

# Identification of *Myosotis* seeds by means of digital image analysis

Otto Brinkkemper · Laurens van der Maaten ·  
Paul Boon

Received: 27 September 2010 / Accepted: 29 July 2011  
© Springer-Verlag 2011

**Abstract** Despite their name, the identification of seeds of *Myosotis* species (forget-me-not) has hitherto received little attention from archaeobotanists. In an attempt to assemble a collection of reliable identification criteria, digital image analysis was applied to photographs of *Myosotis* seeds by means of *Fovea Pro 4.0*. This program computes 23 features that describe the size and shape of the seeds shown in scale-normalized photographs. We computed the features for 1,453 individual seeds, and performed statistical analyses of the resulting data set with Discriminant Analysis, Correspondence Analysis, and t-Distributed Stochastic Neighbour Embedding (t-SNE).

The combination of analyses provides clues as to how most of the seven western European species of *Myosotis* can successfully be distinguished. Using these clues, an identification key was developed for the identification of waterlogged *Myosotis* seeds.

**Keywords** *Myosotis* · Identification · Digital image analysis · Discriminant analysis · Correspondence analysis · t-Distributed stochastic neighbour embedding

## Introduction

Although the name ‘forget-me-not’ for *Myosotis* in many languages urges, that it should not to be forgotten, not much attention has been given to this group by archaeobotanists. Whereas undisputedly complex genera such as *Carex* or families such as Poaceae or Apiaceae have received thorough archaeobotanical attention, the seven species of *Myosotis* in the western European flora have been largely neglected. Even in the 72 pages devoted to identification criteria published by Jacomet et al. (1989, pp 277–348), not a single line is devoted to the identification of *Myosotis*. Knörzer (1970) also simply stated that “all seeds of this genus are very similar”. In 1981, Knörzer mentioned that *M. scorpioides* (listed under the synonym *M. palustris*) has larger seeds (1.3–1.6 mm) than *M. arvensis*, *M. ramosissima* and *M. discolor* (mentioned as *M. versicolor*). Behre (1983) distinguished *M. scorpioides* from the remaining species by a less conspicuous ridge at the transition from the dorsal to the ventral side.

In the Dutch archaeobotanical database RADAR (Van Haaster and Brinkkemper 1995), the taxonomical variation within the 349 records of *Myosotis* is large. Many different combinations of taxa are recognized, practically all

---

Communicated by F. Bittmann.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00334-011-0313-3) contains supplementary material, which is available to authorized users.

---

O. Brinkkemper (✉)  
Cultural Heritage Agency, PO Box 1600, 3800 BP Amersfoort,  
The Netherlands  
e-mail: o.brinkkemper@cultureelerfgoed.nl

L. van der Maaten  
Department of Computer Science, University of California  
San Diego, La Jolla, CA, USA  
e-mail: lvdmaaten@gmail.com

L. van der Maaten  
Pattern Recognition and Bioinformatics Laboratory, Delft  
University of Technology, PO Box 5013, 2600 GA Delft,  
The Netherlands

P. Boon  
Data Archiving and Networked Services (DANS),  
PO Box 93067, 2509 AB The Hague, The Netherlands  
e-mail: paul.boon@dans.knaw.nl

including *M. scorpioides*. This clearly shows the need for streamlining and standardization.

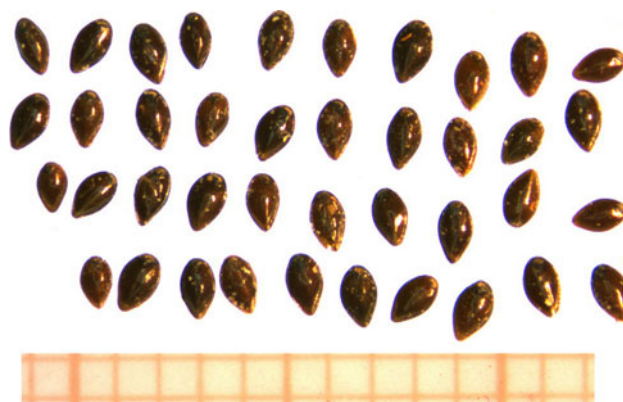
The search for a project to test the software *Fovea Pro* for digital image processing (Russ 2005) thus soon led us to the seven *Myosotis* species. The main research question that this study aims to address is whether digital image processing can supply objective, statistical criteria for the identification of the seven *Myosotis* spp. Such a distinction may be important in archaeological studies, as the seven different species inhabit five distinct biotopes (dunes, woodland, grassland, watersides and arable fields). Hence, specific identification of the seven *Myosotis* spp. may help in the ecological interpretation of archaeobotanical samples.

## Materials and methods

To acquire a representative set of measurements for all species, we collected reference material from at least two locations per species, with at least 100 seeds per location. In some cases, the consulted reference collections of the Dutch Agency for Cultural Heritage (RCE), Leiden Archaeological Centre (IPL), BIAX Consult and Brinkkemper Archaeobotanical Office (BAB) yielded considerably smaller numbers of seeds. In these cases, material from more than two provenances was included in the measurements. Nevertheless, the number of seeds of *M. stricta* remained low, prompting the use of an additional sample that was provided by H.V. Hansen (University of Copenhagen Botanic Garden). The investigated species and the provenance of the material are listed in Table 1 ESM. The samples in the collections of RCE, IPL, BIAX and BAB are samples from one or several plants growing closely together. In most cases, plant material was dried and was supplied to the Dutch State Herbarium to check identifications. For the remaining material, it is more cumbersome to exclude wrong identifications; however, the consistency of the data presented in this study suggests that the identifications of the material are correct. Nevertheless, during the processing of samples we did resolve a small number of incorrect identifications. In particular, IPL4677 contained seeds listed as *M. ramosissima*, but their small size suggested that they belong to *M. discolor* (or the rare *M. stricta*), just as BIAX1306 that was listed as *M. laxa*.

*M. scorpioides* in IPL5548 was much larger than the remaining material of this species, and remarkably pear-shaped besides. This shape was not seen in any other samples, or in other species. This sample was from a Finnish provenance and possibly contains seeds from the montane species *M. scorpioides* ssp. *nemorosa*.

