Visualization of Phylogenetic Tree space

Daniel Berglund

Visualization of Phylogenetic Tree space

Daniel Berglund[*]

Juni 2011

Abstract

When working with phylogenetic trees and Markov chain Monte Carlo (MCMC) methods it is often of interest to visualize the tree space provided by the MCMC algorithm. First, we need to define a metric so we can get an distance between all trees. However the tree space is an high dimensional space so it needs to be reduced to allow visualization. In this thesis we look at two metrics and four dimension reduction methods. The main conclusion is that metric is much more important than the method used to reduced the dimensions. Different metrics can give almost completely different plots while different dimension reduction methods rarely radically changes the plots.

[*]Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: daniel.k.berglund@gmail.com . Supervisor: Sebastian Höhna.

1 Introduction

1.1 Why visualize the Phylogenetic Tree space?

The main reason why you want to visualize the tree space is that a plot in two or three dimensions is much easier to get an overview of than a multidimensional matrix. Using the plot it's possible to check how the MCMC algorithm behave when walking through the tree space. For an example you can see how it behaves with "islands" of higher probability which is important when studying different types of random walks([1]). In [2] they use visualization to compare how many times different types of MCMC algorithms crosses between two islands. It is also important when looking for a good summary tree, i.e. the mean of the trees.

1.2 Outline

In the background section we will start with an introduction to phylogenetics, tree metrics and dimension reduction techniques. All the methods and programs used are also presented in that section. In the next section some plots of trees are presented and discussed. In the last section the results are combined and a conclusion is drawn and discussed.

2 Background

2.1 Phylogenetics and phylogenetic trees

Phylogenetics is the study of evolutionary relationships among groups of species and populations through molecular sequences and morphological data. The goal is to describe the history of evolution for the species or their parts, both living and extinct.

A phylogenetic tree is a branching diagram used to describe the evolutionary relationships. A rooted tree is a directed diagram and has an unique internal node corresponding to the most recent common ancestor. In some rooted trees the length of the edges may bee interpreted as time estimates. The trees can also have different numbers of descendants from an interior node. Binary(bifurcating) trees have two and multifurcating have two or more. All trees in this thesis are binary and rooted.

Formal definition
We define a *phylogenetic tree* T on S as a tree T with nodes $p_1, ... p_n$. To each node p_i we assign a *label* S_i which is an partition of S. Nodes with degree tree or more may have empty labels, but nodes of degree one or two must have non-empty labels. A rooted tree is binary and has a unique internal node with degree two. The set of nodes is P and the set of *edges* is E, T=(P,E). The edges cannot be assigned negative lengths. The set of all trees on S is γ_s. [3]

2.1.1 Newick format
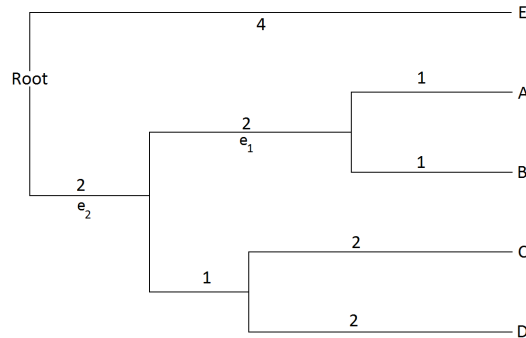
If we have this rooted and binary tree

Figure 1: *A tree with 5 taxa.*

It is represented by (((A:1.0,B:1.0):2.0,(C:2.0,D:2.0):1.0):1.0,E:4.0); in Newick format.

The tree ends with a semicolon. Interior nodes are represented by a pair of matched parentheses. Between them are the nodes that are the descenders from that node, separated by commas. In the example tree the immediate descendants from the root are E and two interior nodes. The two interior nodes are represented by a pair of parentheses enclosing its representation of its immediate descendants. In this case A,B for one interior node and C,D for the other. There can also be new interior nodes which can create further levels of nesting. The names of the taxa can be any string of characters except blanks, colons, semicolons, parentheses, and square brackets.

The number after each node with the colon is the length of the branch leading to that node. It can be any real number, with or without a decimal point. The trees can also be multifurcating at any nodes.[4]

2.2 Tree metrics

To be able to compare the distances between two trees we first need to define a metric on the tree space. In this thesis two metrics are used, Robinson Fould-distance(RF) and weighted Robinson Fould-distance(WRF). They are both based on the splitting of trees by removing an edge and comparison of the subsets that the split creates.

2.2.1 Robinson Foulds metric

Robinson and Foulds introduced a distance metric called Robinson Foulds distance (RF) or partition metric in their paper 1981. It's one of the simplest metrics for trees and is defined as the minimum number of operations needed to transform both trees to the exact same tree using an operation called α.

The operation α and its inverse is defined as:

Let T_1 be phylogenetic tree on set γ_s and let e_1 be an edge of T_1 with corresponding nodes p_1 and p_2. By removing the edge e_1 and merging the two nodes, p_1 and p_2, into a new node we have formed a new tree. The label of the new node is the union of the labels of p_1 and p_2. The inverse of the operation, α^{-1}, is defined as removing one node and replacing it with an edge and two nodes. The edges connecting to the old node may bee assigned to any of the new nodes and the labelling of the new nodes can be chosen in any way as long as the union of the labels from the new nodes is the same as the labels for the removed node. Note that the new tree formed by either α or α^{-1} necessarily don't have to be binary.[3]

