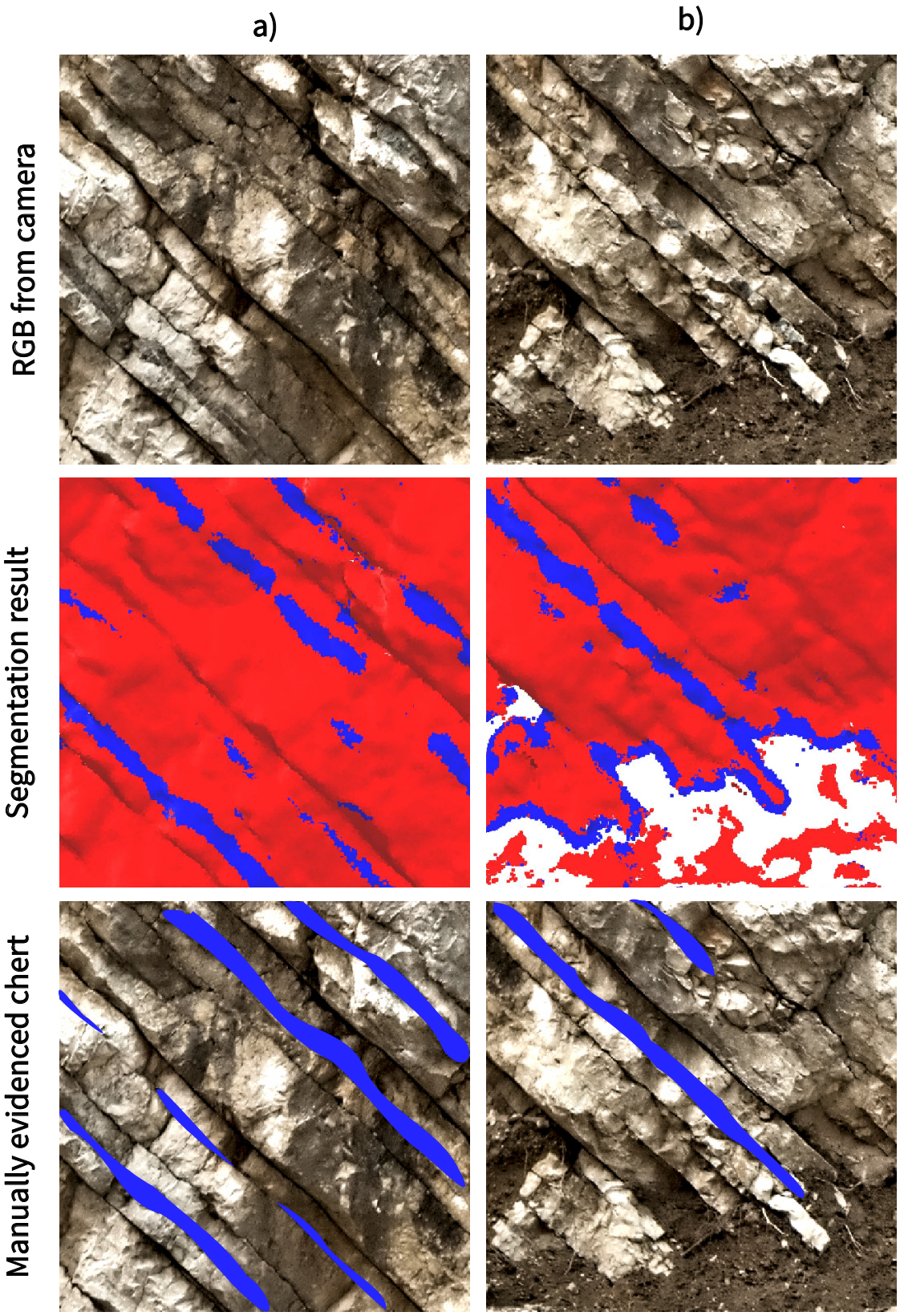


Figure 5.11: Enlargement of insets of Figure 5.9. (top) a picture of the outcrop, (middle) the results of the classification and (bottom) chert highlighted by manual segmentation of the RGB pictures. a) On the rock surface, the classifier correctly recognized chert. b) The classification is not completely satisfying at the limestone-debris contact due to similarity in the feature space of the two materials. Notice that chert (in this specific case) is difficult to identify in pictures due to color similarity with limestone.

5.4 DISCUSSION AND CONCLUSIONS

In this paper we demonstrated that TLS data can be exploited to map chert in natural outcrops. Experiments showed that chert displays low reflectance in the IR band. This absorption appears to be related to a non-selective absorption in the IR domain. Since intensity information alone does not provide enough separability for a satisfactory classification we improved the quality of the automatic mapping of chert using Computer Vision (CV) methods that does not require a radiometric calibration of the intensity data and take advantage of the morphological features of chert (SPIN descriptors and local geometric dispersion indices). Trained SVM classifiers demonstrated high cross-validation scores ($>$ 98 % for both the classification steps) implying good separability of the classes. In the SPIN domain the first three principal components were able to capture this differentiability (Figure 5.12). Considering the whole dataset, Principal Components Analysis (PCA) showed that the the 99% of the variance is achieved within the first 7 Principal Components (PCs), and the first three PCs explain only the 92% of the total variance reflecting the independence of the computed features. These observations suggests that the dataset might be eventually reduced to a minor number of components when dealing with large point clouds (i.e., for faster classifications).

The trained model applied to the whole dataset correctly resolved chert although false positive were detected (Figure 5.11). These wrongly recognized patches are mainly soil at the edges of carbonate rock, which resemble the elongated, low-intensity chert layers in the considered feature domain. These issues could however be easily overcome by excluding irrelevant points of the point-cloud from the analysis or adding ad-hoc SVM classification steps.

The obtained results do not depend on a specific instrument, under the condition that the laser wavelength falls within the separability band. The fact that some TLS instruments operate at 1064 nm or also in visible band should be noted. These instruments cannot be used for chert mapping purposes and, in general, suitability of the device chosen for discrimination purposes should be evaluated depending on materials.

Besides the specific success in chert classification, results of our study show that TLS point clouds contain complex implicit information in the form of spatial characteristics and intensity that can be exploited using feature descriptors together with a supervised non-linear classifier, following a CV-derived pipeline. A-priori knowledge on the nature of a certain object can be thus used to propose an appropriate classification model.

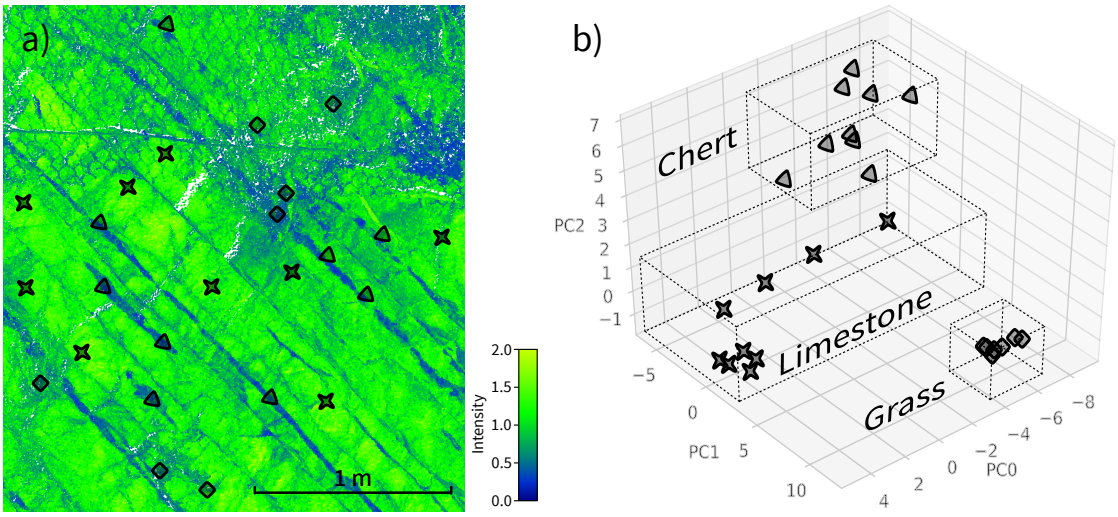


Figure 5.12: a) A subset of the point-cloud, colored by intensity. 6x6 SPIN feature descriptors have been computed using a search radius of 0.1 m. Markers are points of interest subdivided in three categories: chert (△), grass (◇) and limestone (✖). b) To illustrate separability of various materials in the SPIN domain the descriptors have been normalized and reduced to the first 3 principal components via PCA, explaining 91% of the total variance. Bounding boxes help illustrating separability in 3D.

The accurate mapping of chert is of potential interest in many geological applications, from the production of accurate and long records of chert abundance for stratigraphic and paleoclimate studies to the assessment of rock quality when limestones are used for concrete production.