The seeds were digitally photographed together with a strip of millimetre squared graph paper. Ripe, dark brown



**Fig. 1** Example of photograph used for digital image analysis (*M. stricta*) with scale bar in millimetres for calibration

to black seeds were selected from the available samples of recent reference material. From each sample, three photographs with 40 seeds each were taken. All photos were made on a Wild stereo microscope at 6× magnification using a Sony 5.4 megapixel digital camera.

Since the seeds possess a slightly domed dorsal and a rounded ventral side, it was necessary to position them carefully with the ventral side below. By positioning the seeds in this way, their maximum width is captured in the photographs. Besides, the touching of two adjacent seeds on the photo had to be avoided, so that they would not be treated as one (too large) object in the automatic measurements made by the *Fovea Pro* software. Figure 1 shows an example of a photograph used for digital image analysis.

For the digital measurements of parameters from photos, the trial version of *Fovea Pro 4.0* has been used, as developed by John Russ (2005). This program can be installed as a plug-in for image editing software such as *Photoshop*. The enlargement of each individual photo was first calibrated in *Fovea Pro*. To perform the calibration, a length of 13–15 mm of the millimetre paper was measured by selecting the centres of the two end lines. In two cases, a length of 13 mm was measured ten times from one single photo to assess the error of measurement. On a length of 127 and 130 pixels per mm, the standard deviation was 0.10 and 0.12 pixels respectively: a negligible error of less than 0.1%. For all photos, the same strip of millimetre paper has been used, and the length was calibrated once for each individual photo.

To enable the program to measure seed dimensions, the photo first had to be converted into a black and white bitmap. To do so, the background was selected by means of *Photoshop's* magic wand (tolerance 20) and selection lasso tools in an attempt to remove large irregularities such as artefacts that were due to the scratches and dust on the petri dish that contained the seeds. The background and the

millimetre paper were converted into white and the seeds into black. Remaining small black regions (up to a radius of 10 pixels) were removed with the ‘dust and scratches’ filter. This filter also smoothed the frayed margins of the seeds that resulted from the earlier operations. Many seeds still showed a white hole from the reflecting light. The *Fovea Pro* option ‘Fill holes’ prevented the calculation of wrong areas.

Subsequently, all 56 variables available in *Fovea Pro* were measured for all seeds in all photographs. For technical details on how these variables are computed, the reader is referred to the extensive online manual ([www.reindeergraphics.com](http://www.reindeergraphics.com)) and to the illustrated handbook by Russ (2007).

After removing variables from the data not relevant for the identification such as those related to the position in the photograph, 23 variables were left, all of which were related to the size and shape of the seeds. These variables are listed and explained in Table 2 ESM. It is important to note that we only analysed two dimensional pictures, so variables that are related to the thickness of the seeds are therefore not included.

In total, we gathered 1,453 individual series of measurements for the seven species. To distinguish patterns in this large data set (which forms a  $1,453 \times 23$  matrix), three different multivariate statistical analyses were carried out, Discriminant Analysis (DA), Correspondence Analysis (CA) and t-Distributed Stochastic Neighbour Embedding (t-SNE). In DA, the classes of the seeds in the data are used in the statistical analysis (Fisher 1936). Specifically, DA searches for a linear combination of the variables that reduces the variance of the data within each class and, at the same time, maximizes the variance of the data between the different classes. The result of DA is a collection of weights of variables, the norm of which indicates the importance of each variable in distinguishing the different classes. The weights constitute a linear combination of variables that can be used to represent the underlying data in a scatter plot of the two or three most important axes.

In contrast to DA, in CA the class information on the seeds is not used. CA finds a linear combination of the variables that maximizes the variance of the data (Pearson 1901). The norm of the resulting weights indicates how much each variable contributes to the overall variance in the data. Again, the weights can be used to represent the data in a scatter plot. This plot can be inspected for clusters of individual species.

Although DA and CA provide information on the relative importance of the different variables, the linearity of these techniques often result in inferior representations of the data in scatter plots, because the linearity greatly restricts the way in which the data representation is constructed. In techniques such as t-SNE, this problem is

resolved by performing non-linear transformations of the data to construct the scatter plots (van der Maaten and Hinton 2008; van der Maaten 2009). In particular, t-SNE calculates Euclidean distances between each series of measurements and plots the data in a scatter plot in such a way that the distances between each point in this scatter plot resemble these Euclidean distances as closely as possible. The resulting plots often provide much more insight into the structure of the data than the plots constructed by DA or CA. In t-SNE, each variable is given the same weight so that it is not possible to assess which individual variable(s) are decisive for the identification of the individual species (or groups of species).

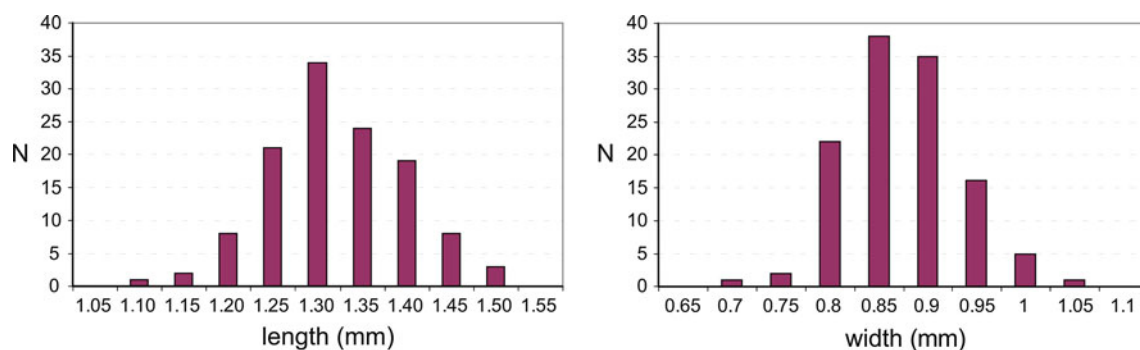
## Results and discussion

### The measurements

An earlier study on measuring seed characteristics using *Fovea Pro* has been published by Rovner and Gyulai (2007). They measured a large range of taxa and, remarkably, found that many of the characteristics of the investigated taxa did not behave according to a Gaussian distribution, with random values that cluster around a single mean value. Instead, this study suggests that many of the characteristics are in fact multimodal. In a comparison between domestic cereals and wild grasses, they found that cultivated taxa yielded data closer to a Gaussian, unimodal distribution than wild ones.