They also define matching edges as the two edges that split the two trees into the same non-empty subsets when removed. Let $(A|B)$ denote the *split* of a tree T into two non-empty sets, A and B. Define the function f on an edge as

$$f(e_1) = (A|B)$$

Edges $e_1 \in E_1$ and $e_2 \in E_2$ are then matched if and only if $f(e_1) = f(e_2)$.
For an example the edge e_1 in figure 2 is matched with the edge e_1 in figure 1 while the edges e_2 don't have any match in the other tree.
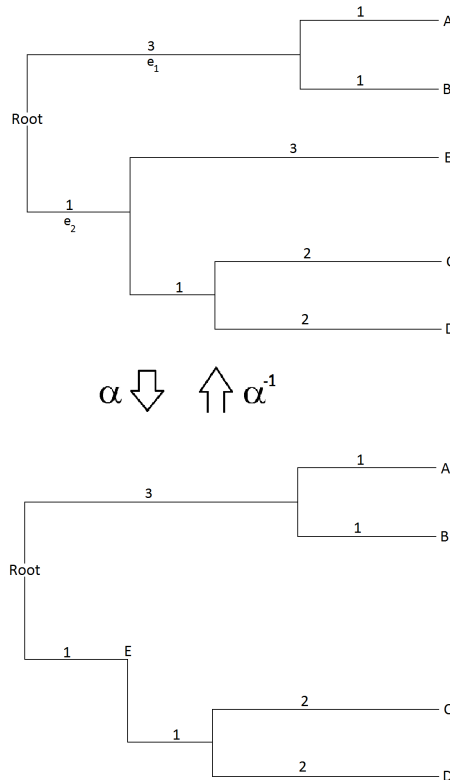

Figure 2: *An example of the operations α and α^{-1}.*

Later in the same paper they show that the distance between two trees T_1, T_2 is defined as follows

If $T_1, T_2 \in \gamma_S$ then

$$d(T_1, T_2) = |E_1/\acute{E}_1| + |E_2/\acute{E}_2|$$

where E_1 and E_2 is the set of all edges on T_1 respectively T_2. $\acute{E}_1$ and $\acute{E}_2$ is defined as

$$\acute{E}_1 = \{e_1 \in E_1 : \exists e_2 \in E_2 s.t. f(e_1) = f(e_2)\}$$
$$\acute{E}_2 = \{e_2 \in E_2 : \exists e_1 \in E_1 s.t. f(e_2) = f(e_1)\}$$

In other words $\acute{E}_1$ is the set of all edges in E_1 that has an match in E_2. So the distance is all the unmatched edges for both trees. For complete proof see Robinson, Foulds, Comparison of phylogenetic trees, 1981.[3]

The distance between the two trees in figure 1 and figure 2 is two. By removing the edge e_2 in both trees one gets the same tree. Remember that the root of the tree has to have a label, otherwise the distance might become shorter than it really is.

It can also be shown that.

Proposition *If $T_1, T_2 \in \gamma_S$ and are binary and rooted then*

$$|E_1/\acute{E}_1| = |E_2/\acute{E}_2|$$

Proof
Binary, rooted trees with the same number of taxa always have the same number of edges and since $\acute{E}_1 = \acute{E}_2$, because it is a one to one correspondence between matched edges, the result follows.

But the metric has its limitations, Steel and Penny [5] write that it is fast to compute but that its Possion distribution is very skewed. So it's only useful if the trees that are being compared are very similar. It also gives longer distances than what is intuitively expected when two taxa far away from each other is the only difference between two trees. This is because it takes a lot of operations to join those two taxa into one node in both trees, while a metric based on grafting and pruning would give a smaller distance. It's even possible for two trees where only one taxa have switched place to have the maximum possible distance between them.[5]

2.2.2 Weigthed Robinson Foulds metric

The weighted Robinson Foulds metric(WRF) is one of the simplest extensions of the metric and adds the length of the edges to the metric. As with RF-distance let $(A|B)$ denote the *split* of a tree T into two non-empty sets, A and B. Removing an edge e from the tree divides the tree into the two sets and thereby creating a split. Let $\sum T$ be the set of all splits in a tree T. The metric is then defined as

For each split $A|B \in \sum T_i$, $i = 1, 2$ we let $w_i(A|B)$ denote the length of the corresponding edge and let $w_i(A|B) = 0$ for all $A|B \notin \sum T_i$

$$d(T_1, T_2) = \sum_{A|B \in \sum T_1 \cup \sum T_2} |w_1(A|B) - w_2(A|B)|$$

However just as RF-distance it can give longer distances than intuitively expected when only one taxa have switched place. The distance between the two example trees in figure 1 and figure 2 is 5.[6][7]

2.3 Dimension reduction

Dimension reduction techniques are methods used for reducing the numbers of dimensions in a dataset and can be split into two groups, feature extraction and feature selection. Feature extraction methods transforms the data into a space with a lower dimension, the transformation can be linear or non-linear. Feature selection methods try to find a subset of the original variables that best describes the dataset. All methods used here are feature extraction methods.[8]

Using these methods the number of dimensions can be reduced to two or three which allows visualization. It is also useful when modelling data since most traditional statistical methods break down when the number of variables is bigger than the number of observations. While some computationally expensive methods([9]) can construct models with high accuracy from high-dimensional data it's often still of interest to reduce the number of dimensions before any modelling.[10]