It has been already pointed out that most TLS work with a single wavelength, limiting the possibility of collecting spectral information about the scanned targets. A solution to this issue has been prospected by (Hartzell et al., 2014) with the combined use of multiple TLS working at different wavelengths. Recent advances in TLS construction tend to the design of multi spectral instruments able to provide intensity information from multiple wavelengths (Hakala et al., 2012). Coupling spectral and geometrical information as proposed in this paper may allow a progressively much wider and successful application of active sensors in the automated discrimination of materials in the near future.

5.5 ACKNOWLEDGMENTS

Authors thank Maria Pelizzo (Institute of Photonics and Nanotechnologies, CNR-Padova) for spectral analyses, Giovanni Rusciadelli (Univ. Chieti) for support in the field, Federico Zorzi (Univ. Padova) for XRPD analysis of samples, Matteo Massironi (Univ. Padova) and Giovanni Gattolini (Univ. Padova) for fruitful discussions. Authors greatly thank Daniel Girardeau-Montaut for his efforts in CloudCompare development. Authors thank two anonymous reviewers for observations that permitted to greatly improve the quality of the manuscript.

6

BUILDING LONG TIME SERIES FROM POINT CLOUDS, THE VISPI QUARRY

Contents

6.1 INTRODUCTION

The Vispi quarry is located in Central Italy near the city of Gubbio, on the east slope of the Contessa Valley. It is an active quarry which extensively exposes part of the Cretaceous Umbria-Marche basinal series, that is, in stratigraphic order: the upper part of the *Maiolica Formation* (Late Tithonian-Early Aptian), the *Marne a Fucoidi* and the *Scaglia Bianca* (Late Albian-Early Turonian). Only the lowermost part of *Scaglia Rossa* is exposed in the section.

The Maiolica Fm. is composed of white to gray pelagic limestones which contains dark gray chert in its uppermost portion. The Marne a Fucoidi Fm. is more rich in shales (including black shales), interlayered by marly limestones, marls and calcareous mudstones. The Scaglia Bianca Fm., which overlays the Marne a Fucoidi, is composed of yellowish to grayish limestones, intercalated with black marlstones and shales (Hu et al., 2006).

Although this section constitutes a continuous and almost undisturbed sedimentary record which may provide in high-quality dataset for investigating the orbital forcing during sedimentation, its high-resolution sampling and logging is difficult due to the extension of the outcrop. The outcrop, being located in an

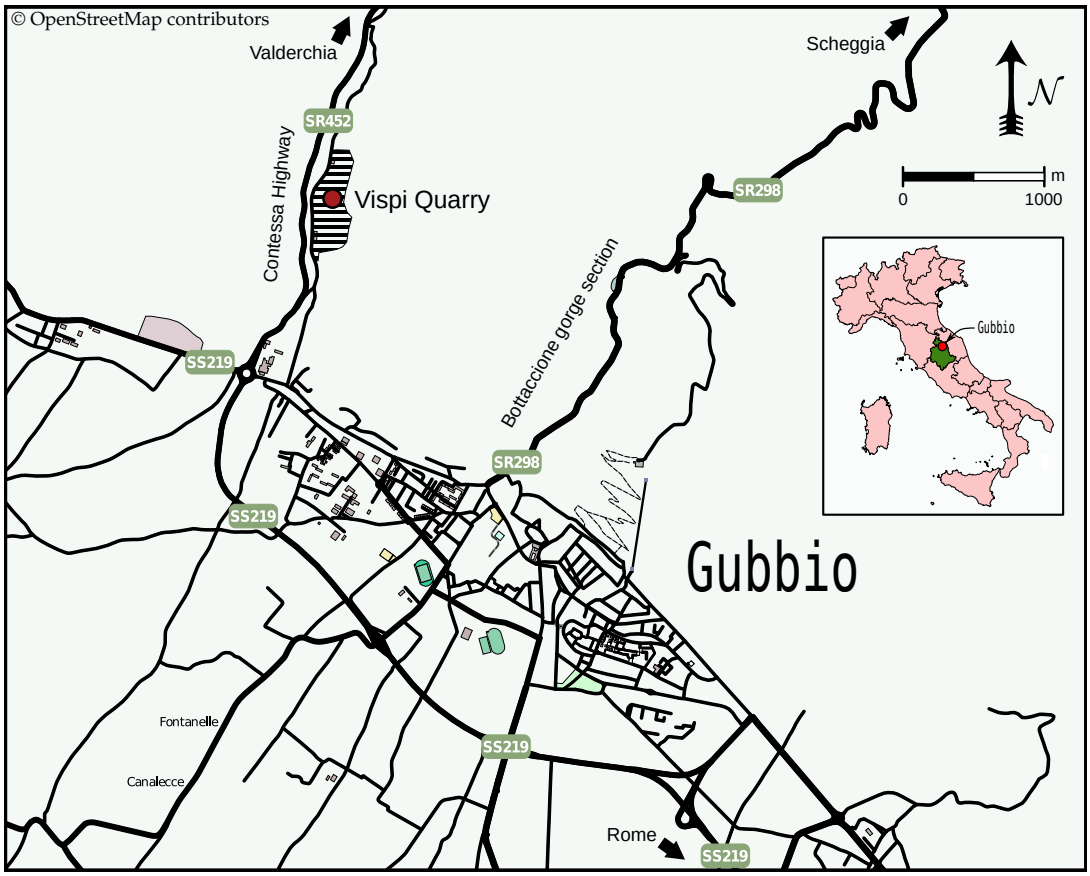


Figure 6.1: Location o the Vispi quarry. Geographic data courtesy of OpenStreetMap contributors.

active quarry, can only be accessed pending the permission from the quarry's manager, which is always for a limited and *short* time span.

In 2011, Franceschi et al. approached the cyclostratigraphic study of part of this quarry using a Terrestrial Laser Scanner (TLS). They were able to reconstruct ~ 50m of the Marne a Fucoidi Fm. by means of intensity series. The spectral analysis thus revealed a strong coherence of their time series compared to the Piobbico core photolog by Grippo (2004), confirming also that orbital forcing drove the depositional system.

That investigation was probably the first *real* application of the TLS intensity. In fact at the time the intensities were raising the interest of many researchers but, basically, their usage as a primary data source for some specific investigation was more speculative than applicative. Still today, the many issues that must be faced contribute in making this kind of data an highly *experimental* data source.

The Vispi quarry has been at the center of other investigations (Hu et al., 2006), but in general, the nearby Bottaccione Gorge and the Contessa Highway have been regarded as easier to be accessed.

On the other hand, this outcrop is instead the ideal situation for using remote sensing methods. Apart from the access issues, it is a well exposed, vegetation-free and extended outcrop that would be difficult to be studied otherwise. At the same time, a favorable average attitude of the strata, together with the right amount of *space* in front of the outcrop highly suggests that, applying the same techniques introduced by Franceschi et al. (2011), it should be possible to obtain extremely long time series, possibly covering the whole quarry.

Simplifying, the series is built by limestones which sometimes tend to become marly, together with frequent shales and nodular and layered cherts. These are the main lithotypes composing the section. The possibility to clearly observe marls-to-limestone alternations had been proven in the past (Franceschi et al., 2009), while gaining a better understanding of the effects of $CaCO_3$ content, shales and cherts which have been demonstrated in this thesis was fundamental to create a sound basis for the study of the Vispi quarry outcrop as a whole.

As illustrated in Chapter 2, the processing of big data requested to improve the original techniques and better establishing the methods employed. This chapter demonstrates those methods in a *practical* case study. We will give insights on the processing of a big dataset of scans, together with some considerations on the issues we encountered during the elaboration, which may be useful for replicating our results. At the end of this chapter a *long time series*, generated from > 20 scans, which represents the stratigraphic series from the Bonarelli level down to the upper part of the Maiolica Fm., will be presented.

6.2 DATA ACQUISITION

The Vispi quarry has been surveyed with an OPTECH TLS, model ILRIS 3D, the same device used by Franceschi et al. (2011) and Penasa et al. (2014) in their investigations.

The technical characteristics of this device are reported in Table 6.1.

As highlighted in Chapter 5 (see Figure 5.4), the intensities come from the device as an integer Digital Number (DN). The intensity recorded for a given point can thus take a restricted number of values. How much this *dynamic range* is wide will depends also on the target's distance. In particular, to obtain good results in terms of *resolution* of intensities the distance should be less than 50

Parameter	Value
Wavelength	1535 nm
Laser Class	eyesafe, class 1
Range (20 % target reflectivity)	800 m
Minimum acquisition distance	3 m
Beam divergence angle	0.17 mrad
Beam divergence at 50 m	21 mm
Resolution at 50 m	17.7 mm
Single point acquisition accuracy at 50 m	7 mm
Modeling accuracy at 50 m	3 mm
Minimum spot spacing at 50 m	1.3 mm
Scan velocity	2000 pts/s

Table 6.1: Technical specifications for the OPTECH ILRIS 3D laser scanner, from Franceschi et al. (2009). Notice that a slightly different wavelength of 1541 nm has been reported by Larsson et al. (2007). This difference is totally negligible for our purposes.

meters, as suggested by Franceschi et al. (2009). In practice, smaller the distance, higher the resolution that can be expected in terms of *intensities*.