In our data set of *Myosotis* seeds, no multimodal distribution of the variables could be observed. For instance, histograms of seed length and width suggest that the underlying distribution is unimodal (Fig. 2). We performed Lilliefors tests on our data set, with the null hypothesis that the underlying data distribution is Gaussian (Lilliefors 1967). For length, the hypothesis was rejected for *M. arvensis*, and for width it was rejected for *M. laxa* at 5% confidence level and all measurements per species lumped. However, for the other species, the length and width measurements can be considered as Gaussian at the 5% level. The variable area is not Gaussian for *M. arvensis* and *M. sylvatica*, but it is Gaussian for the other five species. We did not test for each individual provenance whether its underlying distribution is Gaussian; however, for length, width and area, Gaussian distributions appear to be the norm (compare Fig. 2).

Our results conflict with the results of Rovner and Gyulai (2007), who claimed that the underlying distributions of seed area are non-Gaussian. Possibly, the multimodal distributions for area found by them are the result of the random positioning of the seeds. Their acquisition of photos by means of a video camera might also have



**Fig. 2** Histograms of length and width of *M. ramosissima* (IPL166;  $N = 120$ )

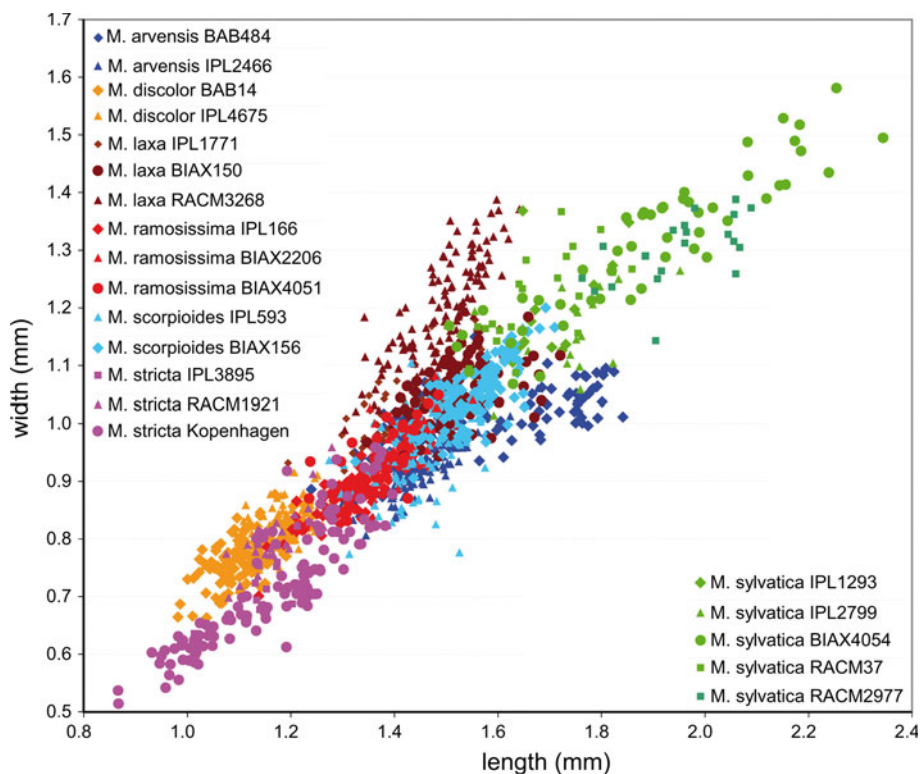
introduced structural differences in different photos. The careful positioning of the seeds in our study presumably explains why our findings suggest that the distributions of most seed characteristics are Gaussian.

From the point of view of easy applicability in routine identifications, it would be ideal if the lengths and widths of the seeds would suffice to distinguish between the various species. To explore the feasibility of such an approach, we constructed a scatter plot that shows length against width for all 1,453 seeds (Fig. 3). This diagram reveals that *M. sylvatica* can easily be separated from the other species because of its larger length and width. All seeds longer than 1.9 mm in our data set belong to this species, and a

width larger than 1.2 mm, combined with a length larger than 1.7 mm also uniquely identifies this species. A width larger than 1.2 mm only occurs in *M. laxa*, in combination with a length less than 1.7 mm. However, a considerable proportion of these seeds is less wide and overlaps with *M. scorpioides*.

In the small range, *M. stricta* and *M. discolor* form a group below a length of 1.2 mm and a width of 0.9 mm which has no overlap with other species. *M. ramosissima* and *M. scorpioides* overlap largely in length and width, and these species also partly overlap in size with *M. arvensis*. A limited number of this latter species have a length larger than 1.7 mm and a width smaller than 1.1 mm, making

**Fig. 3** Scatter diagram of length versus width for all species



them more slender than the remaining species. Taken together, the scatter plot reveals that identification is already possible to some extent based only on length and width, but it also shows that not all species can be identified unequivocally by using these two variables. Therefore, the three mentioned multivariate analyses were applied on the full data set of 23 variables, furthering an attempt to identify other variables that discriminate between species.

