BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF MECHANICAL ENGINEERING
FAKULTA STROJNÍHO INŽENÝRSTVÍ

INSTITUTE OF PHYSICAL ENGINEERING
ÚSTAV FYZIKÁLNÍHO INŽENÝRSTVÍ

PROCESSING OF TOMOGRAPHIC DATA BY PRINCIPAL COMPONENT ANALYSIS METHOD FOR ARCHAEOLOGICAL APPLICATIONS

ZPRACOVÁNÍ TOMOGRAFICKÝCH DAT METODOU ANALÝZY HLAVNÍCH KOMPONENT PRO ARCHEOLOGICKÉ APLIKACE

MASTER'S THESIS
DIPLOMOVÁ PRÁCE

AUTHOR
AUTOR PRÁCE

Bc. David Prokop

SUPERVISOR
VEDOUCÍ PRÁCE

Ing. Pavel Pořízka, Ph.D.

BRNO 2018

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF MECHANICAL ENGINEERING

Specification Master's Thesis

Department:	Institute of Physical Engineering
Student:	**Bc. David Prokop**
Study programme:	Applied Sciences in Engineering
Study field:	Physical Engineering and Nanotechnology
Supervisor:	**Ing. Pavel Pořízka, Ph.D.**
Academic year:	2018/19

Pursuant to Act no. 111/1998 concerning universities and the BUT study and examination rules, you have been assigned the following topic by the institute director Master's Thesis:

Processing of tomographic data by principal component analysis method for archaeological applications

Concise characteristic of the task:

X–Ray Computed Tomography is a standardized method which allows imaging and visualization of studied samples and their further characterization. Typical outputs from XCT measurement are large datasets, those datasets need to be then reconstructed and segmented.
In case of inhomogeneous samples, like archaeological samples – ceramics, other visual inspection of the segmented data could be rather difficult. It is then convenient to look for automatic algorithms that allow fast overview analysis. In the process of segmentation, each individual structure and inclusion could be characterized by additional variables (volume, size, etc.), these variables could be a starting point for further statistical elaboration. Considering the studied problematics, chosen method of multidimensional statistical analysis, namely principal component analysis (PCA), seems like a good alternative.

Goals Master's Thesis:

1) To prepare samples and to analyze them using X–Ray Computed Tomography (XCT).
2) To adopt processes in the segmentation of XCT data.
3) To implement Principal Component Analysis algorithm in the processing of structural data.

Recommended bibliography:

HERMAN, G. T. Fundamentals of computerized tomography: image reconstruction from projections. 2nd edition. New York: Springer. Advances in pattern recognition, 2009. ISBN 978-1-84628-723-7.

MELOUN, M. and MILITKÝ, J. Statistická analýza experimentálních dat. 2nd edition. Praha: Academia, 2004. ISBN 80-200-1254-0.

Deadline for submission Master's Thesis is given by the Schedule of the Academic year 2018/19

In Brno,

L. S.

prof. RNDr. Tomáš Šikola, CSc.	doc. Ing. Jaroslav Katolický, Ph.D.
Director of the Institute	FME dean

ABSTRACT

X-ray computed tomography is an imaging method which enables 3D observations of the inner structure of objects. The microstructure of objects carries essential information that could be used to characterize samples. The here presented research aims at connecting the microCT-derived data with their statistical elaboration in order to achieve the best description of analysed samples. The outcome of the thesis will be the classification of samples based on their microstructure. Based on the information about classification, we can test some hypotheses concerning the origin of the samples. This thesis could serve as a cornerstone for a combination of datasets coming from various sources by the means of statistical methods.

KEYWORDS

X-ray computed tomography, principal component analysis, image processing, segmentation, clustering.

ABSTRAKT

Rentgenová počítačová tomografie je metoda sloužící ke 3D zobrazování vnitřní struktury objektů. Mikrostruktura objektů ukrývá důležité informace, které mohou být použity k jejich charakterizaci. Tato práce podává spojení mezi datasety získanými pomocí rentgenové počítačové mikrotomografie a oblastí statistického zpracování dat. Výstupem metody, pak bude klasifikace vzorků na základě informací o jejich mikrostruktuře. Z výsledků klasifikace vzorků, pak můžeme vyvodit různé hypotézy týkající se původu vzorků. Tato práce by mimo jiné mohla sloužit jako takový nový vhled do problematiky kombinace dat různého původu, pomocí metod statistické analýzy.

KLÍČOVÁ SLOVA

Rentgenová počítačová tomografie, analýza hlavních komponent, zpracování obrazu, segmentace, shlukování.

PROKOP, David *Processing of tomographic data by principal component analysis method for archaeological applications*: master's thesis. Brno: Brno University of Technology, Faculty of Mechanical Engineering, Institute of Physical Engineering, 2019. 53 p. Supervised by Ing. Pavel Pořízka, Ph.D. Externally supervised by Dr. Federico Bernardini.

DECLARATION

I declare that I have elaborated my master's thesis on the theme of "Processing of tomographic data by principal component analysis method for archaeological applications" independently, under the supervision of the master's thesis supervisor and with the use of technical literature and other sources of information which are all quoted in the thesis and detailed in the list of literature at the end of the thesis.

As the author of the master's thesis I furthermore declare that, concerning the creation of this master's thesis, master's thesis, I have not infringed any copyright. In particular, I have not unlawfully encroached on anyone's personal copyright and I am fully aware of the consequences in the case of breaking Regulation § 11 and the following of the Copyright Act No 121/2000 Vol., including the possible consequences of criminal law resulted from Regulation § 152 of Criminal Act No 140/1961 Vol.

Brno ....

(author's signature)

Acknowledgement

I would like to thank my supervisor Doctor Pavel Pořízka for his comments and corrections during the creation of my thesis. I would like to thank the head of the CEITEC research group Professor Jozef Kaiser, for giving me the opportunity to participate in the activities of the research group. Next my gratitude goes to Doctor Federico Bernardini, who was my mentor during my internship at ICTP and was also helping me by giving me valuable feedback concerning my work. Last but not least I would like to thank my family, because without their great support, none of this would have been possible. I acknowledge the support of the Brno University of Technology through grant FSI-S-17-4506. Part of this work was carried out with support of the project CEITEC 2020 (LQ1601) with financial support from the Ministry of Education, Youth and Sports of the Czech Republic under the National Sustainability Programme II and CEITEC Nano Research Infrastructure (MEYS CR, 2016-2019).

Bc. David Prokop

CONTENTS

Introduction **1**

1 X-ray computed tomography **3**
1.1 X-ray sources 4
1.2 X-ray detectors 7
1.3 Reconstruction 8

2 Image analysis **13**
2.1 CT data representation 14
2.2 Segmentation 14
2.2.1 Tresholding 15
2.2.2 Region growing 15
2.2.3 Machine learning 16

3 Principal component analysis **19**
3.1 Derivation 19
3.2 Scaling and centring 20
3.3 Number of used principal components 21
3.4 Geometrical interpretation of the PCA 23
3.5 Outliers detection 24
3.6 Cluster analysis 24
3.6.1 K-means clustering 25
3.6.2 K-medoids clustering 26
3.6.3 Fuzzy clustering 26

4 Experiment **27**
4.1 Origin of the samples 27
4.2 Measurement 29
4.3 Metrics extraction 29
4.4 Data preparation 32

5 Data processing **35**
5.1 Principal component analysis 35
5.2 Cluster analysis 39
5.3 Interpretation 40

Conclusion **45**

Bibliography **47**

List of abbreviations and symbols **53**

INTRODUCTION

Using large datasets is becoming more and more common in every discipline. Today's biggest task is to find a tool which would reduce effectively their dimensionality. By the word effectively here it is meant to keep the loss of information within datasets to a minimum, thus maintaining the interpretability of data at a maximum possible level. An example of such a discipline where handling of large datasets is needed on a daily basis is X-ray computed tomography. X-ray computed tomography has become a standard method for 3D visualization of objects. Today this method covers a whole range of topics. Some examples are the visualization of biological samples, study of the inner structure of advanced materials and coatings or 3D modelling and characterization of archaeological samples. Last of the applications has still a hidden potential that remains to be fully uncovered. One of the current tasks that archaeologists deal with concerns the study of ancient technology, raw materials and possible origin of the artefacts, such as pottery. Such information are useful to reconstruct ancient manufacturing processes, exchange systems and trade. Thus developing a method that could help to solve this objective is of major interest. In this thesis, a big effort has been put in order to present a multi-analytical approach method that could serve as a cornerstone for these types of complex analyses. The main principle of the method will be based on the extraction of important parameters, from a group of pottery samples, for example area of inclusions, inclusions and clay ratios and so on. All these parameters will then be used for a statistical analysis, namely principal component analysis (PCA). The result of this analysis will then be a classification of samples based on their microstructure.

This thesis follows up my recent work done collaborating with local archaeological group at the International Centre for Theoretical Physics (ICTP) in Trieste. The first achievements of our work were already presented at XVIII International Conference on Science, Arts and Culture in Croatia. The publication that recently followed up was included as a supplementary material. The theoretical part of this work can be divided into three parts. In the first part (chapter 1), basics of X-ray computed tomography are explained, with particular emphasis given to the problem of reconstruction. The second part (chapter 2) addresses the problem of CT data representation, together with the presentation of various segmentation methods. In the third part (chapter 3), the main properties of PCA method are presented, with some emphasis placed on data clustering algorithms. The experimental part (chapter 4 and chapter 5) deals with the physical realization of measurement of the samples, together with the presentation of the method that serves to classify the samples. Last but not least, in chapter 5, all achieved results are presented, along with their interpretation and a discussion.

1 X-RAY COMPUTED TOMOGRAPHY

Tomography from the Greek word "tomos" meaning slice or section and "graphien", meaning to write. CT is a non-destructive technique allowing 3D visualisation of objects. The first medical images were performed by Hounsfield in 1972. During this time, the problem of how to handle and store huge amount of data still persisted. Later on in the 1980s, high speed and high memory computers appeared. With this in mind, it was clear that it did not take too long and a whole new range of industrial applications had appeared [2].

The basic set-up consists of detecting the photons which are transmitted through the studied object. When radiation passes through the studied object, the intensity of the transmitted beam is changed. This change in beam intensity which we register at the detector is described by the Beer-Lambert law.

$$I = I_0 e^{-\mu(x)}, \tag{1.1}$$

where I is the detected intensity of the transmitted beam, I_0 defines the intensity of a beam generated by X-ray source, μ is the linear attenuation coefficient and x is the sample thickness.

This formula holds only for a monochromatic beam and for cases when μ is a constant within whole material. In praxis this is not possible, and more general formula must be used. When the material is non-homogeneous, the linear attenuation coefficient μ is no longer a constant.

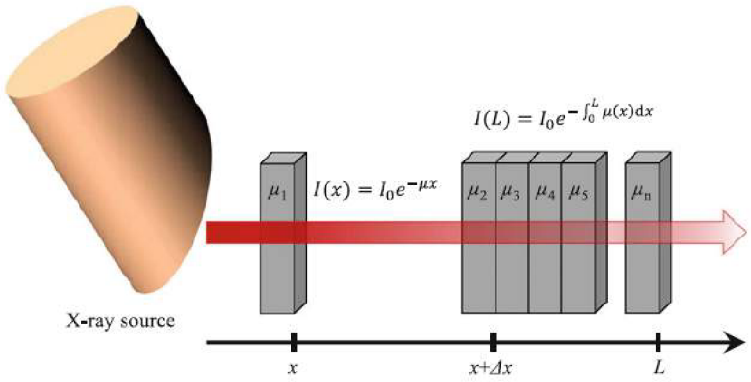

Fig. 1.1: Attenuation of X-rays in the case of (left) a homogeneous object (right) an object composed of different materials with different attenuation coefficients. Adjusted from [8].

In this case, the attenuation coefficient μ is defined as

$$\int \mu(x)dx = -\ln\left(\frac{I}{I_0}\right). \tag{1.2}$$

In addition, to adapt the relation to a polychromatic X-ray beam spectrum, this equation needs to be modified as follows:

$$\int \mu(x)dx = -\ln\left(\frac{\int j_0(E)\exp(\int \mu(x,E)dx)dE}{I_0}\right), \tag{1.3}$$

where $j_0(E)$ is the spectral density and dE is the energy spread of X-ray spectra.

1.1 X-ray sources

Laboratory X-ray source scheme is on Fig.1.2. Here electrons are emitted from a cathode, heated filament, and then the electrons are accelerated towards a metal target, usually thorium, molybdenum or copper.

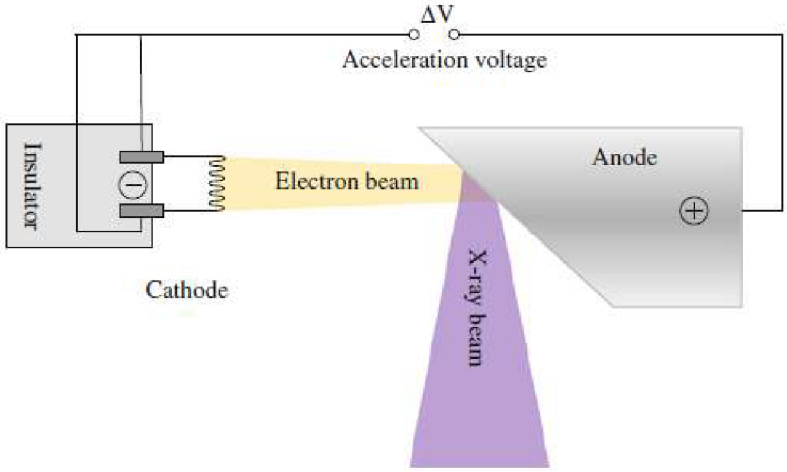

Fig. 1.2: Schematic representation of an X-ray tube. Electrons generated from the cathode are accelerated by applied voltage between cathode and anode. Generation of X-rays occurs when electrons hit the anode [8].

X-ray sources differ in many parameters, such as beam spot size, geometry, energy, etc. But two mechanisms correspond to the X-ray formation process that all these X-ray sources have in common. The first process is related to interactions of a free electron with those bound to an atom. If a charged particle with high energy approaches a solid metal, it will be able to excite or ionise the atoms. Then during this process electrons of another energetic shell can refill the gap and during this transition, they will emit the energy difference as radiation, known as characteristic radiation. The other process occurs due to the presence of electric forces that act on a charged particle when the particle hits a metal anode. These forces will then

change the velocity of a charged particle, and from the Larmor's formula [1] [45],

$$W(t) = \frac{2}{3c^3}\frac{e^2}{4\pi\varepsilon_0}\dot{v}^2, \tag{1.4}$$

it is stated that accelerated particle yields electromagnetic radiation. This radiation is called braking radiation, sometimes referred to as Bremsstrahlung [26]. Both processes are depicted in Fig. 1.3. Unfortunately, the whole process of X-ray generation

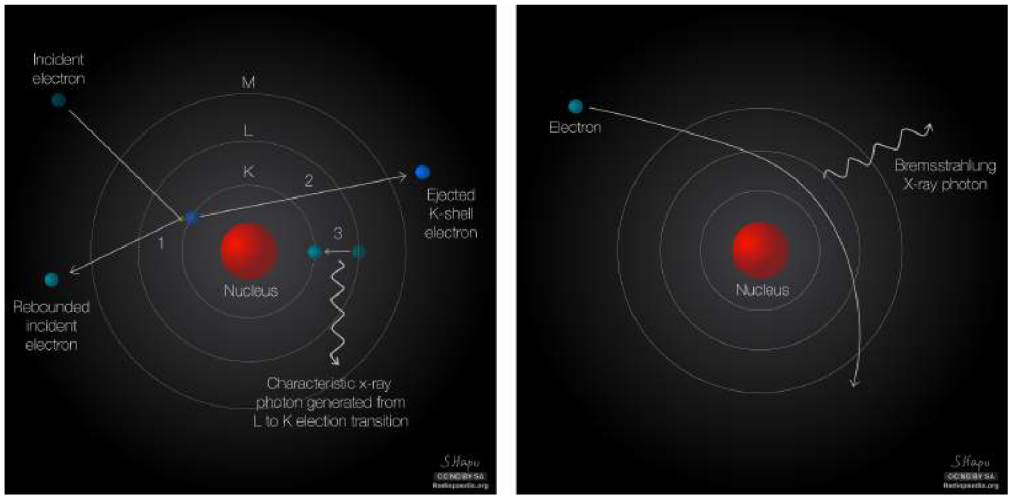

Fig. 1.3: Processes related to X-rays generation. Left: Creation of an characteristic K_α line. Right: Braking radiation (Bremsstrahlung) [3].

is not very efficient. About 99% of electrons that impact the target transfer their energy to processes that are related to heat production [20]. Both quality and quantity of the X-ray beam are then determined by many factors. Most importantly by acceleration voltage, tube current, beam filters and anode material. Acceleration voltage affects both energy interval and amplitude of the radiation. An increase of tube voltage then corresponds to a shift of the X-ray spectra to higher energies. Another parameter that can be adjusted is tube current. Tube current generates a linear increase in X-ray intensity which changes the number of emitted X-ray quanta while maintaining the distribution of X-ray energies. Effect of different tube voltage and current on X-ray spectra is depicted in Fig. 1.4.

Beam filtration is another important way of modifying beam quality. X-ray beam composes of photons with a wide range of energies. Due to the dependency of the attenuation coefficient on wavelength, lower energy photons are easily absorbed. As a consequence, the centre of the polychromatic X-ray is shifted to higher energies, leading to the effect called beam hardening. To prevent beam hardening, physical installation of beam filters is employed. These filters are usually thin metal foils

[1] This formula shows total power radiated by a non relativistic point charge.

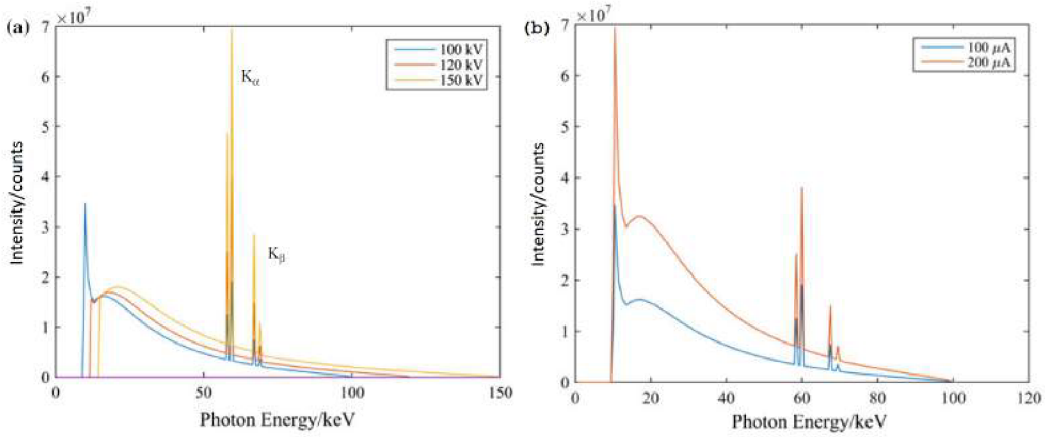

Fig. 1.4: X-ray spectrum at; (a) 100, 120 and 150 kV, (b) 100 and 200 µA for a tungsten target. X-ray spectra were obtained by software SpekCalc [8].

which are inserted nearby the beam output window. Effects of different filters on the resulting beam is described in Fig. 1.5.

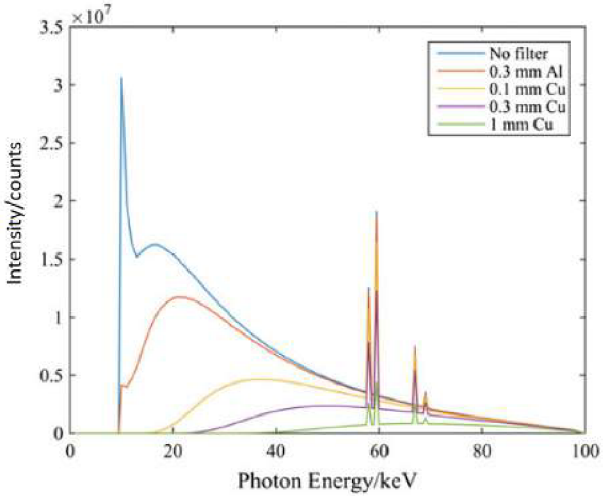

Fig. 1.5: Influence of X-ray filtration by different filter materials and thicknesses on X-ray spectra at an acceleration voltage 100 kV and a tungsten target material. X-ray spectra were obtained by software SpekCalc [8].

The anode material defines characteristic spectrum of an X-ray tube. The target material used in an X-ray anode must have a high atomic number to maximise Bremsstrahlung output. The target material must also tolerate a tremendous heat load. Therefore tungsten is chosen as the target material for many laboratory purposes due to its high atomic number, high melting point, and low vapour pressure [31].

1.2 X-ray detectors

As mentioned in the previous section, X-rays passing through the sample are attenuated due to absorption or scattering processes. The attenuated intensity of X-rays is then measured by an X-ray detector. There are different kinds of X-ray detectors, based on a principle which they use to convert X-ray energy to electrical signals. Two main types of detectors are scintillation (solid state detectors) and gas ionisation detectors [8]. In this section, the main focus will be given to the description of the scintillation detectors, since these detectors are widely used in an industrial CT.

Scintillation detectors as its name suggests consist of a scintillation medium, typically crystal, and a photodiode. In the first step, short-wave X-ray radiation is converted by the scintillation medium into visible light. This light is then directed to a photomultiplier tube, Fig. 1.6. As the light strikes the photocathode, the electrons are emitted via the photoelectric effect. These electrons then cascade through the series of dynodes (electrode which serves as an electron multiplier through secondary emission) maintained at different potentials to result in an output signal [42].

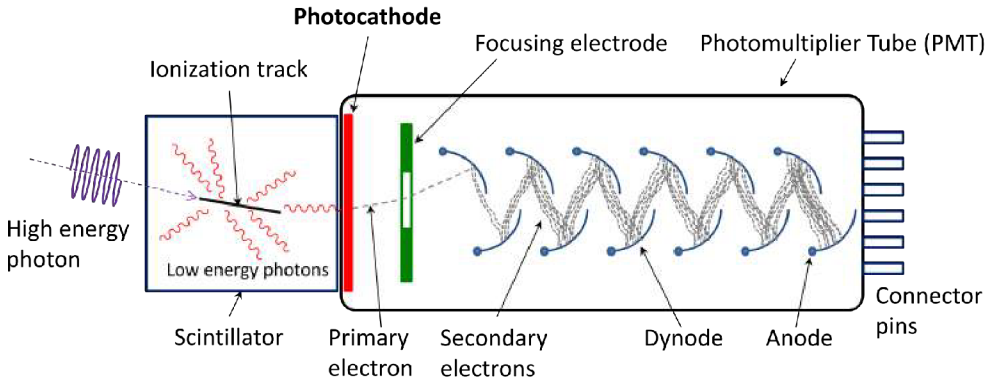

Fig. 1.6: Principle of X-ray detection by scintillation detector [40].

Scintillation detectors are then extendable to multi-array system. These detectors come in two geometries, either flat or curved. Modern µCT detectors used in industry employ flat panel geometry. A typical composition of a flat panel detector manufactured by the General Electric company is presented in Fig. 1.7.

X-ray quanta entering the detector are converted to the visible light in the CsI (caesium iodide) scintillation layer. Needle structure of the CsI then acts as an optical fibre and guides the light to the photodiode. Here the photons are absorbed, and electric charge is produced. Electric charge is then integrated and stored in the detection element until the end of detector exposure. Read-out process is then initialised by the thin-film transistor (TFT), which switches the charge to the read-out

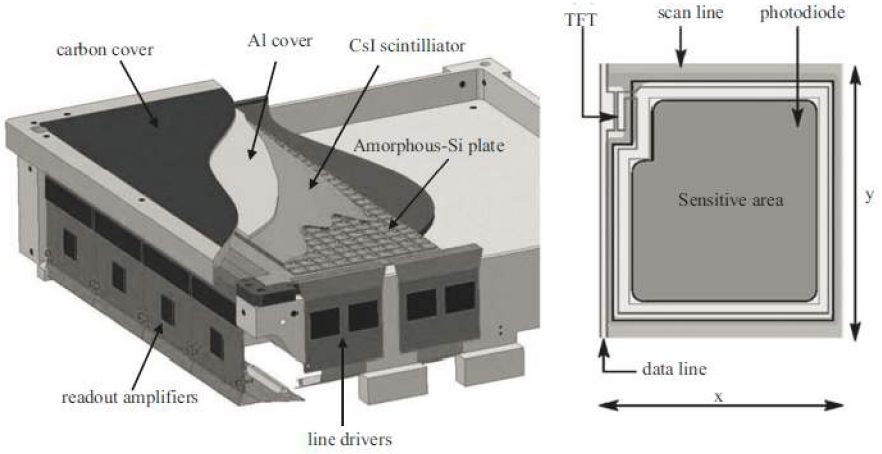

Fig. 1.7: Composition of a digital flat panel detector (courtesy General Electric CT Systems). Outer carbon layer serves as protection from mechanical damage. Al layer serves as a mirror for reflected photons [8].

electronics via the data link. There, an amplification and anologue-digital conversion are performed [6].

1.3 Reconstruction

In CT, the measurements coming from the scanner do not provide a final image, but rather the image is the result of intricate mathematical processing of the measured data. Tomography thus represents an inverse problem, where the unknown parameters of a system need to be estimated from the known reaction of the system to external signals [25].

A known reaction of the system is in case of CT represented by the intensity registered at the detector. The intensity of the X-ray reaching a position of the detector is proportional to the integral of the 2D transparency distribution of the object along the pass. Cross section images are a result of the reconstruction process of the 2D transparency distribution functions from the set of 1D functions obtained by integrals along the lines of various directions. These integrals are called projections [15].

This projection is defined as:

$$\iint_{-\infty}^{+\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - s)dxdy, \tag{1.5}$$

and represents Radon transform of the 2D distribution function $f(x,y)$, Fig. 1.8.

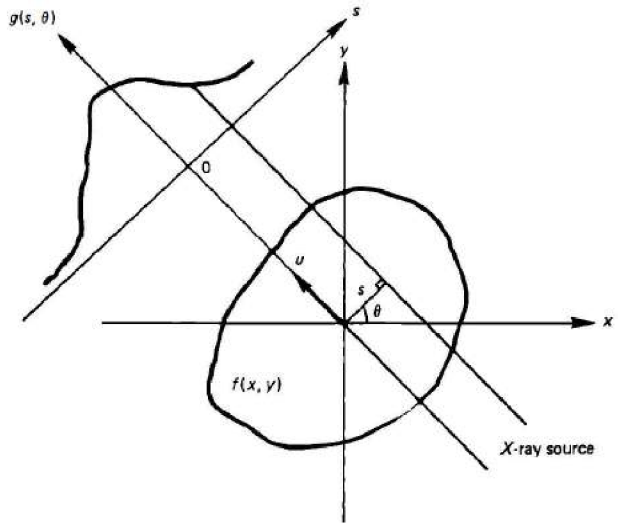

Fig. 1.8: Radon transform of a 2D object as a function of a distance from origin s and angle θ [15].

