Identification of microplastics in soils using 2D geometric shape descriptors

Irada Ismayilova, Tabea Zeyer, Sabine Timpf

Angaben zur Veröffentlichung / Publication details:

Ismayilova, Irada, Tabea Zeyer, and Sabine Timpf. 2021. "Identification of microplastics in soils using 2D geometric shape descriptors." *AGILE: GIScience Series* 2: 32. https://doi.org/10.5194/agile-giss-2-32-2021.

AGILE: GIScience Series, 2, 32, 2021. https://doi.org/10.5194/agile-giss-2-32-2021
Proceedings of the 24th AGILE Conference on Geographic Information Science, 2021.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras.
This contribution underwent peer review based on a full paper submission.

Identification of Microplastics in Soils Using 2D Geometric Shape Descriptors

Irada Ismayilova[a], Tabea Zeyer[b], Sabine Timpf[a]

irada.ismayilova@geo.uni-augsburg.de, tabea.zeyer@geo.uni-augsburg.de, sabine.timpf@geo.uni-augsburg.de
[a]Geoinformatics Group, Institute of Geography, University of Augsburg, Augsburg, Germany
[b]Water and Soil Resources Research, Institute of Geography, University of Augsburg, Augsburg, Germany

Abstract. Microplastics (MP), until now mostly studied in aquatic ecosystems, are also largely polluting terrestrial ecosystems, especially soil systems. Overall, there is a lack of robust and fast methods to identify, separate and eliminate MPs from soils. This paper is a first attempt to use 2D shape descriptors and Random Forest Machine Learning method in order to discriminate soil and MP particles. The results of this study demonstrate promising potential of the Machine Learning approach and shape descriptors in this relatively new scientific field of determining MPs in soils.

Keywords: microplastic, shape descriptors, machine learning

1 Introduction

Due to its appealing characteristics, such as cheap price, water resistance and durability, plastics production has greatly increased since having been introduced in the 1950s (Horton and Dixon, 2018; Geyer et.al., 2017). With increased production and wide usage of plastics, an enormous amount of plastics ends up in our environment. This was first realized for oceans (Zarfl et al., 2011) and generally aquatic ecosystems (Prata et al., 2019), where macro and microplastic is wide-spread. However, an increasing number of studies indicate that specifically microplastics (hereafter MP) substantially pollutes terrestrial ecosystems including soils (Bläsing & Amelung 2018)

Most methods to determine MP in aquatic environments cannot be straightforwardly transferred to terrestrial systems. Especially in case of soil consisting of a matrix of mineral and organic particles it is challenging to determine MP particles and fibres in low concentrations (Möller et al. 2020).

In this paper we analysed the potential of a Random Forest (RF) Machine Learning (ML) approach together with 2D geometric shape descriptors, to determine MP and soil particles from optical images taken with a microscope (VK-X1000, Keyence, Japan).

Identifying objects using their shapes was and remains of interest in many fields including cognitive science, landscape analysis, molecular biology, soil science and others. Shape representation in comparison to using e.g., the colour or texture of an object is semantically more effective (Persoon and Fu, 1977). This interest is especially relevant for digital imaging, when a 3D object is projected to a plane surface (2D) and the information on the third dimension is lost. Even if the research is moving towards the usage of mostly 3D shape descriptors, 2D descriptors remain useful in many practical tasks, such as shape classification. This is mainly due to their high quantitative characteristics and fast computation (Paquet et al., 2000).

Shape descriptors in the existing literature are separated into two groups: contour-based and area-based descriptors. The first one takes into account the boundary information of an object whereas the latter one does not rely on the boundary (it might be incomplete) but rather on all the pixels within the shape region. Area-based descriptors are considered more robust and less sensitive to shape deformations (Baraldi and Soares, 2017). Moreover, choosing either contour or area-based descriptors might not always be sufficient. Therefore, a combination of different descriptors should be utilized in order to adequately describe shapes (Zhang and Lu, 2003).