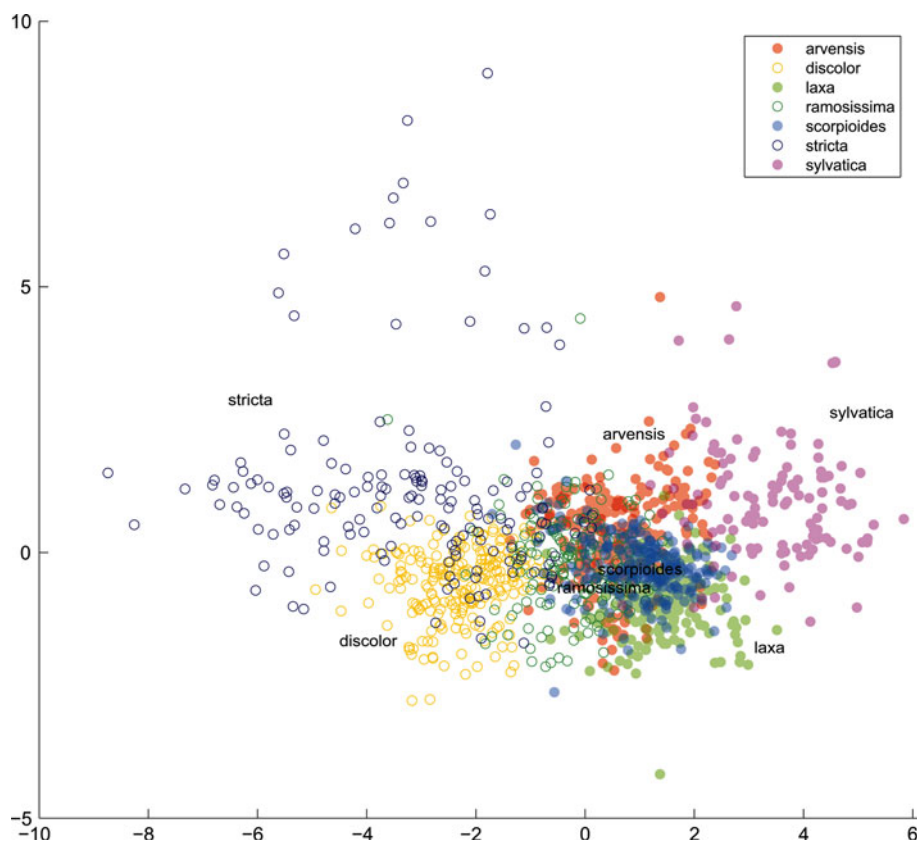
#### Discriminant analysis

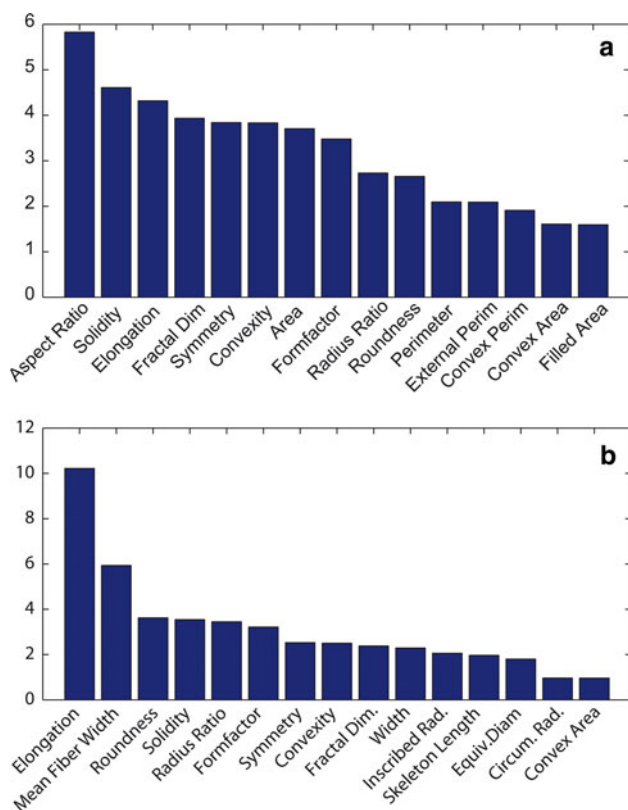
The scatter plot from the DA of all measurements is shown in Fig. 4. The separate species are given their own colours, irrespective of their provenance. The resulting plot shows some species forming their own, separate cluster, while other species became more mixed. This analysis again gives the best separation of *M. sylvatica*, but *M. stricta* also forms a separate cluster that only shows some overlap with *M. discolor*. The locations of *M. arvensis*, *M. laxa*, *M. ramosissima* and *M. scorpioides* show much overlap in the DA plot.

By assessing the contribution of the individual variables to each of the first two axes, we can get insight into the variables that are most relevant in the observed separation

of the species. The weights of the variables of the first (horizontal) axis are plotted in a histogram in Fig. 5a. Aspect ratio (length/width), solidity and elongation are the most important variables on the X-axis. The Y-axis is for a large part determined by elongation (Fig. 5b). From the scatter plot in Fig. 4, we may conclude that these variables can apparently distinguish *M. sylvatica* from the other species and, to a lesser extent, to discriminate between *M. stricta* and *M. discolor*. To be able to judge the usefulness of the variables in distinguishing the seven species, we constructed a scatter plot of the two most important variable of each axis, aspect ratio versus elongation (Fig. 6). This plot reveals that these variables can be used to distinguish *M. laxa* from the other species. At the same time, the clustering of the other seeds into classes is worse than in the length/width scatter. Hence, we conclude that a two-dimensional DA scatter plot does not contribute to a further separation of species if the whole data set is analysed. Having said this, the reader should note that the remaining four DA dimensions may contain additional information on how to distinguish the seven species. We did not further investigate this, since the first two axes account for 97% of the Fisher criterion (the variance within the classes divided by the variance between the averages of the classes).

**Fig. 4** Discriminant analysis scatter plot of all species



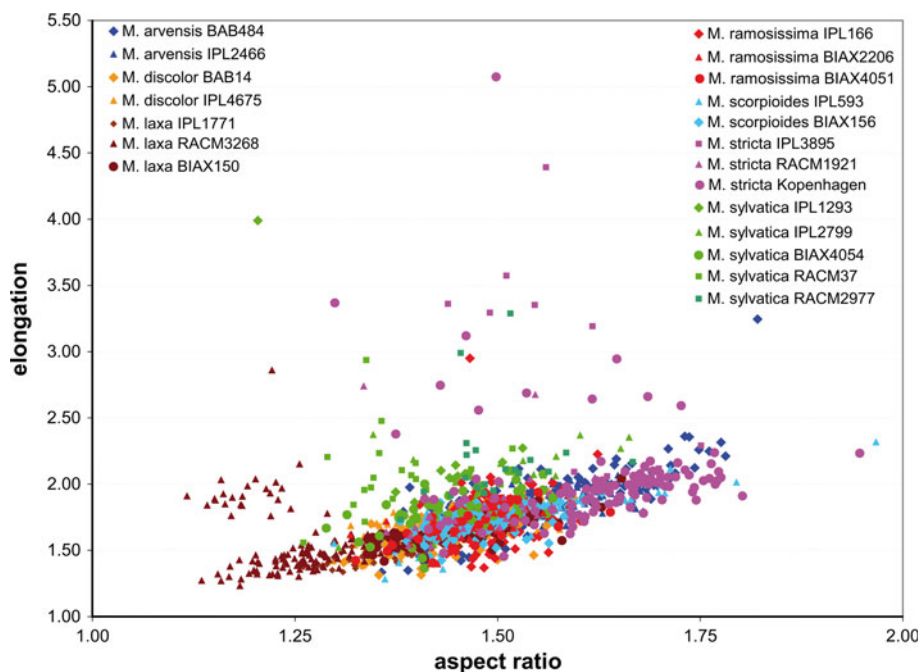


**Fig. 5** Histogram of the weight of the variables; **a** on the first axis of Discriminant Analysis; **b** on the second axis

