

## SOFTWARE METAPAPER

# ScatterJn: An ImageJ Plugin for Scatterplot-Matrix Analysis and Classification of Spatially Resolved Analytical Microscopy Data

Fabian Zeitvogel<sup>1</sup> and Martin Obst<sup>2</sup>

<sup>1</sup> Center for Applied Geosciences, University of Tuebingen, Tuebingen, Germany  
fabian.zeitvogel@uni-tuebingen.de

<sup>2</sup> Bayreuth Center of Ecology and Environmental Research, Bayreuth University, Bayreuth, Germany  
martin.obst@uni-bayreuth.de

We present ScatterJn, an ImageJ (and Fiji) plugin for scatterplot-based exploration and analysis of analytical microscopy data. In contrast to commonly used scatterplot tools, it handles more than two input images (or image stacks, respectively) by creating a matrix of pairwise scatterplots. The tool offers the possibility to manually classify pixels by selecting regions of datapoints in the scatterplots as well as in the spatial domain. We demonstrate its functioning using a set of elemental maps acquired by SEM-EDX mapping of a soil sample. The plugin is available at <https://savannah.nongnu.org/projects/scatterjn>.

**Keywords:** data evaluation; scatterplot matrix; 2D histogram; classification; microscopy; ImageJ plugin  
**Funding statement:** This work was funded by DFG Emmy-Noether program to M.O. (OB 362/1-1).

## (1) Overview

### Introduction

From spatially resolved chemical information of samples such as elemental or species maps, it is possible to derive information about reaction mechanisms and processes [1]. Methods that allow for the spatially resolved chemical analysis of samples on the micron and submicron scales include scanning electron microscopy (SEM) and transmission electron microscopy (TEM), both in combination with energy-dispersive X-ray spectroscopy (EDX) mapping, nanoscale secondary ion mass spectrometry (nanoSIMS), scanning transmission X-ray microscopy (STXM), and confocal laser scanning microscopy (CLSM). These methods typically produce data in the form of sets of images (or image stacks in 3D case) that each represent the spatial distribution e.g. of elements or chemical species. The information contained in such datasets is sometimes difficult to access. A number of evaluation methods exist to extract part of this information. These include direct visual display in colour-coded overlay images; calculation of statistical indicators [2, 3]; and various types of cluster analysis, which aim at detecting areas of similar composition [4–6]. Each of these methods deals with particular aspects of the data and has its specific shortcomings and advantages.

An additional, very useful approach that is often applied to analytical microscopy datasets [3, 7–9] is an analysis based on 2-dimensional histograms or so-called

scatterplots. Displaying image data in scatterplots provides an intuitive view on properties such as variations, correlations, trends, and clustering. This can be combined with tracing back features in the scatterplot (e.g. clusters) to the spatial domain [7]. Because the scatterplots have to be displayed in a coordinate system, this type of analysis is usually limited to datasets consisting of no more than 2 (in a few cases 3) images.

Our tool, ScatterJn, creates a matrix of 2D scatterplots, which allows for extending this approach to datasets consisting of more than 2 or 3 images. While the generation of scatterplot matrices combined with interactive functions is a common tool for visualizing multi-dimensional datasets [10–12], to our knowledge it has not yet been applied to image data. With our tool, we complement existing methods by providing an additional way of data exploration and analysis.

A number of software tools exist that allow for colocalization analysis of image data, most notably the Fiji [13] plugin *Coloc 2* ([http://fiji.sc/Coloc\\_2](http://fiji.sc/Coloc_2)). Apart from generating scatterplots, this tool allows for the calculation of statistical indicators such as the Pearson's coefficient, Manders' coefficients [2], and Li's intensity correlation quotient [14] as well as statistical significance testing [15]. These are very intricate methods for evaluating pairwise image data, which are not implemented in ScatterJn. Instead, ScatterJn focuses on qualitative analysis and exploration of datasets that consist of more than two images.

### Implementation and architecture

The tool ScatterJn is designed as a plugin for ImageJ [16] and Fiji [13]. It is published under the GNU GPL v.3 license and available at <https://savannah.nongnu.org/projects/scatterjn>. The plugin was written in Java and is partly based on ScatterJ [1], a scatterplot tool for pairwise analysis of images.