A similar reasoning can be applied to the geometric resolution, the number of points that will be used to represents a m^2 of outcrop will vary as a function of the distance from the device. Similarly, the *laser footprint* on the outcrop itself increases with the distance due to the *divergence* of the laser beam.

The Vispi outcrop has an horizontal extension of > 500 m with a vertical exposure of > 200 m. Obtaining a full-size TLS scan of this such extended object is not possible considering the limits of intensity resolution we wanted to meet (distances < 50 m). Besides, being impractical, obtaining a complete coverage of the whole quarry it is also not needed: our goal was to provide a complete coverage in terms of stratigraphic exposure.

The steep attitude of the strata, which deeps toward North, completely exposes the stratigraphic series at the base of the quarry. From a frontal position, at the outcrop toe, a set of 29 TLS scans were made from different stations. The TLS was powered using a portable generator which was moved together with the TLS in consecutive stations, which allowed to cover the whole stratigraphy.

The survey started from the Northern sector of the quarry, from the Bonarelli level which is well exposed in the Northern edge of the quarry. Each new scan has been taken so to account for at least a ~ 10% overlap with the next one. A minor number of scans in the southern part of the survey were instead taken without any geometric overlap, but still granting *stratigraphic* overlap between consecutive scans.

Average distance from the outcrop was ~ 25 meters, in a range from 10 to 40 meters from the quarry's walls, each scan covering an horizontal distance of ~ 10 meters.

Contextually with the scan operations, a series of pictures taken from the Contessa roadway were collected with a standard camera (*Nikon D*3100 equipped with fixed 35 mm lens). About 15 photographic stations were set along the road to cover the whole quarry from the highest possible points of view. At each station, a variable number of pictures (5 to 10) were shot to cover the whole outcrop. Those pictures were then used to produce a photogrammetric model.

The entire surveying operation required 2 days of work with just 2 operators (*Marco Franceschi & Luca Penasa*). This time limit was also imposed by the querry's manager which granted us access permission only during *a weekend*, while the mining activity was suspended.

6.3 POINT CLOUDS PREPROCESSING

The scans were processed using the proprietary parser provided by OPTECH. Each scan was exported as a text file composed by the $[x, y, z]$ coordinates and the corresponding intensity as a 16 bit DN, described in the parser's option as *RAW* intensity.

The point clouds were then converted to the PCD format for being loadable by the SPC toolkit. Before any processing, the distance from the sensor of each point was added to each cloud, considering the device as centered in the origin ($[0, 0, 0]$). This scalar field is needed by the SPC toolkit (see Chapter 2) to perform the normalization of intensities.

The photogrammetric reconstruction was performed using the open source *MicMac* toolkit[1] (Pierrot-Deseilligny, 2013; Rosu et al., 2015).

A set of 111 photographs were internally and externally oriented, after homologous points matching, using the tools provided by MicMac. For each homologous point, a 3D point was computed providing a sparse representation

1 *Multi Images Correspondances par Méthodes Automatiques de Corrélation* Available at `http://logiciels.ign.fr/?Telechargement,20`. See also `http://logiciels.ign.fr/?Micmac`

of the whole quarry. A smaller subset of pictures was chosen to perform a dense reconstruction. ~ 10 photographs permitted to obtain dense point clouds for great part of the quarry's extension.

The photogrammetric reconstruction, although being dense, does not have any *scale* information. The unit of measure of the point clouds obtained from photogrammetry does not represent the dimension of the quarry. For solving this problem the TLS point clouds were used in the following way.

The partial overlap of the first 20 TLS scans was exploited to obtain a *rough* registration of the clouds. Each cloud was co-registered with the next one manually defining a set of > 4 homologous points, using the open source *CloudCompare* software (Girardeau-Montaut, 2014). The registration have been then refined with the Iterative Closes Point (ICP) algorithm (Besl and McKay, 1992; Chen and Medioni, 1992) provided by the software. Although this registration being not *optimal*, due to the small overlap between nearby clouds, it permitted to compute a scaling factor for the photogrammetric point cloud. Notice that, as we will discuss later, there is no need to perfectly register all clouds, and in fact we have a subset of clouds for which co-registration is not possible, due to lack of overlap.

The photogrammetric model was registered with the TLS clouds using a version of the ICP algorithm which embeds the scale estimation (Zinßer et al., 2005), implemented directly in *CloudCompare*. The remaining TLS point clouds, which were still not registered, were registered against the photogrammetric model.

This process resulted in all the point clouds, either coming from TLS or from the photogrammetry, being registered in the same reference frame. This process, that is obviously fairly tedious, could be improved by using dedicated targets for granting a semi-automatic registration. Due to the limitations in time for accessing the quarry, the faster in-field method that could be adopted was chosen, and no targets were employed. Furthermore, in many cases, positioning registration targets was also not possible, especially when scanning vertical areas of the quarry.

6.4 TIME SERIES EXTRACTION

The whole time series extraction process was conducted from within CloudCompare, using the *Vombat* plugin, as described in Chapter 2. In this section, some concepts introduced there will be used.

All the point clouds to be processed were loaded in CloudCompare as a Reference On Disk (ROD). This allows to load the point when needed for visualization, and unloading it from the memory when not needed, while still preserving a *reference* of where the point cloud is stored on disk.

For each point cloud, a single Region Of Interest (ROI) was identified, each comprising the portion of the cloud which was intended to be used to generate the time series. The ROI was chosen so to be free of vegetation and any other disturbing element, as humidity and faults presence. In general, the Franceschi et al. (2011) recommendations on point cloud cleaning were adopted.

For each of these ROIs, a time series was generated, but fo being able to do so, the algorithm requires to know the attitude of the bedding into the specific ROI. For each ROI, a single attitude was thus estimated, which was then converted into a local Stratigraphic Reference Frame (SRF).

Summarizing, for each cloud a ROI was defined by the user, subsequently each of them was equipped with a SRF which defines a *direction* in the 3D space which is expected to be related with the *time* (as depicted in Section 6.1) and will serve as a *local ruler* for measuring the section and for producing a time series, representative of the changes of intensities along this direction.

At this point, the time series were computable, but the setup was still not complete. Each SRF at this stage was still *unaware* of the other SRFs, hence each of them was defining its own *stratigraphic zero*. A set of *constraints* were thus defined to link each SRF to the next one.

A first *Sample* (crf. Chapter 2 for the specific interpretation of the term) was placed at the base of the Bonarelli level and assigned to the first SRF. This sample was set to work as sample with *fixed* stratigraphic position, which have been assigned to 0 m.

A second *Sample* was positioned on an easily recognizable stratigraphic feature, present also in the next point cloud. On this cloud a third *Sample* was positioned on the same stratigraphic level. These last two *Samples* together were set up to form a Stratigraphic Tie Constrain (STC) between the two clouds. These constraints are used by the SPC library to create a composite stratigraphic reference frame between different ROIs and clouds. The same procedure was used to create stratigraphical *ties* between the SRFs.

The time series were then extracted as described in Chapter 2 using a sampling step of 1cm, with a bandwidth (if the Kernel Smoother) of 1 cm. The resulting time series are shown in a Figure 6.2 and Figure 6.3.