#### Correspondence analysis

The first two axes of CA account for 92% of the total variation in the data set (the total of their eigenvalues

**Fig. 6** Scatter diagram of aspect ratio versus elongation for all species



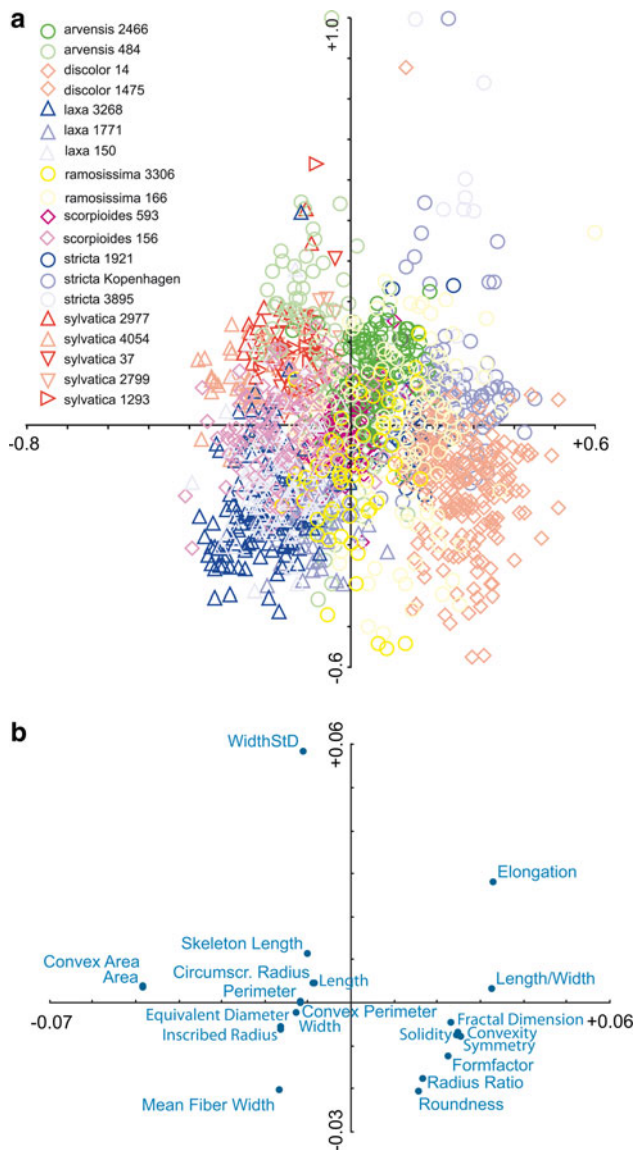
divided by the sum of all eigenvalues in the dataset). The locations of the seven species on these axes are plotted in Fig. 7a. In the resulting plot, *M. discolor* and *M. stricta* are separated from the other species along the first axis (right side of the plot) and *M. laxa* along both axes (lower left side of the plot).

The corresponding weights of the variables are plotted in Fig. 7b. The X-axis is mainly determined by area and convex area, which are more or less identical since the seeds have no appendages. The vertical axis is mainly determined by width standard deviation and to a lesser degree by elongation. A scatter plot of area versus width standard deviation is shown in Fig. 8.

In this plot, *M. sylvatica* is separated from the other species on the right side and *M. discolor* and *M. stricta* on the left. The positions of one of the samples of *M. scorpioides* (IPL593) and one of *M. laxa* (IPL1771) are remarkable. These samples have high width standard deviations, which mean that their widest and smallest widths greatly differ. Since the other samples of these species do not have outlying width standard deviations, this characteristic does not seem to be of general value for identification.

#### t-Distributed stochastic neighbour embedding

The t-SNE scatter plot of the seven species is shown in Fig. 9. Again, the group of *M. discolor* and *M. stricta* can be recognized on the lower left part of the plot. Part of the *M. stricta* seeds, however, has a wider range that partly overlaps with *M. arvensis* and *M. ramosissima*. A large part



**Fig. 7** Scatter plot of Correspondence Analysis; **a** for all species; **b** for the variables

of *M. sylvatica* occurs on the opposite side, but there is also a smaller but dense cluster in the upper left part of the plot. *M. laxa*, *M. arvensis*, *M. ramosissima*, and *M. scorpioides* show much overlap again.

#### Further separation of smaller groups of species

Since *M. laxa* can be separated rather well from the three similar species on the basis of roundness, a separate DA was carried out with *M. arvensis*, *M. scorpioides* and *M. ramosissima*. The resulting scatter plot yields a very poor separation of the individual species (Fig. 10). CA

did not result in an adequate separation of the three species either.