One of the most popular ways to represent projections is through sinogram space. Sinogram is formed by stacking all of the projections acquired at different angles [32]. The example of a sinogram is in Fig. 1.9. The term sinogram comes from the idea that an image can be regarded as a supersposition of single points. If we consider Radon transform of a single point, the 2D plot of all projected values as a function of an angle will be a sine curve [26].

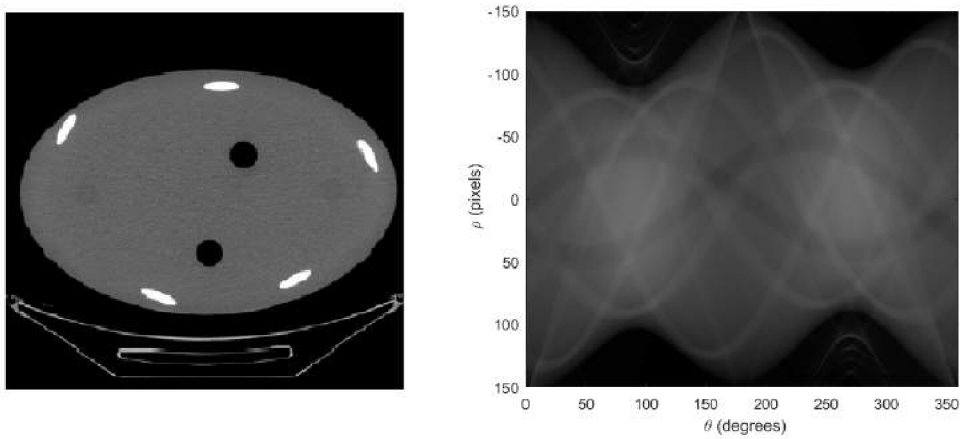

Fig. 1.9: Phantom generated in Matlab (left), its sinogram (right).

The Radon transform of a function $f(x, y)$, in case of CT $f(x, y) = \mu(x, y)$, is defined as its line integral along a line inclined at an angle θ; $0 \leq \theta < \pi$ from the $y-$ axis and at distance s; $-\infty < s < \infty$ from the origin [15].

The act of reconstruction can be described as a process of inverting the Radon transform. Whole process of reconstruction is schematically depicted in Fig. 1.10.

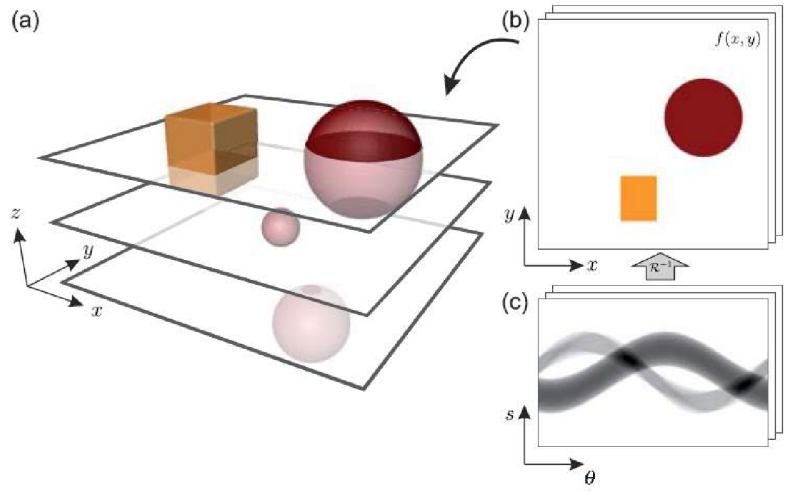

Fig. 1.10: From 2D slices to 3D object. (c) sinogram, ie., the projected profiles plotted as a function of an angle θ. These profiles can be inverted to obtain reconstructed slices,(b). These slices can be then virtually stacked to yield (a) a 3D volume [26].

In [8] it is stated that Fourier transform of a parallel projection of an object $f(x, y)$, obtained at an angle θ describes a radial line, in the Fourier space of the object, taken at the same angle. This statement is known as Central or Fourier slice theorem, see Fig. 1.11, mathematically described as

$$P_\theta(\omega) = F(\theta, \omega), \tag{1.6}$$

and sets a basis for tomographic reconstruction [8].

Although the process then may seem straightforward, there are some difficulties with the numerical implementation of the inverse 2D Fourier transform. FFT (Fast Fourier Transform) algorithm requires data on a Cartesian grid, but after applying 1D FFT, the data are filled on a polar grid. To transform data back to Cartesian grid, interpolation must be performed. This procedure is schematically described in Fig. 1.12.

Due to the problem of data interpolation in the frequency domain, alternative approaches have been developed. The most widely used of them is Filtered Back Projection (FBP). Simple back-projection of Radon transform gives a product of convolution of original function with a point spread function (PSF) $\frac{1}{r}$ which produces blurring in the image. To overcome the limitations set by conventional back-projection a filter back-projection method (FPB) has been developed [15].

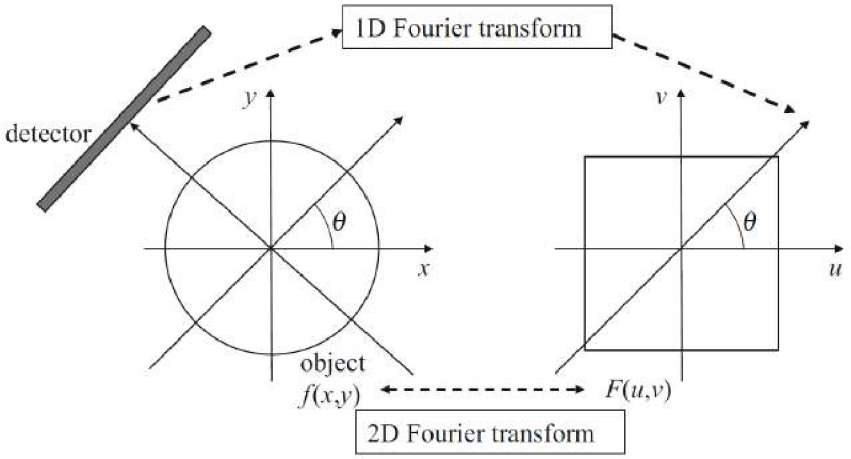

Fig. 1.11: Schematic representation of the Fourier slice theorem [8].

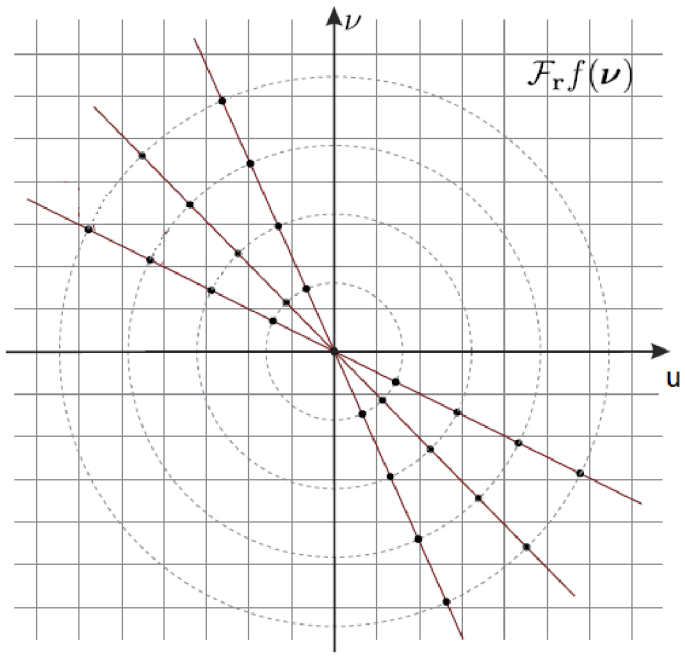

Fig. 1.12: Illustration of the Fourier slice theorem for 2D case. The 1D Fourier transform of a projection corresponds to a slice (red line) in the 2D Fourier transform of the object. If the function is given at discrete points, indicated by the black dots, a sophisticated interpolation in Fourier space is necessary to avoid artifacts in the tomographic reconstruction [26].

2 IMAGE ANALYSIS

In this chapter basics of digital image processing are covered together with the representation of the CT data. Castleman in [9] defines digital image processing as a process of subjecting numerical representation of objects to a series of operations in order to obtain a desired result. A digital image can be described as a matrix of numbers (pixels), where position of each element is identifiable using a Cartesian coordinate system [8].

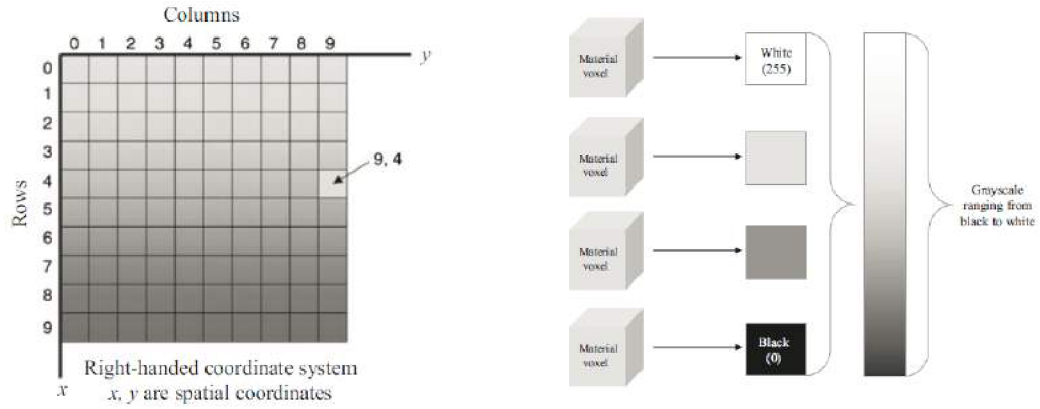

Fig. 2.1: Left: Coordinate system used to describe position of pixels in the digital image. Right: Conversion of information contained inside voxel to a numerical value. A high number represents a high intensity and low number represents a low intensity, adapted from [8].

The parameters to describe image are: matrix, pixels (picture elements), bit depth and voxels [42]. Matrix describes the size of the image. Size S of the matrix of m columns and n rows represented in bits can be calculated as:

$$S = m \times n \times 2^k, \tag{2.1}$$

where k is a bit depth of the image (number of bits per pixel). In most applications it varies from 8-bit to 32-bit.

Voxel is a 3D pixel that represents the information contained in a volume of a material. Voxel size is typically function of a detector pixel size and a magnification $\mathscr{M}$ which depends on the ratio of distances between source and a detector $D_{(SD)}$ and the distance from source to object $D_{(SO)}$ [8].

$$\text{Voxel size} = \frac{\text{pixel size}}{\mathscr{M}}, \quad \mathscr{M} = \frac{D_{SD}}{D_{SO}}. \tag{2.2}$$

2.1 CT data representation

The tomographic data are volumetric data formed up by stacking all the reconstructed greyscale images. These images contain only information about the detected intensity represented by shades of grey. Number of grey values varies with the bit depth. Due to the fact that human eye can resolve only about 40 grey levels, 8 bit images (256 greyscale values) are sufficient enough for any further image analysis. Function that summarizes grey-level distribution in the image is grey-level histogram. This function shows for each grey level, the number of pixels in the image that have that grey level [9]. Example of a histogram can be seen in Fig. 2.2.

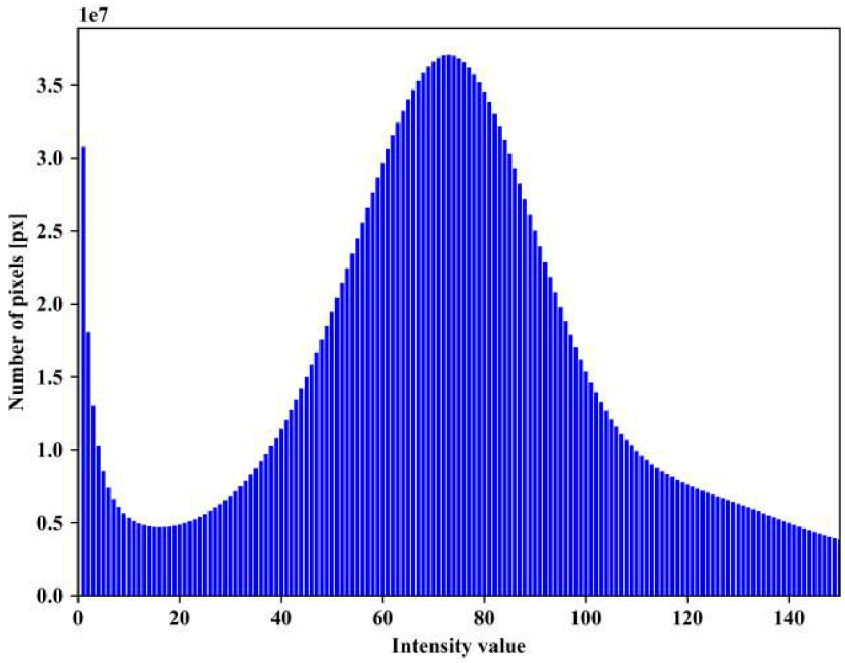

Fig. 2.2: Histogram of greyscale values for the volume of one randomly chosen sample.

2.2 Segmentation

In image processing context, ability to distinguish the objects of interest from the background is highly desired. This task can be realized through segmentation [44]. The image segmentation is a process of partitioning of digital image into several non-overlapping regions. Region is then defined as a connected set of pixels [9]. During the years, many segmentation techniques have been developed. Some rely on

automatic or semi-automatic techniques and some are purely manual. Here we will try to describe some of the most used techniques. Although it may seem that these techniques appear here separately, in many practical cases a combination of several approaches is needed to obtain satisfying results.

2.2.1 Tresholding

One of the simplest segmentation techniques is tresholding. This procedure attempts to determine intensity value, called treshold, which separates the desired regions. The segmentation is then achieved by grouping all pixels with intensities greater than the threshold into one class and all other pixels into another class [37]. There are two types of tresholding methods. Global tresholding and adaptive tresholding. Global tresholding is a method which allows to select only one level of treshold, based on which the segmentation is performed. This could be suitable for cases where image consists of one structure and a background. In practical cases this is not true and adaptive tresholding must be used. Adaptive tresholding consists in analyzing the image intensities around each pixel and selecting an individual threshold for each pixel, taking in consideration the degree of the intensity values in its local neighborhood [44].

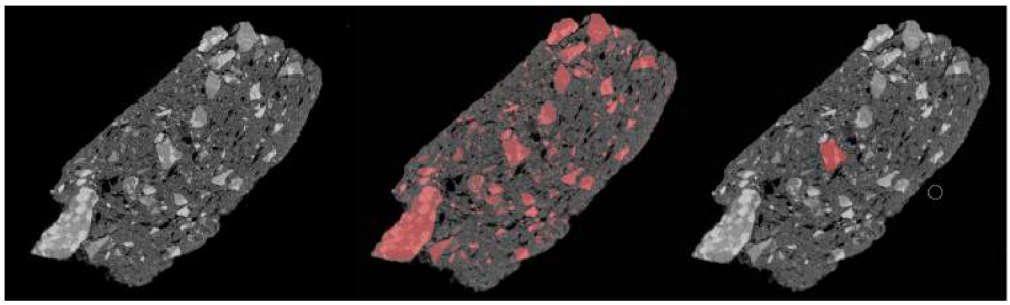

Fig. 2.3: Segmentation using different techniques. Left: Original image. Middle: Segmented image using global tresholding method, red-coloured areas correspond to inclusions captured by one specific treshold. Right: Manual segmentation using brushing tool.

2.2.2 Region growing

Region growing algorithm is a simple pixel-based image segmentation method, which involves the selection of pixels (the seeds), and then growing regions around these seeds, using a homogeneity criteria. If the joining pixels have similar image features as the seed, they are integrated into that region.

An advantage is that region growing algorithms are fast and can perform accurate segmentations of regions that have the same features but are spatially separated. However, they are sensitive to noises and therefore may produce undesired segments, regions with holes or disconnected regions [44]. Example of segmentation using region growing method is in Fig. 2.4.

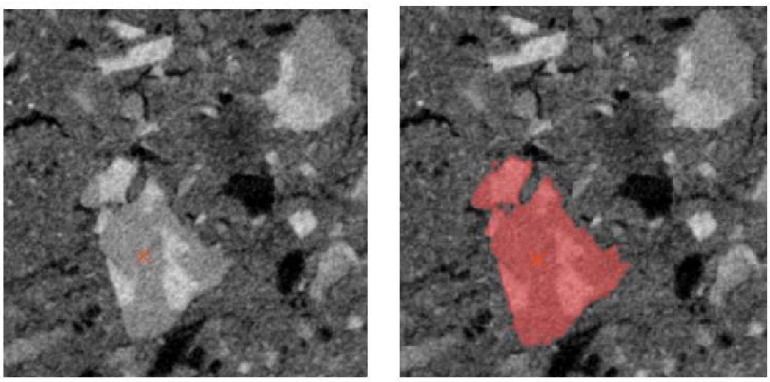

Fig. 2.4: Segmentation using the magic wand tool. This tool allows user to select the position of the seed voxel, here indicated by the red cross. After setting suitable value of treshold it connects respected areas captured by this treshold.

2.2.3 Machine learning

Since 1990, artificial neural networks (ANN) have come to be used as a different approach for image segmentation. Their properties, such as graceful degradation in the presence of noise, their ability to be used in real-time applications and the ease of implementing them with VLSI (very large scale integration) processors, led to a booming of ANN-based methods for segmentation [1]. Machine learning algorithms can be divided into two basic groups, namely supervised and unsupervised machine learning algorithms. Supervised learning techniques learn mapping from input data to output (labels), during a procedure called training. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object and the desired output value. A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples [11]. These techniques have shown great promise in image analysis [46]. Supervised techniques as name suggests require expert human input for segmentation. Usually this means that human experts are carefully selecting training images and manually segmenting them into sub-regions. Each region is assigned with a label and the proposed architecture is trained using the selected images as training

data. The method is then able to segment similar images [1]. Procedure depicted in Fig. 2.5.

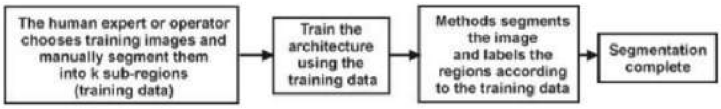

Fig. 2.5: Typical workflow for supervised segmentation methods, while using neural networks.

Unsupervised methods or clustering processes are semi or fully automatic. Big disadvantage of supervised algorithms is that they require large amounts of manually annotated data. Thus, it is challenging for these methods to cope with the growing amount of images. An alternative approaches have been proposed for example in [34], [27].

3 PRINCIPAL COMPONENT ANALYSIS

The starting point for all multivariate analysis methods is a data matrix M. This matrix contains all the measured data with columns corresponding to variables and rows assigned to observations. When looked more formally, it turns out that PCA is a method that decomposes data matrix M into a product of two new matrices, the scores matrix T and loadings matrix P [38]. The loadings can be understood as the weights for each original variable when calculating the principal component (PC), while scores represent the projection of observations from original space onto PC space. The original data matrix M can be represented as:

$$M = TP^T + E, \tag{3.1}$$

where E is the error matrix compensating the deviation of the model data to the original data [13]. The minimisation of the error matrix (also known as the reconstruction error) is another approach to derive PCA and was proposed by Carl Pearson in 1901 and can be found in [35].

3.1 Derivation

In this chapter, the main emphasis is given to the mathematical derivation of PCA method. The whole process of derivation of PCA, including generalized form of equations can be found in [33].

Suppose that X is a vector of m random variables and that the variances of the m random variables and the structure of the covariances or correlations between m variables are of interest. For the purpose of simplicity the derivation of PCA was carried out just for one PC. Lets suppose that we can form a new latent variable y_1 as a linear combination of original variables.

$$y_1 = \sum_{j=1}^{m} V_{1,j} X_{c,j} = V_1^T X_c, \tag{3.2}$$

where $X_c = (x_1 - \mu_1, x_2 - \mu_2, \cdots, x_m - \mu_m)^T$ is a column vector of original centred variables, then variance is from the definition

$$\mathrm{Var}(y_1) = \mathrm{Var}(V_1^T X_c) = E[(V_1^T X_c)(V_1^T X_c)^T] = V_1^T E(X_c X_c^T) V_1 = V_1^T C V_1, \tag{3.3}$$

C is known as the covariance matrix. From this, it can be seen that the value of $\mathrm{Var}(y_1)$ is dependent on the value of vector V_1, so it is clear that, as it stands, the maximum will not be achieved for finite V_1 so a normalization constraint must be imposed. For the purpose of derivation it is convenient to introduce a normalization constraint where $V_1^T V_1 = 1$. [18]. In the next step, we would like to maximise

the entity $\mathrm{Var}(y_1)$. For maximisation or minimisation of a function with respect to constraints, the natural tool is a method of Lagrange multipliers.

$$G_1 = V_1^T C V_1 - l_1(V_1^T V_1 - 1) = \sum_{i=1}^{m}\sum_{j=1}^{m} V_{1,i} C_{i,j} V_{1,j} - l_1(\sum_{i=1}^{m} V_{1,i}^2 - 1), \tag{3.4}$$

where l_1 is a Lagrange multiplier. Then a vector V_1 maximising G_1 must fulfil condition

$$\frac{\partial G_1}{\partial V_1} = 2CV_1 - 2l_1V_1 = \underline{(C - l_1 I)V_1 = 0,} \tag{3.5}$$

where I stands for indentity matrix. Because V_1 is non-zero, then $C - l_1 I$ must be singular. This could be satisfied only when $\det(C - l_1 I) = 0$, this gives in general m-th order polynomial equation with roots $\lambda_1 > \lambda_2 > \cdots > \lambda_m$. If we multiply the underlined part of eq. 3.5 by vector V_1^T, after a short rearrangement, we get

$$V^T C V_1 = l_1 V_1^T V_1 = l_1. \tag{3.6}$$

We have then arrived at a conclusion which states that variance of a first PC is equal to l_1 and will be maximized for $l_1 = \lambda_1$. With this, the derivation of the first PC is complete. For the derivation of other PCs, new constraints must be applied. These constraints are here to satisfy the assumption of orthonormality of PCs. This condition could be mathematically described as $V_i^T V_j = \delta_{i,j}$. General derivation of all other PCs can be seen in [33].

3.2 Scaling and centring

An integral part of the data pre-processing is scaling and centring. In some cases, variables have different scales. The scale of variables usually corresponds to the unit in which they are measured. The goal is then to have a PCA model that would be independent on the scale of variables. The most common type of scaling is scaling to unit variance. This type of scaling is also useful in situations when no prior information about variables is available. Because the scaling procedure affects the covariation matrix, it must be done with caution. Mean-centring is the second part of a usual pre-processing procedure. With mean-centring, the average value of each variable is calculated and then subtracted from the data. This procedure leads to repositioning of the origin of the coordinate system of the original data. Graphical representation of scaling and mean-centring procedure can be seen in Fig. 3.1. The influence of data standardization in case of chemical data has been widely studied in [39].

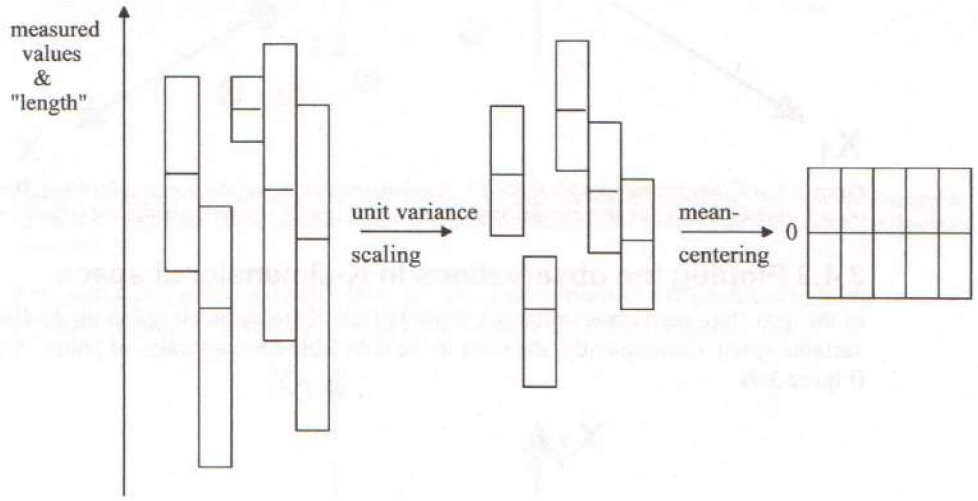

Fig. 3.1: Mean centring and unit variance scaling of all variables. After this procedure all variables will have equal 'length' and mean value zero [13].

3.3 Number of used principal components

Visualization of data sets is crucial in any field. If a given dataset of m variables is being plotted in m-dimensional space, then ability to visualize anything ends with $m = 3$. In many practical applications, where m normally exceeds 10, finding a way to visualize data in q dimensional subspace ($q < m$), without the loss of generic information is desired. The question that stands out now is, how to determine the dimension q of the subspace. When the correct dimensionality of the subspace is not retained we risk either loss of important information (underestimation) or we include noise in the subsequent analysis (overestimation) [36]. During recent years, various methods have been developed. Some were based on a purely statistical and mathematical description, for example Monte Carlo simulations [21], Cross Validation [13], while others were based on empirically derived methods. In this thesis, the number of PCs has been selected by a widely used method based on the cumulative percentage of total variation. The number of selected PCs is then chosen to satisfy the condition when the cumulative percentage exceeds some selected percentage level.

The cumulative percentage level t_q retained by q number of PCs can be defined as:

$$t_q = 100 \frac{\sum_{k=1}^{q} l_k}{\sum_{k=1}^{m} l_k}. \tag{3.7}$$

This level should fall somewhere between 70 % to 90 %. Another way how to look at the variation, is through a plot of eigenvalue vs. PC number known as a scree plot, named after the shape of rocky detritus at the foot of a hill [18]. The Cattell's scree plot was developed by R. B. Cattell in 1966. It is based on the assumption

that relevant information is larger than random noise and that the magnitude of the variation of random noise seems to level off quite linearly with the number of components [7].