In this study, we aim at identifying particles belonging to two groups: soil and MP. To be able to discriminate, we implement a shape descriptor analysis and classify the results using the RF model from ML. We use a combination of descriptors for more accurate shape description (Amanatiadis et al., 2011; Zhang and Lu, 2004). In this paper, we included four simple 2D shape descriptors: Circularity, Aspect Ratio, Solidity and Compactness. We chose these descriptors because they

are rotation invariant and are easy to implement (Sarfraz and Ridha, 2007).

The paper is structured as follows: the second section introduces the data preparation steps and methodology used in this research, i.e., shape analysis and RF. In section three, we present the results of applying the methodology to the data. Section four discusses the results and section five shows our conclusions.

2 Methodology

2.1 Preparation of soil / MP samples

Two types of samples were prepared and analysed with a microscope to test the potential of the method to discriminate MP from soil particles. The first sample used under the microscope consist of loamy sand, with a clay (< 2 µm), silt (2 – 63 µm), and sand (63 – 2000 µm) content of 10 %, 18% and 72% respectively, placed on a paper filter. The second sample was prepared using industrially produced High Density Polyethylene (HDPE) particles with a size range of 250 – 300 µm. Both samples were scanned with an x240 zoom resulting in optical images (RGB) with a raster resolution of 96 dpi (see Figure 1).

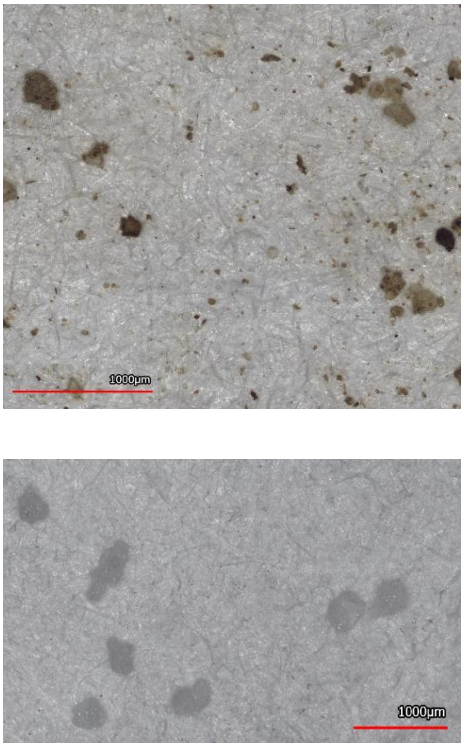

Figure 1: The optical image of soil particles < 2000 µm (top) and microplastic particles 250-300 µm on paper filters (bottom).

2.2 Image pre-processing

From the utilized images, we extract in total 3574 particles for further analysis. Prior to calculating shape descriptors, the scales of the images are calibrated in a software called ImageJ (Rueden et.al., 2017). We then convert them into 8-bit binary images similar to Fig. 2.

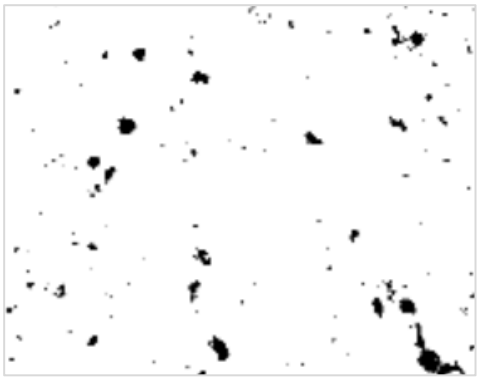

Figure 2: Cut out of an 8-bit binary image produced from the optical image containing soil particles.

Each binary image is prepared slightly differently because of the presence of water spots on the paper filter. This way we can choose the most suitable threshold for each image and thus avoid information loss or false outlining of particles.

We outline every single soil and MP particle on binary images as illustrated in Fig. 3. Then, using ImageJ, we calculate all five descriptors as described in Section 2.3.