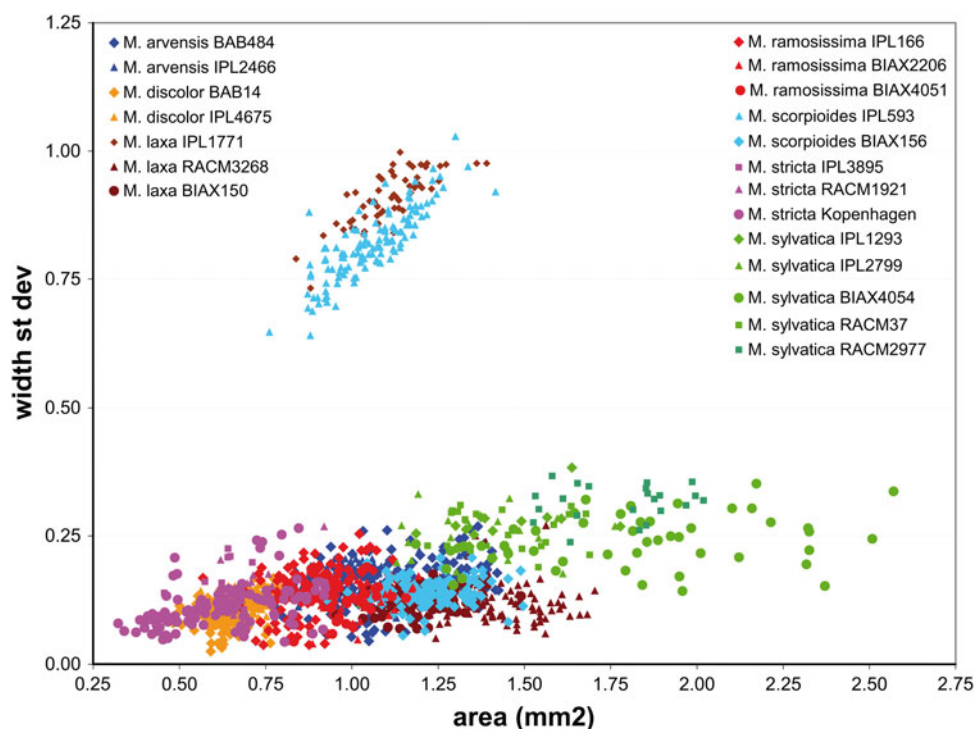
By contrast, the analysis by means of t-SNE did result in the separation of part of the seeds of *M. arvensis* (upper left) and *M. scorpioides* (lower left, in Fig. 11). Apparently, the non-linear combination of the variables provides more clues for identification, suggesting that the information that is necessary to discriminate between the three species is present in the 23 investigated variables. For the identification of subfossil seeds, however, these clues can only be used by measuring the 23 variables for the new seed and then performing a new t-SNE analysis together with the three sets of identified seeds. Alternatively, the t-SNE analysis could be combined with techniques for automatic classification (Duda et al. 2000).

A t-SNE scatter plot of *M. discolor* and *M. stricta* does not produce an adequate separation of these two species (Fig. 12). With DA it is impossible to plot a scatter of these two species, as two groups only yield one dimension to be plotted. However, it is possible to calculate the weight of the variables that provide an optimal separation of the two species on the single available axis (Fig. 13). Aspect ratio and radius ratio have the highest weights, so these two variables contribute most to an optimal separation of the two species. A scatter graph of these two variables indeed produces a promising result (Fig. 14). Small seeds with an aspect ratio (length/width) above 1.60 and/or a radius ratio below 0.62 belong to *M. stricta*. A relatively small part (c. 5%) of *M. discolor* can be separated by an aspect ratio below 1.33 and/or radius ratio above 0.73, the remainder of the seeds overlap with a minority of *M. stricta* in the intermediate ranges.

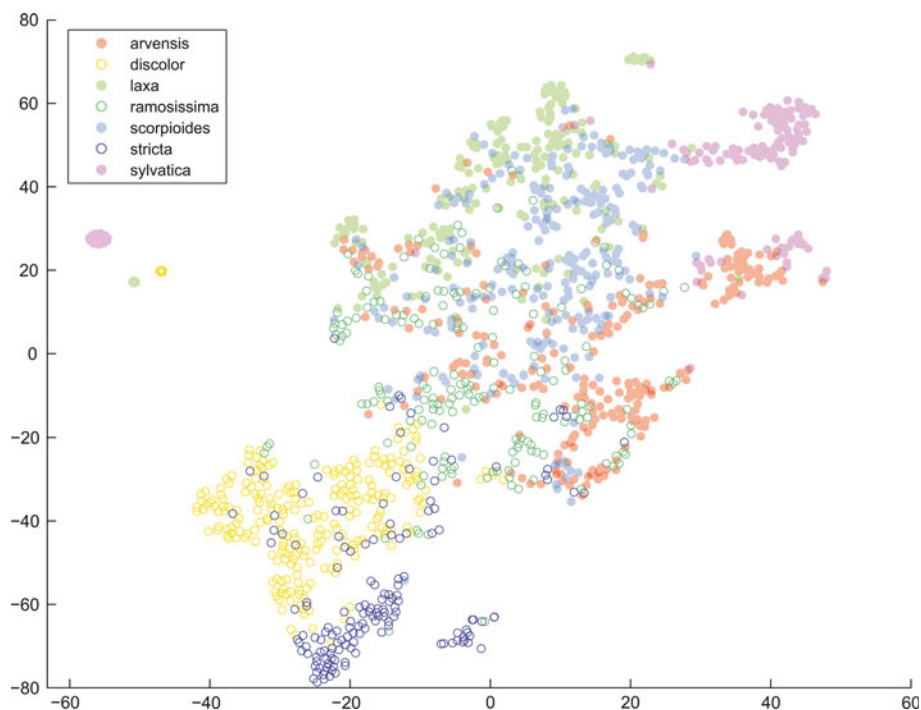
## Conclusions

From the results presented in this study, we conclude that the identification of seeds of various *Myosotis* species can now be performed more precisely than was hitherto published in the archaeobotanical literature. Although it is not feasible to identify each individual seed to species level with certainty on the basis of simple seed metrics such as length and width, these two variables already allow for the discrimination of three groups. *M. sylvatica* has by far the largest seeds, with an average length of 1.8 mm. All seeds larger than 1.9 mm can be attributed to this species with certainty, as well as the seeds exceeding 1.7 mm in length and 1.2 mm in width. Moreover, the seeds of *M. sylvatica* are characterized by a greatest width clearly above half of the height, giving them a shouldered appearance.