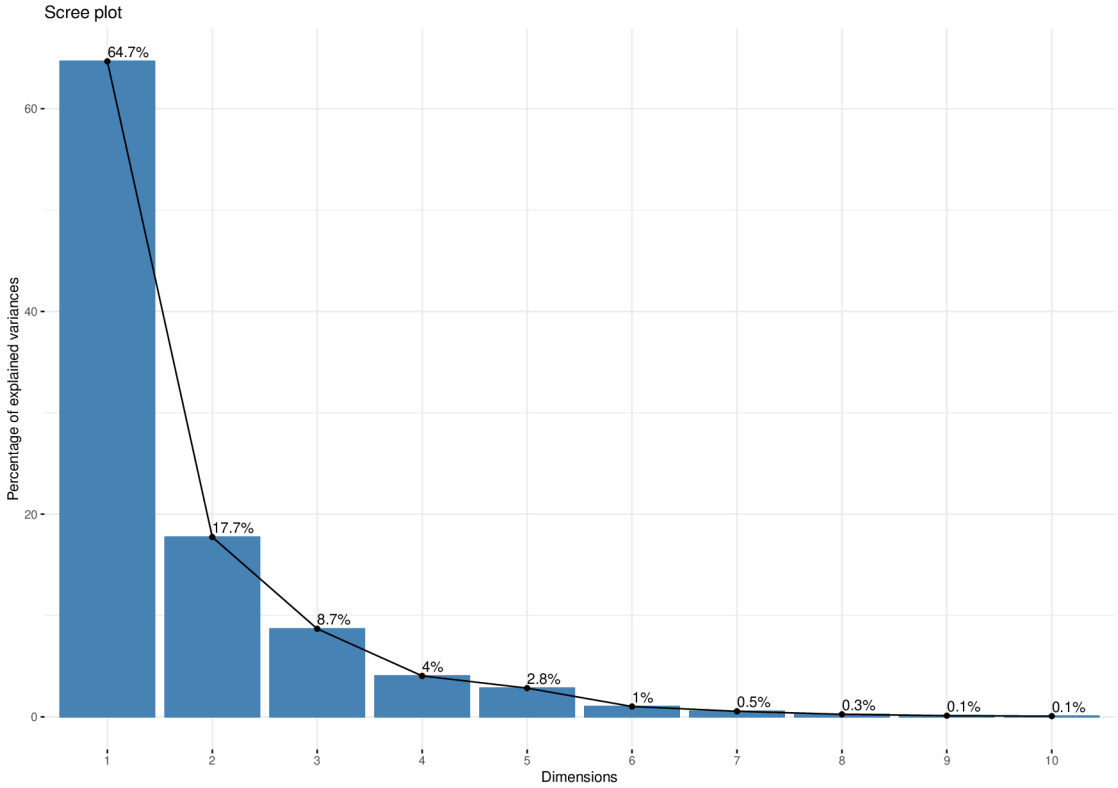

Fig. 3.2: Example of a scree graph. It can be seen that 82.4 % is retained by using only first two PCs.

Cattell (1966) argues that the q should be chosen when the scree graph goes from 'steep' to 'shallow', a point beyond which the scree graph first goes mostly linear.

A more realistic cut off for the eigenvalues is obtained with the so called broken stick rule. In this model a line is added to the scree plot. This line is calculated assuming that random data will follow a so-called broken stick distribution. The broken stick distribution hypothesizes how random variation will partition and uses the analogy of how the lengths of pieces of a stick will be distributed when broken at random places into j pieces [7].

3.4 Geometrical interpretation of the PCA

First PC can be understood as the line in the m-dimensional space that best aproximates the data in the least square sense [13]. The new coordinates (scores) in the component space are calculated as the projections of observations onto this line. Similar procedure could be done for the second PC. This component can be also depicted as a line in the m-dimensional variable space, which passes through the average point while being also orthogonal to the first PC. These two PCs then form together a plane that serves as a window into the m-dimensional variable space. Projection of the original observation onto this plane can be depicted as in Fig.3.3.

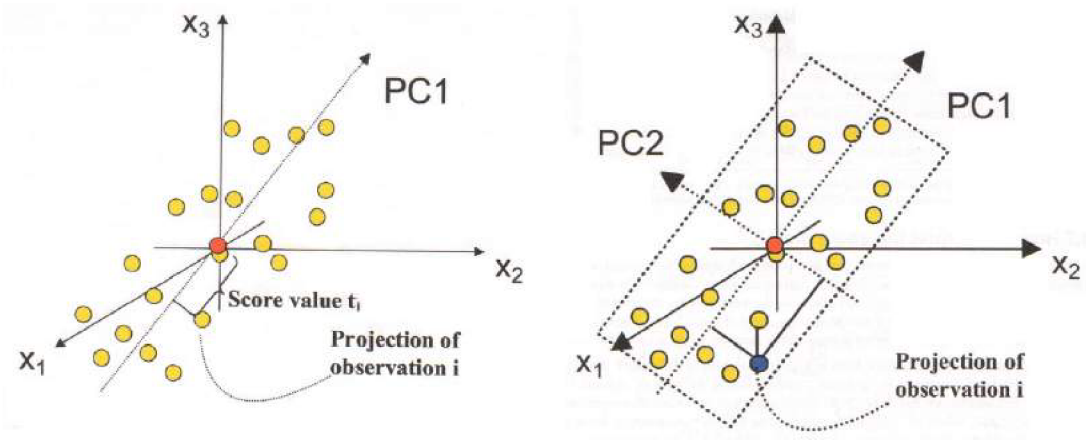

Fig. 3.3: Left: The PC1 is the line which best captures the variability of the point swarm. It represents the maximum variance direction in the data. Projection of observation onto this line is known as score. Right: The PC2 then captures the second largest source if variation in the data, while also being orthogonal to PC1. These two components together form a plane. Each observation can be projected onto this plane, giving score for each [13].

Then it can be also stated that the PC loadings express the orientation of the model plane in the m-dimensional space. The relation between original variables and PCs is expressed by the cosine of the angles $\alpha_{1,2,3}$. This, for first PC is depicted in Fig. 3.4.

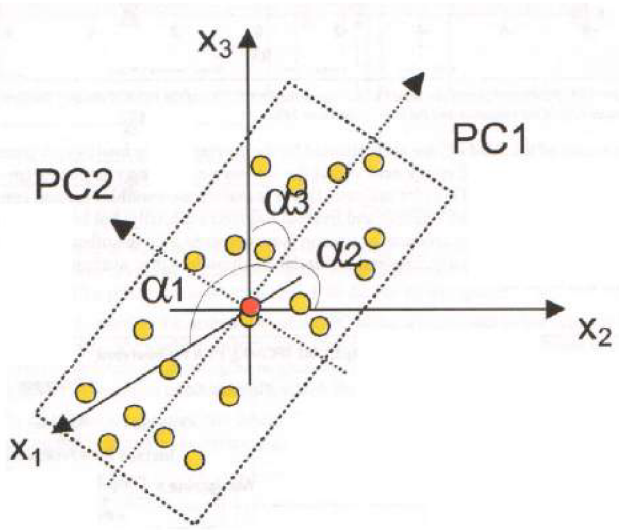

Fig. 3.4: The PC loadings uncover how the PCA model plane is inserted in the variable space [13].

3.5 Outliers detection

Outliers are samples that are somehow disturbing or unusual. Outliers detection is about identifying and handling such observations. Several researchers have explained the concepts of applying PCA as a robust method in detecting these outliers in multivariate settings [41].

Detection of outliers is usually done after a score plot inspection. Removal of outliers is crucial because outliers have a tendency to rotate the component space towards themselves [13]. The effect of adding or removing an outlying observation on direction of the PC can be seen in Fig. 3.5.

Thus, in order to produce a reliable cluster of these data, it is imperative to ensure that the dataset is free from outliers. Due to this, the cluster analysis usually involves a pre-processing step to identify the existence of outliers within a data set [41].

In some situations, the removal of the outliers is not possible, for example when the outlying observation is an integral part of the study.

3.6 Cluster analysis

As mentioned before, PCA can also be only a starting point for another type of analysis. One of the results of the PCA is a score plot in which observations that have similar properties tend to group. In some cases, when categorical variable, for example origin of the samples, is preliminarily known, PCA could be sufficient

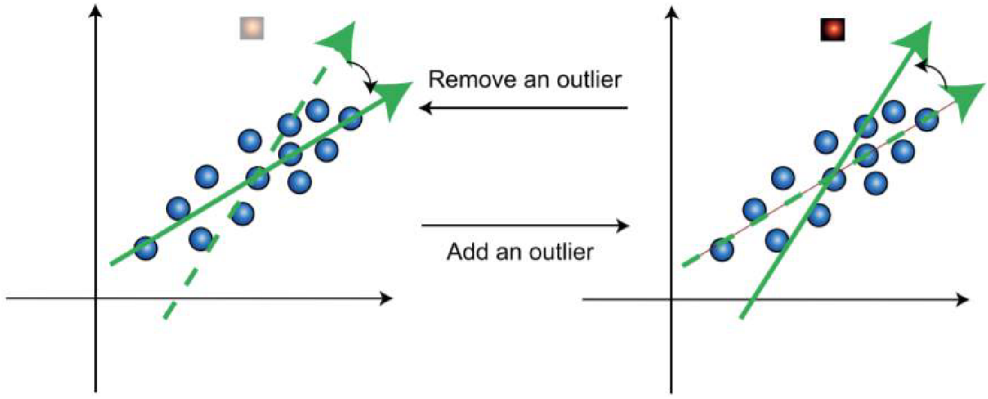

Fig. 3.5: . The effects of adding/removing an outlier on the PC direction. Here clustered blue circles represent normal data instances, the red square denotes an outlier, and the green arrow is the dominant PC direction [28].

enough. Unfortunately, in most of the situations, this is not true, and we have no preliminary information about the sample groups. In these cases and also when the distinction between newly formed groups is not clear, cluster analysis can be performed to improve the cluster separation. Cluster analysis then tries to find a natural grouping of a data set which would minimise the variation within the clusters and maximise the variation between the clusters [13].

3.6.1 K-means clustering

One of the oldest algorithms used for cluster analysis is the K-means algorithm. In its simplest form, this algorithm tries to find K non-overlapping clusters. These clusters are represented by their centroids. A centroid is calculated as a mean value of the data points within the cluster. Cluster analysis then follows this procedure. In the first step, a number of K clusters is selected by the user. Then each data point is assigned to its closest centroid. This collection of points then forms a cluster. The addition of a new point into a cluster leads to a recalculation of the centroid. This process is then repeated until no point changes its clusters [50]. In the case of euclidean distance this could be expressed by term [12].

$$J_k = \arg\min \sum_{k=1}^{K} \sum_{i \in C_k} (M_i - m_k)^2, \tag{3.8}$$

where $(M_1, \cdots, M_n) = M$ is the data matrix and $m_k = \sum_{i \in C_k} x_i / n_k$ is the centroid of the cluster C_k and n_k is the number of points in C_k.

3.6.2 K-medoids clustering

Due to the fact that mean is easily influenced by extreme values, another type of clustering algorithm has been designed. Medoid is the most centrally located object of the cluster, with a minimum sum of distances to other points [16]. Thus the partitioning method can still be performed based on the principle of minimizing the sum of the dissimilarities between each object and its corresponding reference point [49]. Comparison between using K-mean clustering and K-medoids clustering can be found in Fig. 3.6.

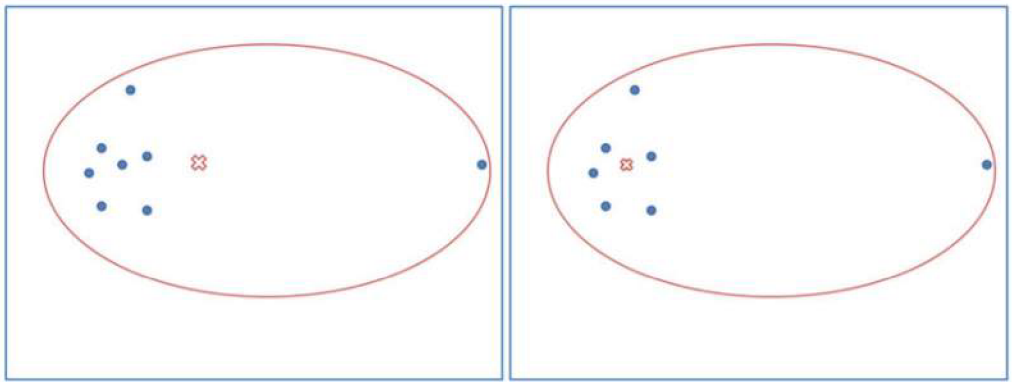

Fig. 3.6: Mean vs medoid in 2D space; (a) Mean centroid, (b) Median centroid. The red cross represents the center found by mean or medoid calculation, [17].

3.6.3 Fuzzy clustering

A big disadvantage of previous algorithms is their "hard clustering"nature. These algorithms force each data point to belong to a specific group with the same pattern. This procedure in some cases distorts the original data and could lead to confusing results. For this purpose, fuzzy clustering algorithms have been proposed to fight the hard clustering nature of previous algorithms [30]. A fuzzy set is described as a group of anything that cannot be precisely defined. Fuzzy clustering is a sophisticated method for handling data which are unlabeled, contain outliers or include unusual patterns, [24].

4 EXPERIMENT

4.1 Origin of the samples

In the present study 18 samples have been analyzed using a multi-analytical approach in order to study their microstructure. The information about composition of different samples can then help us test the hypothesis about different origin of the samples.

The acquired samples were a product of an extensive investigation carried out between 2015 and 2018 at the UNESCO site of Al Khutm (Bat, Oman), Fig 4.1.

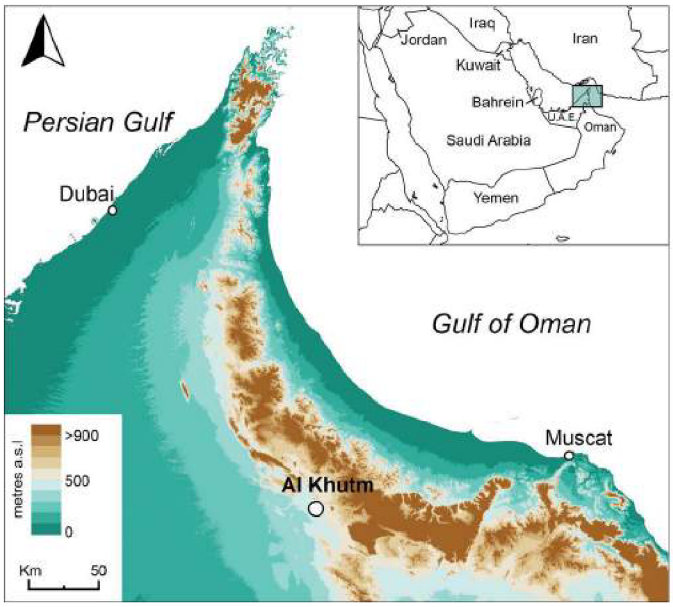

Fig. 4.1: Position of the UNESCO site of Al Khutm.

This work has allowed us to uncover an important and well preserved Bronze Age complex [10]. The main achievement of the excavation is the exposure of a 4 m high circular tower, an imposing 2 m high perimeter wall and several structures located inside the tower, including several entrances, a deep well found in the central corridor and two rows of rooms on both sides, Fig. 4.2.

Fig. 4.2: The protohistoric tower at the UNESCO site of Al Khutm after the last excavation campaign in 2018.

Fig. 4.3: An example of selected pottery samples.

4.2 Measurement

The samples (selection of them in Fig. 4.3) were imaged by X-ray microCT at the Multidisciplinary Laboratory of the Abdus Salam International Centre for Theoretical Physics (Trieste, Italy), using a system specifically designed for the study of archaeological and palaeoanthropological materials [5].

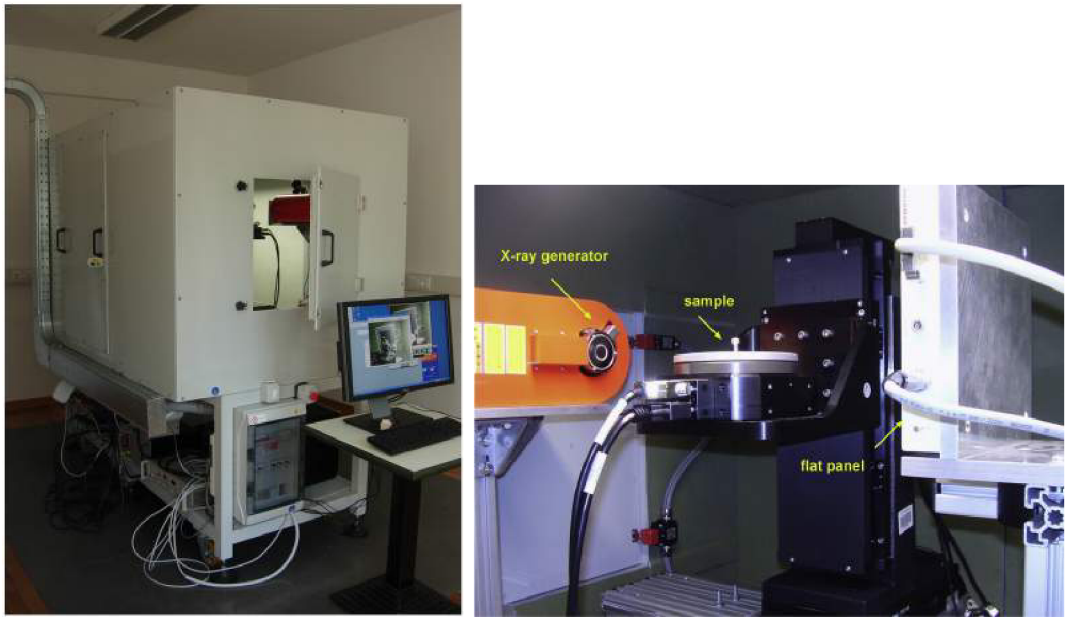

Fig. 4.4: (a) CT station at ICTP, with a lead-shielded cabinet used to operate the system. (b) Inner part of the CT system with highlighted particular components. Both stage system and detector are mounted on a flexible mechanical set-up [48].

The microCT acquisitions of specimens were carried using a sealed X-ray source (Hamamatsu L8121-03) at a voltage of 110 kV, at a current of 90 µA and with a focal spot size of 5 µm. The X-ray beam was filtered by a 0.1 mm thick copper absorber. A set of 1440 projections of the artefacts was recorded over a total scan angle of 360° by a flat panel detector (Hamamatsu C7942SK-25; pixel size of 50 µm). The resulting microCT slices were reconstructed using the commercial software Digi XCT (Digisens) in 32-bit format and obtaining an isotropic voxel size 12 µm.

4.3 Metrics extraction

The resulting volumetric data have then been segmented using Avizo. Avizo is a very useful software in terms of segmentation. It offers a large set of segmentation

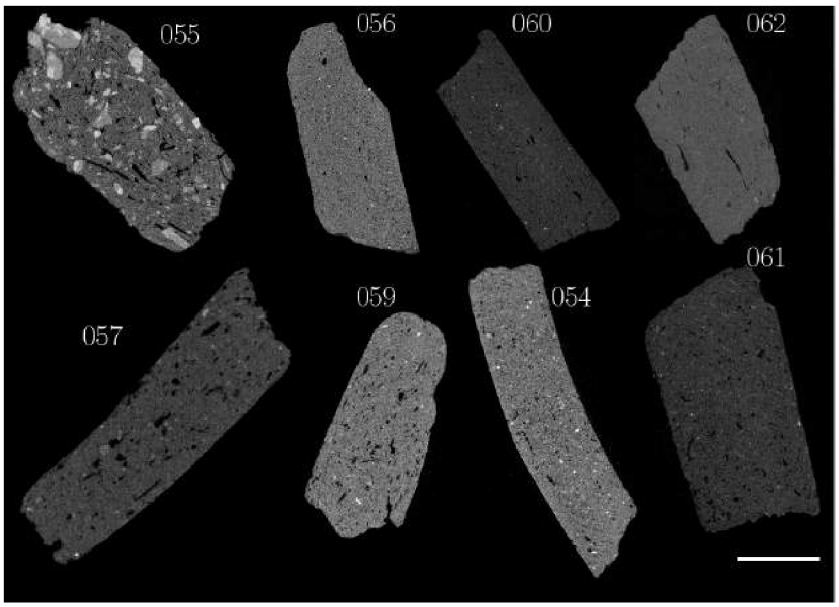

Fig. 4.5: CT cross sections of selected samples. Scale bar corresponds to 0.5 cm.

tools, ranging from purely manual to fully automatic: brush (painting), lasso (contouring), magic wand (region growing), thresholding, intelligent scissors, contour interpolation, extrapolation, etc. [47]. The segmentation of selected samples was done mainly using region growing techniques with a combination of different manual techniques to achieve satisfying results. After the segmentation process, it is important to find some parameters based on which the samples could be compared. There are certain possibilities offered by the commercial software which allow extraction of various parameters that could be used for samples classification. One of the software that gives the opportunity to extract different metrics from the segmented image is Avizo.

There are many parameters that could be exported from the Avizo. For example information about width, length and area of each single inclusion. From this, ratio between the total areas of different pottery components can be calculated. Information about the ratio between the total areas of different pottery components can give interesting insights on how the paste of the vessels was produced. Anyway, samples with a similar lithic inclusion/clay ratio could correspond to quite different pastes: the same ratio could be obtained considering a sample with very rare and big lithic grains or considering a vessel characterised by very abundant and small lithic inclusions. In order to overcome this bias, we have calculated, using Avizo v.8 software, the number, the area, the maximum length and the maximum width of every single lithic inclusion within the three virtual sections selected for all samples [5].

Selection of 2D segmentantion instead of 3D made the whole process of segmen-

tation more time efficient, while the results were comparable with those obtained through 3D process [4]. The process of segmentation is shown in Fig. 4.6.

After that, we have summed up the results for each sample, putting together the data obtained from the segmentation and analysis of the three virtual sections.

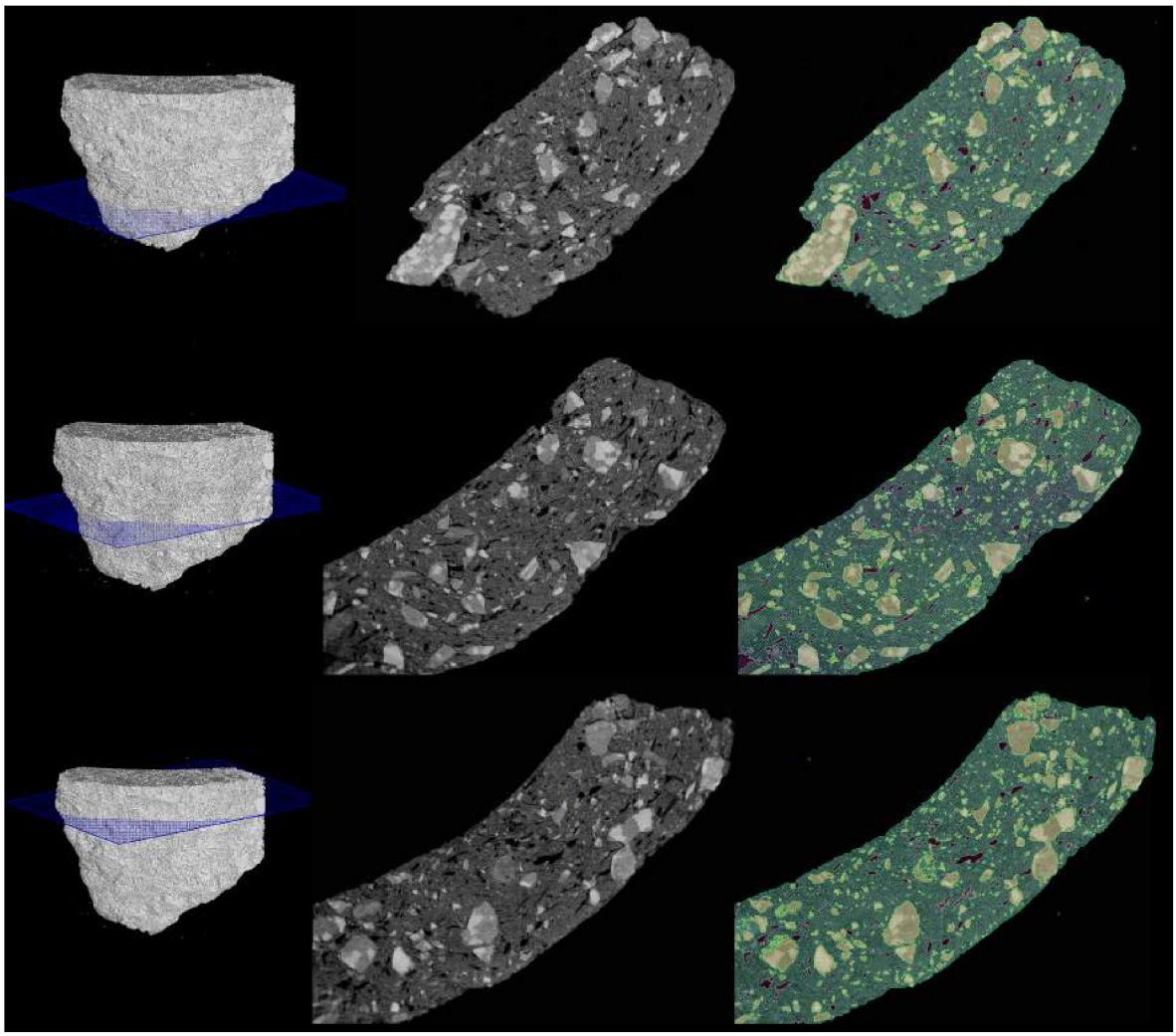

Fig. 4.6: Left: 3D model of a selected sample; Middle: Virtual cross-section, position of the slice is indicated by the blue plane in the 3D model; Right: Segmentation of different pottery components of the sample. Green-clay, yellow-inclusions, purple-voids.

4.4 Data preparation

To manage the resulting huge amount of data, we have divided our datasets (values for area, maximum length and width of all inclusions) into four intervals using Matlab R2016b. The size of the intervals has been selected by taking into account the statistical distribution of each variable. This procedure was important to have the best representation of dimensional parameters describing the lithic inclusions within the paste, allowing a good separation between samples showing lithic grains of different sizes [5].

As far as the area is concerned, the quartile values are 325 µm^2, 1137 µm^2 and 4714 µm^2 and the intervals have been set as follows: very small inclusions, 0 – 325 µm^2; small inclusions, 325 – 1137 µm^2; medium inclusions, 1137 – 4714 µm^2; and big inclusions, bigger than 4714 µm^2. As far as the length is concerned, the quartile values are 28 µm, 53 µm and 108 µm and the intervals have been set as follows: very small, 0 - 28 µm; small, 28 - 53 µm; medium, 53 – 108 µm; and big, bigger than 108 µm. As far as the width is concerned, the quartile values are 12 µm, 38 µm and 66 µm and the intervals have been selected as follows: very small, 0 – 12 µm; small, 12 – 38 µm; medium, 38 – 66 µm; and big, bigger than 66 µm. The result of this procedure is a table showing the number of inclusions falling into its respective interval. To overcome situations when there were too big differeces in number of inclusions within sample, the percentages of lithic inclusions falling within the defined size intervals have then been calculated. The result is summarized in Tab.4.1. Before the PCA the data were centred and scaled. The aim was to create a model that would be size and unit independent.

Tab. 4.1: Table showing the percentage of lithic inclusions falling into its respective interval. Considering the area, length and width of each single lithic inclusion. As additional parameter, inclusion clay ratio has been calculated for each sample.