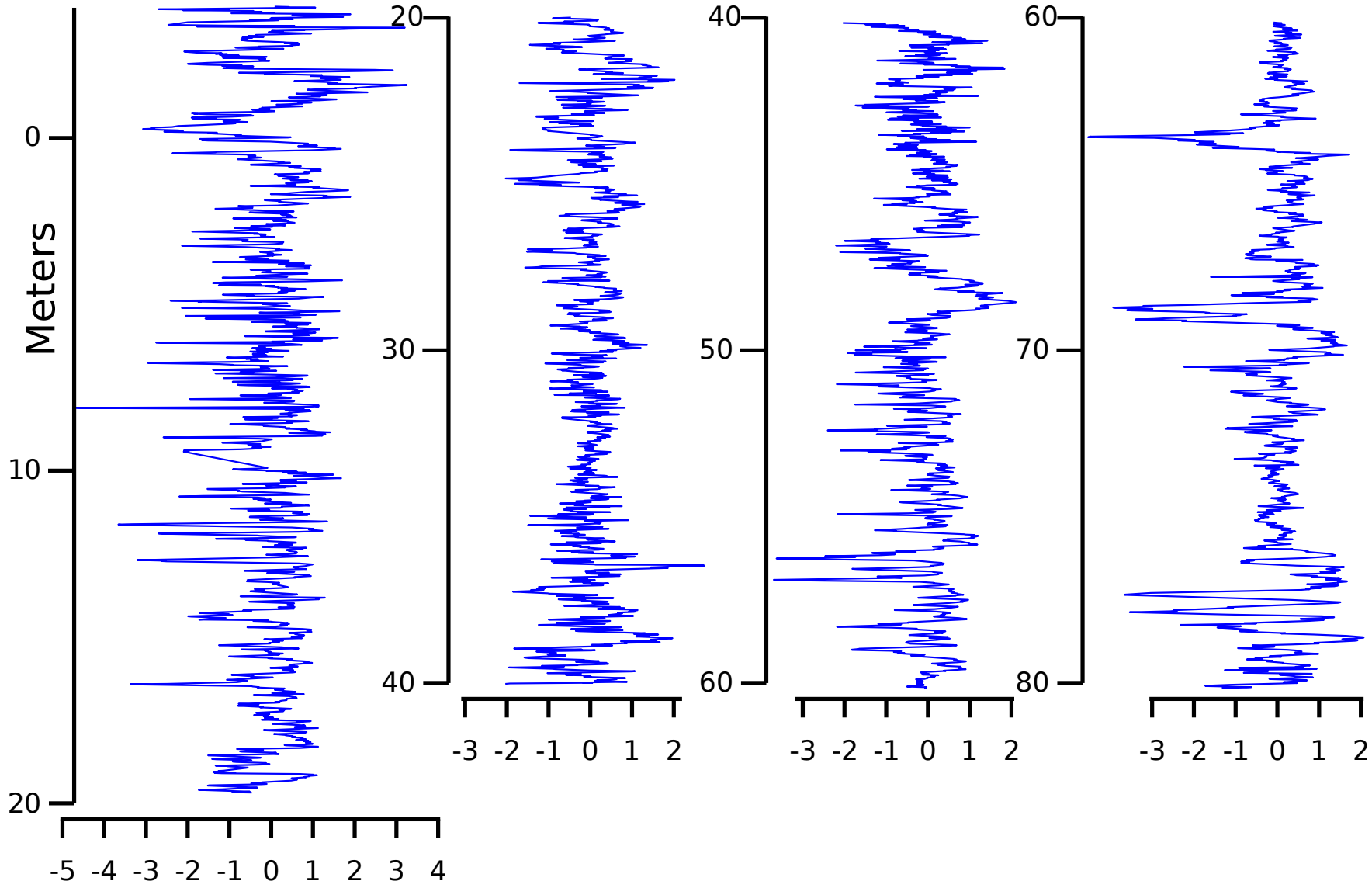


Figure 6.2: First part of the time series, from 0 (Bonarelli level) to 80 meters. Higher stratigraphic positions correspond to older strata.

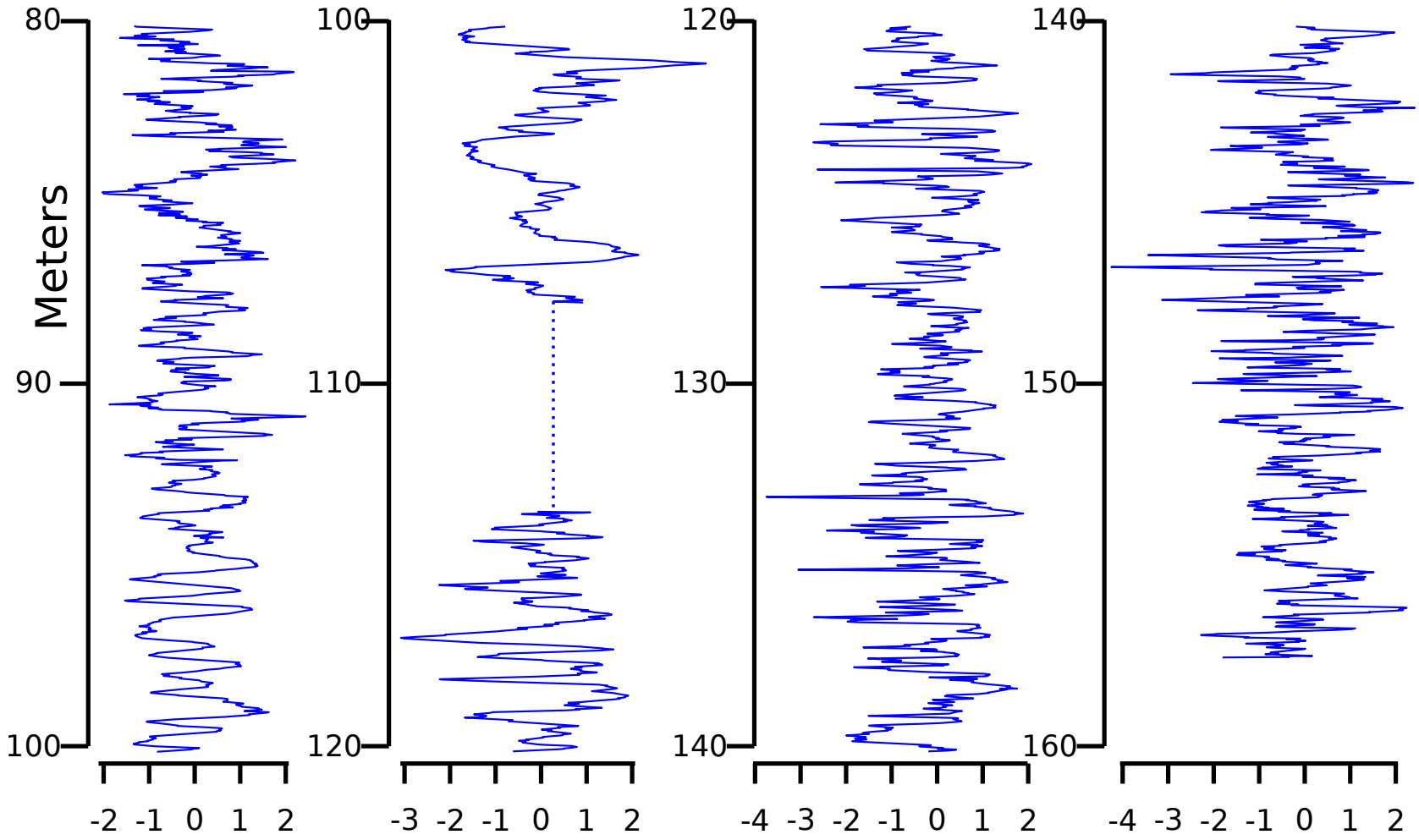


Figure 6.3: Second part of the time series, from 80 to 160 meters.

7

CONCLUSIONS

The aim of this work was to study the applicability of the method proposed by Franceschi et al. (2009) as a data source able to provide *long* time series for cyclostratigraphic investigations.

Previous investigations proved the feasibility of converting the intensity values from TLSs into a *proxy* for the reflectance of the materials under investigation. Transforming an *intensity* measure into a real *reflectance* is a problem that has been faced in many different contexts and unfortunately it still does not have an unique and simple solution.

Most of intensity *calibration* methods which have been proposed in literature require fairly complex experimental setup to collect a dataset of *observations* that can be used to estimate the parametrs of the correction model.

We demonstrated that these observations can be potentially extracted *directly* from the point clouds themselves, exploiting overlapping clouds. Thanks to some extra information provided by the user about the materials in the scene these observation can be used to calibrate corrections models for the intensity. We used a simple *normalization* model intended to remove the effects of distance from the sensor and the incidence angle from intensity to demonstrate the feasibility of this approach.

The use of intensities as *proxy* for lithological properties was investigated with the final goal of gaining a better understanding of the *meaning* of the intensity in the specific context of typical basinal lithotypes. We followed a *mixed*-type investigation based both on experimental observations, obtained from laboratory measures, and on in-field data collection on real stratigraphic cases. This dual approach was intended to provide also real dataset that could potentially undergo cyclostratigraphical analysis.

We proposed a model for interpreting stratigraphic series made of limestone, clay and chert, some of the most common lithotypes in the Umbria-Marche basin. This stratigraphic sequence was explicitly chosen because of its potential

importance in the cyclostratigraphic context. It is a well known series, with many *ideal* conditions for testing remote sensing methods.

The TLS method was applied on a calcareous homogenites. The comparison of the intensity with calimetric analysis showed how the time series generated with TLS can be compared to more classical time series, as a $CaCO_3$ wt.% curve, demonstrating that the intensity can be used as a proxy for $CaCO_3$ also in highly homogeneous sequences.

Chert, on the other hand, demonstrated to be characterized by *low reflectances* which produces low intensity values when observed with a TLS. This fact, together with their common shape was exploited to build an original classification method, that is able to automatically extract chert from a point clouds.

Creating time series from point clouds require a fairly large amount of data processing. For this an original solution was implemented. Although it is mainly targeted to TLS point clouds, the main concept can be easily translated to any other 3D data source, and easily integrated with data coming from other devices (e. g. passive hyperspectral cameras).