The goal is to reduce the dimensions with as little information loss as possible. However it is generally not possible to reduce the data with zero loss, even when reducing to the "true" number of dimensions due to errors in the original measurements.
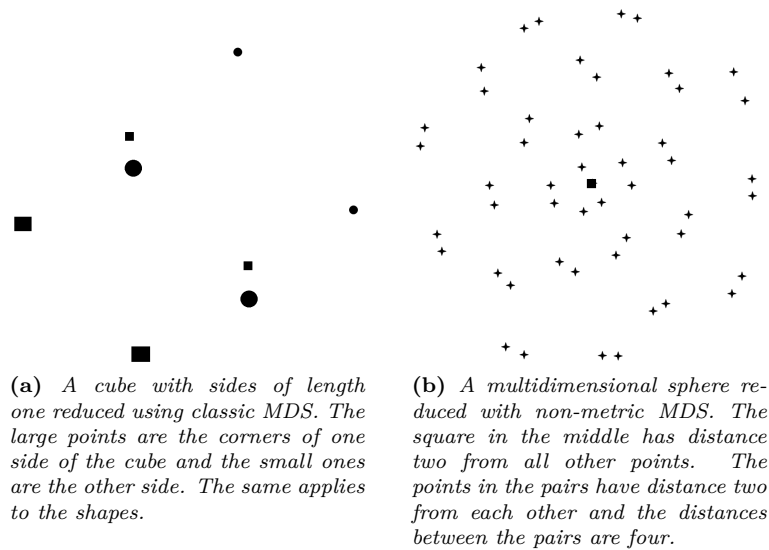
(a) *A cube with sides of length one reduced using classic MDS. The large points are the corners of one side of the cube and the small ones are the other side. The same applies to the shapes.*

(b) *A multidimensional sphere reduced with non-metric MDS. The square in the middle has distance two from all other points. The points in the pairs have distance two from each other and the distances between the pairs are four.*

Figure 3: *Two examples of distortion*

To examine the distortion of the distances in the visualization we can visualize figures that we already know how they look and what their true distances are. Then it's easier to see how the distances get changed. Two appropriate figures shown here are a cube and a multidimensional sphere. In the cube the small circle and large square next to each other are actually further away from each other than the two large shapes. In the sphere the distance between the points in a pair seems much shorter than from the pair to the middle point even though the distance is the same. We will take a closer look on them using all dimension reduction methods in the results and discussion section. Because of the distortion and that the scales and orientation are arbitrary it's important to remember that you cannot compare any distances between reduced datasets. On the other hand it's possible to overcome this problem when dealing with trees by considering all possible solutions and then plot the results of the analysis in that space. However this only possible with small datasets because of the large number of possible trees. It is not possible either with a continuous metric like the weighted RF-distance([1]). Also many of these methods are iterative which means they can get stuck in local optimums so it is important to do multiple runs with different starting points.

The methods used here are multidimensional scaling(MDS) and principal component analysis(PCA). MDS is used to find a spatial configuration of objects when you have measure of their similarity or dissimilarity. It has its origins from a problem where one have distances from a map and want to know how the map looks. PCA uses an orthogonal transformation to transform a set of observations into a set of values of uncorrelated variables which are called principal components. The number of components are less or equal to the number of variables in the original dataset. Either you can use the PCs that describe the dataset best or use them for feature selection.[11][10]

2.3.1 Multidimensional scaling

Multidimensional scaling (MDS) are methods for multivariate and exploratory data analysis. The goal is to find a spatial configuration of the objects represented as points when the information known is a measure of their similarity or dissimilarity. These points are arrange so that their distances corresponds to the similarities of the objects, similar objects have points closer to each other and dissimilar objects by points that are far apart.[11]

Most software divide the methods it into two groups, metric and non-metric. The classical method is a type of metric MDS which assumes that the distance matrix displays metric properties like distances on a map. The distances in the reduced matrix therefore preserves the intervals and ratios as good as possible. It minimizes a loss function called strain. The other metric methods are an extension of the classical MDS with generalized optimization with weights, different loss functions and so on, but the transformation is still linear. In those methods the loss function is called stress and often minimized using stress majorization. But sometimes the metric assumption will be too strong or that the distances between points is of no interest, only their order is. That is where non-metric MDS is used, it only assumes that the order of the distances is important and uses a monotonic transformation of the distance matrix. So it will try to keep the order as good as possible while intervals and ratios are ignored. [11][12]

Metric MDS only have to optimize the points in the configuration while non-metric is a twofold optimization. First the monotonic transformation of the input distances has to be optimized. Then the points in the configuration will be arranged in such away that their distances match the scaled input distances as close as possible.[11]

The most common stress function in MDS is Kruskal Stress, also called Stress1. This is the stress function used here in non-metric and metric MDS. Let $f(x_{ij})$ denote the transformation, d_{ij} is the distance between the points i and j and x_{ij} is the distances given in the input matrix. For the metric MDS we have $f(x_{ij}) = x_{ij}$. Then the Kruskal Stress function is defined as

$$\sqrt{\frac{\sum_{ij}(f(x_{ij}) - d_{ij})^2}{\sum_{ij} d_{ij}^2}}$$

The stress doesn't tell everything about the goodness of fit and another useful tool is a Shepard plot. It displays the relationship between the true distances and the distances in the point configuration by plotting them against each other. In metric and classic MDS the points should be close to a straight increasing line and in non-metric MDS the should be close to the monotonically increasing line of the transformed true distances, called disparities. Shepard plots can also be used on other methods to see how the new distances looks compered to the true distances.[11]

2.3.2 Principal component analysis

Principal component analysis(PCA) is the best linear dimension reduction technique in the mean squared error sense. It reduces the data by finding orthogonal linear combinations, called principal components (PC), that explains the most variance. Their are as many PCs as there are variables in the dataset. The first PC, s_1 explains the most variance, the second PC, s_2 explains the second most variance and is ortogonal to the first PC and so on. It is usually possible to just keep the first PCs and discard the rest with little loss in explained variance. These components can then be used to reduce the data, either you use a number of the first components themselves or use them for feature selection. We have $s_1 = X^T w_1$, where X is an n x n matrix and the coefficient vector(n x 1) w_1 solves

$$w_1 = \arg\max_{||w||=1} Var(X^T w)$$

The second PC is the linear combination with the second largest variance and orthogonal to the first PC, and so on. The variables in the dataset should be normalized to have mean zero and the same variance since PCA is sensitive to scaling, because the variance depends on the scale.[10]

2.4 Procedures and programs

In this section we will go through how the distance matrices can be calculated, reduced and plotted. The programs used are Matlab, R, Weka[13], Spotfire[14] and two programs written for this thesis.