	Area				Length				Width				Material
Name	**Very small_A**	**Small_A**	**Medium_A**	**Big_A**	**Very small_L**	**Small_L**	**Medium_L**	**Big_L**	**Very small_W**	**Small_W**	**Medium_W**	**Big_W**	**Incl_Clay**
049	5,10%	17,05%	32,26%	45,59%	7,00%	19,98%	27,16%	45,86%	9,29%	24,78%	21,08%	44,85%	41,55%
050	8,33%	11,81%	40,28%	39,58%	10,42%	17,36%	40,97%	31,25%	10,42%	26,39%	24,31%	38,89%	0,70%
051	9,49%	23,36%	42,34%	24,82%	13,87%	23,36%	41,61%	21,17%	14,23%	32,12%	29,56%	24,09%	0,18%
052	6,48%	17,82%	43,29%	32,41%	7,41%	23,15%	37,96%	31,48%	11,34%	27,31%	28,94%	32,41%	0,63%
053	4,38%	9,16%	31,87%	54,58%	4,78%	13,94%	30,68%	50,60%	5,58%	15,54%	25,10%	53,78%	0,84%
054	7,23%	11,85%	35,94%	44,98%	7,83%	14,86%	35,54%	41,77%	8,63%	22,49%	23,90%	44,98%	1,68%
055	6,37%	12,35%	31,85%	49,43%	8,30%	15,75%	28,00%	47,95%	9,14%	20,10%	22,72%	48,05%	24,76%
056	39,31%	45,03%	13,57%	2,09%	50,22%	32,23%	14,69%	2,86%	60,93%	31,32%	5,36%	2,39%	0,80%
057	7,24%	22,39%	38,05%	32,32%	9,26%	25,59%	32,66%	32,49%	12,63%	30,13%	25,08%	32,15%	1,37%
058	6,73%	13,76%	38,99%	40,52%	8,56%	17,74%	35,93%	37,77%	10,09%	23,09%	27,98%	38,84%	0,71%
059	41,99%	26,10%	24,28%	7,63%	45,70%	27,08%	20,36%	6,86%	50,87%	27,99%	13,72%	7,42%	0,73%
060	1,90%	5,70%	43,04%	49,37%	2,53%	8,23%	47,47%	41,77%	3,16%	13,92%	29,75%	53,16%	0,44%
061	35,81%	20,17%	26,53%	17,49%	32,72%	25,33%	24,21%	17,73%	35,73%	31,30%	14,99%	17,98%	2,93%
062	63,64%	19,48%	11,76%	5,12%	54,31%	26,73%	13,55%	5,41%	60,06%	28,13%	6,44%	5,37%	0,83%
063	6,72%	8,40%	36,97%	47,90%	6,72%	12,61%	34,45%	46,22%	6,72%	19,33%	26,89%	47,06%	0,41%
064	13,87%	20,80%	41,18%	24,16%	10,92%	30,46%	34,66%	23,95%	14,92%	34,87%	27,52%	22,69%	0,45%
065	5,59%	19,08%	36,18%	39,14%	5,59%	25,66%	34,21%	34,54%	9,87%	24,67%	26,32%	39,14%	0,52%
066	3,93%	9,09%	22,36%	64,62%	3,93%	12,53%	22,36%	61,18%	6,39%	13,51%	16,95%	63,14%	1,91%

5 DATA PROCESSING

5.1 Principal component analysis

In the next step, PCA has been performed using the programming language R [22]. The input dataset for PCA was table 4.1. Before the PCA, the data were centred and scaled (column-wise scaling to unit variance). The first outcome of the PCA method is a scree plot, showing the statistical significance of each PC.

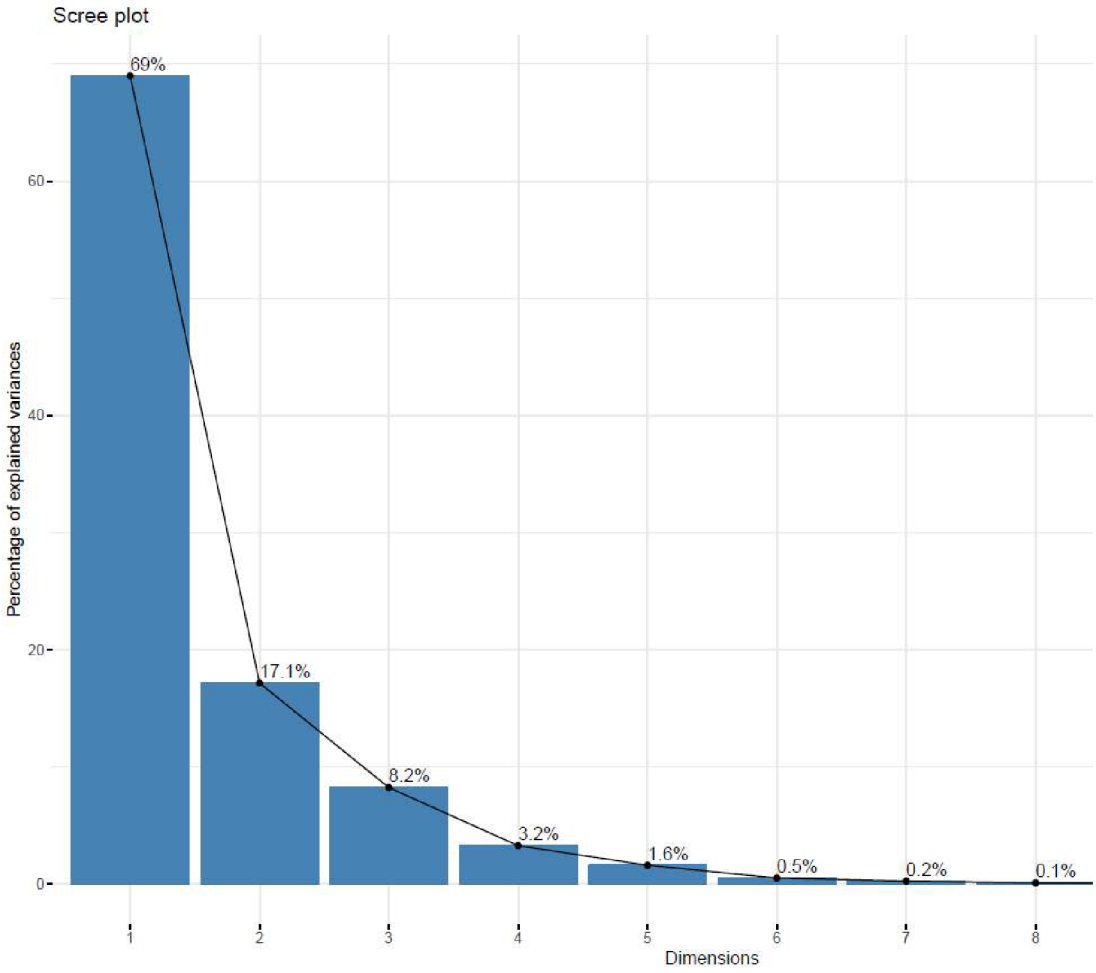

Fig. 5.1: Scree plot showing the variance captured by corresponding PC.

If the first two or three PCs capture the most of the variation present in the data, the scree plot is a steep curve that bends quickly and flattens out.

From the scree plot inspection, we can clearly see that 86.1% of variance is retained by using only first two PCs. Due to this, we have decided to use only first two PCs for the next calculations.

Fig 5.2, called loadings plot, shows how strongly each variable influences a PC. Their projected values on each PC then show how they are important in describing the differences between investigated samples. Another thing that can be deduced from the loadings plot is the relationship between variables and the quality of their representation. Variables positively correlated are grouped together, with small angle between them. While those negatively correlated are located in opposite quadrants. Variables which are not correlated are perpendicular to each other. The quality of representation can be expressed by a value of square cosine (cos2). The distance between the end points of variables measures the quality of the variables. The best representation of a variable is given when its end lies on the correlation circle, a condition corresponding to the maximum square cosine value, that is 1.

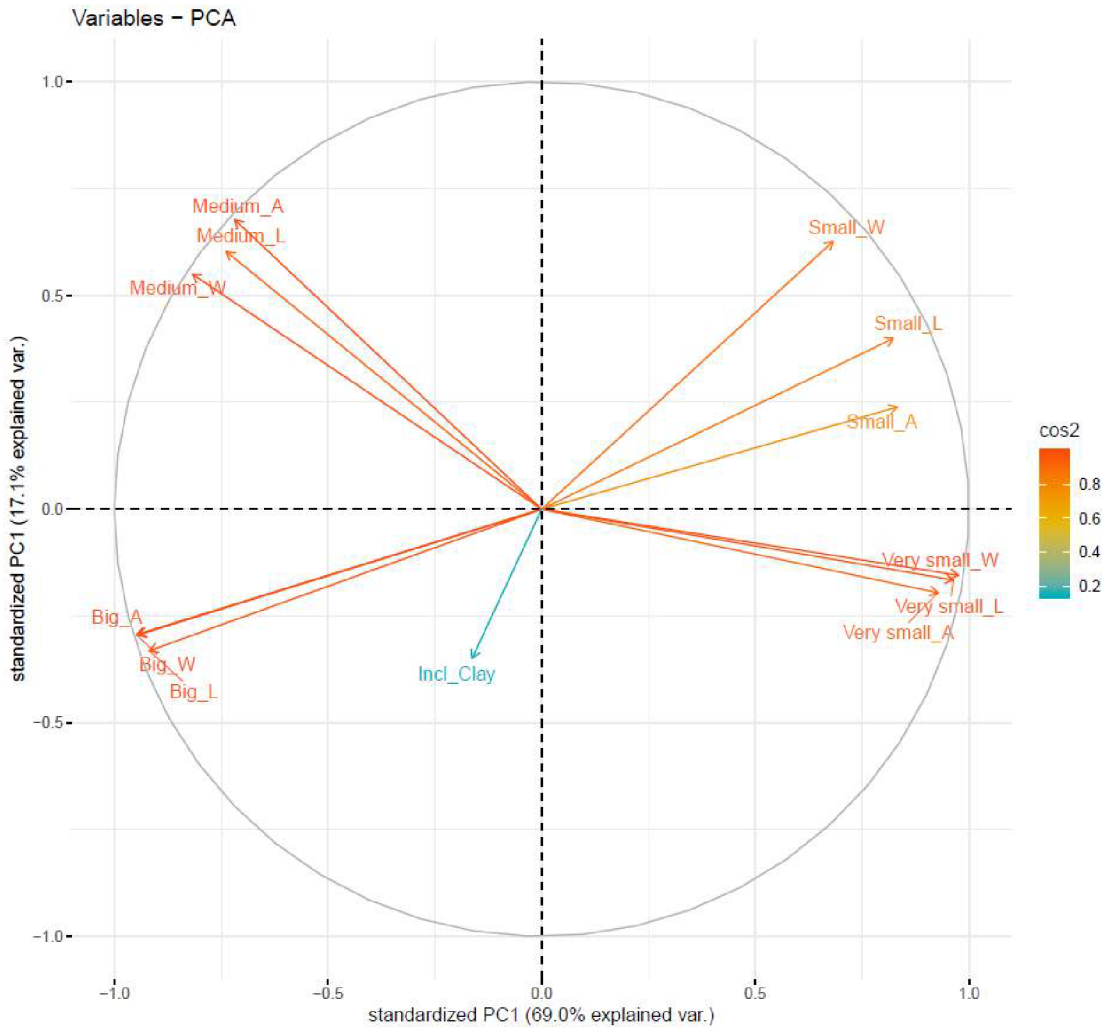

Fig. 5.2: Loadings plot showing the correlation of variables. Here can be seen that by selecting four intervals we were able to divide our variables to four quadrants in the loadings plot. A, area; L, length; W, width; Incl_Clay, lithic inclusion/clay ratio.

A compact way to visualize the results of PCA is through biplot. In the bivariate plot (scores plot) in Fig. 5.3, we can observe the position of the samples with respect to PC1 and PC2 and the variables. When an observation falls close to a given variable, it has a high value of such a variable, while its value is low it is located on the opposite side of the same variable.

Basically it follows common rule, when the points are far away from each other in the scores plot, they are different. If points are close together, they are more similar.

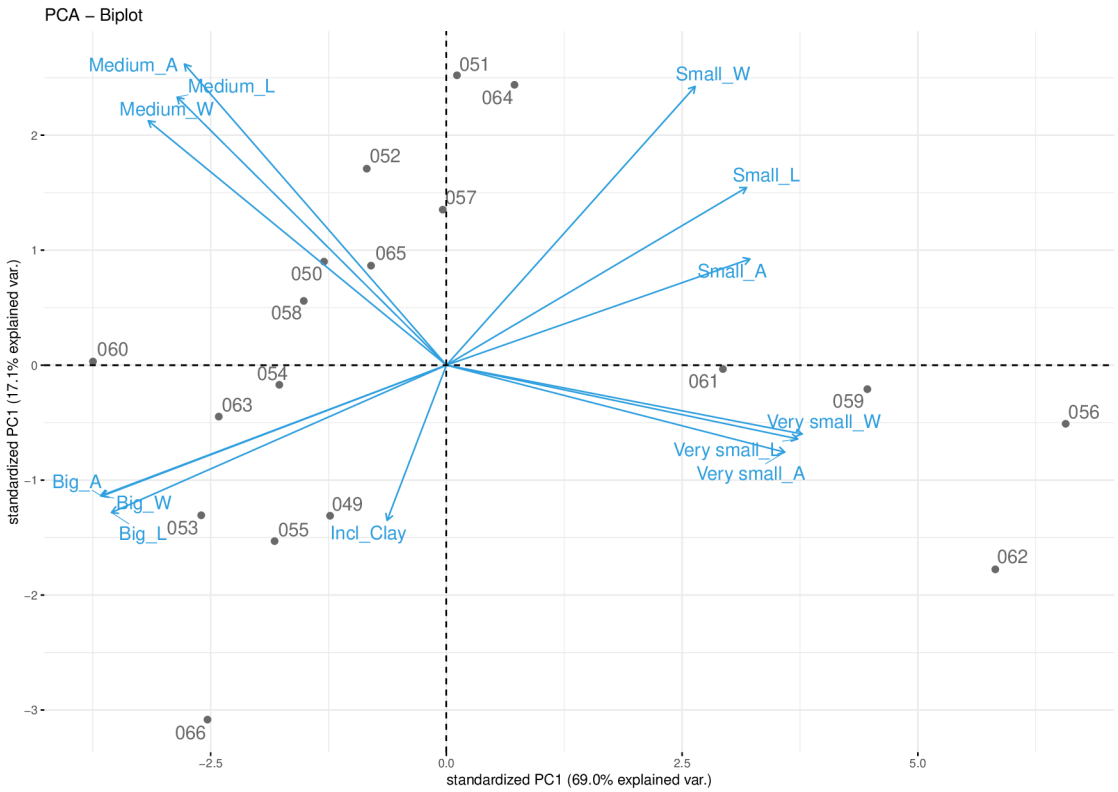

Fig. 5.3: Bivariate plot showing the variables (blue arrows) and the position of investigated samples/observations (grey dots). Horizontal axis correponds to PC1, with 69.0% explained variance. Vertical axis belongs to PC2, with 17.1% explained variance.

In Fig. 5.3 we can already observe some formation of groups of samples. Just by visual inspection of the biplot three main groups of samples can be identified. Samples 061 and 059 can be classified as the samples with very small inclusions. While on the other hand samples 053, 055 and 049 form a group of samples with big inclusions. There are also many samples, that would be hard to classify, just by looking. For example samples 052, 050, 058, etc. There are also some outlying observations, for example sample 066, 062, 056. But the detection of outliers is not an easy task and thus we need to proceed carefully. An intuitive way, how to detect

the outliers, can be done by calculating the Euclidean distance of each observation to the origin of the coordinate system. The result is shown in Fig. 5.4.

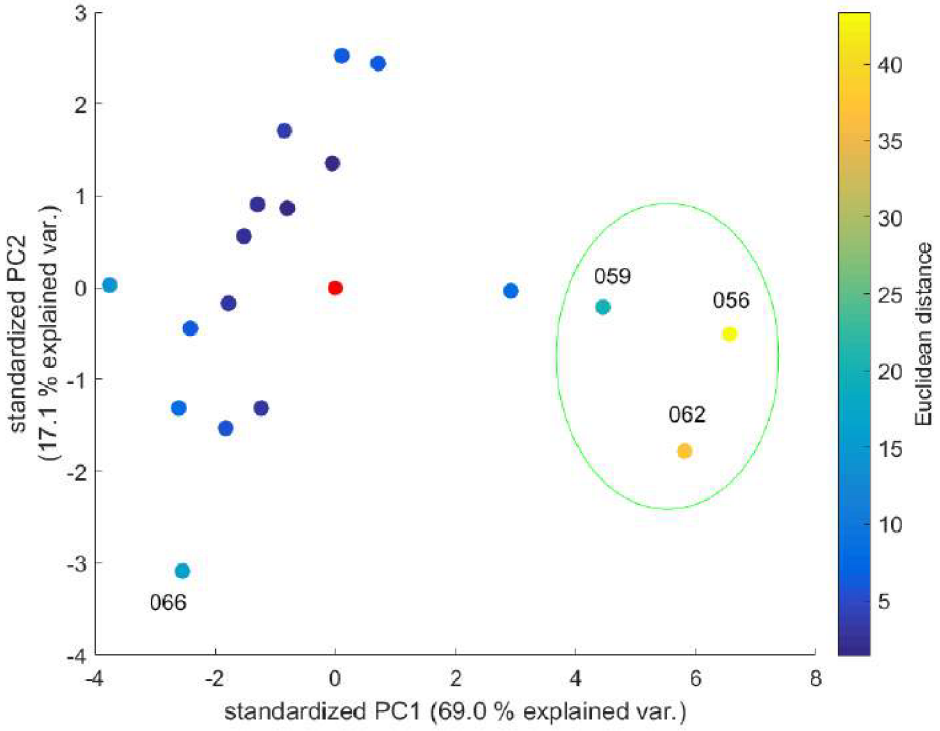

Fig. 5.4: Detected outlier using a standard Euclidean distance. Red dot corresponds to the origin of the PC space.

Technically speaking, here we have empirically discovered four outlying observations, just by visual inspection. Now we have to provide more quantitative analysis to support our hypothesis. For this reason we have calculated robust Mahalanobis distance [29] between individual observations and the geometric median, taking into account all observations. Result shown in Fig. 5.5.

From this it is clear that samples 056, 062, 059, 061 are outliers. But in our case we were not allowed to remove these observations. The reason to do so was, that they were an important part of the analysis and removal of these samples would lead to losing one whole classified group of samples, i.e. samples with small inclusions. That is why we have decided to extend our analysis and take a different approach.

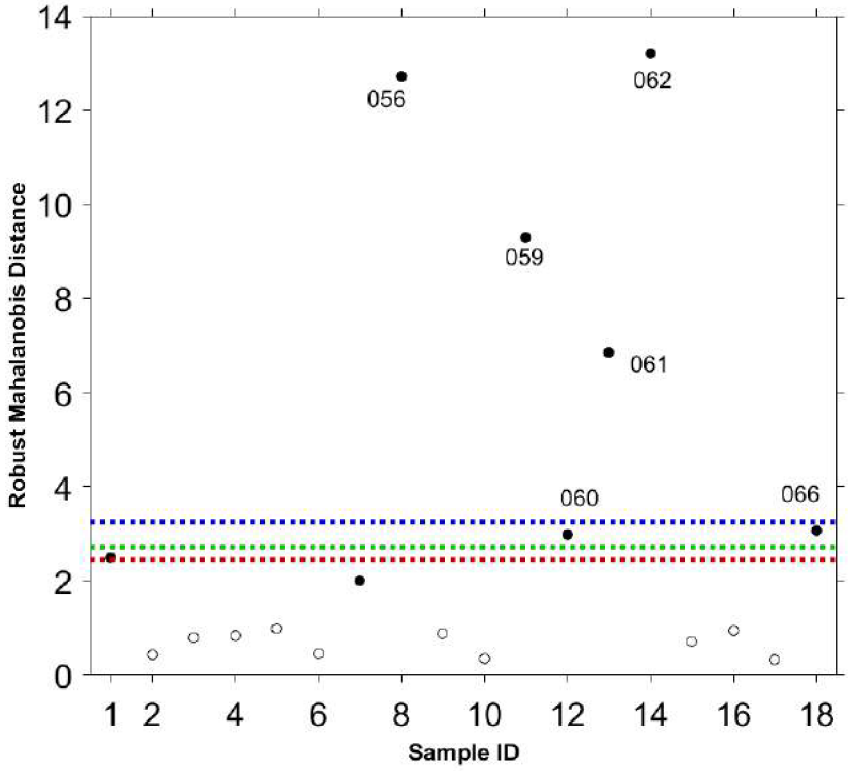

Fig. 5.5: Detected outliers using a robust Mahalanobis distance [43]. Horizontal lines correspond to different signicance levels of the χ^2 test, used in [14]. Red line corresponds to $\alpha = 0.1$, green line to $\alpha = 0.05$ and blue line to $\alpha = 0.01$.

5.2 Cluster analysis

In our case the separation between the observations is not very clear. Due to this, we have decided to do cluster analysis on the result of the PCA, to improve the separation of observations. In this part, performance of different clustering algorithms has been tested. Due to the fact that all algorithms gave identical results, we have decided to randomly select one of these algorithms. The algorithm that has been selected was K-means clustering algorithm. The optimal number of clusters has been selected empirically by testing different number of clusters. Three clusters have been chosen due to the fact that in case of four clusters, one cluster was consumed by only one observation (sample 066). This observation is clearly an outlier, see Fig. 5.6. Because of this, we have decided to perform PCA again, but now without sample 066. Result of the cluster analysis, is shown in Fig. 5.7. This time we have decided to form four clusters.

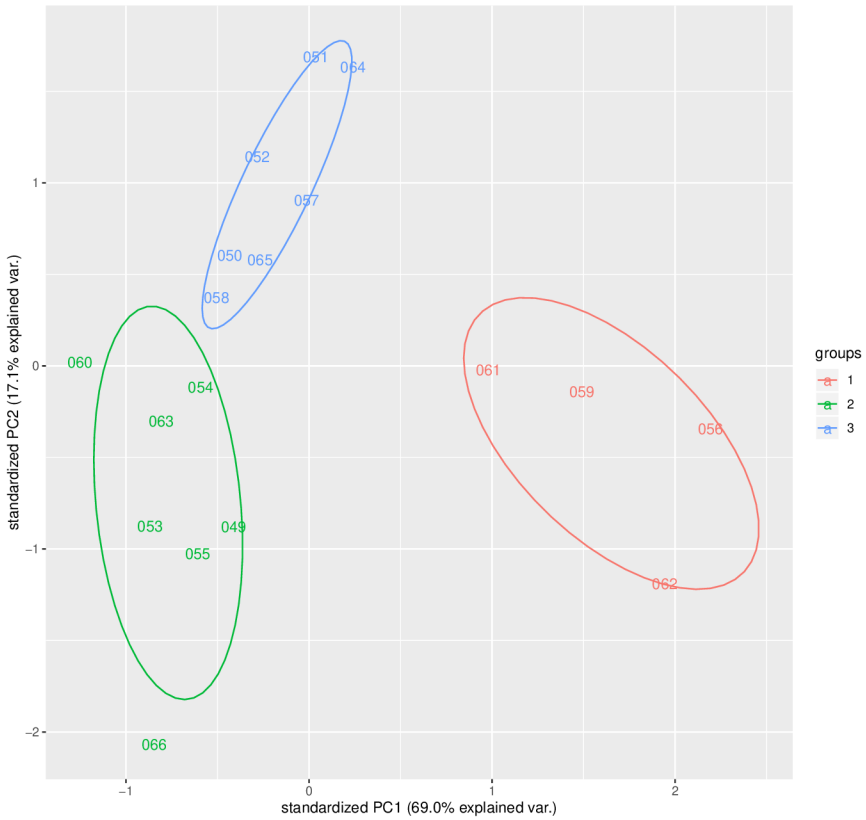

Fig. 5.6: Result of the K-means cluster analysis for all observations. The ellipses around each group correspond to 95 % confidence interval.

5.3 Interpretation

Among all the samples, four main groups of samples have been identified. The first group (blue) corresponds to samples with prevalent number of very small or small inclusions (samples: 061, 059, 056, 062). Second group (violet) of samples are samples with small and medium-sized inclusions (samples: 051, 064, 052, 057, 065). Third group (red) is here as a some kind of intermission phase between medium-sized samples and samples with big inclusions (samples: 050, 058, 054, 063, 060). The last group (green) that has been identified belongs to the samples that could be characterized as samples with big inclusions (samples: 049, 055), with a slight exception of sample 053, whose internal structure corresponds more to samples with medium-sized inclusions. Among all these observations, it can be seen that sample 066 does not belong to any of the groups. The closest group, where this sample could belong is group of samples with high percentage of big inclusions. After this I have tried to confront the result of the cluster analysis with the data obtained by microCT imaging.

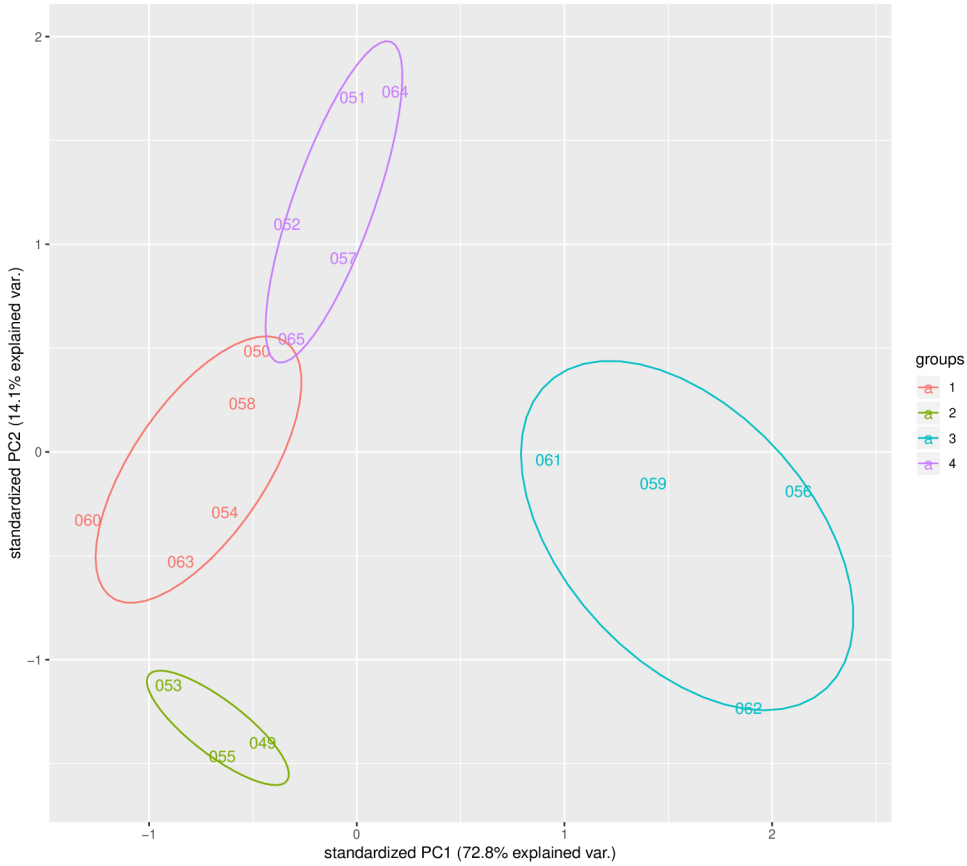

Fig. 5.7: Result of the K-means cluster analysis. Analysis was performed on the result of the PCA with removal of sample 066 which was identified as an outlier.