### Input data

As input, ScatterJn requires a set of two or more 8-bit greyscale images that have to be loaded in ImageJ. All input images must have the same pixel dimensions and must be spatially aligned in a way that they represent exactly the same area of the sample. The plugin handles 2D images, representing sample surfaces or projections, and 3D image stacks, representing sample volumes. Any kind of pre-processing of the images, such as alignment or conversion to 8-bit greyscale type, which depends on the type of data, has to be done by the user with other tools. In order to use the plugin, the input images are selected by the user.

### Generation of scatterplots

The different images are assumed to represent compositional maps of the same sample area or volume, respectively. Corresponding pixels (i.e. pixels with the same coordinates) in the individual images thus correspond to the same position on the sample. The grey value of each pixel thereby represents the result of a quantitative intensity measurement that, for example, represents the local concentration of a chemical species.

A set of datapoints is formed out of the grey values of pixels at the same position in the different input images. Therefore, a datapoint contains the results of all measurements taken at the corresponding position. For a set of  $n$  input images, a datapoint has  $n$  coordinates. The number of datapoints is equal to the number of pixels in one input image. Each datapoint corresponds to a pixel position in the spatial domain.

The datapoints are displayed in a scatterplot matrix. For each combination of two images, a 2-dimensional scatterplot is created. In this scatterplot, the grey values of the individual pixels in image 1 are plotted against the grey values of the equivalent pixel in image 2. This corresponds to a projection of the  $n$ -dimensional cloud of datapoints onto a 2-dimensional plane. The scatterplots are arranged in a matrix, which is displayed in a new image window. To account for the spatial position corresponding to the datapoints, a spatial-domain map is created that is displayed in a second image window.

### User-defined display settings

As grey values in 8-bit images are integers within the [0, 255] interval, a large number of datapoints will typically have the same coordinates in the scatterplot and thus be plotted at the same position in the coordinate system. Therefore, the density of datapoints in the scatterplots is visualized using an adaptable colour scale. For a more detailed description see [1]. It is possible for the user to adjust display parameters such as the colour scale used for the scatterplots and labels in the scatterplot matrix image. A factor can be entered for histogram binning, which leads to smaller scatterplots.

### Interactive classification functions

The tool offers functions for binary and gradual classification based on the position of datapoints in the scatterplot matrix and/or in the spatial-domain map.

For binary classification, the user can select areas in any of the scatterplots using ImageJ's ROI tools. The selected datapoints are then highlighted in all scatterplots and in the spatial-domain map. It is possible to simultaneously select up to one ROI per scatterplot and/or one ROI in the spatial-domain map. The selection in the spatial-domain map can also be defined using a mask generated from any greyscale image of the same region and dimensions as the input images.

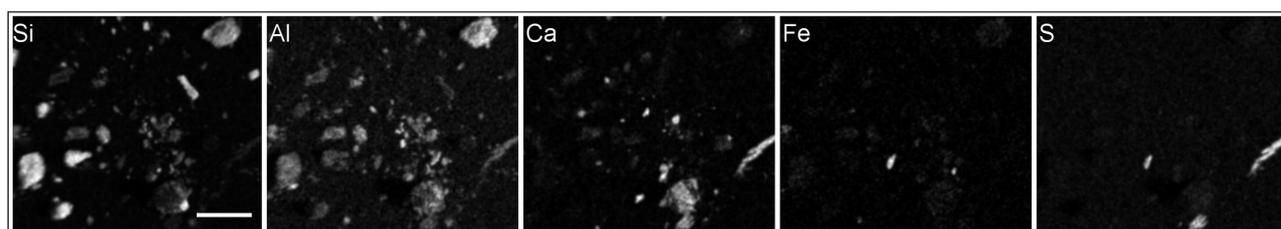
As a method of gradual classification visualizing possible linear trends, datapoints can be classified according to their angular distance from the x-axis in one of the scatterplots. The angular distance values are displayed using a colour scale.

### Example Dataset

Instead of merely presenting instructions for the usage of the plugin, here we demonstrate its basic functions using an example dataset and illustrate the results. An extensive manual, however, is included in the plugin itself and can be accessed via the "Info" button.