2.4.1 Calculating the distances

For this thesis two programs were written in C++ to calculate the distances. To shorten the time for calculating the RF-distance the result $|E_1/\acute{E}_1| = |E_2/\acute{E}_2|$ was used, that means only one of them needs to be calculated in $d(T_1, T_2) = |E_1/\acute{E}_1| + |E_2/\acute{E}_2|$. The distances were calculated on five datasets, shown in a table below. Each dataset contains 2000 trees. However not all trees were used since the program crash if the trees are to many. Trees were selected in two ways, every tenth tree for a total of 200 trees and the first 300 trees.

Dataset	Number of species sampled	Number of species missing
Acanthiza	12	0
Aegotheles	7	1
Ficedula	25	0
Myioborus	19	13
Tringa	12	1

Table 1: *A table showing the datasets that were used in this thesis.[15]*

2.4.2   Reducing the dimensions

Here we will look at R, Matlab and Weka[13] and which commands to use to run different MDS and PCA methods.

Both R and Matlab can do classical and non-metric MDS, however only Matlab can perform metric MDS. Also R doesn't return the disparities for non-metric MDS. Classical MDS is performed with the command cmdscale in R and with mdscale using the option 'strain' as 'criterion' in Matlab and they give the same result. Metric and non-metric MDS in Matlab is performed with the same command but with the option 'metricstress' and 'stress' instead of 'strain' as the 'criterion' option. In R non-metric MDS is performed using the command metamds in the mass package or in this case isomds. Isomds is from the vegan package[16] and uses metamds several times to find the best solution. Non-metric MDS in R and Matlab don't give exactly the same results. However it's possible that it's only the orientation that is different since they both use the same stress function, Kruskal Stress. In this thesis we used R, since the program doesn't stop if it has problems converging, as for Matlab.

Weka was used for PCA, it can be found under unsupervised filters. It normalize the data and outputs a set number of PCs or enough PCs to cover a set percentage of the information. However there are commands in both R and Matlab that can give the PCs. It's important though to make sure that methods used normalizes the data.

2.4.3   Visualization

The Shepard plots are created by using the distances from the original distance matrix and calculating all the transformed distances from the point configuration. When using Matlab for non-metric MDS it's also possible to plot the the disparities. Spotfire can be used to create the plots of the tree space and to colour the trees. Spotfire is a program for visualization and can create different types of plots and change colours or shapes according to variables in the dataset.[14]

3   Results and discussion

We will here go through all results starting with the cube and the sphere which were briefly talked about in the background section to see how diffrent methods distort the distances. Then we will look at how good the fit is. Finally we will look at how groups of trees behave when changing metric or dimension reduction method.

3.1   The cube and the sphere

The cube and sphere is useful for testing how different methods distort the distances since we know how they should look and what their true distances are. The cube has sides of length one. The black points are the corners of one side of the cube and the red ones are the corners on the other side, the same applies to the shapes. The multidimensional sphere has a red point in the middle of the sphere and all the other points are on the surface of the sphere. The red point in the middle has distance two from all other points. The points in the pairs have distance two from each other and the distances between the pairs are four.

The cube looks almost the same on all plots but with different orientation of the two squares. However metric MDS stands out as it puts the squares on top of each other. Most of the transformed distances are close to the original ones, except in metric MDS. The non-metric plot is especially good at giving a good representation of the distances, although it's hard to see that it's a cube. The sphere is much more distorted than the cube and gave a result that was not expected. The sphere reduced with PCA has the middle point on the left side and all the other points in an vertical line on the right side as shown in the plot below. Each point on the right side is actually both points in the pair being placed in the exact same spot. The sphere reduced with metric MDS has the same strucutre but with the the line going around the middle point instead of being in an vertical line. Both methods get the distance to the middle point close to the true distance while most of the other distances are to short. All the plots of the cube and sphere can be found in appendix 5.1.
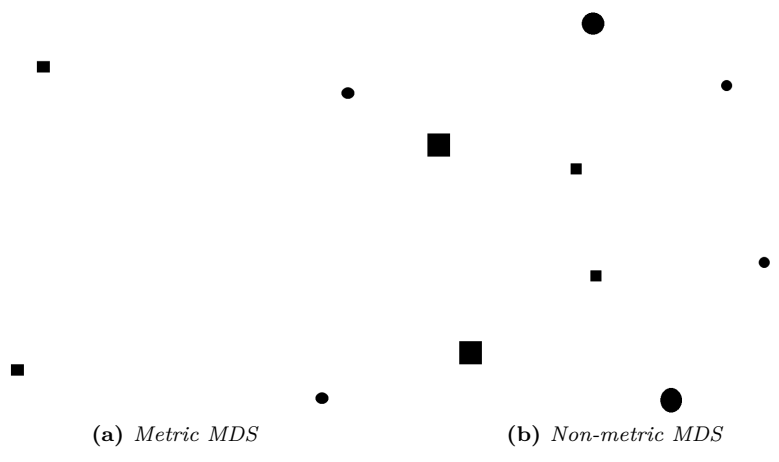
(a) *Metric MDS* **(b)** *Non-metric MDS*