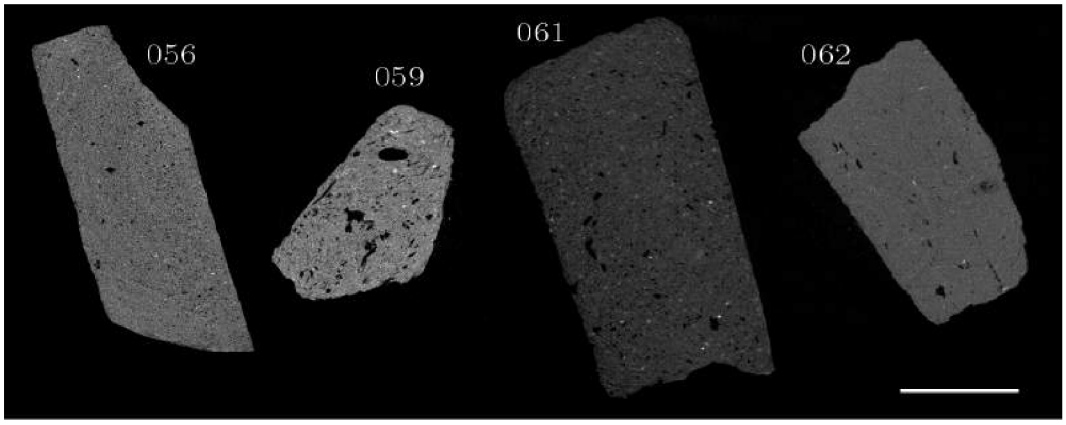

Fig. 5.8: CT cross sections of samples that were categorized as samples with very small/small inclusions. Bright areas correspond to dense lithic inclusions, probably calcite. Scale bar corresponds to 0.5 cm.

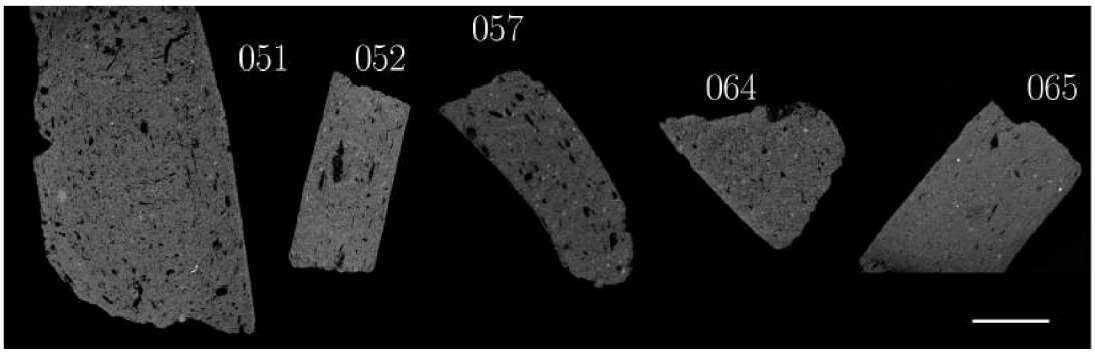

Fig. 5.9: CT cross sections of samples that fell in the category of samples with medium-sized inclusions. Scale bar corresponds to 0.5 cm.

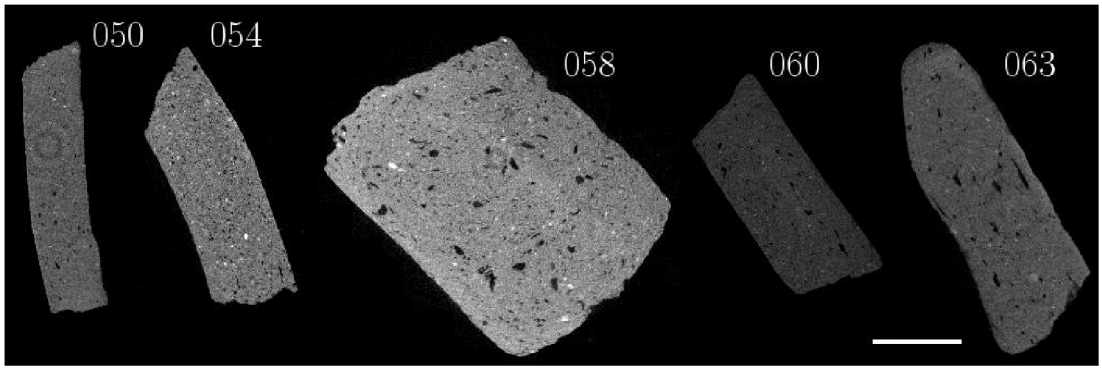

Fig. 5.10: CT cross sections of samples that represent group that could be characterized as an intermission phase between medium and big inclusions. with high percentage of medium-sized inclusions. Scale bar corresponds to 0.5 cm

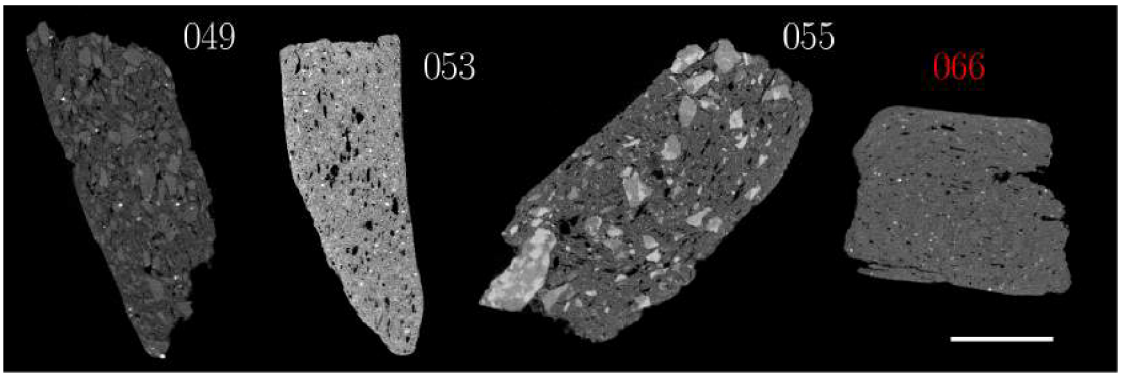

Fig. 5.11: CT cross sections of samples that were characterized as samples with prevalance of big inclusions. Sample 066 that was discovered to be an outlier was highlighted in red. Scale bar correponds to 0.5 cm.

MicroCT results have revealed that most samples show small-medium grain-size, while only a few ones show a coarse-grained paste with variably dense lithic inclusions, which supports the results obtained through PCA. One thing that came out clear after visual inspection of microCT data was that sample 066 should belong to the samples with medium-sized inclusions. Due to the fact that it has been categorized as a sample with big inclusions, it gave us a first suspicion that the sample 066 was something different. The conclusion of microCT analysis was that all the analyzed samples were locally produced with the exception of sample 066. This results were confirmed by prompt gamma activation analysis (PGAA), where as the result of the analysis was that all the samples share similiar chemical composition, with exception of sample 066.

CONCLUSION

This thesis addresses the implementation of protocols to extract structural information from microCT data through segmentation procedures and the development of various statistical methods useful to classify archeological samples in order to study their technology and possible origin.

The theoretical part of the thesis gives an insight into basic CT principles and some image processing methods. Later, the PCA method is presented, together with the derivation and geometrical interpretation of principal components.

The experimental part of the thesis describes the experiment performed at the Multidisciplinary Laboratory of the International Centre for Theoretical Physics (Trieste, Italy), using a system specifically designed for the study of archaeological and paleoanthropological materials. The classification algorithm is presented in the next part. The classification algorithm was tested on pottery samples from the archaeological site of Al Khutm (Bat, Oman). In the first step, the samples were measured using a CT machine. As the next step, the reconstructed volumetric data were segmented and then some parameters of interest, describing the pottery matrix, were exported using Avizo v. 8. These parameters then served as an input for the PCA method, later followed by a cluster analysis. The outcome of the method was the identification of four main groups of samples, based on the size of their inclusions. Another significant result was the identification of one outlying observation (sample 066), where preliminary results indicated that it had been probably imported from the Indus valley (present-day Pakistan).

The presented method was based on data coming from segmentation, thus bound to one specific technique, i.e. CT. However, the use of statistical principles for the purpose of classification or characterization of samples is limitless and combination of datasets coming from various sources could offer whole new types of analyses in the future.

BIBLIOGRAPHY

[1] AMZA, C. A review on neural network-based image segmentation techniques," De Montfort University, Mechanical and Manufacturing Engg., The Gateway Leicester, LE1 9BH, United Kingdom, pp. 1-23, 2012. Available from: https://www.researchgate.net/publication/228873725.

[2] BARUCHEL, José. X-ray tomography in material science. Paris: Hermes Science, 2000. ISBN 27-462-0115-1.

[3] BELL, Daniel, Prashant MUDGAL. Characteristic radiation. In: Radiopaedia.org [online]. [cit. 2019-05-14]. Available from: https://radiopaedia.org/articles/characteristic-radiation.

[4] BERNARDINI, F., A. VECCHIET, A. DE MIN, et al. Neolithic pottery from the Trieste Karst (northeastern Italy): A multi-analytical study. Microchemical Journal. 2016(124), 600-607. DOI: 10.1016/j.microc.2015.09.019. ISSN 0026265X. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0026265X15002209.

[5] BERNARDINI, Federico, Elena LEGHISSA, David PROKOP, et al. X-ray computed microtomography of Late Copper Age decorated bowls with cross-shaped foots from central Slovenia and the Trieste Karst (North-Eastern Italy): technology and paste characterisation. Archaeological and Anthropological Sciences. 2019. DOI: 10.1007/s12520-019-00811-w. ISSN 1866-9557. Available from: http://link.springer.com/10.1007/s12520-019-00811-w.

[6] BUZUG, Thorsten M. Computed tomography: from photon statistics to modern cone-beam CT. Berlin: Springer, c2008, p. 50-60. ISBN 978-3-540-39408-2.

[7] BRO, Rasmus a Age K. SMILDE. Principal component analysis. Anal. Methods. 2014, 6(9), 2812-2831. DOI: 10.1039/C3AY41907J. ISSN 1759-9660. Available from: http://xlink.rsc.org/?DOI=C3AY41907J.

[8] CARMIGNATO, Simone. Industrial x-ray computed tomography. New York, NY: Springer Berlin Heidelberg, 2017. ISBN 978-3-319-59571-9.

[9] CASTLEMAN, Kenneth R. Digital image processing. Englewood Cliffs, N.J.: Prentice Hall, c1996, p. 71-90. ISBN 978-0132114677.

[10] CATTANI M. et al. 2017, Proceedings of the 47th Seminars for Arabian Studies 47, July 2017. ISBN: 978-1784915209.

[11] DASGUPTA, A NATH, A. Classification of Machine Learning Algorithms. International Journal of Innovative Research in Advanced Engineering (IJIRAE). 2016. ISSN: 2349-2763. 3. 6-11. 10.6084/M9.FIGSHARE.3504194.V1.

[12] DING, Chris a Xiaofeng HE. K -means clustering via principal component analysis. Twenty-first international conference on Machine learning - ICML '04. New York, New York, USA: ACM Press, 2004. DOI: 10.1145/1015330.1015408. ISBN 1581138285. Available from: http://portal.acm.org/citation.cfm?doid=1015330.1015408.

[13] ERIKSSON, L. Multi- and megavariate data analysis: basic principles and applications. 3rd, rev. ed. Malmö: Umetrics Academy, 2013, p. 33-54. ISBN 91-973730-5-2.

[14] HADI, Ali S. Identifying Multiple Outliers in Multivariate Data. Journal of the Royal Statistical Society: Series B (Methodological). 1992, 54(3), p. 761-771. DOI: 10.1111/j.2517-6161.1992.tb01449.x. ISSN 00359246. Available from: http://doi.wiley.com/10.1111/j.2517-6161.1992.tb01449.x

[15] JAIN, Anil K. Fundamentals of digital image processing. Englewood Cliffs, NJ: Prentice Hall, c1989, p. 431-460. ISBN 978-0133361650.

[16] JIN X., HAN J. (2011) K-Medoids Clustering. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. Available from: https://link.springer.com/referenceworkentry/10.1007/978-0-387-30164-8_ 426.

[17] JIN X., HAN J. (2016) Mean Shift. In: Sammut C., Webb G. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA. Available from: http://link.springer.com/10.1007/978-1-4899-7502-7_532-1.

[18] JOLLIFFE, Ian T. Principal component analysis. 2nd ed. New York: Springer, c2002. Springer series in statistics. ISBN 03-879-5442-2.

[19] JOLLIFFE, Ian T. a Jorge CADIMA. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016, 374(2065). DOI: 10.1098/rsta.2015.0202. ISSN 1364-503X. Available from: http://rsta.royalsocietypublishing.org/lookup/doi/10.1098/rsta.2015.0202.

[20] KALASOVÁ, Dominika. Využití fázového kontrastu v rentgenové počítačové tomografii. Brno, 2016. Diplomová práce. Brno University of Technology, Faculty of Mechanical Engineering. Vedoucí práce: Tomáš ZIKMUND.

[21] KANYONGO, Gibbs Y. Determining The Correct Number Of Components To Extract From A Principal Components Analysis: A Monte Carlo Study Of The Accuracy Of The Scree Plot. Journal of Modern Applied Statistical Methods. 2005, 4(1), 120-133. DOI: 10.22237/jmasm/1114906380. ISSN 1538-9472. Available from: http://digitalcommons.wayne.edu/jmasm/vol4/iss1/13.

[22] KASSAMBARA, Alboukadel. Practical Guide To Principal Component Methods in R [online]. STHDA, 2017 [cit. 2018-11-04]. ISBN 1975721136. Available from: http://www.sthda.com/english/wiki/practical-guide-to-principal-component-methods-in-r.

[23] KNIPE, Henry a Yeung J. Bremsstrahlung. In: Radiopaedia.org [online]. [cit. 2019-05-14]. Available from: https://radiopaedia.org/articles/bremsstrahlung.

[24] KUMARMALHOTRA, Virender, Harleen KAUR a M. AFSHAR ALAM. An Analysis of Fuzzy Clustering Methods. International Journal of Computer Applications. 2014, 94(19), 9-12. DOI: 10.5120/16497-6578. ISSN 09758887. Available from: http://research.ijcaonline.org/volume94/number19/pxc3886578.pdf.

[25] KUCHMENT, Peter. The radon transform and medical imaging. Philadelphia: Society for Industrial and Applied Mathematics, [2014]. CBMS-NSF regional conference series in applied mathematics, 85. ISBN 978-1611973280.

[26] KRENKEL, M. (2015). Cone-beam x-ray phase-contrast tomography for the observation of single cells in whole organs. Göttingen Series in X-Ray Physics. doi:10.17875/gup2015-908.

[27] LEE, Tong Hau, Mohammad Faizal Ahmad FAUZI a Ryoichi KOMIYA. Segmentation of CT brain images using unsupervised clusterings. Journal of Visualization. 2009, 12(2), 131-138. DOI: 10.1007/BF03181955. ISSN 1343-8875. Available from: http://link.springer.com/10.1007/BF03181955.

[28] LEE, Yuh-Jye, Yi-Ren YEH a Yu-Chiang Frank WANG. Anomaly Detection via Online Oversampling Principal Component Analysis. IEEE Transactions on Knowledge and Data Engineering. 2013, 25(7), 1460-1470. DOI: 10.1109/TKDE.2012.99. ISSN 1041-4347. Available from: http://ieeexplore.ieee.org/document/6200273/.

[29] LEYS, Christophe, Olivier KLEIN, Yves DOMINICY a Christophe LEY. Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. Journal of Experimental Social Psychology. 2018, 74,

150-156. DOI: 10.1016/j.jesp.2017.09.011. ISSN 00221031. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022103117302123.

[30] LI, Jiamin a Harold W. LEWIS. Fuzzy Clustering Algorithms — Review of the Applications. 2016 IEEE International Conference on Smart Cloud (SmartCloud). IEEE, 2016, (INSPEC Accession Number: 16575407), 282-288. DOI: 10.1109/SmartCloud.2016.14. ISBN 978-1-5090-5263-9. Available from: http://ieeexplore.ieee.org/document/7796188/.

[31] MCCOLLOUGH, C H. The AAPM/RSNA physics tutorial for residents. X-ray production. RadioGraphics. 1997, 17(4), 967-984. DOI: 10.1148/radiographics.17.4.9225393. ISSN 0271-5333. Available from: http://pubs.rsna.org/doi/10.1148/radiographics.17.4.9225393.

[32] MEDEIROS, Luciano Frontino de, Hamilton Pereira da SILVA a Eduardo Parente RIBEIRO. Tomographic Image Reconstruction Of Fan-Beam Projections With Equidistant Detectors Using Partially Connected Neural Networks. Learning and Nonlinear Models. 2003, 1(2), 122-130. DOI: 10.21528/LNLM-vol1-no2-art5. ISSN 1676-2789. Available from: http://abricom.org.br/lnlm/publicacoes/vol1-no2/vol1-no2-art5.

[33] MELOUN, M. a Jiří MILITKÝ. Statistická analýza experimentálních dat. Vyd. 2., upravené a rozšířené. Praha: Academia, 2004. ISBN 978-802-0012-548.

[34] MORIYA, Takayasu, Holger R. ROTH, Shota NAKAMURA, et al. Unsupervised segmentation of 3D medical images based on clustering and deep representation learning. Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging. SPIE, 2018, 2018-3-12, 2018(10578). DOI: 10.1117/12.2293414. ISBN 9781510616455. Available from: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10578/2293414/Unsupervised-segmentation-of-3D-medical-images-based-on-clustering-and/10.1117/12.2293414.full.

[35] PEARSON, K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, Vol. 2, No. 6. (1901), 559-572. DOI: 10.1080/14786440109462720. ISSN 1941-5982. Available from: https://www.tandfonline.com/doi/full/10.1080/14786440109462720.

[36] PERES-NETO, Pedro R., Donald A. JACKSON a Keith M. SOMERS. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics and Data Analysis. 2005,

49(4), 974-997. DOI: 10.1016/j.csda.2004.06.015. ISSN 01679473. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0167947304002014.

[37] PHAM, Dzung L., Chenyang XU a Jerry L. PRINCE. Current Methods in Medical Image Segmentation. Annual Review of Biomedical Engineering. 2000, 2(1), 315-337. DOI: 10.1146/annurev.bioeng.2.1.315. ISSN 1523-9829. Available from: http://www.annualreviews.org/doi/10.1146/annurev.bioeng.2.1.315.

[38] POŘÍZKA, P. Využití spektrometrie laserem buzeného plazmatu (LIBS) pro materiálovou analýzu. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2014. 105 s. Vedoucí disertační práce prof. Ing. Jozef Kaiser, Ph.D.

[39] POŘÍZKA, P. Impact of Laser-Induced Breakdown Spectroscopy data normalization on multivariate classification accuracy. JOURNAL OF ANALYTICAL ATOMIC SPECTROMETRY 32(2), p. 277 - 288, 2017, doi: 10.1039/c6ja00322b.

[40] Photomultiplier tube. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001 [cit. 2019-05-14]. Available from: https://commons.wikimedia.org/wiki/File:PhotoMultiplierTubeAndScintillator.svg.

[41] SAHA, Pritam, Nabanita ROY, Deotima MUKHERJEE a Ashoke Kumar SARKAR. Application of Principal Component Analysis for Outlier Detection in Heterogeneous Traffic Data. Procedia Computer Science. 2016, 83, 107-114. DOI: 10.1016/j.procs.2016.04.105. ISSN 18770509. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1877050916301284.

[42] SEERAM, Euclid. Computed tomography: physical principles, clinical applications, and quality control. Fourth edition. St. Louis, Missouri: Elsevier, [2016], p. 26-46. ISBN 978-0323312882.

[43] SEMECHKO, Anton. Multivariate-Outliers, (2019), GitHub repository, https://github.com/AntonSemechko/Multivariate-Outliers.

[44] STOLOJESCU-CRISAN, C. a S. HOLBAN. A Comparison of X-Ray Image Segmentation Techniques. Advances in Electrical and Computer Engineering. 2013, 13(3), 85-92. DOI: 10.4316/AECE.2013.03014. ISSN 1582-7445.

[45] ŠTOLL, Ivan, Jiří TOLAR a Igor JEX. Klasická teoretická fyzika. Praha: Univerzita Karlova, nakladatelství Karolinum, 2017, p. 293-298. ISBN 978-80-246-3545-3.

[46] TESAŘOVÁ, Markéta. Využití průmyslové rentgenové počítačové mikrotomografie ve vývojové biologii. Brno, 2018. 71 s. Diplomová práce. Vysoké učení technické v Brně. Fakulta strojního inženýrství. Vedoucí práce Marie NOVOTNÁ.

[47] Thermo Fisher Scientific, FEI SAS. Thermo Scientific Avizo Software 9. User's Guide. ZIB, Germany, 2018.

[48] TUNIZ, C., F. BERNARDINI, A. CICUTTIN, et al. The ICTP-Elettra X-ray laboratory for cultural heritage and archaeology. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2013, 711, 106-110. DOI: 10.1016/j.nima.2013.01.046. ISSN 01689002. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168900213001381.

[49] VELMURUGAN, T. Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points. Journal of Computer Science. 2010, 6(3), 363-368. DOI: 10.3844/jcssp.2010.363.368. ISSN 1549-3636. Available from: http://www.thescipub.com/abstract/10.3844/jcssp.2010.363.368.

[50] WU, Junjie. Advances in k-means clustering. New York: Springer, 2012, p. 2-20. ISBN 978-3642298066.

LIST OF ABBREVIATIONS AND SYMBOLS

CT	computed tomography
PCA	principal component analysis
ICTP	International Centre for Theoretical Physics
CEITEC	Central European Institute of Technology
c	speed of light
e	elementary charge
ϵ_0	vacuum permeability
Al	Aluminium
Cu	Copper
keV	kiloelectronvolt
CsI	caesium iodide
TFT	thin-film transistor
FFT	fast fourier transform
FBP	filtered back projection
$\delta()$	Dirac delta distribution
r	radial distance in the Fourier space
ANN	artificial neural networks
VLSI	very lage scale integration
M	data matrix
T	scores matrix
P	loadings matrix
E	errror matrix
C	covariance matrix

Archaeological and Anthropological Sciences
https://doi.org/10.1007/s12520-019-00811-w

ORIGINAL PAPER

X-ray computed microtomography of Late Copper Age decorated bowls with cross-shaped foots from central Slovenia and the Trieste Karst (North-Eastern Italy): technology and paste characterisation

Federico Bernardini[1,2] · Elena Leghissa[3] · David Prokop[4,5] · Anton Velušček[3] · Angelo De Min[6] · Diego Dreossi[7] · Sandro Donato[8,9] · Claudio Tuniz[1,2] · Francesco Princivalle[6] · Manuela Montagnari Kokelj[10]

Received: 20 December 2018 / Accepted: 15 February 2019

Abstract

About 20 Late Copper Age bowls with cross-shaped foots from Deschmann's pile dwellings (Ljubljansko barje, central Slovenia) and Trieste Karst (North-Eastern Italy) have been investigated using X-ray computed microtomography (microCT) in order to study the vessel-forming technique, to characterise their pastes and to test the hypothesis that some Karst bowls could have been imported from nowadays central Slovenia or even more distant regions. In three selected virtual slices per sample, clay, lithic inclusions and pores have been segmented and quantified. In addition, the area, maximum length and width of each lithic inclusion have been calculated. Then, the microCT-derived results have been statistically analysed by principal component analysis (PCA). The orientation of pores and disjunctions in microCT volumes show that the basins of the bowls were built using mainly the coiling technique, while the base was shaped starting from a central piece, to which a layer of clay was added and then reshaped in order to produce the foots. The Slovenian bowls include both medium/coarse-grained and very fine- or fine-grained vessels mainly tempered with carbonate inclusions. The pastes of the Karst bowls are considerably heterogeneous. One bowl was most likely imported to the Karst but not from central Slovenia as it shows peculiar components, shape and decoration. The other two imported vessels show a very fine-grained paste comparable to the one of several samples from Deschmann's pile dwellings. Such technological similarity is confirmed by PCA of microCT data and petrographic observations. Our study confirms the existence of strong cultural connections between central Slovenia and the northernmost Adriatic coast during the Late Copper Age.

Keywords Late Copper Age decorated bowls · Central Slovenia · Trieste Karst (North-Eastern Italy) · X-ray computed microtomography · PCA of microCT-derived data · Technology · Paste characterisation

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12520-019-00811-w) contains supplementary material, which is available to authorized users.

✉ Federico Bernardini
fbernard@ictp.it

[1] Centro Fermi, Museo Storico della Fisica e Centro di Studi e Ricerche "Enrico Fermi", Piazza del Viminale 1, 00184 Rome, Italy

[2] Multidisciplinary Laboratory, The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy

[3] Institute of Archaeology, ZRC SAZU, Novi Trg 2, 1000 Ljubljana, Slovenia

[4] Central European Institute of Technology (CEITEC), Purkyňova 123, 61200 Brno, Czech Republic

[5] Institute of Physical Engineering, Brno University of Technology, Brno, Czech Republic

[6] Department of Mathematics and Geosciences, University of Trieste, Via Weiss 8, 34127 Trieste, Italy

[7] Elettra-Sincrotrone Trieste S.C.p.A., Strada Statale 14 km 163.5 in AREA Science Park, 34149 Trieste, Basovizza, Italy

[8] Department of Physics, University of Trieste, Via Valerio 2, 34127 Trieste, Italy

[9] INFN Sezione di Trieste, Padriciano 99, 34149 Trieste, Italy

[10] Department of Humanistic Studies, University of Trieste, Via Principe di Montfort 3, 34124 Trieste, Italy

Published online: 21 March 2019

Introduction

The three Deschmann's pile dwellings (*slo. Dežmanova kolišča*), located near Ig in the Ljubljansko barje (central Slovenia), are one of the main third millennium BC archaeological evidence in the south-eastern Alpine region (Fig. 1). In 2011, they were inscribed on the World Heritage List of UNESCO in the transnational serial property *prehistoric pile dwellings around the Alps*. The well-preserved ceramic finds from these sites have been included in various typo-chronological analyses to define the cultural development in this area during the Late Copper and Early Bronze ages (e.g. Dimitrijević 1956, 1966, 1979; Korošec 1958–1959; Bóna 1965; Ecsedy 1977; Parzinger 1984; Govedarica 1989; Maran 1998; Leghissa 2017). According to the latest typological analyses and a few absolute dates (Leghissa 2017, 2018), two main phases have been identified in Deschmann's pile dwellings. The oldest one (twenty-eighth–twenty-sixth century BC) is mainly characterised by pottery typical of the Late Copper Age Vučedol culture, widespread from the Balkans to south-eastern Central Europe, and pottery pointing to connections with other cultures, located especially in Central Europe. The younger phase (twenty-sixth century BC to twenty-fifth century BC) is defined by finds of the Ljubljana culture that has developed under the influences of Corded Ware (and Globular Amphora) as well as Somogyvár-Vinkovci and Bell Beaker cultures.

The Ljubljana culture is mostly attested in Deschmann's pile dwellings, but some typical elements are present also in many caves of the Trieste Karst (North-Eastern Italy; Fig. 1) and eastern Adriatic coast (Montagnari Kokelj 1981; Govedarica 1989; Gilli and Montagnari Kokelj 1993, 1994, 1996; Montagnari Kokelj and Crismani 1997; Montagnari Kokelj et al. 2002; Leghissa 2017, Fig. 181). Contacts between the two areas investigated in this contribution are already attested during the oldest phase of Deschmann's pile dwellings, but they undoubtedly increased during the development and spread of the Ljubljana culture (see Leghissa 2017 and the literature quoted there).