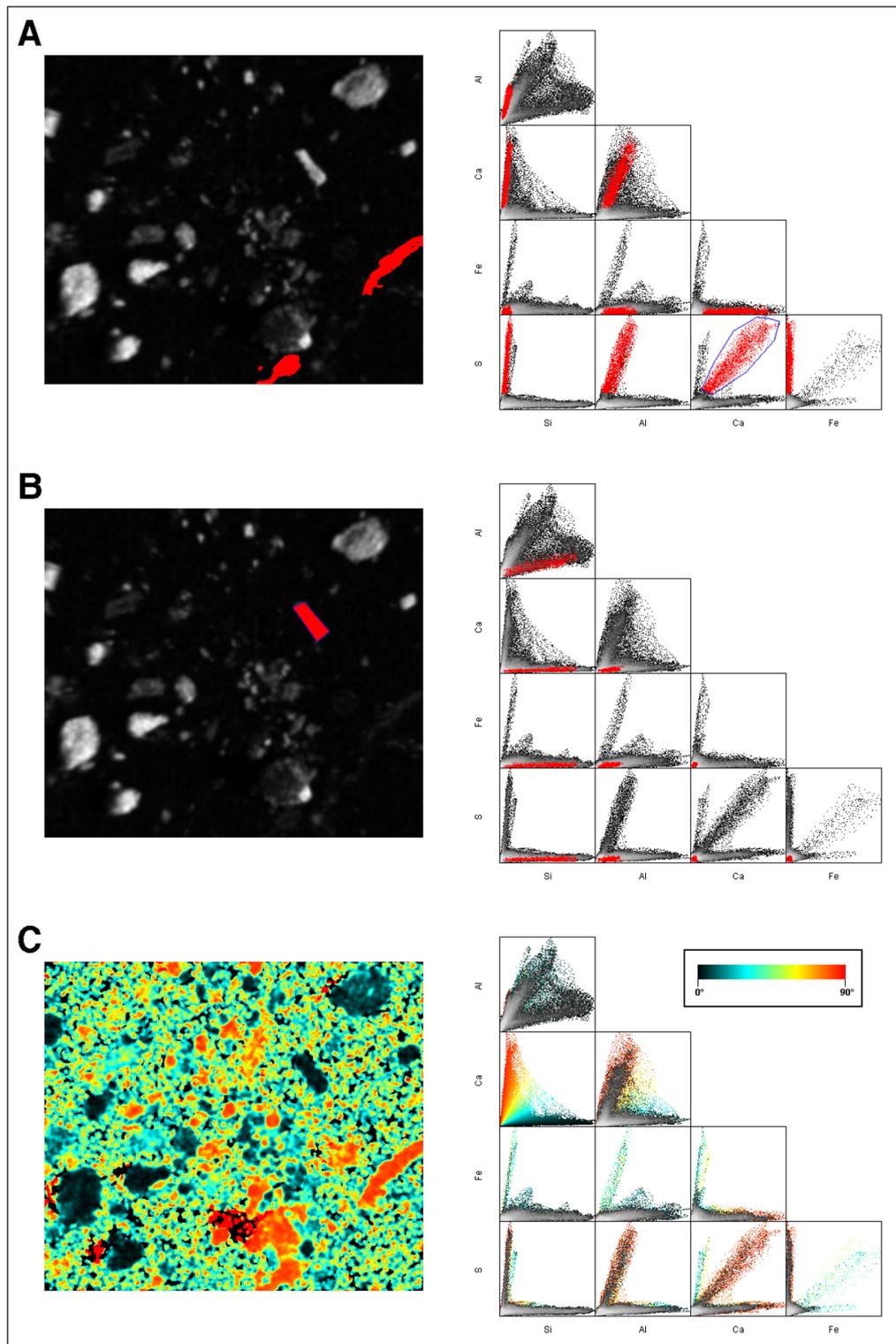
### Example dataset: acquisition and pre-processing

The example dataset is shown in **Fig. 1**. It consists of a set of elemental maps obtained by SEM-EDX mapping of a soil sample. The sample was washed in deionized water, dried in air, spread on an adhesive carbon pad attached to an aluminium sample holder stub, and coated with 8 nm of carbon using a BAL-TEC SCD 005 sputter coater equipped with a BAL-TEC CEA 035 (BAL-TEC, Balzers, Liechtenstein). Maps of  $512 \times 448$  pixels were acquired using a LEO 1450 VP instrument (now Zeiss, Oberkochen, Germany) equipped with an Oxford Inca X-sight EDS



**Figure 1:** Elemental maps of a soil sample used as demonstration dataset. Images were processed as described in the text. Scale bar 50  $\mu\text{m}$ .





**Figure 3:** Demonstration of classification using ScatterJn. (A) Selection of datapoints in the scatterplot matrix; (B) selection of pixels in the spatial-domain map; (C) angular distance map in the Ca/Si scatterplot – the colour scale in the inset shows the angle and was added manually for illustration purposes.