Figure 4: *The cube reduced with metric and non-metric MDS*
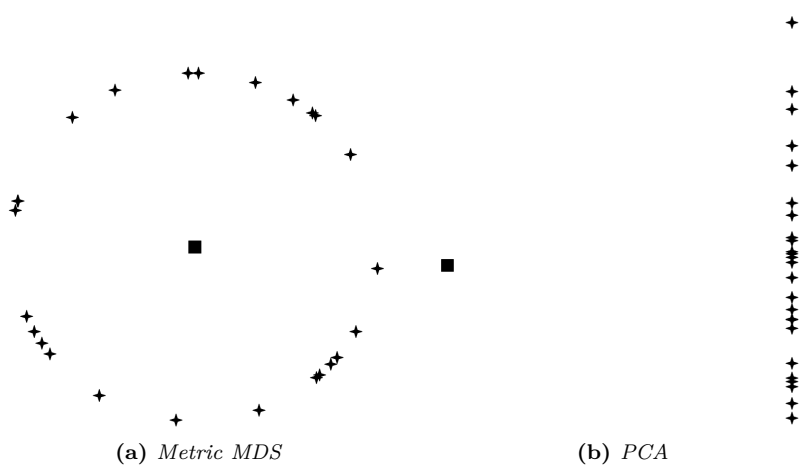

(a) *Metric MDS* **(b)** *PCA*

Figure 5: *The sphere reduced with metric MDS and PCA*
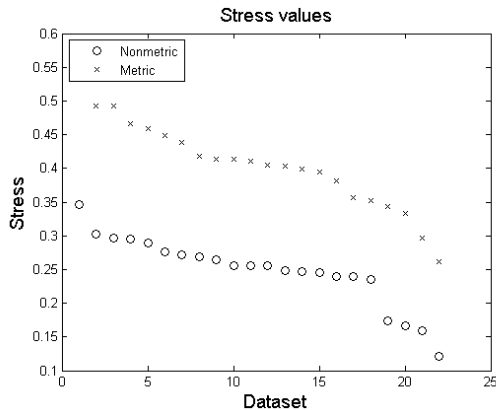
3.2 Goodness of fit

Figure 6: *A plot of stress values for all the datasets, the order is the same as in the table in appendix 5.3*

To examine goodness of fit we will use stress values and Shepard plots, however only metric and non-metric MDS give stress values. Both methods use stress1 which means that their stress values can be compared. As expected non-metric MDS always has lower stress than metric MDS. It has around 0.1-0.15 lower stress, which in some cases can be almost 50% lower. However non-metric MDS has problems converging on all datasets except the cube and Aegotheles, WRF, 300 first trees. Those two datasets along with both Myioborus RF datasets are the only ones with a stress value below 0.17. It's hard to say though what good and bad stress values are since it's much like judging R^2 in multiple linear regression, it depends on the situation. Since the cube reduced with non-metric MDS has good preserved distances and has a stress value of 0.16 it's likely that stress values around 0.16 or lower can be considered as good. There is another downside as well, stress doesn't seem to tell everything about the fit. If we look at some of the plots of the tree space reduced with metric MDS and non-metric MDS we see that they are not that different from each other despite the sometimes big difference in stress. We will now look at some of the Shepard plots to see if they can tell us more about the fit.

When looking at the Shepard plots we quickly see that PCA has the worst performance. They have a cone like structure with a wide spread of possible values for the longer distances. The spread is far too wide to be considered a good fit, especially when compared to metric and non-metric MDS which have a more narrow spread. Classic MDS also tends to look like a cone, but not as much as PCA. Metric and non-metric look more like they should, that is a narrow increasing field and in the case of metric MDS close to the 1:1 line. The plots also look different depending on the metric because RF-distance is discrete while WRF-distance is continuous. On the WRF datasets all methods transforms some of the long distances into to distances that are to short but relatively few distances becomes too long. RF-distance has an more even spread between too short and too long distances, but still a few more become too short than too long.
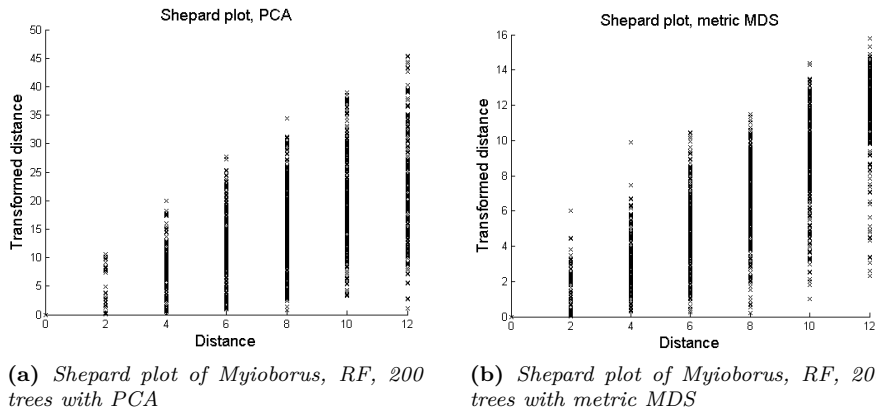
(a) *Shepard plot of Myioborus, RF, 200 trees with PCA*