Decorated bowls with cross-shaped foots are reported from central Slovenia and the Trieste Karst. This type of vessel is common in the Vučedol culture, spread from the Balkans to

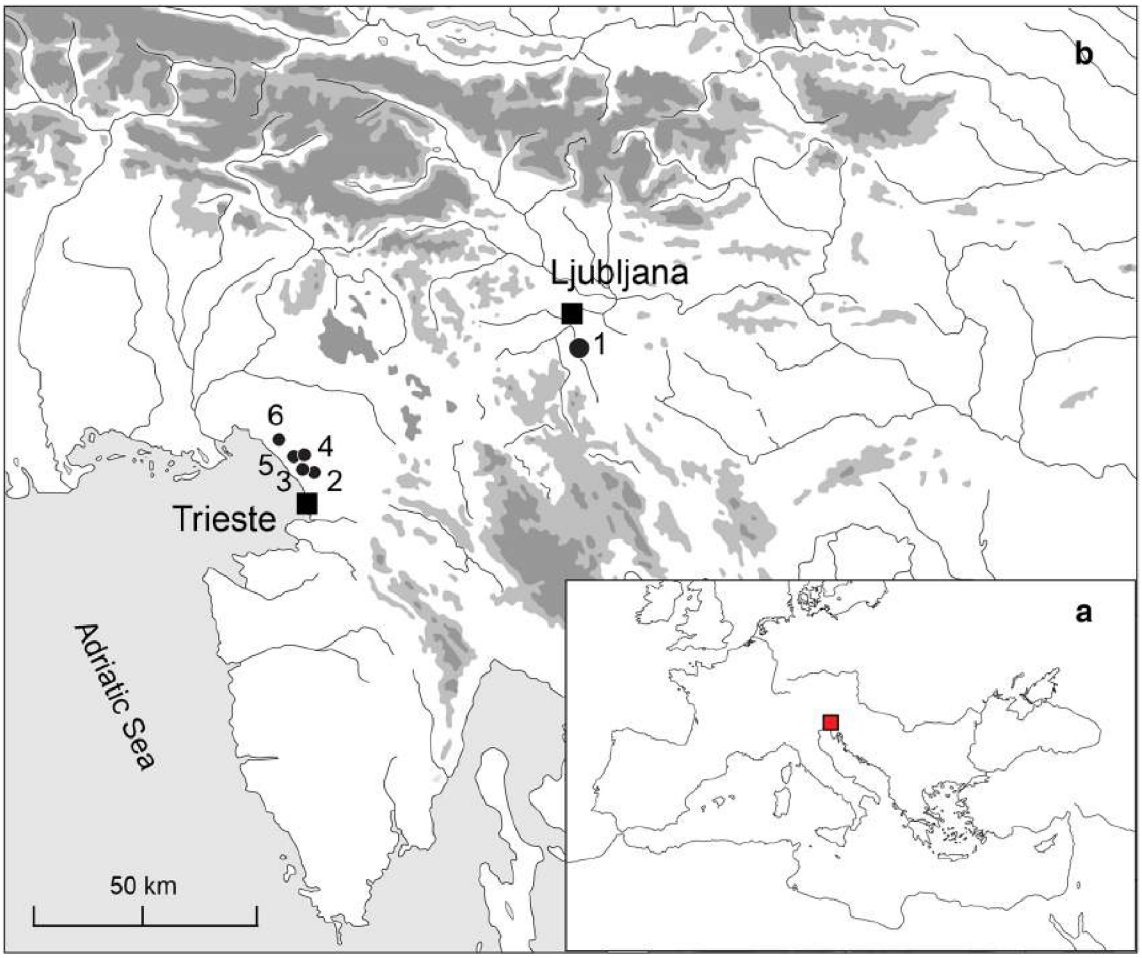

Fig. 1 Position of the archaeological sites where the investigated bowls have been discovered. 1, Deschmann's pile dwellings (Slovenia); 2, Ciclami cave; 3, Zingari cave; 4, Cotariova cave; 5, Pettine cave; 6, Edera cave. All cave sites are located in the Trieste Karst

south-eastern Central Europe, but it is also reported in other contemporaneous cultures in Central Europe. In Deschmann's pile dwellings, where this type is abundant, several variants are known, while only a few bowls were discovered in Trieste Karst. Despite the small number, the latter are heterogeneous as to shape and decoration.

Most of the known bowls from the Trieste Karst and a selection of bowls from Deschmann's pile dwellings (Fig. 1) have been studied using X-ray computed microtomography (hereinafter microCT) to define their production technology and to characterise their pastes. MicroCT has already been applied to the study of ancient pottery providing information about vessel-forming techniques and provenance (Kahl and Ramminger 2012; Bernardini et al. 2013, 2015, 2016; Sanger 2016; Machado et al. 2017). When the artefacts under consideration cannot be sampled, such as in our case, microCT is a precious tool to perform fabric characterisation of pottery, allowing the visualisation and quantification of lithic inclusions, clay matrix, pores and disjunctures generally related to firing shrinkage, loss of organic temper or the result of how the vessels were shaped into their final form (Bernardini et al. 2016).

We applied a similar approach to study the Copper Age bowls from Slovenia and the Trieste Karst, and we performed a multivariate statistical analysis of microCT-derived data. The obtained results have been used to discuss the origin of the few vessels found in the Karst area. In fact, they could have been imported to the Karst or they could have been locally produced following a model common in Deschmann's pile dwellings and beyond.

Materials and methods

In the present study, 23 bowls with cross-shaped foots have been studied: 14 from Deschmann's pile dwellings and 9 from five caves of the Trieste Karst (Table 1; Fig. 2). These vessels are among the most abundant ceramic forms in Deschmann's pile dwellings. Most of them can be typologically ascribed to the Vučedol culture phase and just a few ones to the later Ljubljana culture. They are generally richly decorated on the exterior, the interior and the rim with stab-and-drag, simple incised lines or rarely with impressions of cords wrapped around thin plates (i.e. the typical decoration technique of the Ljubljana culture; Leghissa 2015). In the Trieste Karst, a few bowls of this type were found. Despite their small number, they are outstanding for their rich and complex decorations showing incised lines or impressions of a twisted double cord, a technique mainly reported from contemporary Central European cultures (e.g. *Schnurkeramische Kultur* or Corded Ware culture; see Buchvaldek 1967; Furholt 2003).

Macroscopic observation

The surface of all samples has been observed using a stereomicroscope and/or a magnifying glass in order to identify visible lithic grains. Most vessels contain quite abundant limestone/calcite inclusions, while white mica and quartz grains have been recognised in a few bowls (Table 2).

Microfocus X-ray computed tomography

The vessels were imaged by X-ray microCT at the Multidisciplinary Laboratory of the Abdus Salam International Centre for Theoretical Physics (Trieste, Italy), using a system (Tuniz et al. 2013) specifically designed for the study of archaeological and palaeoanthropological materials (e.g. Bernardini et al. 2012, 2016; Tuniz et al. 2012, 2013; Bernardini et al. 2017; Duches et al. 2018).

The microCT acquisitions of most specimens were carried out by using a sealed X-ray source (Hamamatsu L8121-03) at a voltage of 110 kV, at a current of 90 μA and with a focal spot size of 5 μm. A few very dense vessels were analysed using a voltage of 140 kV, a current of 200 μA and a focal spot size of 20 μm. The X-ray beam was filtered by a 0.1-mm-thick copper absorber. A set of 1440 or 1800 projections of the artefacts was recorded over a total scan angle of 360° by a flat panel detector (Hamamatsu C7942SK-25; pixel size of 50 μm). The resulting microCT slices were reconstructed using the commercial software Digi XCT (Digisens) in 32-bit format and obtaining an isotropic voxel size from about 20 to 40 μm (Table 1).

Most of the samples have been analysed two times with different resolutions. Data sets with a resolution of about 20 μm have been used to perform the segmentation of the paste components, while the data sets with a lower resolution, including the whole vessels or a larger part of them, have been mainly used to study technological production traces.

Segmentations and microCT-derived data analysis

Using Avizo v.8 software, three virtual sections for each sample, taken at the centre and the edges of each data set, have been segmented in order to separate the clay matrix from the lithic temper materials and the pores following a procedure already applied by Bernardini et al. (2016). In the same work, a comparison between results from 2D segmentation of selected slices and 3D segmentation of extracted sub-volumes of archaeological pottery has shown that the much faster 2D segmentation generally gives comparable results (for details, see Bernardini et al. 2016).

After the segmentation process, we have first calculated, using the same software, the total area of clay matrix, pores and lithic inclusions (including both temper material and lithic components within the raw material) for all three virtual

Table 1 List of the studied Copper Age bowls with cross-shaped foots from central Slovenia and the Trieste Karst (North-Eastern Italy)

Inventory number	Site	Cultural attribution	Methods	Voxel size	References
B1482	Deschmann's pile dwellings	L. b. variant of V. c.	μCT	41.01; 81.42	Korošec and Korošec 1969, T. 38: 1a, b; Leghissa 2017, T. 91: 1
B1963	Deschmann's pile dwellings	L. b. variant of V. c.	μCT	21.42; 43.15	Korošec and Korošec 1969, T. 44: 9a–c; Leghissa 2017, T. 107: 2
B1984	Deschmann's pile dwellings	L. b. variant of V. c.	μCT	39.71	Korošec and Korošec 1969, T. 46: 12; Leghissa 2017, T. 95: 1
B1965	Deschmann's pile dwellings	L. b. variant of V. c.	μCT	21.42; 39.71	Korošec and Korošec 1969, T. 44: 11a–c; Leghissa 2017, T. 102: 2
B1939	Deschmann's pile dwellings	L. b. variant of V. c.	μCT	21.42; 39.71	Korošec and Korošec 1969, T. 47: 6a, b; Leghissa 2017, T. 93: 2
B1994	Deschmann's pile dwellings	L. b. variant of V. c.	μCT	21.42; 39.71	Korošec and Korošec 1969, T. 45: 6a, b; Leghissa 2017, T. 95: 5
B1505	Deschmann's pile dwellings	L. b. variant of V. c.	μCT	21.42; 39.71	Korošec and Korošec 1969, T. 42: 3a–c; Leghissa 2017, T. 96: 2
B5009	Deschmann's pile dwellings	L. b. variant of V. c.	μCT	21.42; 44.68	Korošec and Korošec 1969, T. 40: 4a, b; Leghissa 2017, T. 97: 3
NI19	Deschmann's pile dwellings	L. b. variant of V. c.	μCT	21.42; 39.71	Korošec and Korošec 1969, T. 39: 5; Leghissa 2017, T. 97: 4
B1479	Deschmann's pile dwellings	L. b. variant of V. c.	μCT	21.42; 39.71	Korošec and Korošec 1969, T. 40: 2a, b; Leghissa 2017, T. 103: 4
B1972	Deschmann's pile dwellings	L. c.	μCT	21.42; 39.71	Korošec and Korošec 1969, T. 47: 9; Leghissa 2017, T. 104: 3
B1973	Deschmann's pile dwellings	L. c.	μCT	39.71	Korošec and Korošec 1969, T. 49: 4; Leghissa 2017, T. 99: 2
B1490	Deschmann's pile dwellings	L. c.	μCT	21.42; 39.71	Korošec and Korošec 1969, T. 41: 7a, b; Leghissa 2017, T. 41: 7 a, b
B1497	Deschmann's pile dwellings	L. c.	μCT	39.71	Korošec and Korošec 1969, T. 38: 2a, b; Leghissa 2017, T. 109: 3
3469	Zingari cave	Influences of V. c.	μCT	30.95	Gilli and Montagnari Kokelj 1996, Fig. 34: 202
20591	Ciclami cave	Influences of V. c.	μCT	21.42; 42.19	Gilli and Montagnari Kokelj 1993, Fig. 51: 497
20592	Ciclami cave	Influences of V. c.	μCT	21.42; 38.99	Gilli and Montagnari Kokelj 1993, Fig. 37: 355
20419	Cotariova cave	Influences of V. c.	μCT	21.42; 43.33	Montagnari Kokelj et al. 2002, T. 27: 244
139461	Cotariova cave	Influences of V. c.	μCT	21.42	Montagnari Kokelj et al. 2002, T. 27: 245
SN	Cotariova cave	L.c.	μCT	21.42	Montagnari Kokelj et al. 2002, T. 27: 246
139462	Pettine cave	Influences of V. c.	μCT; OM	21.42	Marzolini 1983, Fig. 1: 22, 25
139463	Pettine cave	Influences of V. c.	μCT; OM	21.42	Marzolini 1983, Fig. 1: 11
139464	Edera cave	Influences of V. c.	μCT	21.42	Marzolini 1970, Fig. 2/1

L. b. Ljubljansko barje, *V. c.* Vučedol culture, *L. c.* Ljubljana culture

sections selected for all the artefacts. The ratio between the total areas of different pottery components can give interesting insights on how the paste of the vessels was produced (e.g. Bernardini et al. 2016).

Anyway, samples with a similar lithic inclusion/clay ratio could correspond to quite different pastes: the same ratio could be obtained considering a sample with very rare and big lithic grains or considering a vessel characterised by very abundant and small lithic inclusions. In order to overcome this bias, we have calculated, using Avizo v.8 software, the number, the area, the maximum length and the maximum width of every single lithic inclusion within the three virtual sections selected for all samples. After that, we have summed up the results for each vessel, putting together the data obtained from the segmentation and analysis of the three virtual sections.

To manage the resulting huge amount of data, we have divided our data sets (values for area, maximum length and width of all inclusions) into four intervals using Matlab R2016b.

The size of the intervals has been selected by taking into account minimum and maximum values of each variable and its distribution. This procedure was important to have the best

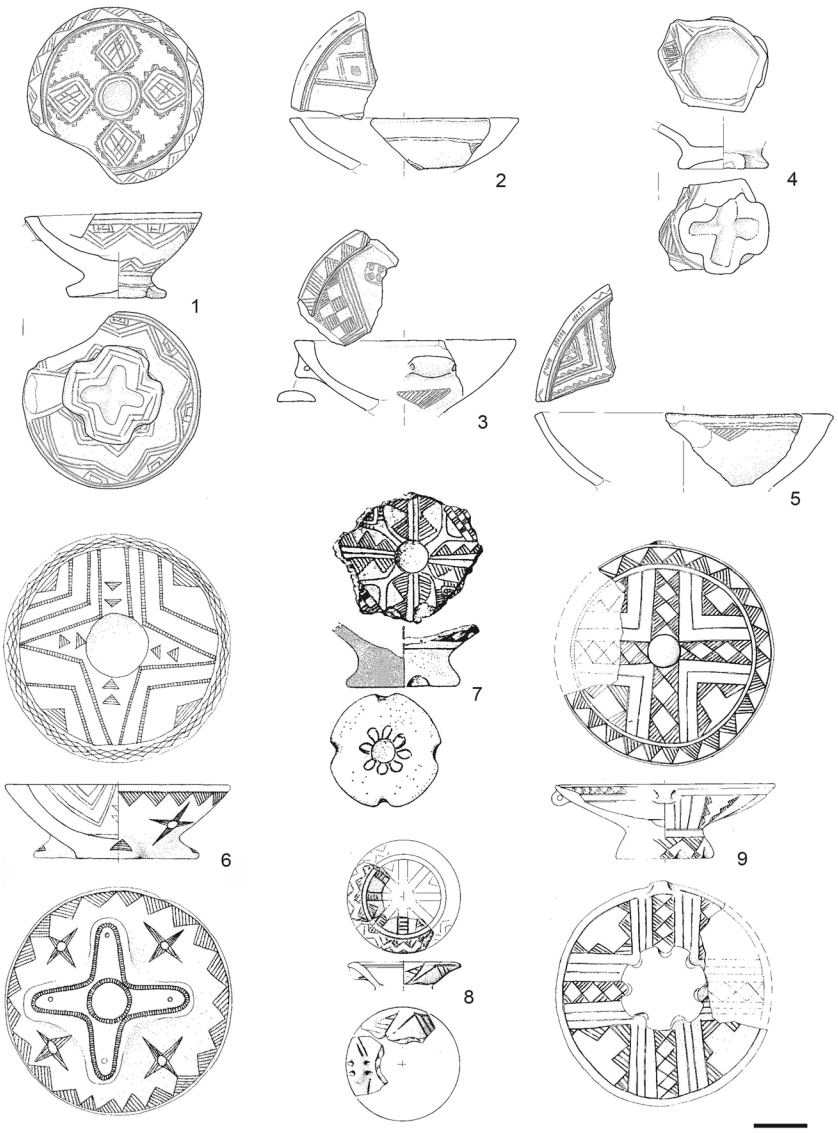

Fig. 2 Drawings of selected Copper Age bowls. 1, Deschmann's pile dwelling B1482; 2, Deschmann's pile dwelling B1984; 3, Deschmann's pile dwelling B1939; 4, Deschmann's pile dwelling B1505; 5, Deschmann's pile dwelling B1963; 6, Ciclami cave 20591; 7, Cotariova cave 20419; 8, Pettine cave 139463; 9, Ciclami cave 20592. Scale bar = 3 cm; for the references, see Table 1

representation of dimensional parameters describing the lithic inclusions within the paste, allowing a good separation between samples showing lithic grains of different sizes. As far as the area is concerned, the minimum value is 458 μm^2, the maximum value is 12,300,000 μm^2 and the intervals have been set as follows: very small inclusions, 0–50,000 μm^2; small inclusions, 50,000–100,000 μm^2; medium inclusions, 100,000–200,000 μm^2; and big inclusions, bigger than 200,000 μm^2. As far as the length is concerned, the minimum value is 29 μm, the maximum value is 5457 μm and the intervals have been set as follows: very small, 0–500 μm; small, 500–1000 μm; medium, 1000–1500 μm; and big, bigger than 1500 μm. As far as the width is concerned, the minimum value is 21 μm, the maximum value is 3653 μm and the intervals have been selected as follows: very small, 0–200 μm; small, 200–300 μm; medium, 300–400 μm; and big, bigger than 400 μm. The percentages of lithic inclusions falling within the defined size intervals have then been calculated.

In order to extract the maximum information possible from the data, principal component analysis (PCA) (e.g. Jolliffe 2002) has been performed using the programming language R (Kassambara 2017) and considering as variables inclusion/clay ratio and percentages of area, maximum length and maximum width. PCA is a multivariate analysis tool to reduce the dimensions of a given data set. It is based on the idea of rotating the original coordinate system to a new one. The direction of the new axes is then chosen to explain the maximum variance within the data set.

Optical microscopy

Small samples have been extracted from the two fragmented bowls of Pettine cave (139462, 139463) located in the Trieste

Table 2 Minerals identified through a stereomicroscope and/or a magnifying glass

Inventory number	Calcite	Limestone	Quartz	White mica	Observations
B1482	/	+++	/	/	/
B1963	+	/	/	/	/
B1984	/	/	/	/	Very fine-grained; no visible inclusions
B1965	/	/	/	+	/
B1939	+++	/	?	/	/
B1994	/	/	/	/	Very fine-grained; no visible inclusions
B1505	/	/	/	/	Very fine-grained; no visible inclusions
B5009	+++	/	/	/	/
NI19	+++	/	+	/	/
B1479	+++	/	/	/	/
B1972	/	+	/	/	/
B1973	+	/	/	/	/
B1490	++	?	/	/	Lithic fragments
B1497	+++	+++	/	/	Possible feldspar
3469	+++	+	/	/	/
20591	/	/	/	+++	/
20592	+++	/	/	/	/
20419	+	+	/	/	/
139461	++	+	/	/	/
SN	/	/	/	/	Very fine-grained; no visible inclusions
139462	/	/	/	/	Very fine-grained; no visible inclusions
139463	/	/	/	/	Very fine-grained; no visible inclusions
139464	+	+	/	/	/

+ = present; ++ = common; +++ = abundant

Karst, and they have been used to produce thin sections at the University of Padua. The thin sections have been observed via a polarising microscope at the Department of Mathematics and Geosciences of the University of Trieste. Unfortunately, we have not been allowed to take samples from the other artefacts.

Results

Pastes

The observation of the surfaces of the artefacts has allowed to identify lithic inclusions in many samples (Table 2) showing a

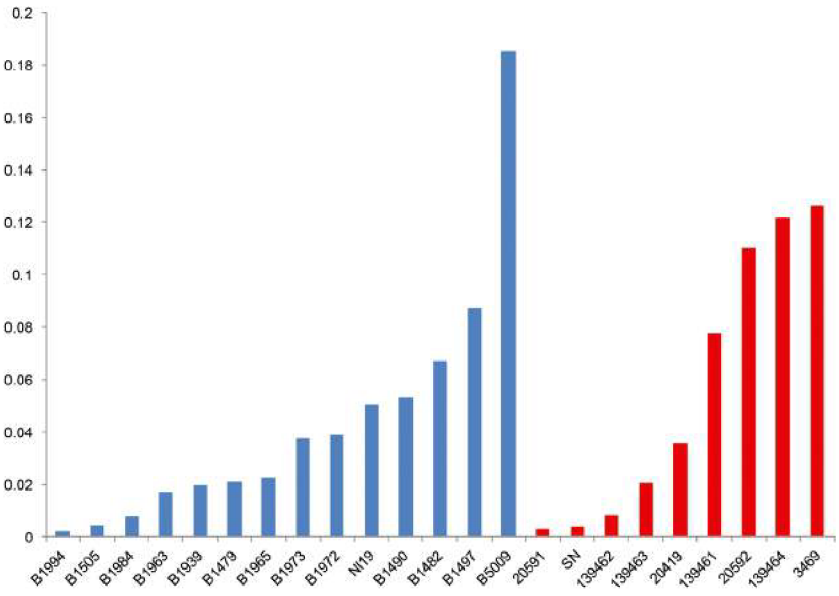

Fig. 3 Lithic inclusion/clay ratio for the Slovenian (light blue) and Italian samples (red)

coarse-grained paste. Among the identified inclusions, calcite and fragments of limestones are the most abundant, while quartz and white mica, probably muscovite, are rare. It is worth to mention bowl 20591 from the Trieste Karst because it differs from all other samples for the presence of abundant white mica. A few crystals of the same mineral have been recognised in sample B1965 from Deschmann's pile dwellings.

The segmentation of microCT data sets has allowed the virtual separation of the clay matrix, lithic inclusions and voids in 2D and their quantification. According to observations of microCT-derived virtual slices and the resulting lithic inclusion/clay ratios, fine- and coarse-grained pastes have been recognised in both Slovenian and Italian samples but no sharp distinctions can be made between sample groups (Fig. 3). The coarse-grained samples show abundant temper material with a size up to very fine gravel (from < 1 to 2–3 mm; lithic inclusion/clay ratio from about 0.04 up to 0.2), while the fine-grained vessels are variably tempered with lithic inclusions whose size is generally smaller than 1 mm (lithic inclusion/clay ratio lower than 0.04).

MicroCT data have revealed the presence of calcite, identifiable through the rhombohedral shape of crystals; clay pellets; grog fragments; rounded concentric inclusions often with a lighter centre, probably derived from the transformation of primary minerals during firing; low-density inclusions; and bone remains. The components identified in each sample are summarised in Table 3. It is not totally clear what the low-density inclusions correspond to, but their density is lower than that of clay, and they could be identified as charcoal fragments, macroscopically observed in some specimens.

Among Slovenian samples, some (Fig. 4; e.g. samples B1994, B1505, B1984, B1963, B1939, B1479 and B1965) are characterised by very fine- or fine-grained pastes with a lithic inclusion/clay ratio between 0.002 and 0.03 (Fig. 3). Small calcite crystals, clay pellets, rounded dense inclusions and bone remains have been identified in some samples (Fig. 4, Table 3). In sample B1939, a few small fish vertebrae have been identified and one of them has been virtually extracted (Fig. 5).

In medium- and coarse-grained vessels, calcite crystals have been often recognised (Table 3, Fig. 6).

Despite the small number of bowls discovered in the Karst area, they show quite heterogeneous features (Table 3, Fig. 7). Sample 20591 is different from all the other ones due to its very fine-grained (lithic inclusion/clay ratio 0.002) and dense paste where clay pellets and very rare and dense small inclusions have been imaged. A probable fish vertebra has been identified in sample 139464.

Table 3 Paste components identified through microCT. *D.p.d.* Deschmann's pile dwellings

Inventory number	Site	Grain size	Calcite	Grog	Clay pellets	Bones	Concentric inclusions	Low-density inclusions
B1482	Dpd	Coarse	x	/	/	/	/	/
B1963	Dpd	Fine	x	/	x	/	x	/
B1984	Dpd	Very fine	?	x	x	/	x	x
B1965	Dpd	Fine	?	/	/	/	x	/
B1939	Dpd	Fine	?	/	x	Fish vertebrae	/	x
B1994	Dpd	Very fine	/	/	/	/	x	/
B1505	Dpd	Very fine	?	/	x	/	x	x
B5009	Dpd	Coarse	x	/	x	/	/	/
NI19	Dpd	Coarse	x	/	x	/	/	/
B1479	Dpd	Fine	x	x	x	/	/	/
B1972	Dpd	Fine	x	/	x	/	x	/
B1973	Dpd	Fine	?	/	x	x	/	/
B1490	Dpd	Medium	x	/	x	x	x	/
B1497	Dpd	Coarse	x	/	/	?	/	x
3469	Zingari cave	Coarse	x	/	/	/	x	/
20591	Ciclami cave	Very fine	/	/	x	/	/	/
20592	Ciclami cave	Coarse	x	/	x	/	/	/
20419	Cotariova cave	Coarse	x	/	x	x	/	/
139461	Cotariova cave	Fine	x	/	/	/	/	/
SN	Cotariova cave	Fine	/	?	x	/	x	/
139462	Pettine cave	Fine	?	/	x	/	x	/
139463	Pettine cave	Fine	x	x	x	/	x	/
139464	Edera cave	Coarse	x	/	/	Fish vertebra?	x	/

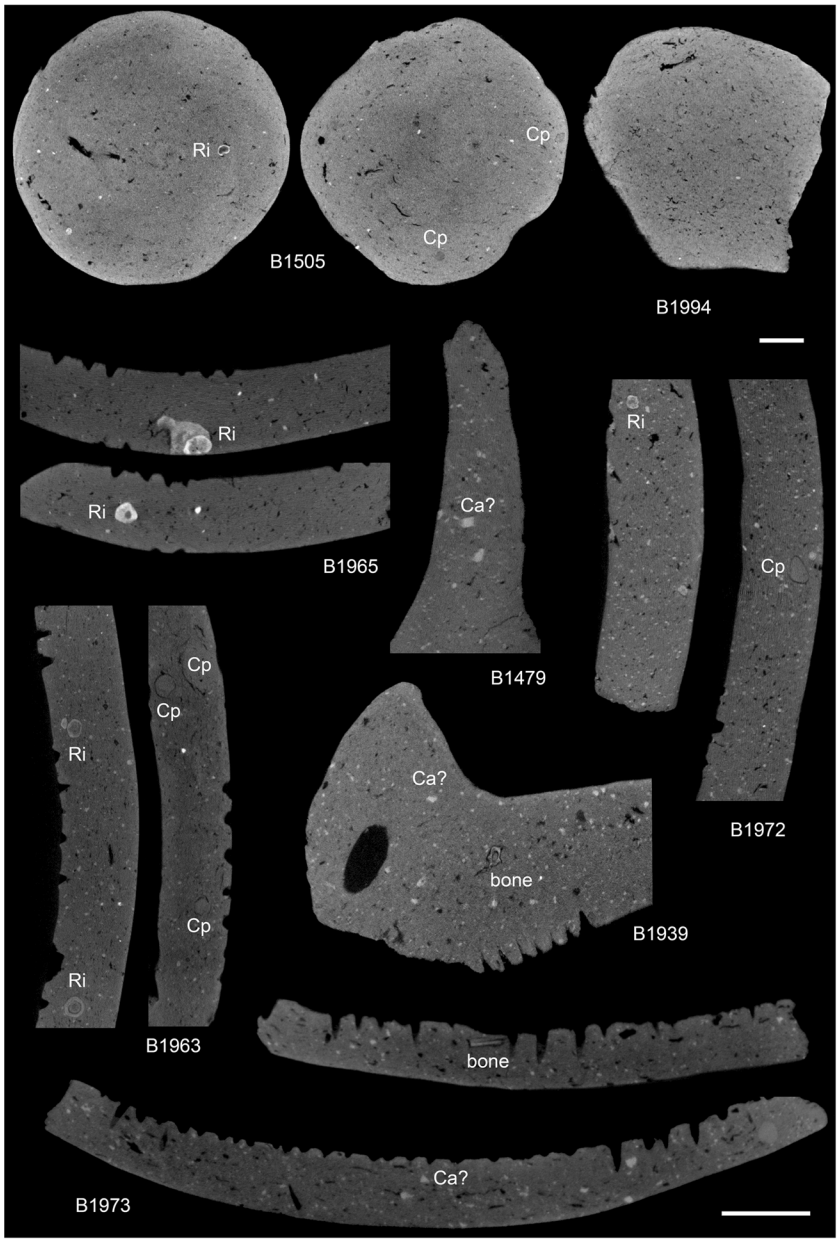

Fig. 4 Virtual sections of selected fine-grained bowls from Deschmann's pile dwellings. Ca, calcite; Cp, clay pellet; Ri, rounded inclusion; scale bars = 1 cm

Only the two bowls from Pettine cave were sampled to produce thin sections. Sample 139462 shows a brown fabric characterised by quite abundant angular quartz with a grain size from silt to medium sand, quite abundant flint fragments up to 0.5 mm large and abundant grog fragments. The grog fragments present different fabric and colours, suggesting they were obtained by crushing different vessels. In addition, a quartzite fragment containing muscovite crystals has been identified (Fig. 8). Sample 139463 shows a reddish paste containing similar components, such as quartz, angular flint grains, muscovite and grog fragments, but also rare calcite. It is worth mentioning that some grog fragments contain abundant flint fragments (Fig. 8).