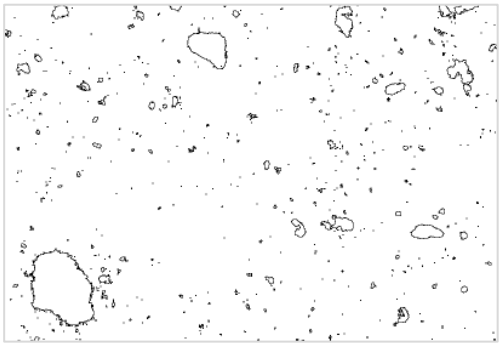

Figure 3: The outlines of each single soil particle derived from the 8-bit binary image.

2.3 Shape analysis

In order to differentiate MP particles from soil particles we chose the following 2D shape descriptors:

- Circularity
- Aspect Ratio
- Solidity
- Compactness

Additionally, we take into consideration the perimeter of the particles in order to observe if any of the 2D shape descriptors will reliably identify studied particles. In this study, MPs (HDPE 250 - 300 µm) are analysed within a certain size range and therefore, perimeter will be used as an additional control parameter in the RF model.

2.3.1 Circularity

The circularity of an object's shape is calculated based on the following equation:

$$Circularity = 4\pi * \frac{[Area]}{[Perimeter]^2} \quad (1)$$

Where area is calculated based on the following equation:

$$Area = \pi * r^2 \quad (2)$$

Circularity is calculated in a range from 0.0 to 1.0 where, the closer the value to 1 the more circular is the shape. Whereas, the more the value approaches 0, the more elongated is the shape.

2.3.2 Aspect Ratio

The aspect Ratio of a particle is calculated based on the following equation:

$$AR = \frac{[Major\ Axis]}{[Minor\ Axis]} \quad (3)$$

Aspect Ratio indicates a ratio between the width to the height of the particle.

2.3.3 Solidity

The solidity of the objects shape is calculated based on the following equation:

$$\text{Solidity} = \frac{[Area]}{[Convex\ hull\ area]} \quad (4)$$

The convex hull is determined based on the Gift Wrap Algorithm (Jarvis, 1973).

2.3.4 Compactness

The compactness of an object's shape is calculated based on the following formula:

$$Compactness = \frac{\left({}^{4}/_{\pi}\right)^2}{Major\ axis} \quad (5)$$

Compactness is describing how compact the shape of an object is. It takes a value of 1 for a perfect circle while a square has a compactness of $\frac{\pi}{4}$. The measure of compactness will decrease as a shape gets more irregular.

2.3.5 Perimeter

The perimeter of an objects shape is defined as the length of the outside boundary of the object.

2.4 Random Forest

For classification of MP and soil particles, we use the RF algorithm proposed by Breiman in 2001. RF is an ensemble learning method that can be applied for both classification and regression tasks. We chose RF due to its robustness against noise and correlation between variables as well as its capability to reach a high classification accuracy (Samuel et al., 2019; Breiman, 2001).

Trees in the RF are built by answering yes/no questions. In comparison to earlier versions of tree classifiers, RF uses the best of the random variables and bootstrap sampling in order to split at each node (Breiman, 2001).

RF is implemented in many software packages and working with a RF model is relatively easy. With the help of Hyperparameter tuning (process of adjusting one or many parameters of RF model) a better classification result can be achieved. The most frequently tuned parameters of RF are e.g., the number of trees grown in the forest and the number of variables used to split the internal node. The performance of the RF model, in a classification problem, can be verified using the Confusion Matrix (CM). The CM shows predicted values versus actual values and is built based on the 'Out Of Bag' error (OOB) and an independent error assessment (Prajwala, 2015). Another commonly used accuracy indicator is 'Area under the Receiver Operating Characteristics' (AUROC) (Breiman, 2001). With AUROC it is possible to explain the goodness of predicted classes. It plots true positive rate versus false positive rate. Thus, the larger the area under the curve, the better the different classes may be distinguished (Golkarian et.al., 2018).