Our investigations proved that long time series, potentially long-enough to observe long-period Milankovitch-like cycles, can be provided by remote sensing methods, when a multi disciplinar approach is pursued. 3D data sources intrinsically have a great potential as data source for stratigraphic investigations but require the development of specific methods, which allows for the transformation of such a complex dataset into something meaningful for the stratigrapher.

BIBLIOGRAPHY

Antonio Abellán, Thierry Oppikofer, Michel Jaboyedoff, Nicholas J. Rosser, Michael Lim, and Matthew J. Lato. Terrestrial laser scanning of rock slope instabilities. *Earth Surface Processes and Landforms*, 39(1):80–97, January 2014. ISSN 01979337. doi: 10.1002/esp.3493. URL http://doi.wiley.com/10.1002/esp.3493.

E. Ahokas, S. Kaasalainen, J. Hyyppä, and J. Suomalainen. Calibration of the Optech ALTM 3100 laser scanner intensity data using brightness targets. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(Part 1):1Á6, 2006. URL http://www.isprs.org/proceedings/XXXVI/part1/Papers/T03-11.pdf.

Walter Alvarez, Roberto Colacicchi, and Alessandro Montanari. Synsedimentary slides and bedding formation in Apennine pelagic limestones. *Journal of Sedimentary Research*, 55(5), 1985.

K. Anttila, S. Kaasalainen, A. Krooks, H. Kaartinen, A. Kukkoa, T. Manninen, P. Lahtinen, and N. Siljamo. Radiometric calibration of tls intensity: Application to snow cover change detection. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(5/W12):6, 2011. URL http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXVIII-5-W12/175/2011/isprsarchives-XXXVIII-5-W12-175-2011.pdf.

Julia Armesto-González, Belén Riveiro-Rodríguez, Diego González-Aguilera, and M. Teresa Rivas-Brea. Terrestrial laser scanning intensity data applied to damage detection for historical buildings. *Journal of Archaeological Science*, 37 (12):3037–3047, December 2010. ISSN 03054403. doi: 10.1016/j.jas.2010.06.031. URL http://linkinghub.elsevier.com/retrieve/pii/S0305440310002281.

A.M. Baldridge, S.J. Hook, C.I. Grove, and G. Rivera. The ASTER spectral library version 2.0. *Remote Sensing of Environment*, 113(4):711–715, April 2009. ISSN 00344257. doi: 10.1016/j.rse.2008.11.007. URL http://linkinghub.elsevier.com/retrieve/pii/S0034425708003441.

Mathilde A.F. Balduzzi, Dimitry Van der Zande, Jan Stuckens, Willem W. Verstraeten, and Pol Coppin. The Properties of Terrestrial Laser System Intensity

for Measuring Leaf Geometries: A Case Study with Conference Pear Trees (Pyrus Communis). *Sensors*, 11(12):1657–1681, January 2011. ISSN 1424-8220. doi: 10.3390/s110201657. URL http://www.mdpi.com/1424-8220/11/2/1657/.

S. J. Batenburg, A. Montanari, M. Sprovieri, F. J. Hilgen, R. Coccioni, and A. S. Gale. Astronomical tuning of black cherts in the Cenomanian Scaglia Bianca as precursors of the Bonarelli level (OAE2) at Furlo, Italy. In A. Abbasi and N. Giesen, editors, *EGU General Assembly Conference Abstracts*, volume 14 of *EGU General Assembly Conference Abstracts*, page 6185, April 2012.

F. Bektas, T. Topal, M.C. Goncuoglu, and L. Turanli. Evaluation of the alkali reactivity of cherts from Turkey. *Construction and Building Materials*, 22(6):1183–1190, 2008. ISSN 0950-0618. doi: 10.1016/j.conbuildmat.2007.02.002.

J.A. Bellian, C. Kerans, and D.C. Jennette. Digital Outcrop Models: Applications of Terrestrial Scanning Lidar Technology in Stratigraphic Modeling. *Journal of Sedimentary Research*, 75(2):166–176, March 2005. ISSN 1527-1404. doi: 10.2110/jsr.2005.013. URL http://jsedres.sepmonline.org/cgi/doi/10.2110/jsr.2005.013.

P.J. Besl and Neil D. McKay. A method for registration of 3-D shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(2):239–256, February 1992. ISSN 0162-8828. doi: 10.1109/34.121791.

Sam. Boggs. *Petrology of sedimentary rocks*. Cambridge University Press, Cambridge, 2009. ISBN 9780511516429 0511516428 9780511515149 0511515146 9780511719332 0511719337 9780511626487 0511626487. URL http://dx.doi.org/10.1017/CBO9780511626487.

T. Brandtberg. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5):325–340, 2007.

Nicolas Brodu and Dimitri Lague. 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68:121–134, 2012. URL http://www.sciencedirect.com/science/article/pii/S0924271612000330.

Darrin Burton, Dallas B. Dunlap, Lesli J. Wood, and Peter P. Flaig. Lidar Intensity as a Remote Sensor of Rock Properties. *Journal of Sedimentary Research*, 81

(5):339–347, 2011. ISSN 1527-1404. doi: 10.2110/jsr.2011.31. URL http://jsedres.sepmonline.org/cgi/doi/10.2110/jsr.2011.31.

Nélia Castro and Børge J. Wigum. Assessment of the potential alkali-reactivity of aggregates for concrete by image analysis petrography. *Cement and Concrete Research*, 42(12):1635–1644, 2012. ISSN 0008-8846. doi: 10.1016/j.cemconres.2012.08.009.

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011. URL http://dl.acm.org/citation.cfm?id=1961199.

J. E. T. Channell, B. d'Argenio, and F. Horvath. Adria, the African promontory, in Mesozoic Mediterranean palaeogeography. *Earth-Science Reviews*, 15(3):213–292, 1979.

Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and vision computing*, 10(3):145–155, 1992.

Yuwei Chen, Esa Räikkönen, Sanna Kaasalainen, Juha Suomalainen, Teemu Hakala, Juha Hyyppä, and Ruizhi Chen. Two-channel Hyperspectral LiDAR with a Supercontinuum Laser Source. *Sensors*, 10(7):7057–7066, July 2010. ISSN 1424-8220. doi: 10.3390/s100707057. URL http://www.mdpi.com/1424-8220/10/7/7057/.

Roger N Clark and Ted L Roush. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 89(B7):6329–6340, 1984.

Rodolfo Coccioni, Giuseppe Bancalà, Rita Catanzarit, Eliana Fornaciari, Fabrizio Frontalini, Luca Giusberti, Luigi Jovane, Valeria Luciani, Jairo Savian, and Mario Sprovieri. An integrated stratigraphic record of the Palaeocene-lower Eocene at Gubbio (Italy): new insights into the early Palaeogene hyperthermals and carbon isotope excursions: Early Palaeogene integrated stratigraphy at Gubbio. *Terra Nova*, 24(5):380–386, October 2012. ISSN 09544879. doi: 10.1111/j.1365-3121.2012.01076.x. URL http://doi.wiley.com/10.1111/j.1365-3121.2012.01076.x.

S. Cresta, S. Monechi, and G. Parisi. *Stratigrafia del mesozoico e cenozoico nell'area Umbro-Marchigiana: itinerari geologici sull'Appennino umbro-marchigiano (Italia)*, volume 39. Istituto Poligrafico e Zecca dello Stato, 1989.

James K. Crowley. Visible and near-infrared spectra of carbonate rocks: Reflectance variations related to petrographic texture and impurities. *Journal of Geophysical Research: Solid Earth*, 91(B5):5001–5012, 1986. ISSN 2156-2202. doi: 10.1029/JB091iB05p05001.

Jérôme Demantké, Clément Mallet, Nicolas David, and Bruno Vallet. Dimensionality based scale selection in 3D lidar point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Laser Scanning*, 2011, 2011.

D. J. DeMaster. North Carolina State University, Raleigh, NC, USA. *Sediments, Diagenesis, and Sedimentary Rocks: Treatise on Geochemistry*, 7:87, 2005. URL http://books.google.com/books?hl=en&lr=&id=K8M3FOL3MZgC&oi=fnd&pg=PA87&dq=%22effects+(cm2+of+sediment%22+%22as+diatoms+commonly+are+used+to%22+%22making+this+phase+second+only%22+%22several+1,000+mM+(Siever,+1992).+As%22+%22commonly+precipitate+their%22+%22Hildebrand+(2000),+and+Wetherbee+et%22+&ots=FqmiVOzXYL&sig=mDZY1MTrA9k5ACEBExTmeZIegHk.

Jochen Erbacher, Jürgen Thurow, and Ralf Littke. Evolution patterns of radiolaria and organic matter variations: a new approach to identify sea-level changes in mid-Cretaceous pelagic environments. *Geology*, 24(6):499–502, 1996. URL http://geology.gsapubs.org/content/24/6/499.short.