Technology

Discontinuities and pores within the paste can give information about the bowl-shaping techniques. Lindahl and Pikirayi (2010) have demonstrated that the orientation of pores within the ceramic body, from one wall surface to the other, can be used to distinguish vessel-forming techniques, in particular between U and N coiling and modelling techniques. The U coiling technique gives curved convex sub-parallel discontinuities, while N coiling technique produces sub-parallel pores with a diagonal orientation. In vessels produced by modelling techniques, the discontinuities are parallel to the wall surfaces.

According to virtual cross sections of the rims of the investigated bowls, the upper part of the vessels was produced

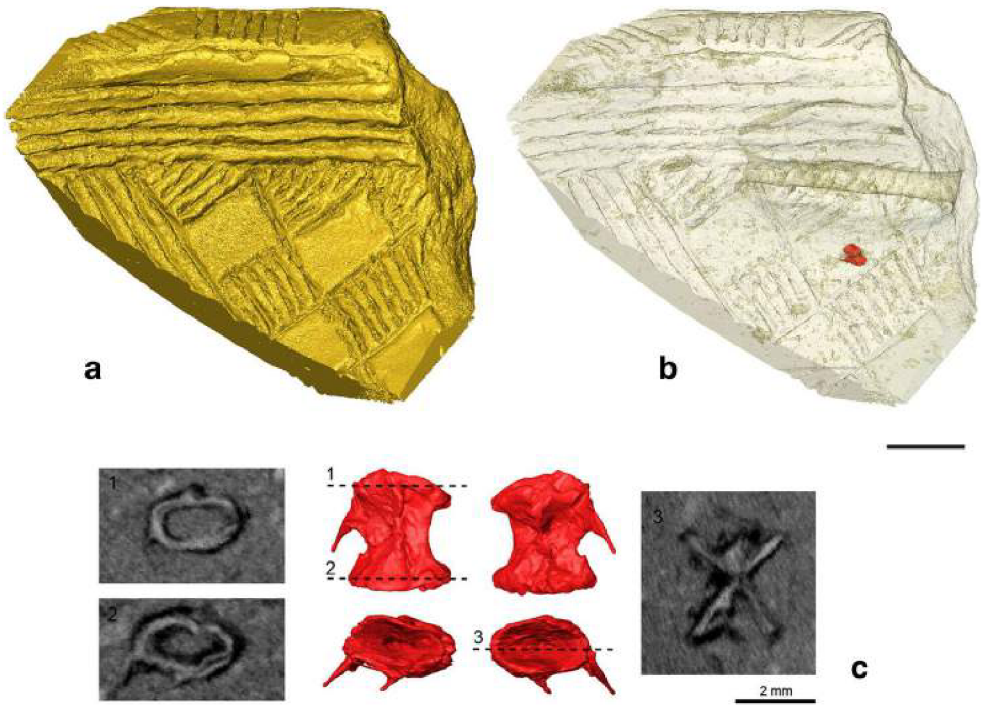

Fig. 5 Virtual extraction of a fish vertebra from sample B1939. **a** Virtual rendering of the artefact. **b** Virtual rendering of the artefact with the paste in transparency and the vertebra in red. **c** Extraction of the vertebra and selected virtual sections. Where not specified, the scale bar corresponds to 1 cm

using both the coiling and the modelling techniques. Samples B1994, B1490 and 139464 show pores with diagonal orientation and were therefore produced by N coiling technique, while sample 3469 shows pores parallel to the surfaces of the vessel, produced by a modelling technique (Fig. 9).

In virtual sections of both Italian and Slovenian samples, the core of the base is surrounded by a circular concentric layer of paste. This suggests the base was shaped starting from a central piece, to which a layer of clay was added and then reshaped in order to produce the foots (Fig. 10). The upper part of the bowl was probably separately produced and then joined to the base as it is suggested by the discontinuities in longitudinal virtual sections of sample B1482 (Fig. 11).

Analysis of microCT-derived data

After the segmentation of clay, lithic inclusions and pores in three slices per sample (Supplementary Table 1), we calculated the area, the maximum length and the maximum width of every single lithic inclusion. Table 4 shows the number of lithic inclusions divided into four size intervals considering their area, maximum length and maximum width.

In Fig. 12, we have plotted the percentages of inclusions falling within the different size intervals considering their area, maximum length and maximum width. Comparing Table 4 and Fig. 12, the differences between samples can be observed.

When the area of inclusions is considered (Fig. 12), several samples (B1479, B1939, B1963, B1965, B1972, B1973 and 139461–139464, SN) show a higher percentage of very small inclusions and a gradual decrease in the percentage of small, medium and big inclusions. Sample 20591 has a peculiar paste characterised by rare very small and small inclusions without any medium and big ones. Other samples, such as samples B1497, N19 and 3469, show a completely different distribution of inclusions with a prevalence of very big lithic grains. Finally, other samples show similar percentages of inclusions belonging to all four classes or a prevalence of very small inclusions coupled with a similar lower percentage of inclusions belonging to other classes (20419 and 20592).

Considering the maximum length of inclusions, a prevalence of very small inclusions (< 500 μm) has been detected in most samples. However, in some samples with a coarse grain size, the percentage of small inclusion (500–1000 μm) is similar to that one of very small inclusions (Fig. 12). The distribution of inclusions considering their maximum width is quite well comparable to that one obtained, taking into consideration the areas (Fig. 12).

To visualise and summarise the pottery fabric information given by microCT-derived inter-correlated data, PCA has then been performed, considering percentages of lithic inclusions, area, length, width and inclusion/clay ratio as variables. Figure 13a shows the relationships between variables and the quality of their representation. Variables positively correlated are grouped together, while those negatively correlated are located in opposite quadrants. The quality of representation can be expressed by a value of square cosine (cos^2). The distance between the end points of variables measures the quality of the variables. The best representation of a variable is given when its end lies on the correlation circle, a condition corresponding to the maximum square cosine value, that is 1. In Fig. 13a, variables are well represented and three main groups of variables are positively

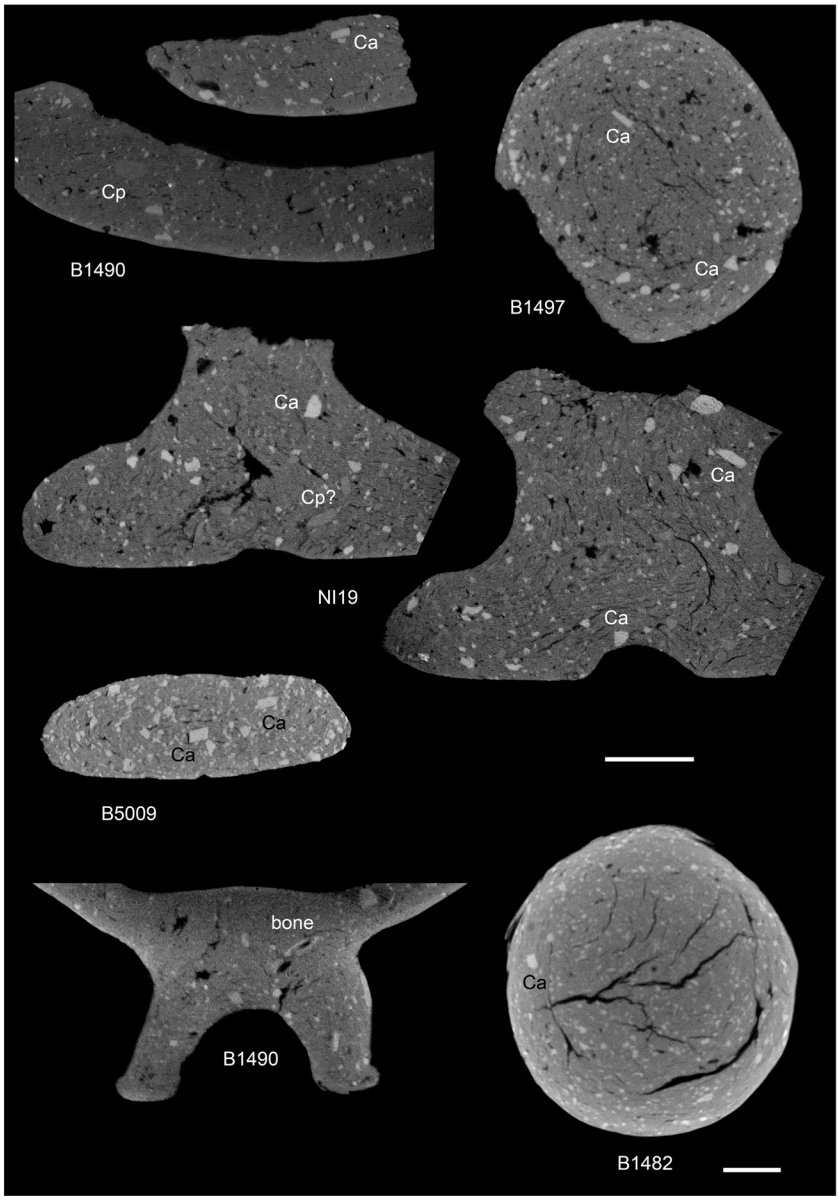

Fig. 6 Virtual sections of selected coarse-grained bowls from Deschmann's pile dwellings. Ca, calcite; Cp, clay pellet; scale bars = 1 cm

correlated: very small area-very small length-very small width, small area-small width and small length-medium area-medium length-medium width-big area-big length-big width.

In the bivariate plot of Fig. 13b, we can observe the position of the samples with respect to principal components 1 and 2 and the variables. When a vessel falls close to a given variable, it has a high value of such a variable, while its value is low if it is located on the opposite side of the same variable.

The samples falling in the top right quadrant are characterised by very fine-grained fabric with prevalent very small inclusions. Most of them are from the Karst with exception of sample B1963, that is very close to the bottom right quadrant. In this quadrant, there are mainly fine-grained Slovenian samples and two bowls from Pettine cave (Trieste Karst). This group of materials is characterised by a high total clay area because they lie on the opposite side of the lithic inclusion/clay ratio variable. Samples with medium and big inclusions and a big lithic inclusion/clay ratio mainly fall in the top left quadrant. Even if this diagram simply provides a description of pottery fabric and lithic inclusions, it is interesting to note that in the bottom quadrants, all the samples, with the exception of two Italian bowls, are from Deschmann's pile dwellings.

Discussion and conclusions

The orientation of pores and disjunctions in microCT volumes has allowed recognising the bowl-forming process in several

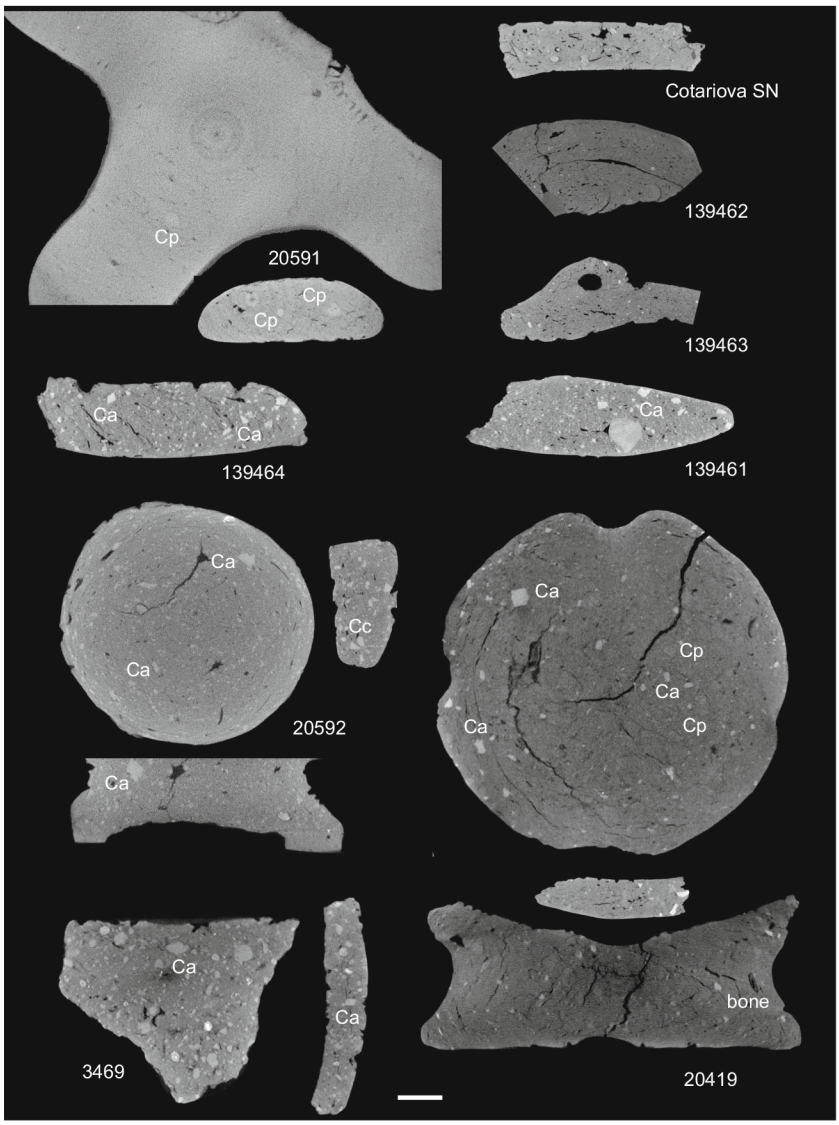

Fig. 7 Virtual sections of all analysed bowls from the Karst. Ca, calcite; Cp, clay pellet; scale bar = 1 cm

artefacts. The basin was produced using mainly the N coiling technique, recognised in the rim of three artefacts (B1994, B1490 and 139464), but also the modelling one (3469; Fig. 9). The base was shaped starting from a central piece, to which a layer of clay was added and then reshaped in order to produce the foots.

MicroCT analysis of the bowls with cross-shaped foots from Deschmann's pile dwellings has showed the use of two main paste types. The first one includes bowls with a medium or coarse-grained paste tempered with abundant and generally poorly sorted carbonate/calcite inclusions, identified through both macroscopic and microCT observations (B1482, B1490, B1497, B5009, N19). The lithic temper material has a size up to very fine gravel (from <1 to 2–3 mm), and the paste lithic inclusion/clay ratio spans from about 0.04 up to 0.2. The big lithic inclusions (> 200,000 μm^2) are prevalent (B1497, N19) or abundant (B1490, B5009; Figs. 3, 6 and 12). In the PCA bivariate plot, such vessels fall relatively close to each other in the top left quadrant, with the exception of sample B1482 (Fig. 13).

The second group is larger and includes very fine- or fine-grained vessels. Some of them show a paste almost without temper (B1994, B1505, B1984), while the others generally contain well-sorted fine calcite/limestone grains with a prevalence of very small inclusions (< 50,000 μm^2). Consequently, the inclusion/clay ratio is very low and spans between 0.002 and 0.03. Grog, clay pellets, rounded dense inclusions and bone remains have been identified in some of them (Figs. 3, 4 and 12, Table 3). A few small fish vertebrae have been identified in sample B1939. Most of the samples fall in the bottom right quadrant of PCA bivariate plot (Fig. 13).

The two main pastes were used regardless of the bowl size and quality of decorations and manufacture. Calcite and limestone fragments, used as tempered material, are already reported from pottery assemblages of the Ljubljansko barje dated to

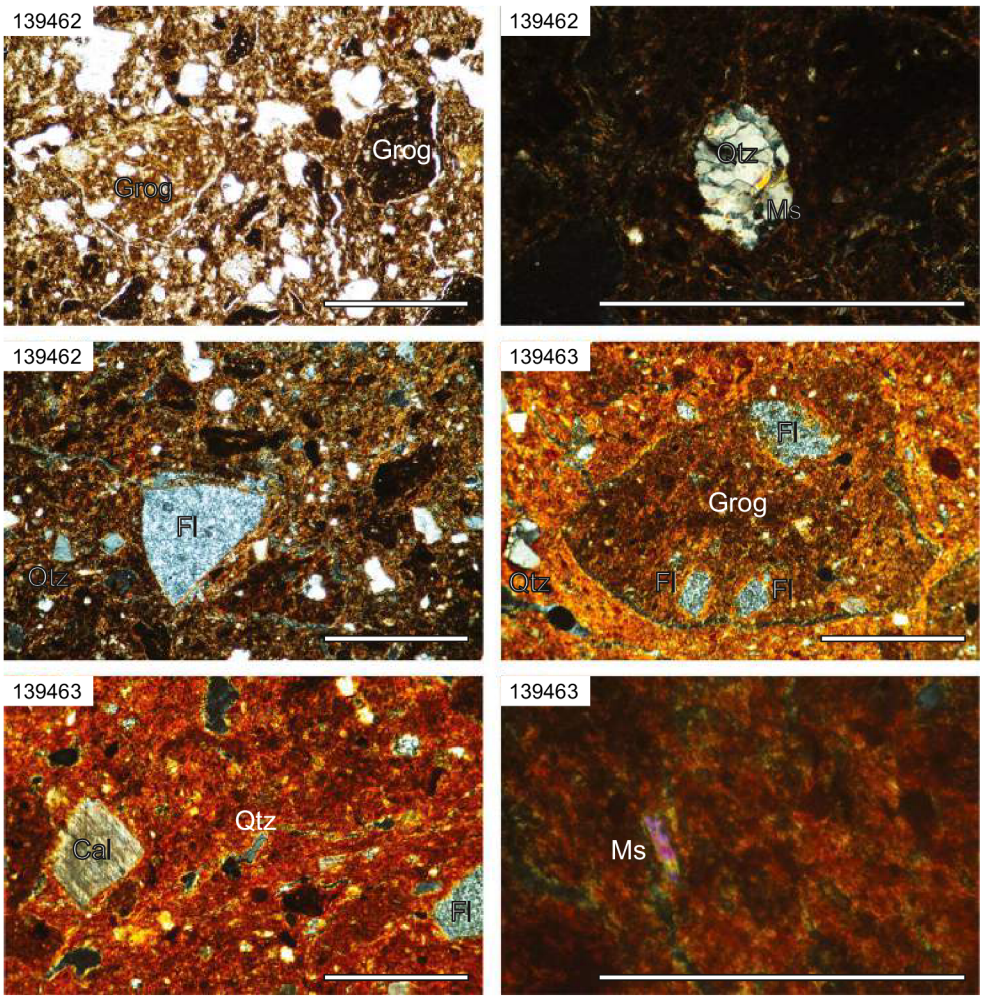

Fig. 8 Microphotographs of thin sections produced from bowls 139462 and 139463 from Pettine cave. The top left image was taken using plane-polarised light while all the others using crossed polars; scale bars = 0.5 mm. Cal, calcite; Fl, flint; Ms, muscovite; Qtz, quartz

the fifth and fourth millennia BC, and they could be easily gathered from the karst outcrops south of the Ljubljansko barje (Žibrat Gašparič 2013). The presence of fish remains within the paste would suggest that the clay raw material of sample B1939 was likely collected from local lake deposits.

The results of microCT analysis support the hypothesis that the bowls were locally produced using different recipes. This is not surprising if we consider that the investigated material come from three pile dwellings in use for a few centuries. Non-destructive chemical analysis by prompt-gamma activation analysis (PGAA) of the same vessels and natural clay samples from Ljubljansko barje is in progress to confirm such interpretation.

Even if the group of bowls from Trieste Karst includes only nine vessels, they show very heterogeneous pastes.

Sample 20591 from Ciclami cave is different from all the investigated Slovenian and Italian artefacts. Its decoration shape and technique are not reported from the bowls with cross-shaped foots from Deschmann's pile dwellings. Its paste is rich of muscovite mica, very dense and without carbonate

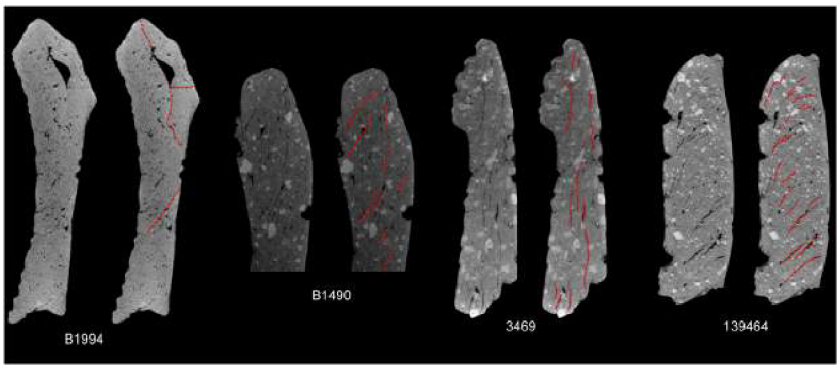

Fig. 9 Virtual transversal sections of the selected rims of investigated bowls showing technological traces highlighted by dotted red lines

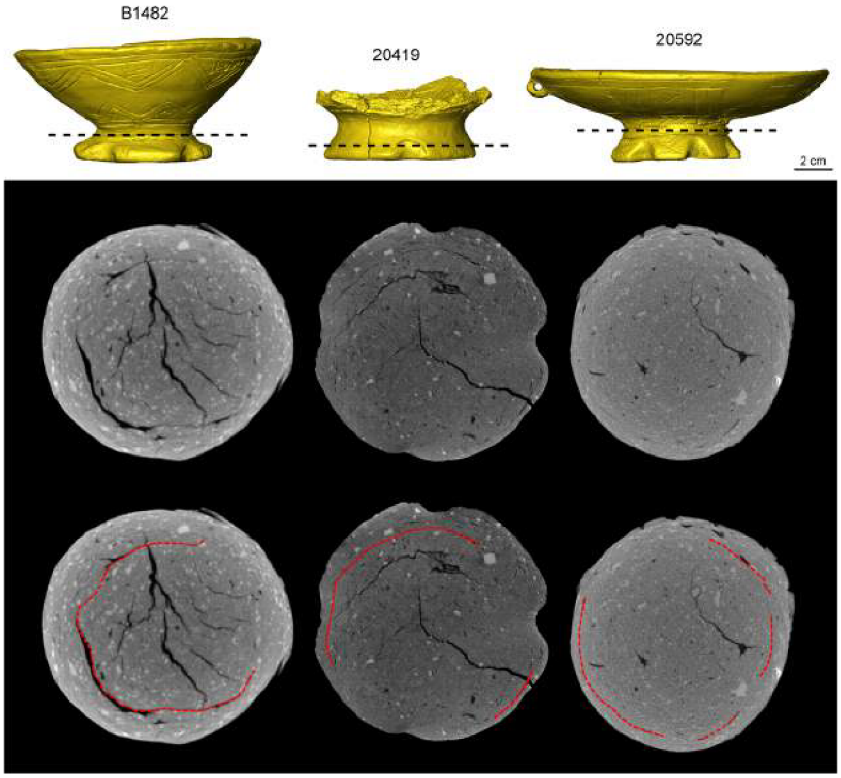

Fig. 10 Virtual transversal sections of the base of selected bowls from Slovenia (B1482) and Italy (20419 and 20592). Technological traces are highlighted by dotted red lines

inclusions and contains only a few dense very small and small lithic grains (lithic inclusion/clay ratio 0.002). Neolithic pottery from the Trieste Karst is very rich in calcite inclusions, and abundant muscovite has not been reported so far (Spataro 1999; Bernardini et al. 2016). For all these reasons, bowl 20591 has likely been imported to the Karst area. According