(b) *Shepard plot of Myioborus, RF, 200 trees with metric MDS*

Figure 7: *Shepard plots*

The Shepard plots for the dataset Aegotheles, WRF, 300 first trees stands out compered to the rest, two of them are shown below. They have big gaps in the Shepard plots because the dataset has two trees that are far away from all other trees which creates the gaps in all the Shepard plots. They all look like they should with a linear increasing line except the one reduced with non-metric MDS. It has a big jump in transformed distances where the gap is. At first glance this might look worse than it really is if disparities are not plotted. However we can use Matlab on this particular dataset to get the disparities since there is no problem with convergence on this dataset. Also on this dataset the difference between the Shepard plots from R and Matlab are neglectable. In the middle near the jump it follows the red line poorly so in that area it has a very bad fit, but everywhere else it follows the line well. This shows how important it is to plot the disparities.
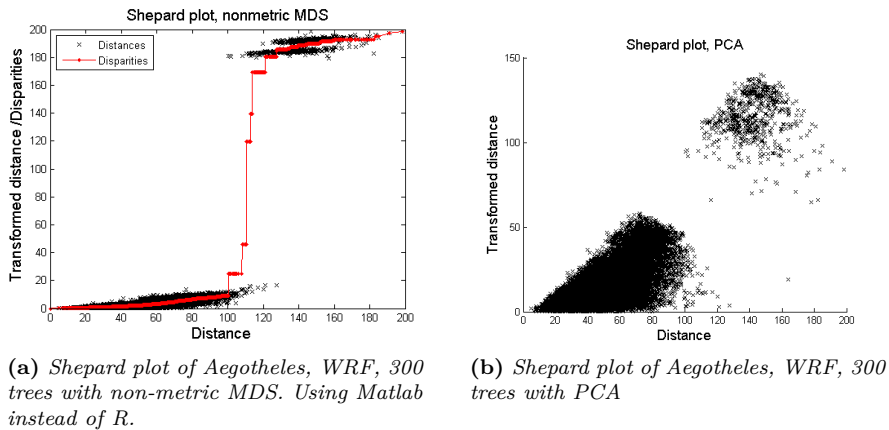
There are more Shepard plots in appendix 5.2.

(a) *Shepard plot of Aegotheles, WRF, 300 trees with non-metric MDS. Using Matlab instead of R.*

(b) *Shepard plot of Aegotheles, WRF, 300 trees with PCA*

Figure 8: *Shepard plots*

3.3 Examining groups in the plots

We will mark different groups of trees in a plot and then mark the exact same trees in another plot to see how the groups behave when changing metric or dimension reduction method. One example is shown below were the plot of Acanthiza has been split into two groups. By looking at several plots like these with different groups mark the conclusion is that generally the groups stay together when changing method, but not when changing metric. However, the groups can be slightly drawn together when changing method. This often often happens when changing from a MDS method to PCA. What usually happens when changing metric is that the groups mix with each other and can sometimes be distributed over large areas of tree space.
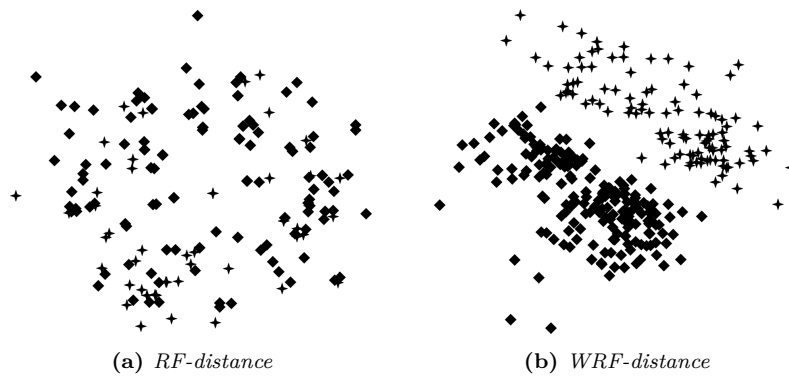

(a) *RF-distance*   **(b)** *WRF-distance*

Figure 9: *Acanthiza, 300 trees reduced with metric MDS and different metrics. The groups in the WRF plot were first marked according to two subgroups, squares and stars, then the same trees got marked in the other plot.*
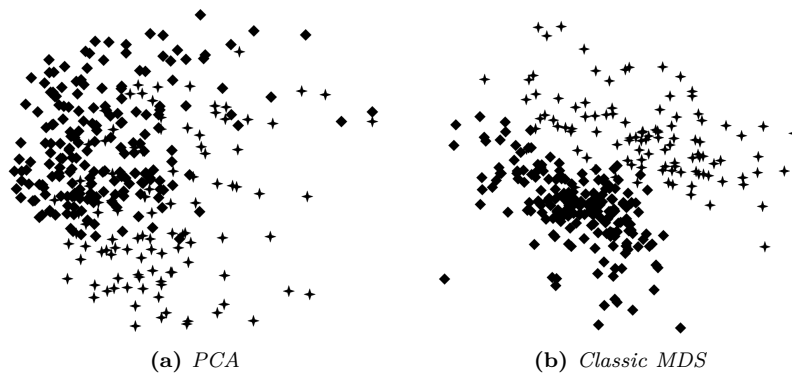

(a) *PCA*   **(b)** *Classic MDS*

Figure 10: *Acanthiza, WRF-distance, 300 trees reduced with PCA and classical MDS. Same groups as in the plot above.*

4 Conclussion and future work

So what is the best metric and dimension reduction method? The colouring of the plots clearly showed that the metric has the most impact on the plots. Dimension reduction method seems to have little impact on how groups of trees behave but it can affect the placing of those groups.