Wei Fang, Xianfeng Huang, Fan Zhang, and Deren Li. Intensity Correction of Terrestrial Laser Scanning Data by Estimating Laser Transmission Function. *IEEE Transactions on Geoscience and Remote Sensing*, 53(2):942–951, February 2015. ISSN 0196-2892, 1558-0644. doi: 10.1109/TGRS.2014.2330852. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6849466.

A. Ferraz, F. Bretar, S. Jacquemoud, G. Gonçalves, L. Pereira, M. Tomé, and P. Soares. 3-D mapping of a multi-layered Mediterranean forest using ALS data. *Remote Sensing of Environment*, 121:210–223, 2012.

A.M. Ferrero, G. Forlani, R. Roncella, and H.I. Voyat. Advanced geostructural survey methods applied to rock mass characterization. *Rock Mechanics and Rock Engineering*, 42(4):631–665, 2009.

Marco Franceschi, Giordano Teza, Nereo Preto, Arianna Pesci, Antonio Galgaro, and Stefano Girardi. Discrimination between marls and limestones using intensity data from terrestrial laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6):522–528, November 2009. ISSN 09242716.

doi: 10.1016/j.isprsjprs.2009.03.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0924271609000446.

Marco Franceschi, Nereo Preto, Linda A. Hinnov, Chunju Huang, and Giovanni Rusciadelli. Terrestrial laser scanner imaging reveals astronomical forcing in the Early Cretaceous of the Tethys realm. *Earth and Planetary Science Letters*, 305(3-4):359–370, May 2011. ISSN 0012821X. doi: 10.1016/j.epsl.2011.03.017. URL http://linkinghub.elsevier.com/retrieve/pii/S0012821X11001610.

Marco Franceschi, Luca Penasa, Jan Smit, Rodolfo Coccioni, Jérôme Gattacceca, Antonio Cascella, Sandro Mariani, and Alessandro Montanari. Comparison of carbonate content, facies ranking and Terrestrial Laser Scanner intensity as proxies for the cyclostratigraphic analysis of the Ypresian-Lutetian pelagic section of Smirra (Umbria-Marche Basin, Italy). *Palaeogeography, Palaeoclimatology, Palaeoecology*, submitted, 2015.

D. García-Sellés, O. Falivene, P. Arbués, O. Gratacos, S. Tavani, and J.A. Muñoz. Supervised identification and reconstruction of near-planar geological surfaces from terrestrial laser scanning. *Computers & Geosciences*, 37(10):1584–1594, 2011. ISSN 0098-3004. doi: 10.1016/j.cageo.2011.03.007.

M. Ghil, M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou. *Advanced Spectral Methods for Climatic Time Series*. 2001.

Daniel Girardeau-Montaut. Cloudcompare, a 3D point cloud and mesh processing free software. EDF R&D, Telecom ParisTech., 2014. URL http://cloudcompare.org/.

H. González-Jorge, D. Gonzalez-Aguilera, P. Rodriguez-Gonzalvez, and P. Arias. Monitoring biological crusts in civil engineering structures using intensity data from terrestrial laser scanners. *Construction and Building Materials*, 31:119–128, June 2012. ISSN 09500618. doi: 10.1016/j.conbuildmat.2011.12.053. URL http://linkinghub.elsevier.com/retrieve/pii/S095006181100732X.

Alessandro Grippo. Cyclostratigraphy and chronology of the Albian stage (Piobbico core, Italy). 2004.

A. Guarnieri, A. Vettore, F. Pirotti, M. Menenti, and M. Marani. Retrieval of small-relief marsh morphology from Terrestrial Laser Scanner, optimal spatial filtering, and laser return intensity. *Geomorphology*, 113(1-2):12–20,

December 2009. ISSN 0169555X. doi: 10.1016/j.geomorph.2009.06.005. URL http://linkinghub.elsevier.com/retrieve/pii/S0169555X09002360.

Teemu Hakala, Juha Suomalainen, Sanna Kaasalainen, and Yuwei Chen. Full waveform hyperspectral LiDAR for terrestrial laser scanning. *Optics express*, 20 (7):7119–7127, 2012. URL http://www.opticsinfobase.org/abstract.cfm?uri=oe-20-7-7119.

Preston Hartzell, Craig Glennie, Kivanc Biber, and Shuhab Khan. Application of multispectral LiDAR to automated virtual outcrop geology. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88:147–155, February 2014. ISSN 09242716. doi: 10.1016/j.isprsjprs.2013.12.004. URL http://linkinghub.elsevier.com/retrieve/pii/S0924271613002967.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*, volume 1. Springer Series in Statistics New York, 2 edition, 2009. URL http://statweb.stanford.edu/~tibs/book/preface.ps.

Gideon M. Henderson. New oceanic proxies for paleoclimate. *Earth and Planetary Science Letters*, 203(1):1–13, 2002. URL http://www.sciencedirect.com/science/article/pii/S0012821X02008099.

L.A. Hinnov and F.J. Hilgen. Cyclostratigraphy and Astrochronology. In *The Geologic Time Scale*, pages 63–83. Elsevier, 2012. ISBN 9780444594259. URL http://linkinghub.elsevier.com/retrieve/pii/B9780444594259000044.

David Hodgetts. Laser scanning and digital outcrop geology in the petroleum industry: A review. *Marine and Petroleum Geology*, 46:335–354, September 2013. ISSN 02648172. doi: 10.1016/j.marpetgeo.2013.02.014. URL http://linkinghub.elsevier.com/retrieve/pii/S0264817213000494.

Bernhard Höfle and Norbert Pfeifer. Correction of laser scanning intensity data: Data and model-driven approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(6):415–433, December 2007. ISSN 09242716. doi: 10.1016/j.isprsjprs.2007.05.008. URL http://linkinghub.elsevier.com/retrieve/pii/S0924271607000603.

Bernhard Höfle, Michael Vetter, Norbert Pfeifer, Gottfried Mandlburger, and Johann Stötter. Water surface mapping from airborne laser scanning using signal intensity and elevation data. *Earth Surface Processes and Landforms*, 34 (12):1635–1649, September 2009. ISSN 01979337, 10969837. doi: 10.1002/esp.1853. URL http://doi.wiley.com/10.1002/esp.1853.

M. R. House. Orbital forcing timescales: an introduction. *Geological Society, London, Special Publications*, 85(1):1–18, January 1995. ISSN 0305-8719. doi: 10.1144/GSL.SP.1995.085.01.01. URL http://sp.lyellcollection.org/cgi/doi/10.1144/GSL.SP.1995.085.01.01.

Xiumian Hu, Luba Jansa, and Massimo Sarti. Mid-Cretaceous oceanic red beds in the Umbria–Marche Basin, central Italy: Constraints on paleoceanography and paleoclimate. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 233(3-4): 163–186, April 2006. ISSN 00310182. doi: 10.1016/j.palaeo.2005.10.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0031018205005742.

Masayuki Ikeda, Ryuji Tada, and Hironobu Sakuma. Astronomical cycle origin of bedded chert: A middle Triassic bedded chert sequence, Inuyama, Japan. *Earth and Planetary Science Letters*, 297(3-4):369–378, September 2010. ISSN 0012821X. doi: 10.1016/j.epsl.2010.06.027. URL http://linkinghub.elsevier.com/retrieve/pii/S0012821X10004024.

Jordi Inglada. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(3):236–248, 2007. ISSN 0924-2716. doi: 10.1016/j.isprsjprs.2007.05.011.

B. Jutzi and H. Gross. Normalization of Lidar Intensity Data Based on Range and Surface Incidence Angle. In *In proceedings of Laserscanning 09*, volume 38, pages 213–218, September 2009.

S. Kaasalainen, J. Hyyppä, P. Litkey, H. Hyyppä, E. Ahokas, A. Kukko, and H. Kaartinen. Radiometric calibration of ALS intensity. *Int. Arch. Photogramm. Remote Sens*, 36:201–205, 2007a. URL http://www.isprs.org/proceedings/XXXVI/3-W52/final_papers/Kaasalainen_2007.pdf.

S. Kaasalainen, H. Hyyppa, A. Kukko, P. Litkey, E. Ahokas, J. Hyyppa, H. Lehner, A. Jaakkola, J. Suomalainen, A. Akujarvi, M. Kaasalainen, and U. Pyysalo. Radiometric Calibration of LIDAR Intensity With Commercially Available Reference Targets. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2): 588–598, February 2009a. ISSN 0196-2892, 1558-0644. doi: 10.1109/TGRS.2008.2003351. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4689353.

S. Kaasalainen, H. Niittymaki, A. Krooks, K. Koch, H. Kaartinen, A. Vain, and H. Hyyppa. Effect of Target Moisture on Laser Scanner Intensity. *IEEE*

Transactions on Geoscience and Remote Sensing, 48(4):2128–2136, April 2010a. ISSN 0196-2892, 1558-0644. doi: 10.1109/TGRS.2009.2036841. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5357425.