**Fig. 8** Scatter diagram of area versus width standard deviation for all species



**Fig. 9** Scatter plot of t-SNE Analysis for all species

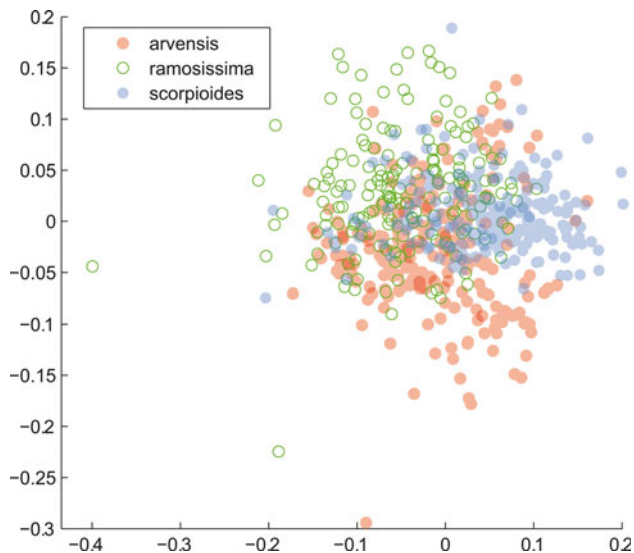


In the small range, *M. discolor*/*M. stricta* can be separated, with a length smaller than 1.2 mm and/or a width smaller than 0.9 mm. Within this group, more than half of the *M. stricta* seeds can be distinguished because their length/width ratio is larger than 1.60 mm and/or their radius ratio is smaller than 0.62 mm. A few of the seeds of

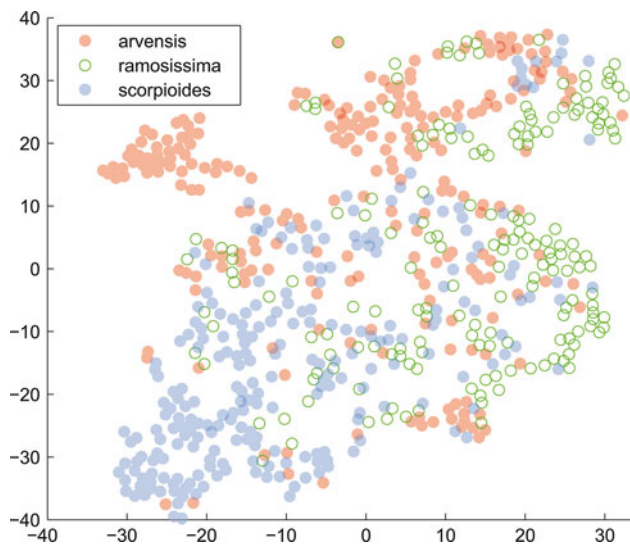
*M. discolor* have a length/width ratio smaller than 1.33 mm and/or a radius ratio larger than 0.73 mm.

In length and width, the remaining four species occupy a position between these two extremes. From these four, *M. laxa* has a relatively small length/width ratio. Values below 1.3 mm suggest this species (only few *M. sylvatica*





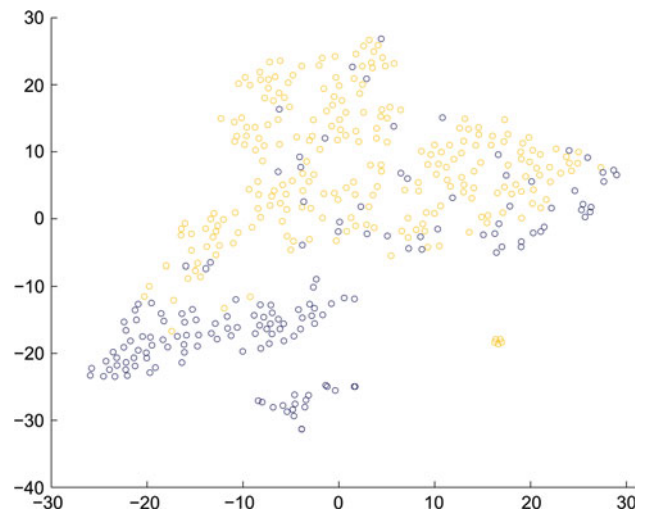
**Fig. 10** Scatter plot of Correspondence Analysis for *M. arvensis*, *M. ramosissima* and *M. scorpioides*



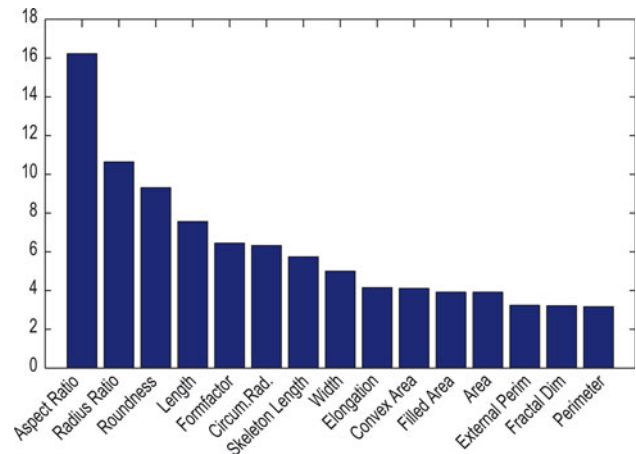
**Fig. 11** Scatter plot of t-SNE Analysis for *M. arvensis*, *M. ramosissima* and *M. scorpioides*

have these values as well, but these are much longer). This criterion serves to split off c. 40% of the seeds of *M. laxa*.

*M. arvensis*, *M. ramosissima* and *M. scorpioides* cannot be separated by means of DA or CA. Using t-SNE, however, we showed that it is possible to successfully identify c. 20% of *M. arvensis* and c. 25% of *M. scorpioides*. In routine analysis, these three species cannot be identified on the basis of volumetric data.