If we compare the dimension reduction methods we can rule out PCA and classical MDS first because of their bad Shepard plots. Even though non-metric MDS has problems converging it gives smaller stress values and slightly better Shepard plots than metric MDS. But the plots of the trees still look a bit similar with the only difference often being that non-metric MDS gives more well defined groups of trees while metric MDS gives a plot with more evenly spread trees. It probably doesn't matter so much if one chooses non-metric or metric MDS but non-metric might be considered slightly better. When testing different methods for dimension reduction Shepard plots proved to be much more useful than the stress values. The stress values add little information beyond what is already present in the Shepard plots. They are also harder to interpret and not all methods give stress values.

There is no clear best metric of the two used. As described in the background ssection the RF-distance has some limitations and the WRF-distance as well. The spread of trees of trees in the plots are diffferent, RF-distance gives plots that has more evenly spread trees than WRF plots. So if we are trying to find groups of trees for which the WRF-distance is preferable. However there is no way to say which one is the best without going more in-depth about their distribution and other properties. Since the metrics give so different plots it's likely a good idea to use different metrics. Two trees that are close to each other in one metric might be far from each other when using another metric.

For future work in this subject the focus should be on the metrics since those affect the plots most. However it might still be interesting to try more non-linear methods. There is for an example two non-linear extensions of PCA, Principal curves and Vector quantization. These examples are described in more detail in [10], together with some other methods.

5   Appendix

5.1   Cube and sphere plots

The cube has sides with length one. The big points are the corners of one side of the cube and the small ones are on the other side. The same applies to the shapes.

The sphere consists of one middle point(square) and 25 pairs(star). The distance to the middle is two and the distances between the pairs are four. The distance between the points within the pairs is 2.
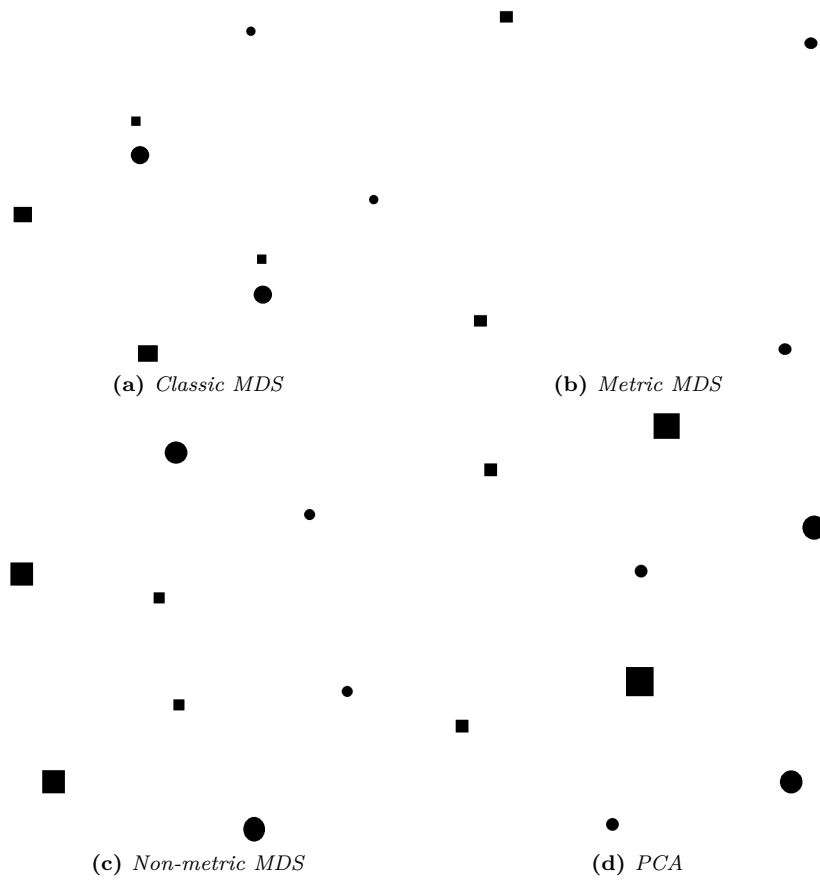
(a) *Classic MDS*