Sanna Kaasalainen, Tomi Lindroos, and Juha Hyyppa. Toward Hyperspectral Lidar: Measurement of Spectral Backscatter Intensity With a Supercontinuum Laser Source. *IEEE Geoscience and Remote Sensing Letters*, 4(2):211–215, April 2007b. ISSN 1545-598X. doi: 10.1109/LGRS.2006.888848. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4156158.

Sanna Kaasalainen, A. Vain, Anssi Krooks, and Antero Kukko. Topographic and Distance Effects in Laser Scanner Intensity Correction. In *Laserscanning 2009, IAPRS*, volume XXXVIII, pages 219–223, 2009b.

Sanna Kaasalainen, Juha Suomalainen, Teemu Hakala, Yuwei Chen, E. Raikkonen, Eetu Puttonen, and Harri Kaartinen. Active hyperspectral LIDAR methods for object classification. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, pages 1–4. IEEE, 2010b. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5594863.

Sanna Kaasalainen, Anttoni Jaakkola, Mikko Kaasalainen, Anssi Krooks, and Antero Kukko. Analysis of Incidence Angle and Distance Effects on Terrestrial Laser Scanner Intensity: Search for Correction Methods. *Remote Sensing*, 3 (12):2207–2221, October 2011. ISSN 2072-4292. doi: 10.3390/rs3102207. URL http://www.mdpi.com/2072-4292/3/10/2207/.

Klaas Klasing, Daniel Althoff, Dirk Wollherr, and Martin Buss. Comparison of surface normal estimation methods for range sensing applications. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3206–3211. IEEE, 2009. doi: 10.1109/ROBOT.2009.5152493. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5152493.

B. Koch, C. Straub, M. Dees, Y. Wang, and H. Weinacker. Airborne laser data for stand delineation and information extraction. *International Journal of Remote Sensing*, 30(4):935–963, 2009.

A. Krooks, S. Kaasalainen, T. Hakala, and O. Nevalainen. Correction of Intensity Incidence Angle Effect in Terrestrial Laser Scanning. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-5/W2:145–150, October 2013. ISSN 2194-9050. doi: 10.5194/isprsannals-II-5-W2-145-2013.

URL http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/II-5-W2/145/2013/.

T. H. Kurz, S. J. Buckley, and J. A. Howell. Close range hyperspectral imaging integrated with terrestrial LiDAR scanning applied to rock characterization at centimeter scale. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, 34:417–422, 2012a. URL http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXIX-B5/417/2012/isprsarchives-XXXIX-B5-417-2012.pdf.

T.H. Kurz, S.J. Buckley, and J.A. Howell. Close-range hyperspectral imaging for geological field studies: Workflow and methods. *International Journal of Remote Sensing*, 34(5):1798–1822, 2013.

Tobias H. Kurz, Julie Dewit, Simon J. Buckley, John B. Thurmond, David W. Hunt, and Rudy Swennen. Hyperspectral image analysis of different carbonate lithologies (limestone, karst and hydrothermal dolomites): the Pozalagua Quarry case study (Cantabria, North-west Spain): Hyperspectral image analysis of carbonate lithologies. *Sedimentology*, 59(2):623–645, February 2012b. ISSN 00370746. doi: 10.1111/j.1365-3091.2011.01269.x. URL http://doi.wiley.com/10.1111/j.1365-3091.2011.01269.x.

Dimitri Lague, Nicolas Brodu, and Jérôme Leroux. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing*, 82:10–26, August 2013. ISSN 09242716. doi: 10.1016/j.isprsjprs.2013.04.009. URL http://linkinghub.elsevier.com/retrieve/pii/S0924271613001184.

J.-F. Lalonde, N. Vandapel, D.F. Huber, and M. Hebert. Natural terrain classification using three-dimensional ladar data for ground robot mobility. *Journal of Field Robotics*, 23(10):839–861, 2006.

Hans Petter Langtangen, editor. *Python Scripting for Computational Science*, volume 3 of *Texts in Computational Science and Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-73915-9, 978-3-540-73916-6. URL http://link.springer.com/10.1007/978-3-540-73916-6.

Håkan Larsson, Tomas Chevalier, and Frank Gustafsson. *3-D Structure and Reflectance Measurements: A System Analysis of Lidar Optech ILRIS-3D*. FOI-R. Sensor systems, Defence research agency (FOI), 2007.

J. Laskar, P. Robutel, F. Joutel, M. Gastineau, A. C. M. Correia, and B. Levrard. A long-term numerical solution for the insolation quantities ofï¿œtheï¿œEarth. *Astronomy and Astrophysics*, 428(1):261–285, December 2004. ISSN 0004-6361, 1432-0756. doi: 10.1051/0004-6361:20041335. URL http://www.edpsciences.org/10.1051/0004-6361:20041335.

J. Laskar, A. Fienga, M. Gastineau, and H. Manche. La2010: a new orbital solution for the long-term motion of the Earth. *Astronomy & Astrophysics*, 532:A89, August 2011. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/201116836. URL http://www.aanda.org/10.1051/0004-6361/201116836.

Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. A sparse texture representation using affine-invariant regions. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–319. IEEE, 2003. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1211486.

Derek D. Lichti and Sonam Jamtsho. Angular resolution of terrestrial laser scanners. *The Photogrammetric Record*, 21(114):141–160, 2006. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1477-9730.2006.00367.x/full.

Lorraine E. Lisiecki and Philip A. Lisiecki. Application of dynamic programming to the correlation of paleoclimate records: DYNAMIC PROGRAMMING SIGNAL CORRELATION. *Paleoceanography*, 17(4):1–1–1–12, December 2002. ISSN 08838305. doi: 10.1029/2001PA000733. URL http://doi.wiley.com/10.1029/2001PA000733.

H.-G. Maas and G. Vosselman. Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3):153–163, 1999. URL http://www.scopus.com/inward/record.url?eid=2-s2.0-0032783294&partnerID=40&md5=83d80061176f0fb994ee9ebfbef1d0cf. cited By (since 1996)182.

Michael E Mann and Jonathan M Lees. Robust estimation of background noise and signal detection in climatic time series. *Climatic change*, 33(3):409–445, 1996.

L. Marchegiani, G. Bertotti, G. Cello, G. Deiana, S. Mazzoli, and E. Tondi. Pre-orogenic tectonics in the Umbria-Marche sector of the Afro-Adriatic continental margin. *Tectonophysics*, 315(1-4):123–143, 1999.

F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, 2004. doi: 10.1109/TGRS.2004.831865. cited By (since 1996)554.

A. Mian, M. Bennamoun, and R. Owens. On the Repeatability and Quality of Keypoints for Local Feature-based 3D Object Retrieval from Cluttered Scenes. *International Journal of Computer Vision*, 89(2-3):348–361, September 2010. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-009-0296-z. URL http://link.springer.com/10.1007/s11263-009-0296-z.

Niloy J. Mitra, An Nguyen, and Leonidas Guibas. Estimating surface normals in noisy point cloud data. *International Journal of Computational Geometry & Applications*, 14(04n05):261–276, 2004. URL http://www.worldscientific.com/doi/abs/10.1142/s0218195904001470.

Alan C. Mix, Sara E. Harris, and Thomas R. Janecek. Estimating lithology from nonintrusive reflectance spectra: Leg 138. In *Proceedings of the Ocean Drilling Program, Scientific Results*, 1995. URL http://ir.library.oregonstate.edu/jspui/handle/1957/13644.

Eustoquio Molina, Laia Alegret, Estibaliz Apellaniz, Gilen Bernaola, Fernando Caballero, Jaume Dinarès-Turell, Jan Hardenbol, Claus Heilmann-Clausen, Juan C. Larrasoana, Hanspeter Luterbacher, and others. The Global Stratotype Section and Point (GSSP) for the base of the Lutetian Stage at the Gorrondatxe section, Spain. *Episodes*, 34(2):86–108, 2011. URL http://wzar.unizar.es/isps/pdf/Molina2011Episodes.pdf.

L. Monika Moskal and Guang Zheng. Retrieving Forest Inventory Variables with Terrestrial Laser Scanning (TLS) in Urban Heterogeneous Forest. *Remote Sensing*, 4(12):1–20, December 2011. ISSN 2072-4292. doi: 10.3390/rs4010001. URL http://www.mdpi.com/2072-4292/4/1/1/.

Giorgos Mountrakis, Jungho Im, and Caesar Ogole. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247–259, May 2011. ISSN 09242716. doi: 10.1016/j.isprsjprs.2010.11.001. URL http://linkinghub.elsevier.com/retrieve/pii/S0924271610001140.