2.5 Classification modelling

We start analysis by implementing a correlation analysis to check if there is a correlation between considered variables. We then move on to building a RF model following the procedure shown in Fig. 4

We join the calculated descriptors for both soil and MP particles into a single dataset. We then assign a value of '0' to every soil particle and a value of '1' to every MP particle so that we can perform a binary classification task.

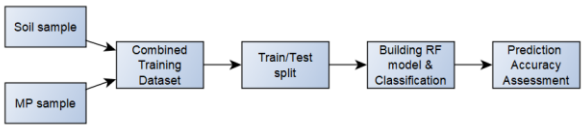

Figure 4: Flowchart of each step in the RF modelling process.

The total number of particles extracted from both images sums up to 3574. We then split the data using a 70/30 proportion into train/test dataset.

Using R software (R Core Team, 2014), we create a RF model. We train it using the training dataset and then test using the testing dataset. Parameters of the RF are not changed and the default values are utilized, where number of trees (n_{tree}) is 500 and m_{try} is $\sqrt{p}$ (p is the number of the predictor variables). RF model outputs accuracy, sensitivity and specificity values. Further, ROC curve and AUC values can also be calculated from the model results. All these predictors are used to assess the prediction accuracy.

3 Results

Fig. 5 illustrates Pearson's correlation for each variable included as descriptive parameter into RF modelling.

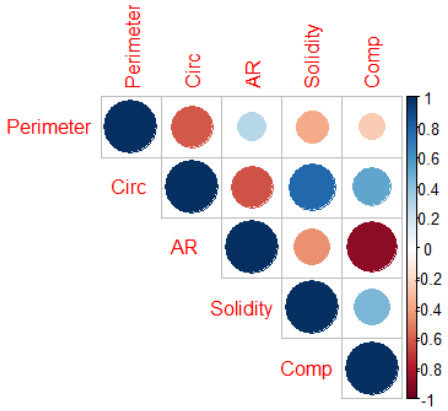

Figure 5: Pearson correlation matrix for all the parameters considered in RF model including Perimeter, Circularity (Circ), Aspect Ratio (AR), Solidity and Compactness (Cop).

There is a strong negative correlation (-0.61) between Perimeter and circularity (Circ) as well as strong positive correlation (0.89) between Compactness and Aspect Ratio (AR). We do not expect these correlation results to affect the definition of variable importance by the RF model.

We run the RF model two times, each time with a different combination of the predictor variables. When using all the five descriptors (trial 1), we reach a prediction accuracy of 96%. We then check the contribution importance of each descriptor variable to the model. We find that the particle perimeter has the highest variable importance. Therefore, on the second try, we remove the perimeter from the analysis and leave only four shape descriptors. This helps us to observe if any of the shape descriptors will reliably identify particles. In the second round, prediction accuracy falls from 96% to 86 % with an especially high number of false positives. In both trials, the circularity of an object's shape does not show any variable importance.

The variable importance of each predictor variable for each trial is given in Fig. 6a and b.

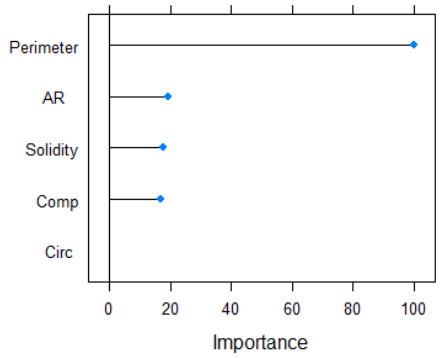

Figure 6a: Trial 1: Feature importance of Perimeter, Aspect Ratio (AR), Solidity, Circularity (Circ) and Compactness (Comp).

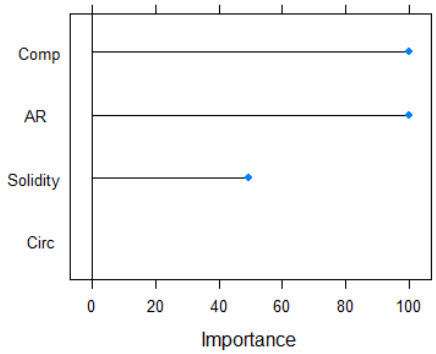

Figure 6b: Trial 2: Feature importance of Solidity, Aspect Ratio (AR), Circularity (Circ) and Compactness (Comp).