#### ***Gradual classification***

In the Ca/Si scatterplot, no clear trends or clusters can be distinguished. Instead, the datapoints appear in a fan-shaped point cloud. Using the “Angular distance map” function, the datapoints were classified according to the angle of their position vector with the x-axis in the Ca/

Si scatterplot. The angle is represented on a continuous colour scale. This function yields more abstract, but at the same time more diversified, information than binary classification. In the spatial-domain map, particles represented in a certain colour have the Ca/Si ratio which is displayed in the in the same colour in the scatterplot. In

scatterplots other than the Ca/Si one, only pixels onto which only one datapoint is projected can be unambiguously assigned a colour; other pixels are depicted in the greyscale colour scale defined for the scatterplots.

#### Quality control

Testing was done on Windows 7, Mac OS X (10.9.5), and Ubuntu 14.04.2.

## (2) Availability

#### Operating system

ScatterJn was written as a plugin for ImageJ; we expect it to run on any operating system that will run a compatible version of ImageJ or Fiji.

#### Programming language

Java 1.5.

#### Additional system requirements

#### Dependencies

The plugin is expected to run with any reasonably up-to-date version of ImageJ or Fiji. The lowest version that was used for testing was ImageJ 1.47q.

#### List of contributors

Fabian Zeitvogel, Martin Obst.

#### Software location

##### Archive

**Name:** ScatterJn

**Persistent identifier:** doi.org/10.5281/zenodo.22167

**Licence:** GNU GPL version 3

**Publisher:** Fabian Zeitvogel

**Date published:** 30/07/15

##### Code repository

**Name:** Savannah

**Identifier:** https://savannah.nongnu.org/projects/scatterjn

**Licence:** GNU GPL version 3

**Date published:** 29/07/15

#### Language

English.

## (3) Reuse potential

With ScatterJn, we complement existing evaluation methods by providing a versatile tool for data exploration. Its intended use is for exploratory analysis of the previously described types of analytical microscopy datasets. This includes a variety of analytical methods such as SEM-EDX, TEM-EDX, STXM, CLSM, nanoSIMS, and other spatially resolved analytical data that are – at least to some extent – quantitative. Such approaches are currently used in many different fields of research. Beyond this, the plugin can potentially be of use in any situation that requires exploration of quantitative image data.

#### Competing Interests

The authors declare that they have no competing interests.

## Acknowledgements

The need for a tool offering the functions of ScatterJn arose within our working group; we would therefore like to thank G. Schmid, P. Ingino, I.J. Adaktylou and L. Hao for their help and discussion. We would also like to thank an anonymous reviewer for his/her helpful suggestions.

## References

1. **Zeitvogel, F, Schmid, G, Hao, L, Ingino, P and Obst, M** 2014 ScatterJ: An ImageJ Plugin for the Evaluation of Analytical Microscopy Datasets. *Journal of Microscopy*, 0: 1–9. DOI: <http://dx.doi.org/10.1111/jmi.12187>
2. **Manders, E M M, Verbeek, F J and Aten, J A** 1993 Measurement of Co-Localization of Objects in Dual-Colour Confocal Images. *Journal of Microscopy*, 169(3): 375–382. DOI: <http://dx.doi.org/10.1111/j.1365-2818.1993.tb03313.x>
3. **Bolte, S and Cordelières, F P** 2006 A Guided Tour into Subcellular Colocalization Analysis in Light Microscopy. *Journal of Microscopy*, 224(3): 213–232. DOI: <http://dx.doi.org/10.1111/j.1365-2818.2006.01706.x>
4. **Cutrona, J, Bonnet, N, Herbin, M and Hofer, F** 2005 Advances in the Segmentation of Multi-Component Microanalytical Images. *Ultramicroscopy*, 103(2): 141–152. DOI: <http://dx.doi.org/10.1016/j.ultramic.2004.11.005>
5. **Lerotic, M, Jacobsen, C, Schafer, T and Vogt, S** 2004 Cluster Analysis of Soft X-Ray Spectromicroscopy Data. *Ultramicroscopy*, 100(1–2): 35–57. DOI: <http://dx.doi.org/10.1016/j.ultramic.2004.01.008>
6. **Ward, J, Marvin, R, O'Halloran, T, Jacobsen, C and Vogt, S** 2013 Rapid and Accurate Analysis of an X-Ray Fluorescence Microscopy Data Set through Gaussian Mixture-Mased Soft Clustering Methods. *Microscopy and Microanalysis*, 19(5): 1281–1289. DOI: <http://dx.doi.org/10.1017/S1431927613012737>
7. **Bright, D S and Newbury, D E** 1991 Concentration Histogram Imaging – A Scatter Diagram Technique for Viewing 2 or 3 Related Images. *Analytical Chemistry*, 63(4): A243–A250. DOI: <http://dx.doi.org/10.1021/ac00004a731>
8. **French, A P, Mills, S, Swarup, R, Bennett, M J and Pridmore, T P** 2008 Colocalization of Fluorescent Markers in Confocal Microscope Images of Plant Cells. *Nature Protocols*, 3(4): 619–628. DOI: <http://dx.doi.org/10.1038/nprot.2008.31>
9. **Obst, M, Wang, J and Hitchcock, A P** 2009 Soft X-Ray Spectro-Tomography Study of Cyanobacterial Biomineral Nucleation. *Geobiology*, 7(5): 577–591. DOI: <http://dx.doi.org/10.1111/j.1472-4669.2009.00221.x>
10. **Martin, A R and Ward, M O** 1995 High Dimensional Brushing for Interactive Exploration of Multivariate Data. In *Proceedings of the 6th Conference on Visualization'95*. IEEE Computer Society. DOI: <http://dx.doi.org/10.1109/VISUAL.1995.485139>
11. **Keim, D** 2002 Information Visualization and Visual Data Mining. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1): 1–8. DOI: <http://dx.doi.org/10.1109/2945.981847>