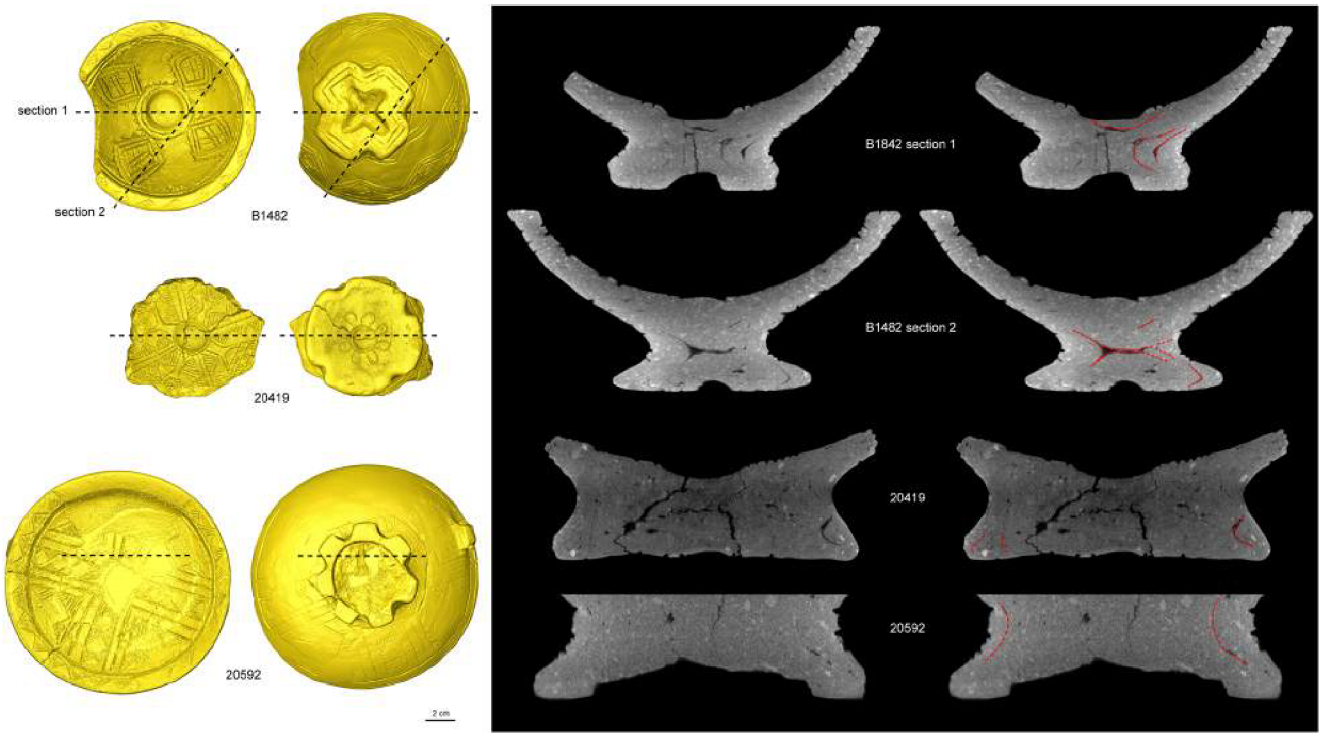

Fig. 11 Virtual longitudinal sections of the base of selected bowls from Slovenia (B1482) and Italy (20419 and 20592). Technological traces are highlighted by dotted red lines

Table 4 Number of lithic inclusions divided into four size intervals considering their area, maximum length and maximum width

Name	Area				Maximum length				Maximum width			
	Very small 0–50,000 μm²	Small 50,000–100,000 μm²	Medium 100,000–200,000 μm²	Big > 200,000 μm²	Very small 0–500 μm	Small 500–1000 μm	Medium 1000–1500 μm	Big > 1500 μm	Very small 0–200 μm	Small 200–300 μm	Medium 300–400 μm	Big > 400 μm
B1505	30	32	18	10	60	27	3	0	20	30	24	16
B5009	220	155	180	206	345	288	86	42	136	184	150	291
B1482	237	309	306	157	509	449	47	4	156	313	249	291
B1490	73	66	75	70	134	113	25	12	49	69	53	113
B1479	78	68	32	20	130	59	7	2	55	55	48	40
B1939	126	99	46	20	218	70	2	1	82	96	73	40
B1963	255	54	16	7	309	22	1	0	185	109	25	13
B1965	24	17	4	7	40	9	2	1	13	21	7	11
B1972	257	118	44	13	353	73	4	2	183	148	67	34
B1973	101	87	58	30	198	68	4	6	80	89	62	45
B1984	1	7	9	7	8	12	1	3	0	6	9	9
B1994	3	5	1	2	7	3	1	0	2	6	0	3
B1497	193	185	339	453	413	545	162	50	173	141	257	599
NI19	66	94	106	116	145	165	52	20	34	83	84	181
3469	59	80	65	102	130	126	29	21	29	72	69	136
20591	86	9	0	0	89	6	0	0	77	14	3	1
20592	28	85	185	489	82	347	239	119	22	35	103	627
20419	10	30	46	109	33	92	45	25	3	27	32	133
139461	562	74	40	28	626	66	8	4	478	143	31	52
SN	33	4	6	1	38	6	0	0	25	11	5	3
139462	54	27	18	15	79	31	4	0	34	36	22	22
139463	116	47	16	10	153	31	4	1	95	60	19	15
139464	345	140	69	68	457	131	22	12	264	169	87	102

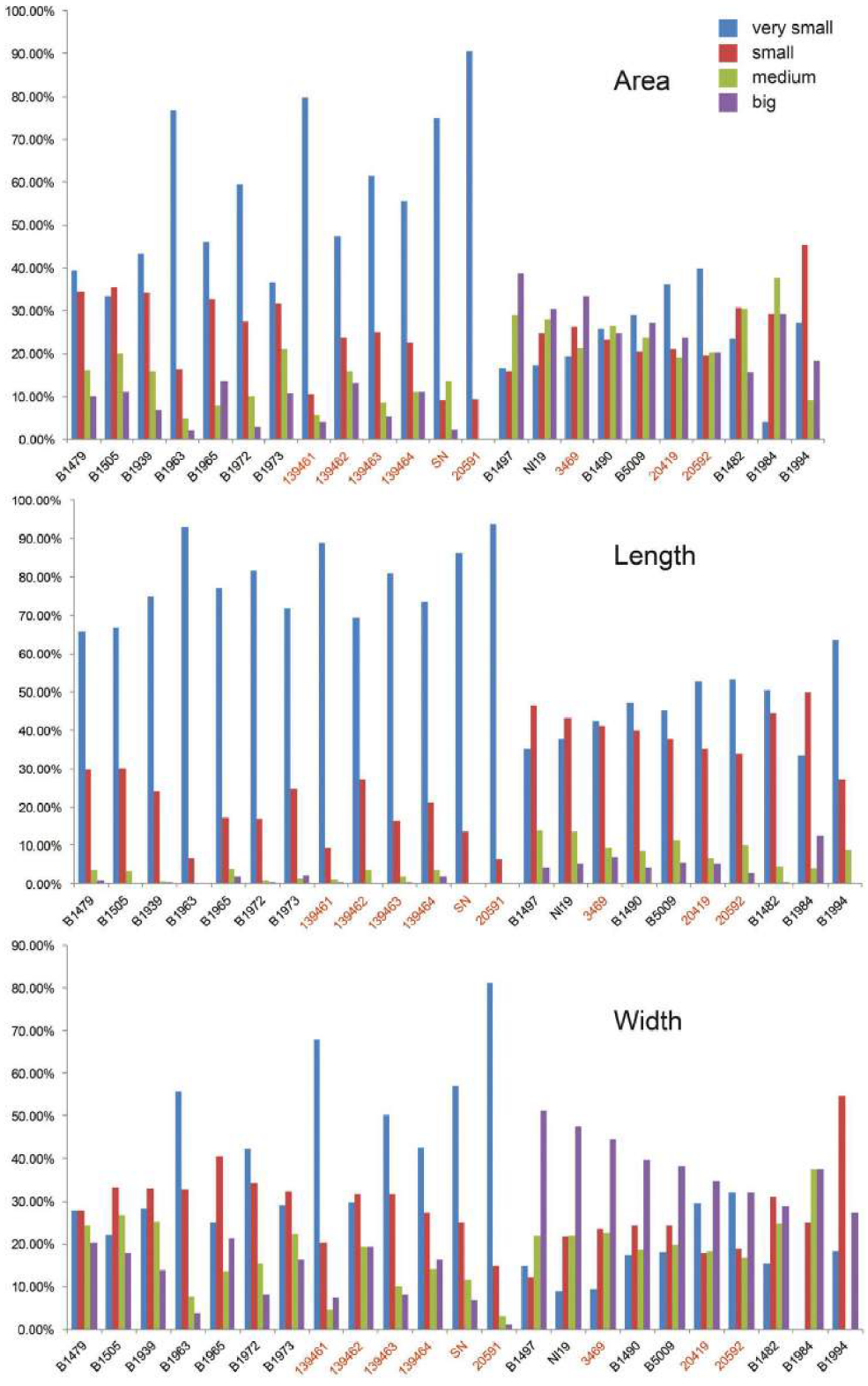

Fig. 12 Histograms showing the percentages of very small, small, medium and big lithic inclusions considering their area, length and width. Red labels, Karst samples; black labels, Slovenian samples

to available data, its technological features are different to those so far identified in the analysed artefacts from Deschmann's pile dwellings too. Bowls with similar decorations are known from Austria, Czech Republic, Slovakia, Hungary, Serbia, etc., as far as Ukraine (see Leghissa 2017, p. 172). The most similar bowls are some vessels from Moravia in the Czech Republic (Medunová-Benešová 1977, T. 47: 1a–c, 5a, b, 6a, b; 48: 1a–c; 49: 1a–c; 48). Nevertheless,

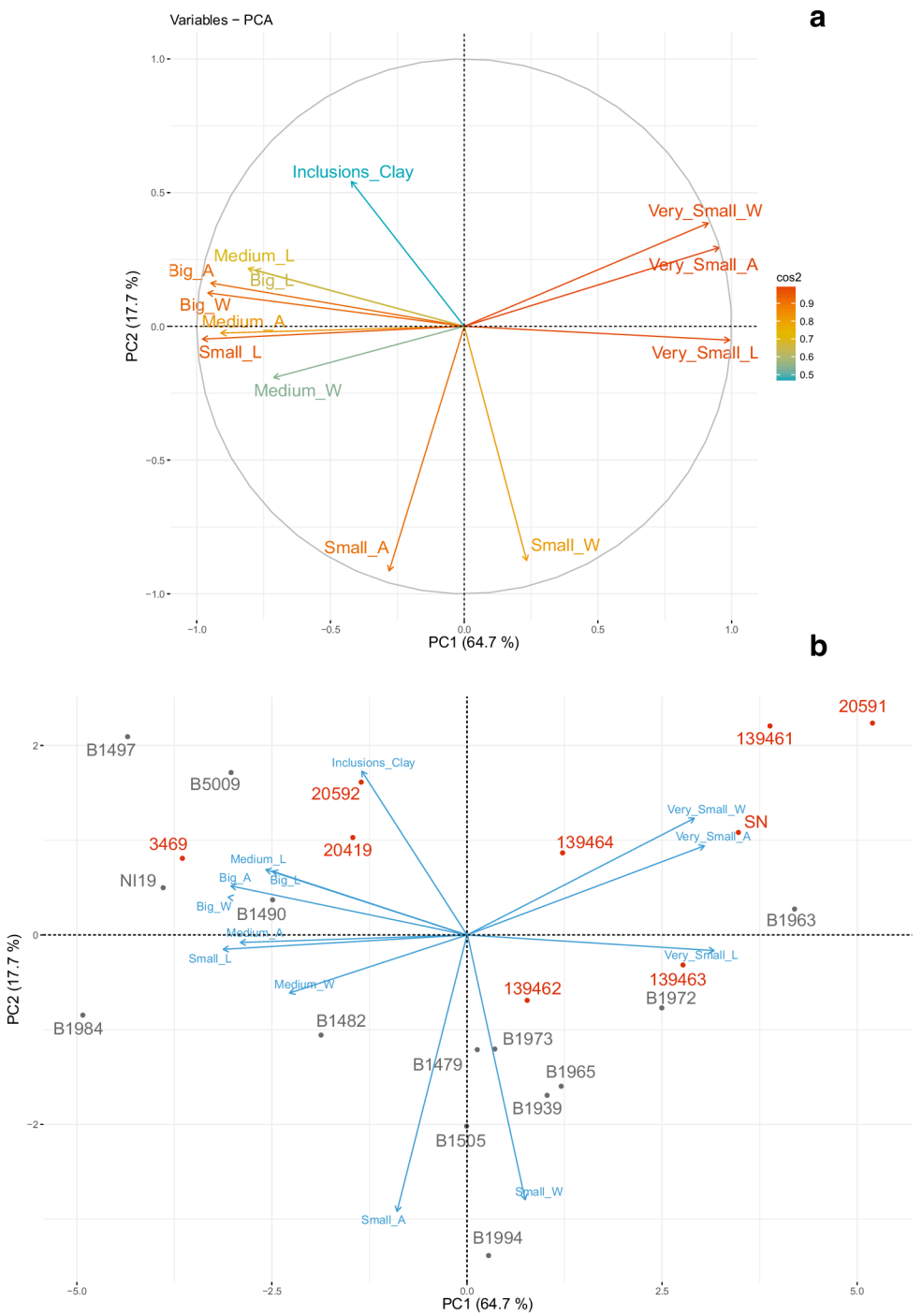

Fig. 13 **a** Variable correlation plot showing the relationships among variables and the quality of their representation. **b** Bivariate plot showing the variables and the position of investigated samples from Slovenia (black labels) and Trieste Karst (red labels). A, area; L, length; W, width; Inclusions_Clay, lithic inclusion/clay ratio

the precise origin of bowl 20591 from Ciclami cave cannot be precisely identified without further investigations.

According to microCT results, the bowls 139462 and 139463 from Pettine cave show a fine-grained paste (lithic inclusion/clay ratios 0.008 and 0.02, respectively) with a prevalence of very small lithic inclusions (Fig. 12) comparable to that of several samples from Deschmann's pile dwellings (Fig. 3). In vessel 139463, calcite crystals and grog fragments have been identified, such as in the Slovenian samples B1479 and B1984. The two bowls fall into the bottom right quadrant of PCA plot, very close to the fine-grained group of Slovenian samples, and they are the only two Italian artefacts included in the two bottom quadrants of the same diagram, indicating a similitude with a large group of bowls from Deschmann's pile dwellings. These data could suggest a possible origin from central Slovenia or surrounding areas, and optical microscopy supports this hypothesis. Abundant grogs and angular flint grains, quartzite and muscovite crystals are a pottery

component that is rare or so far not reported in the prehistoric ceramics of the Trieste Karst (Spataro 1999; Bernardini et al. 2016), while they have been identified in those of Ljubljansko barje. Among the pottery fabric groups of Resnikov prekop, a fifth millennium Neolithic site very close to Deschmann's pile dwellings, fabric groups 3 and 4 are rich in flint and muscovite (Žibrat Gašparič 2013).

Nevertheless, it is worth stressing that the decoration (i.e. impressions of twisted cord) of the two bowls from Pettine cave is not reported among the bowls with cross-shaped foots from Deschmann's pile dwellings but only, very rarely, on other types of vessels. For this reason, we cannot exclude that the Pettine bowls could have been imported from a different region, as supposed for the Ciclami bowl 201591. Chemical data obtained by PGAA (in progress) could help in answering to this question.

The coarse-grained samples (3469, 20419 and 20592) are mainly calcite-tempered pottery with lithic inclusion/clay ratios from about 0.035 to 0.12 and a prevalence or high percentage of big inclusions (> 200,000 μm^2) (Fig. 12). They are technologically similar to the Slovenian bowls with a medium- or coarse-grained paste. This is confirmed by the PCA plot where Karst samples fall into the top left quadrant (Fig. 13). However, sample 3469 is the only one where the orientation of pores suggests it was shaped using a modelling technique.

The last three bowls from the Trieste Karst (SN, 139461 and 139464) contain prevalent very small inclusions but the calcite-tempered 139461 and 139464 artefacts show high lithic inclusion/clay ratios (0.07 and 0.12, respectively) due to the presence of a few very large inclusions. Conversely, the bowl SN does not contain calcite and has a low lithic inclusion/clay ratio (0.03). They fall into the top right quadrant of PCA plot. The possible presence of a fish vertebra in sample 139464 would suggest that the clay raw material was gathered from a lacustrine environment, such as the one of Ljubljansko barje at that time.

The origin and function of the bowls with cross-shaped foots is still debated (see e.g. Kulcsár 2009; Kaiser 2013), but the small number and beautiful decoration of those from the Trieste caves suggest that they could have been used as valuable ceremonial items. The probable origin of part of them from central Slovenia or even more distant regions would indicate that sometimes not only aesthetic and technological models but also the objects themselves moved from other regions to the Karst. The strong connections between central Slovenia and the Karst during the first half of the third millennium BC are confirmed by other classes of artefacts, such as copper and polished stone axes (Bernardini et al. 2014a, 2014b; Bernardini 2018 and the references quoted there).

From a methodological point of view, the present paper provides one of the first examples of statistical analysis of ceramic microCT-derived data, allowing a non-destructive objective description of pottery fabrics.

Acknowledgments We are grateful to Margherita Di Giovannantonio for the language review of the paper.

Funding information David Prokop has been financially supported by the Ministry of Education, Youth and Sports of the Czech Republic under the project CEITEC 2020 (LQ1601) and by the project CEITEC Nano Research Infrastructure (MEYS CR, 2016–2019). Sandro Donato is partially supported by Consorzio per la Fisica Trieste. This research has been supported by Fondo Ricerca Ateneo 2018 of Trieste University in the framework of the project SCAN (Studio mineralogico, petrografico e strutturale di Ceramiche ANtiche tramite metodologie convenzionali e innovative).

References

Bernardini F (2018) Polished stone axes in Caput Adriae from the Neolithic to the Copper Age. J World Prehist 31(4):485–514

Bernardini F, Tuniz C, Coppa A, Mancini L, Dreossi D, Eichert D, Turco G, Biasotto M, Terrasi F, De Cesare N, Hua Q, Levchenko V (2012) Beeswax as dental filling on a neolithic human tooth. PLoS One 7(9):e44904

Bernardini F, Sgambati A, Montagnari Kokelj M, Zaccaria C, Micheli R, Fragiacomo A, Tiussi C, Dreossi D, Tuniz C, De Min A (2013) Airborne LiDAR application to karst areas: the example of Trieste Province (North-Eastern Italy) from prehistoric sites to Roman forts. J Archaeol Sci 40:2152–2160

Bernardini F, De Min A, Lenaz D, Kasztovszky Z, Turk P, Velušček A, Szilágyi V, Tuniz C, Montagnari Kokelj E (2014a) Mineralogical and chemical constraints about the provenance of Copper Age polished stone axes of "Ljubljana type" from Caput Adriae. Archaeometry 56(2):175–202

Bernardini F, De Min A, Lenaz D, Kasztovszky Z, Turk P, Velušček A, Tuniz C, Montagnari Kokelj E (2014b) Petrographic and geochemical comparison between the Copper Age "Ljubljana type" axes and similar lithotypes from Eisenkappler Diabaszug complex (southern Austria). J Archaeol Sci 41:511–522

Bernardini F, Vinci G, Horvat J, De Min A, Forte E, Furlani S, Lenaz D, Pipan M, Zhao W, Sgambati A, Potleca M, Micheli R, Fragiacomo A, Tuniz C (2015) Early Roman military fortifications and the origin of Trieste (Italy). Proc Natl Acad Sci U S A 112:1520–1529

Bernardini F, Vecchiet A, De Min A, Lenaz D, Mendoza Cuevas A, Gianoncelli A, Dreossi D, Tuniz C (2016) Neolithic pottery from the Trieste Karst (Northeastern Italy): a multi-analytical study. Microchem J 124:600–607

Bernardini F, Sibilia E, Kasztovszky Z, Boscutti F, De Min A, Lenaz D, Turco G, Micheli R, Tuniz C, Montagnari Kokelj M (2017) Evidence of open-air late prehistoric occupation in the Trieste area (North-Eastern Italy): dating, 3D clay plaster characterization and obsidian provenancing. Archaeol Anthropol Sci 10:1933–1943. https://doi.org/10.1007/s12520-017-0504-7

Bóna I (1965) The peoples of southern origin of the Early Bronze Age in Hungary I-II. Alba Regia 4–5:17–63

Buchvaldek M (1967) Die Schnurkeramik in Böhmen. Acta Universitatis Carolinae XIX, Philosophica et Historica Monographia XIX, Prague

Dimitrijević S (1956) Prilog daljem upoznavanju vučedolske kulture / Ein Beitrag zur weiteren Kenntnis der Vučedoler kultur. Opusc Archaeol 1:79–85

Dimitrijević S (1966) Rezultati arheoloških iskopavanja na području Vinkovačkog muzeja od 1957–1965 godine. Acta Musei Cibalensis 1, Vinkovci

Dimitrijević S (1979) Vučedolska kultura i vučedolski kulturni kompleks. In: Benac A (ed) Praistorija jugoslovanskih zemalja III. Eneolit, Sarajevo, pp 267–341

Duches R, Nannini N, Fontana A, Boschin F, Crezzini J, Bernardini F, Tuniz C, Dalmeri G (2018) Archeological bone injuries by lithic backed projectiles: new evidence on bear hunting from the Late Epigravettian site of Cornafessa rock shelter (Italy). Archaeol Anthropol Sci. https://doi.org/10.1007/s12520-018-0674-y

Ecsedy I (1977) Adatok a Somogyvár-Vinkovci kultúra kérdéséhez (Angaben zur Frage der Somogyvár-Vinkovci Kultur). Janus Pannonius Múzeum 22:185–194

Furholt M (2003) Die absolutchronologische Datierung der schnurkeramik in Mitteleuropa und Südskaninavien. Universitätsforschungen zur Prähistorischen Archäologie, Bonn, p 101

Gilli E, Montagnari Kokelj E (1993) La grotta dei Ciclami nel Carso Triestino (Materiali degli scavi 1959-1961). Atti Soc Preist Friuli-Venezia Giulia 7:65–162

Gilli E, Montagnari Kokelj E (1994) La Grotta delle Gallerie nel Carso Triestino. Atti Soc Preist Friuli-Venezia Giulia 8:121–194

Gilli E, Montagnari Kokelj E (1996) La Grotta degli Zingari nel Carso Triestino (Materiali degli scavi 1961-1965). Atti Soc Preist Friuli-Venezia Giulia 9:63–126

Govedarica B (1989) Rano bronzano doba na području istočnog Jadrana (L'age du bronze ancien dans la region de l'Adriatique de l'est). Djela ANUBiH 67/7, Sarajevo

Jolliffe IT (2002) Principal component analysis, 2nd edn. Springer, New York

Kahl W-A, Ramminger B (2012) Non-destructive fabric analysis of prehistoric pottery using high-resolution X-ray microtomography: a pilot study on the late Mesolithic to Neolithic site Hamburg-Boberg. J Archaeol Sci 39:2206–2219

Kaiser E (2013) Import, imitation and interaction: a critical review of the chronology and significance of cross footed bowls of the third millennium BC in Southeastern and Eastern Europe. In: Heyd V, Kulcsár G, Szeverényi V (eds) Transitions to the Bronze Age interregional interaction and socio-cultural change in the third millennium BC Carpathian Basin and neighbouring regions. Archaeolingua 49, Budapest, pp 139–152

Kassambara A (2017) Practical guide to principal component methods in R, 1st edn. STHDA

Korošec P (1958–1959) Kulturna opredelitev materialne kulture na koliščih pri Igu. Arh vest 9–10:94–107

Korošec P, Korošec J (1969) Najdbe s koliščarskih naselbin pri Igu na Ljubljanskem barju / Fundgut der Pfahlbausiedlungen bei Ig am Laibacher Moor. Arheološki katalogi Slovenije 3, Ljubljana

Kulcsár G (2009) The Beginnings of the Bronze Age in the Carpathian Basin. The Makó–Kosihy–Čaka and the Somogyvár–Vinkovci cultures in Hungary. Varia archaeologica Hungarica 23, Budapest

Leghissa E (2015) Način okraševanja keramike ljubljanske kulture in pramenaste keramike – eksperimentalna arheologija (Decorating the pottery of the Ljubljana culture and the Litzen pottery—an experimental archaeology case study). Arh vest 66: 275–292

Leghissa E (2017) Dežmanova kolišča pri Igu in njihovo mesto v pozni bakreni in zgodnji bronasti dobi (The Dežman (Deschmann) pile dwellings near Ig in the Late Copper and Early Bronze ages). Unpublished doctoral thesis, University of Ljubljana, Faculty of Arts, Department of archaeology, Ljubljana

Leghissa E (2018) Od prve fotografije do prve radiokarbonski datacije. Vučedolska koliščarska naselbina na Ljubljanskem barju (From the first photo to the first radiocarbon date. A Vučedol pile dwelling site in the Ljubljansko barje). In: Črešnar M, Vinazza M (eds) Srečanja in vplivi v raziskovanju bronaste in železne dobe na Slovenskem: Zbornik prispevkov v čast Bibi Teržan. Znanstvena založba Filozofske fakultete Univerze v Ljubljani, Ljubljana, pp 39–52

Lindahl A, Pikirayi I (2010) Ceramics and change: an overview of pottery production techniques in northern South Africa and eastern Zimbabwe during the first and second millennium AD. Archaeol Anthropol Sci 2:133–149

Machado AS, Oliveira DF, Gama Filho HS, Latini R, Bellido AVB, Assis JT, Anjosb MJ, Lopes RT (2017) Archeological ceramic artifacts characterization through computed microtomography and X-ray fluorescence. X-Ray Spectrom 46:427–434. https://doi.org/10.1002/xrs.2786

Maran J (1998) Kulturwandel auf dem griechischen Festland und den Kykladen im späten 3. Jahrtausend v. Chr. Universitätsforschungen zur Prähistorischen Archäologie 53, Bonn

Marzolini G (1970) La Grotta dell'Edera. Annali del gruppo Grotte dell'Associazione XXX Ottobre 4:19–35

Marzolini G (1983) La Grotta del Pettine di Gabrovizza (Carso Triestino). Atti Soc Preist Friuli-Venezia Giulia 4:33–43

Medunová-Benešová A (1977) "Palliardiho hradisko" Eine äneolitische Höhensiedlung bei Vysočany, Bez. Znojmo, Katalog der Funde. Fontes Archaeologiae Moravicae IX, Brno

Montagnari Kokelj E (1981) Appunti per una storia degli studi sulla "cultura di Lubiana". Atti dei Civici musei di storia ed arte di Trieste 12(1):13–36

Montagnari Kokelj E, Crismani A (1997) La Grotta del Mitreo nel Carso triestino. Atti Soc Preist Friuli-Venezia Giulia 10:7–98

Montagnari Kokelj E, Greif T, Presello E (2002) La Grotta Cotariova nel Carso triestino (Italia nord-orinetale). Materiali ceramici degli scavi 1950-70 (The Grotta Cotoriova in the Trieste Karst (North-Eastern Italy). The pottery of the 1950-70 excavations). Aquileia Nostra 73: 37–190

Parzinger H (1984) Die Stellung der Uferrandsiedlungen bei Ljubljana im äneolitischen und frühbronzezeitlichen Kultursystem der mittleren Donauländer (Mesto kolišč Ljubljanskega barja v eneolitiku in zgodnji bronasti dobi srednjega Podonavja). Arh vest 35:13–75

Sanger MC (2016) Investigating pottery vessel manufacturing techniques using radiographic imaging and computed tomography: studies from the Late Archaic American southeast. J Archaeol Sci Rep 9: 586–598

Spataro M (1999) La caverna dell'Edera di Aurisina (TS): archeometria delle ceramiche. Atti Soc Preist Friuli-Venezia Giulia 11:63–90

Tuniz C, Bernardini F, Turk I, Dimkaroski L, Mancini L, Dreossi D (2012) Did neanderthals play music? X-ray computed microtomography of the Divje babe 'flute'? Archaeometry 54:581–590

Tuniz C, Bernardini F, Cicuttin A, Crespo ML, Dreossi D, Gianoncelli A, Mancini L, Mendoza Cuevas A, Sodini N, Tromba G, Zanini F, Zanolli C (2013) The ICTP-Elettra X-ray laboratory for cultural heritage and archaeology. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 711:106–110

Žibrat Gašparič A (2013) A new look at old material: ceramic petrography and neo/eneolithic pottery traditions in the eastern Ljubljansko barje, Slovenia. DocPraeh 40:147–164