Furthermore, all the accuracy information for all the trials is shown in Tab. 1.

Trial	Accuracy	Sensitivity	Specificity	AUC
1	0.9646	0.9964	0.8435	0.9336
2	0.8667	0.9777	0.4414	0.722

Table 1: Accuracy values for each trial of RF model.

As Table 1 shows, specificity is generally low in both trials. Specificity shows the ratio of the true positives to the total number of positive cases. Moreover, no matter which descriptor variable is removed, sensitivity always remains high.

We also refer to the Receiver Operating Characteristics (ROC) curves in order to evaluate the performance of the RF model for both trial cases (Fig. 7). As the area under the curve increases, the better the RF model performs, indicating that the first trial outperforms the second trial.

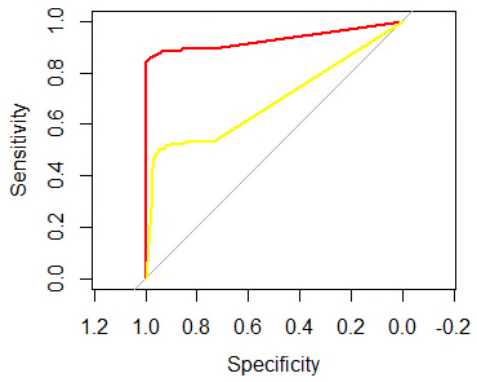

Figure 7: Receiver Operating Characteristics (ROC) curves for both RF model tries. Red line indicates trial 1 and yellow line indicates trial 2.

4 Discussion

In this paper, we introduce a new approach to discriminate soil and MP particles by utilizing 2D shape descriptors and RF classification method. We build the RF model two times by removing parameters from the model one-at-a-time. This helps us to conclude which parameters play a more prominent role in identifying particles and whether the four shape descriptors are capable of differentiating between particles.

Our analysis show that perimeter is the most dominant parameter for differentiating between particles. However, perimeter is strongly affected by particle size. MP and soil particles, used in this study, belong to distinctively different size ranges (MP: 250-300 µm; soil: 0-2000 µm). Therefore, we need to remove perimeter from analysis and repeat the classification task. We observe that prediction accuracy decreases to 86%. This means, that our four chosen shape descriptors still have a high potential to discriminate soil and MP particles.

No matter which parameter combination we use, the circularity of a shape does not influence the decision-making process of the model. This might again be due to the used particles. There is a need for testing different soil and MP types in order to be able to conclude that circularity is or is not a relevant descriptor to discriminate these two particles.

In the setup we used, RF can better identify soil particles than MP particles. This is probably due to the training dataset in which there is a higher prevalence of soil particles. Nevertheless, there is a need to include more MP and soil particles into the analysis in order to have a sufficiently large training dataset that also includes all ranges of different particles.

5 Conclusions and Future Work

In this study, we implement a RF model and simple 2D shape descriptors in order to classify particles into the two classes “soil” and “microplastics”. Our analysis shows that three out of the four chosen shape descriptors are capable of discriminating between our specific particles. Yet, circularity of an object’s shape is not contributing in any way to the results of the models. With a prediction accuracy of 86 %, with only four shape descriptors, we might feel complacent. However, there is a need of extending the training datasets with various soil and MP types and sizes. In this respect, this preliminary study shows great potential of discriminatory power and thus of identifying MP in soils.

In future work we will introduce other, more complex, shape descriptors into the analysis in addition to varying soil and MP types. Currently, a study is under way by one of the co-authors to add further descriptive parameters such as texture and colour to the analysis.

References

Amanatiadis, A., Kaburlasos, V. G., Gasteratos, A., and Papadakis, S. E.: Evaluation of shape descriptors for shape-based image retrieval, IET Image Processing, 5(5), 493-499, https://doi.org/10.1049/iet-ipr.2009.0246, 2011.