12. **Melançon, G, Jourdan, F, Paris, A and Koenig, P-Y** 2006 Multiscale Scatterplot Matrix for Visual and Interactive Exploration of Metabonomics Data. In *Visual Information Expert Workshop*. Springer. DOI: [http://dx.doi.org/10.1007/978-3-540-71027-1\\_18](http://dx.doi.org/10.1007/978-3-540-71027-1_18)
13. **Schindelin, J, Arganda-Carreras, I, Frise, E, Kaynig, V, Longair, M, Pietzsch, T, Preibisch, S, Rueden, C, Saalfeld, S, Schmid, B, Tinevez, J-Y, White, D J, Hartenstein, V, Eliceiri, K, Tomancak, P and Cardona, A** 2012 Fiji: An Open-Source Platform for Biological-Image Analysis. *Nature Methods*, 9(7): 676–682. DOI: <http://dx.doi.org/10.1038/nmeth.2019>
14. **Li, Q, Lau, A, Morris, T J, Guo, L, Fordyce, C B and Stanley, E F** 2004 A Syntaxin 1, Gαo, and N-type Calcium Channel Complex at a Presynaptic Nerve Terminal: Analysis by Quantitative Immunocolocalization. *The Journal of Neuroscience*, 24(16): 4070–4081. DOI: <http://dx.doi.org/10.1523/JNEUROSCI.0346-04.2004>
15. **Costes, S V, Daelemans, D, Cho, E H, Dobbin, Z, Pavlakis, G and Lockett, S** 2004 Automatic and Quantitative Measurement of Protein-Protein Colocalization in Live Cells. *Biophysical Journal*, 86(6): 3993–4003. DOI: <http://dx.doi.org/10.1529/biophysj.103.038422>
16. **Abràmoff, M D, Magalhães, P J and Ram, S J** 2004 Image Processing with ImageJ. *Biophotonics International*, 11: 36-41.
17. **Muehe, E M, Adaktylou, I J, Obst, M, Zeitvogel, F, Behrens, S, Planer-Friedrich, B, Kraemer, U and Kappler, A** 2013 Reducing Conditions and Organic Carbon Input Cause Cadmium Immobilization in a pH-Neutral Soil. *Environmental Science & Technology*, 47: 13430–13439. DOI: <http://dx.doi.org/10.1021/es403438n>

**How to cite this article:** Zeitvogel, F and Obst, M 2016 ScatterJn: An ImageJ Plugin for Scatterplot-Matrix Analysis and Classification of Spatially Resolved Analytical Microscopy Data. *Journal of Open Research Software*, 4: e5, DOI: <http://dx.doi.org/10.5334/jors.89>

**Submitted:** 31 July 2015      **Accepted:** 05 January 2016      **Published:** 03 February 2016

**Copyright:** © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.