(b) *Metric MDS*

(c) *Non-metric MDS*

(d) *PCA*

Figure 11: *The cube reduced with different methods*

(a) *Classic MDS*   **(b)** *Metric MDS*

(c) *Non-metric MDS*   **(d)** *PCA*

Figure 12: *The sphere reduced with different methods*

5.2 Shepard plots of Myioborus

(a) *Classic MDS*

(b) *Metric MDS*
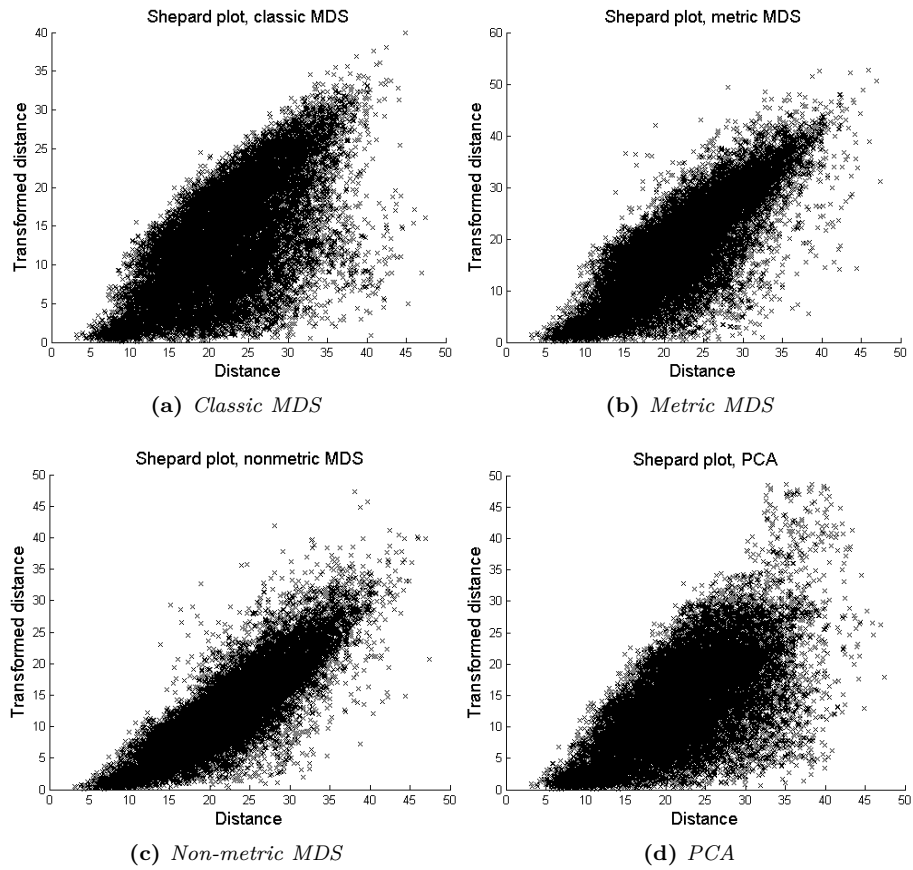
(c) *Non-metric MDS*

(d) *PCA*

Figure 13: *Shepard plots of Myioborus, weighted rf-distance, 200 trees reduced with different methods*

5.3   Stress values

Dataset	Metric	Trees	Non-metric	Metric
Sphere	-	-	0.346	0.553
Ficedula	WRF	First 300	0.302	0.493
Ficedula	WRF	Every tenth tree, 200 trees	0.296	0.492
Ficedula	RF	First 300	0.295	0.466
Ficedula	RF	Every tenth tree, 200 trees	0.290	0.459
Acanthiza	WRF	Every tenth tree, 200 trees	0.268	0.439
Acanthiza	WRF	First 300	0.276	0.449
Acanthiza	RF	Every tenth tree, 200 trees	0.272	0.418
Acanthiza	RF	First 300	0.265	0.405
Tringa	WRF	First 300	0.256	0.411
Tringa	RF	First 300	0.255	0.414
Tringa	WRF	Every tenth tree, 200 trees	0.255	0.413
Tringa	RF	Every tenth tree, 200 trees	0.248	0.403
Aegotheles	RF	First 300	0.247	0.356
Myioborus	WRF	First 300	0.240	0.381
Aegotheles	WRF	Every tenth tree, 200 trees	0.246	0.399
Myioborus	WRF	Every tenth tree, 200 trees	0.239	0.394
Aegotheles	RF	Every tenth tree, 200 trees	0.235	0.343
Myioborus	RF	Every tenth tree, 200 trees	0.174	0.297
Cube	-	-	0.167	0.333
Myioborus	RF	First 300	0.159	0.262
Aegotheles	WRF	First 300	0.121	0.352

Table 2: *A table showing the stress values, sorted in descending order by non-metric MDS.*

Note: The non-metric MDS had problems converging, only the dataset Aegotheles WRF First 300 and the cube converged completly.

References

[1] D. M. Hillis, T. A. Heath, and K. S. John, "Analysis and visualization of tree space," *Systematic Biology*, vol. 54, no. 3, pp. 471–482, 2005.

[2] C. Lakner, P. van der Mark, J. P. Huelsenbeck, B. Larget, and F. Ronquist, "Efficiency of markov chain monte carlo tree proposals in bayesian phylogenetics," *Systematic Biology*, vol. 57, no. 1, pp. 86–103, 2008.

[3] D. F. Robinson and L. R. Foulds, "Comparison of phylogenetic trees," *Mathematical Biosciences*, vol. 53, no. 1-2, pp. 131 – 147, 1981.

[4] J. Archie, W. H.E. Day, W. Maddison, C. Meacham, F. J. Rohlf, D. Swofford, and J. Felsenstein, "The newick tree format," 1986.

[5] M. A. Steel and D. Penny, "Distributions of tree comparison metrics-some new results," *Systematic Biology*, vol. 42, no. 2, pp. pp. 126–141, 1993.

[6] D. Robinson and L. Foulds, "Comparison of weighted labelled trees," in *Combinatorial Mathematics VI* (A. Horadam and W. Wallis, eds.), vol. 748 of *Lecture Notes in Mathematics*, pp. 119–126, Springer Berlin / Heidelberg, 1979. 10.1007/BFb0102690.

[7] D. Bryant, "The splits in the neighborhood of a tree," *Annals of Combinatorics*, vol. 8, pp. 1–11, 2004. 10.1007/s00026-004-0200-z.

[8] H. Samet, *Foundations of multidimensional and metric data structures*. Morgan Kaufmann, 2006.

[9] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001. 10.1023/A:1010933404324.

[10] I. K. Fodor, "A survey of dimension reduction techniques," 2002.

[11] F. Wickelmaier, "An introduction to mds," 2003.

[12] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005.

[13] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "WEKA–experiences with a java open-source project," *Journal of Machine Learning Research*, vol. 11, pp. 2533–2541, 2010.

[14] C. Ahlberg, "Spotfire: an information exploration environment," *SIGMOD Rec.*, vol. 25, pp. 25–29, December 1996.

[15] A. B. Phillimore and T. D. Price, "Density-dependent cladogenesis in birds.," *PLoS Biol*, vol. 6, p. e71, Mar. 2008.

[16] J. Oksanen, F. G. Blanchet, R. Kindt, P. Legendre, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner, *vegan: Community Ecology Package*, 2011. R package version 1.17-10.