**Fig. 12** Scatter plot of t-SNE Analysis for *M. discolor* and *M. stricta*

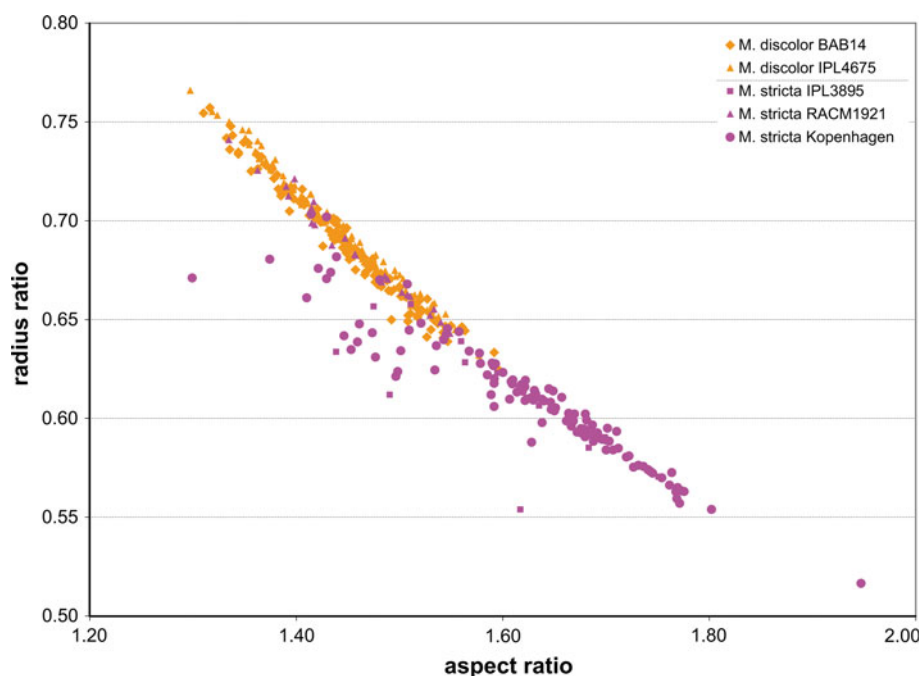


**Fig. 13** Histogram of the weight of the variables on the axis of Discriminant Analysis for *M. discolor* and *M. stricta*

However, the unsharp transition between the dorsal and ventral side of *M. scorpioides* as put forward by Behre (1983) is a useful criterion to further split this group of three species without the use of volumetric data. This criterion is not expressed in one of the variables measured in *Fovea Pro*.

An identification key for waterlogged *Myosotis* seeds is presented below. It is evident that seeds that lack part of their original seed coat yield deviating volumetric values, as a result of which they cannot be readily compared to intact reference seeds. Carbonization also causes differential shrinkage of seeds. Therefore, the key presented below can only be applied to waterlogged subfossil seeds with intact seed coats.

**Fig. 14** Scatter diagram of aspect ratio versus radius ratio for *M. discolor* and *M. stricta*



### Identification key for waterlogged *Myosotis* seeds (with intact seed coat)

- 1a. Seeds with largest width distinctly below half of the height. Seeds generally longer than 1.9 mm or longer than 1.7 mm AND wider than 1.2 mm ..... *M. sylvatica*
- 1b. Seeds smaller ..... 2
- 2a. Seeds with maximum length of 1.2 mm and/or maximum width of 0.9 mm ..... 3
- 2b. Seeds larger ..... 5
- 3a. Aspect ratio (length/width) above 1.6 and/or radius ratio below 0.62 ..... *M. stricta* (p.p.)
- 3b. Otherwise ..... 4
- 4a. Aspect ratio below 1.33 and/or radius ratio above 0.73 ..... *M. discolor* (p.p.)
- 4b. Otherwise, or these digital image analysis data unavailable ..... *M. discolor/stricta*
- 5a. Seeds relatively wide, aspect ratio below 1.25 ..... *M. laxa* (p.p.)
- 5b. Seeds more slender ..... 6
- 6a. Seeds with smooth transition between dorsal and ventral side, not ridged ..... *M. scorpioides*
- 6b. Seeds with ridged transition between dorsal and ventral side ..... *M. arvensis/laxa* (p.p.)/*ramosissima*

**Acknowledgments** Without the reference material that was graciously offered to us by various sources, this article could never have been written. We thank Wim Kuijper (IPL), Henk van Haaster and Wouter van der Meer (BIAX) for their cooperation in this respect. Jan Bastiaens (VIOE, Belgium) kindly arranged the contact with H.V. Hansen (Copenhagen Botanic Garden), who supplied the majority of the seeds of *M. stricta* included in this study. This work was performed while Laurens van der Maaten and Paul Boon were at the Cultural Heritage Agency (RCE). Laurens van der Maaten and Paul Boon were financially supported by the Netherlands Organization for Scientific Research (NWO) program CATCH, project RICH (project number 640.002.401). Laurens van der Maaten is also supported by NWO grant no. 680.50.0908, and by EU-FP7 Social Signal Processing (SSPNet).

### References

- Behre K-E (1983) Ernährung und Umwelt der wikingerzeitlichen Siedlung Haithabu. Die Ergebnisse der Untersuchungen der Pflanzenreste. (Die Ausgrabungen in Haithabu 8) Wacholtz, Neumünster
- Duda RO, Hart PE, Stork DG (2000) Pattern classification. Wiley, New York
- Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Ann Eugenics* 7:179–188
- Jacomet S, Brombacher C, Dick M (1989) Archäobotanik am Zürichsee. Ackerbau, Sammelwirtschaft und Umwelt von neolithischen und bronzezeitlichen Seeufersiedlungen im Raum Zürich. Füssli, Zürich

- Knörzer K-H (1970) Novaesium IV. Römerzeitliche Pflanzenfunde aus Neuß. (Limesforschungen 10) Mann, Berlin
- Knörzer K-H (1981) Römerzeitliche Pflanzenfunde aus Xanten. (Archaeo-Physika 11) Köln
- Lilliefors H (1967) On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J Am Stat Assoc* 62:399–402
- Pearson K (1901) On lines and planes of closest fit to systems of points in space. *Philos Mag* 2:559–572
- Rovner I, Gyulai F (2007) Computer-assisted morphometry: a new method for assessing and distinguishing morphological variation in wild and domestic seed populations. *Econ Bot* 61:154–172
- Russ J (2005) Fovea Pro 4.0. Computer software, Reindeer Graphics
- Russ J (2007) *The image processing handbook*, 5th edn. CRC Press, Boca Raton
- Van der Maaten LJP (2009) Feature extraction from visual data. Tilburg (Doctoral thesis)
- Van der Maaten LJP, Hinton GE (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9:2579–2605
- Van Haaster H, Brinkkemper O (1995) RADAR, a relational archaeobotanical database for advanced research. *Veget Hist Archaeobot* 4:117–125