Richard J. Murphy and Sildomar T. Monteiro. Mapping the distribution of ferric iron minerals on a vertical mine face using derivative analysis of hyperspectral imagery (430–970nm). *ISPRS Journal of Photogrammetry and Remote Sensing*,

75:29–39, January 2013. ISSN 09242716. doi: 10.1016/j.isprsjprs.2012.09.014. URL http://linkinghub.elsevier.com/retrieve/pii/S0924271612001839.

Giovanni Muttoni and Dennis V. Kent. Widespread formation of cherts during the early Eocene climate optimum. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 253(3-4):348–362, September 2007. ISSN 00310182. doi: 10.1016/j.palaeo.2007.06.008. URL http://linkinghub.elsevier.com/retrieve/pii/S0031018207003033.

Joanna M. Nield, Giles F.S. Wiggs, and Robert S. Squirrell. Aeolian sand strip mobility and protodune development on a drying beach: examining surface moisture and surface roughness patterns measured by terrestrial laser scanning. *Earth Surface Processes and Landforms*, 36(4):513–522, April 2011. ISSN 01979337. doi: 10.1002/esp.2071. URL http://doi.wiley.com/10.1002/esp.2071.

Karl Pearson and Louis Napoleon George Filon. Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, pages 229–311, 1898.

Luca Penasa and Dimitri Lague. Self-normalization of terrestrial lidar intensity using scene-available targets. *in preparation*, 2015.

Luca Penasa, Marco Franceschi, Nereo Preto, Giordano Teza, and Vanessa Polito. Integration of intensity textures and local geometry descriptors from Terrestrial Laser Scanning to map chert in outcrops. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93:88–97, July 2014. ISSN 09242716. doi: 10.1016/j.isprsjprs.2014.04.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0924271614000902.

Arianna Pesci and Giordano Teza. Effects of surface irregularities on intensity data from laser scanning: an experimental approach. *Annals of Geophysics*, 51 (5-6):839–848, 2008. ISSN 2037-416X. doi: 10.4401/ag-4462. URL http://www.annalsofgeophysics.eu/index.php/annals/article/viewArticle/4462.

Arianna Pesci, Giordano Teza, Guido Ventura, and others. Remote sensing of volcanic terrains by terrestrial laser scanner: preliminary reflectance and RGB implications for studying Vesuvius crater (Italy). *Annals of Geophysics*, 2008. URL http://www.earth-prints.org/handle/2122/3989.

Norbert Pfeifer, Bernhard Höfle, Christian Briese, Martin Rutzinger, and Alexander Haring. Analysis of the backscattered energy in terrestrial laser scanning data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, 37:1045–1052, 2008.

M Pierrot-Deseilligny. MicMac, Apero, Pastis and Other Beverages in a Nutshell, 2013. URL `http://logiciels.ign.fr/IMG/pdf/docmicmac.pdf`.

S. Pu, M. Rutzinger, G. Vosselman, and S. Oude Elberink. Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6 SUPPL.):S28–S39, 2011.

Michael R. Rampino, Andreas Prokoph, and Andre Adler. Tempo of the end-Permian event: High-resolution cyclostratigraphy at the Permian-Triassic boundary. *Geology*, 28(7):643–646, 2000. URL `http://geology.gsapubs.org/content/28/7/643.short`.

R.A. Viscarra Rossel and A.B. McBratney. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. *Geoderma*, 85(1):19–39, 1998. ISSN 0016-7061. doi: 10.1016/S0016-7061(98)00023-8.

Ana-Maria Rosu, Marc Pierrot-Deseilligny, Arthur Delorme, Renaud Binet, and Yann Klinger. Measurement of ground displacement from optical satellite image correlation using the free open-source software MicMac. *ISPRS Journal of Photogrammetry and Remote Sensing*, 100:48–59, February 2015. ISSN 09242716. doi: 10.1016/j.isprsjprs.2014.03.002. URL `http://linkinghub.elsevier.com/retrieve/pii/S0924271614000653`.

Radu Bogdan Rusu and Steve Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 2011. URL `http://pointclouds.org/`.

M. Santantonio. Pelagic carbonate platforms in the geologic record: their classification, and sedimentary and paleotectonic evolution. *AAPG Bulletin-American Association of Petroleum Geologists*, 78(1):122–141, 1994.

J.R. Schoenberg, A. Nathan, and M. Campbell. Segmentation of dense range information in complex urban scenes. *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings*, pages 2033–2038, 2010.

Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001. ISBN 0262194759.

Jie. Shan and Charles K. Toth. *Topographic laser ranging and scanning: principles and processing*. CRC Press/Taylor & Francis Group, Boca Raton, 2009. ISBN 9781420051421 1420051423.

I. Sims and P. Nixon. RILEM recommended test method AAR-1: detection of potential alkali-reactivity of aggregates—petrographic method. *Materials and Structures*, 36(7):480–496, 2003. URL http://link.springer.com/article/10.1007/BF02481528.

Paolo Sterzai, Michela Vellico, Matteo Berti, Franco Coren, Alessandro Corsini, Alberto Rosi, Paolo Mora, Franco Zambonelli, and Francesco Ronchetti. LIDAR and Hyperspectral data integration for landslide monitoring: the test case of Valoria landslide. *Italian Journal of Remote Sensing*, 42(3):89–99, 2010. URL http://server-geolab.agr.unifi.it/public/completed/2010_RIT_VOL42(3)_089_099_Sterzai.pdf.

Juha Suomalainen, Teemu Hakala, Harri Kaartinen, Esa Räikkönen, and Sanna Kaasalainen. Demonstration of a virtual active hyperspectral LiDAR in automated point cloud classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(5):637–641, September 2011. ISSN 09242716. doi: 10.1016/j.isprsjprs.2011.04.002. URL http://linkinghub.elsevier.com/retrieve/pii/S0924271611000529.

Tang C. and Gérard Medioni. Curvature-augmented tensor voting for shape inference from noisy 3D data. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 24(6):858–864, June 2002. doi: 10.1109/TPAMI.2002.1008395.

David J. Thomson. Spectrum Estimation and Harmonic Analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982. URL http://www.bibsonomy.org/bibtex/29ad5f929a13d06e6c42eed053c66eb62/lran022.

Freek van der Meer. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 8(1):3–17, January 2006. ISSN 03032434. doi: 10.1016/j.jag.2005.06.001. URL http://linkinghub.elsevier.com/retrieve/pii/S030324340500053X.

Robert Vautard, Pascal Yiou, and Michael Ghil. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D: Nonlinear Phenomena*, 58(1–4):95 – 126, 1992. ISSN 0167-2789. doi: http://dx.doi.org/10.1016/0167-2789(92)90103-T. URL http://www.sciencedirect.com/science/article/pii/016727899290103T.

V. Verma, R. Kumar, and S. Hsu. 3D building detection and modeling from aerial LIDAR data. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2213–2220, 2006.

Chi-Kuei Wang and Yao-Yu Lu. Potential of ILRIS3D Intensity Data for Planar Surfaces Segmentation. *Sensors*, 9(7):5770–5782, July 2009. ISSN 1424-8220. doi: 10.3390/s90705770. URL http://www.mdpi.com/1424-8220/9/7/5770/.

Graham P. Weedon. *Time-series analysis and cyclostratigraphy examining stratigraphic records of environmental cycles*. Cambridge University Press, Cambridge, U.K.; New York, 2003. ISBN 0511065965 9780511065965 9780511535482 0511535481. URL http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=120287.

Zhi Xing, Anne-Lise Beaucour, Ronan Hebert, Albert Noumowe, and Béatrice Ledesert. Influence of the nature of aggregates on the behaviour of concrete subjected to elevated temperature. *Cement and Concrete Research*, 41(4):392–402, 2011. ISSN 0008-8846. doi: 10.1016/j.cemconres.2011.01.005.

Bisheng Yang and Zhen Dong. A shape-based segmentation method for mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 81(0):19–30, 2013. ISSN 0924-2716. doi: 10.1016/j.isprsjprs.2013.04.002.

B.S. Yang, Z. Wei, Q. Li, and J. Li. Semiautomated building facade footprint extraction from mobile LiDAR point clouds. *IEEE Geoscience and Remote Sensing Letters*, 10(4):766–770, 2013.

Timo Zinßer, Jochen Schmidt, and Heinrich Niemann. Point set registration with integrated scale estimation. In *Vortrag: Eighth International Conference on Pattern Recognition and Image Processing, Belarusian State University of Informatics and Radioelectronics, International Association for Pattern Recognition, Minsk, Belarus*, volume 18, 2005. URL http://publications.yjschmidt.de/pdf/jschmidt_prip_2005.pdf.

This thesis was typeset with LaTeX 2ε using Hermann Zapf's *Palatino* and *Euler* type faces. The listings are typeset in *Bera Mono*, originally developed by Bitstream, Inc. as "Bitstream Vera".

All the sites addresses were reachable at the time of the final version.

Final Version as of February 1, 2015 (`classicthesis` version 4.1).