Baraldi, A., and Soares, J. V.: Multi-objective software suite of two-dimensional shape descriptors for object-based image analysis, https://arxiv.org/ftp/arxiv/papers/1701/1701.01941.pdf, 2017.

Breiman, L.: Random forests, Machine learning, 45(1), 5-32, https://doi.org/10.1023/A:1010933404324, 2001.

Geyer, R., Jambeck, J. R., and Law, K. L.: Production, use, and fate of all plastics ever made, Science advances, 3(7), https://doi.org/10.1126/sciadv.1700782, 2017.

Horton, A. A., and Dixon, S. J.: Microplastics: An introduction to environmental transport processes, Wiley Interdisciplinary Reviews, Water, 5(2), 1268, https://doi.org/10.1002/wat2.1268, 2018.

Jarvis, R. A.: On the identification of the convex hull of a finite set of points in the plane, Information Processing Letters, 2, 18–21, https://doi.org/10.1016/0020-0190(73)90020-3, 1973.

Paquet, E., Rioux, M., Murching, A., Naveen, T., and Tabatabai, A.: Description of shape information for 2-D and 3-D objects, Signal processing: Image communication, 16(1-2), 103-122, https://doi.org/10.1016/S0923-5965(00)00020-5, 2000.

Persoon, E., and Fu, K. S.: Shape discrimination using Fourier descriptors, IEEE Transactions on systems, man, and cybernetics, 7(3), 170-179, https://doi.org/10.1109/TSMC.1977.4309681, 1977.

Prajwala, T. R.: A comparative study on decision tree and random forest using R tool, International journal of advanced research in computer and communication engineering, 4(1), 196-199, https://doi.org/0.17148/IJARCCE.2015.414, 2015.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical, Computing, Vienna, Austria, http://www.R-project.org/, 2014.

Rueden, C. T., Schindelin, J. and Hiner, M. C.: ImageJ2: ImageJ for the next generation of scientific image data, BMC

Bioinformatics 18,529, https://doi.org/10.1186/s12859-017-1934-z, 2017.

Sarfraz, M., and Ridha, A.: Content-based image retrieval using multiple shape descriptors, In 2007 IEEE/ACS International Conference on Computer Systems and Applications,730-737, https://doi.org/10.1109/AICCSA.2007.370714, 2007.

Wenwen, L. M., Goodchild, F., and Church, R. L.: An efficient measure of compactness for 2D shapes and its application in regionalization problems. International Journal of Geographical Info Science, 1-24, https://doi.org/10.1080/13658816.2012.752093, 2013.

Zhang, D., and Lu, G.: Evaluation of MPEG-7 shape descriptors against other shape descriptors, Multimedia Systems, 9(1), 15-30, https://doi.org/10.1007/s00530-002-0075-y, 2003.

Zhang, D., and Lu, G.: Review of shape representation and description techniques, Pattern recognition, 37(1), 1-19, https://doi.org/10.1016/j.patcog.2003.07.008, 2004.

Blasing, M. and Amelung, W.: Plastics in soil: Analytical methods and possible sources, Science of the Total Environment, 612, 422-435, https://doi.org/10.1016/j.scitotenv.2017.08.086, 2018.

Möller, J.N., Löder, M.G.J. and Laforsch, C.: Finding microplastics in soils: A review of analytical methods, Environmental Science & Technology, 54(4), 2078-2090, https://doi.org/10.1021/acs.est.9b04618, 2020.

Prata, J. C., da Costa, J. P., Duarte, A. C., & Rocha-Santos, T.: Methods for sampling and detection of microplastics in water and sediment: A critical review, TrAC Trends in Analytical Chemistry, 110, 150-159, https://doi.org/10.1016/j.trac.2018.10.029, 2019.

Zarfl, C., Fleet, D., Fries, E., Galgani, F., Gerdts, G., Hanke, G., & Matthies, M.: Microplastics in oceans, Marine Pollution Bulletin, 62, 1589-1591, https://doi.org/10.1016/j.marpolbul.2011.02.040, 2011.