MODELLING ASYMMETRIC SIMILARITY WITH PROMINENCE

Mikael Johannesson

Lund University Cognitive Science
Kungshuset, Lundagård
S-222 22 LUND, Sweden
and
Department of Computer Science
University College of Skövde, Box 408
S-541 28 SKÖVDE, Sweden

mikael.johannesson@ida.his.se

Abstract: This paper aims to introduce and discuss a geometrically based model, the relative prominence model, which is inspired by Tversky's (1977) finding that a factor behind asymmetric similarity seems to be "relative prominence". The model proposes that the experienced directed similarity from I to J is proportional to some symmetric similarity measure between I and J, and the quotient between the "prominences" for J and I. Analysis of empirical data from different areas shows that it is possible for a procedure to estimate the parameters of the model quite well. The paper is concluded with a discussion of the differences between the relative prominence model and related models that handle asymmetry in terms of "stimulus bias".

INTRODUCTION

One way to model similarity is to represent objects in a space consisting of a number of dimensions, each corresponding to some quality (see e.g. Palmer, 1978). Objects, or mental objects, then, are represented by coordinates in such a space. A common assumption is that there is an inversely functional correspondence between distance and similarity for points (objects) in the space. This assumption underlies all geometric models of similarity, including the multidimensional scaling (MDS) family of algorithms (see e.g. Kruskal & Wish, 1978; Shepard, 1962).

When the goal is to describe experienced distances or similarities rather than ideal ones problems may occur because the former are not always adequately described by metric distance functions. All metric distance functions must satisfy three basic axioms: *minimality and equal self-similarity* (Eq. (1)), *symmetry* (Eq. (2)), and *triangle inequality* (Eq. (3))

$$d(i, i) = d(j, j) \leq d(i, j) \tag{1}$$

$$d(i, j) = d(j, i) \tag{2}$$

$$d(i, j) + d(j, k) \geq d(i, k) \tag{3}$$

for all objects i, j and k, where $d(i, j)$ is the distance between objects i and j.

There exists empirical evidence of violations against each of the three axioms, but the focus in this paper will be on the violations against the symmetry axiom (Eq. (2)), i.e. on asymmetry. In his well-known 1977 paper "*Features of Similarity*", Tversky presents empirical results from a wide range of domains suggesting that proximity data sometimes reveal significant and systematic asymmetries, contradicting distance based models of similarity.

However, there exist geometric models of similarity which take asymmetry into account. Nosofsky (1991) shows that asymmetric proximity data on many occasions reflect properties of individual items. He re-

Lund University Cognitive Studies – LUCS 55. ISSN 1101–8453.

viewed what he referred to as *the additive similarity and bias model,* originally proposed by Holman (1979)*,* a descriptive model of asymmetric proximity that incorporates similarity and bias. In this model, the proximity of stimulus i to stimulus j, $p(i, j)$, is given by

$$p(i, j) = F[s(i, j) + r(i) + c(j)] \tag{4}$$

where F is an increasing function, $s(i, j)$ is a symmetric similarity function and r (row) and c (column) are bias functions on the individual objects.

Nosofsky point out that a number of well-known models for asymmetric proximity data are closely related to the additive similarity and bias model (e.g. Tversky's (1977) feature matching model and Krumhansl's (1978) distance-density model).

In Nosofsky's terms, asymmetry may often be characterised in terms of stimulus bias, i.e. a characteristic pertaining to an individual object[1]. This is in line with Tversky's (1977) work. Tversky found that the direction of asymmetries appears to be determined by the relative *prominence* of the stimuli. Prominence, Tversky argued, seems to be related to salience, intensity, frequency, familiarity, goodness in form and informational content. The general pattern observed is that less prominent objects often are experienced as being more similar to more prominent objects than the other way around.

Now, if Tversky is correct, it might be the case that a traditional geometric model could be augmented with the notion of relative prominence and so increase its descriptive and predictive power. In the rest of this paper such a model, termed *the relative prominence model* (RPM), will be presented and discussed.

THE RELATIVE PROMINENCE MODEL

In the relative prominence model, the proximity from stimulus i to stimulus j is given by

$$p(i, j) = s(i, j) \cdot (j_p / i_p) \tag{5}$$

where j_p and i_p are prominences (biases) of j and i.

This model is essentially a special case of the additive similarity and bias model in that it could be reformulated as

$$\log[p(i, j)] = \log[s(i, j)] + \log[j_p] - \log[i_p] \tag{6}$$

1. However, see (Nosofsky, 1991) for cases when asymmetry cannot be characterised in terms of stimulus bias.

The relative prominence model was originally proposed to describe, and possibly also predict, the subset of similarity/dissimilarity data collected in direct rating experiments. Such data are less likely to differ in terms of self-similarities compared to confusability data, or data collected by some indirect method. Differences in self-similarity are something that the model clearly cannot handle in its present form, and will thus be beyond the scope of this paper. Empirically correct or not, the model is purposely designed so that the minimality axiom is not violated.

In order to describe experienced directed similarity between objects, the prominence of the objects needs somehow to be quantified. It is not assumed here that *absolute* quantifications of prominence are meaningful, but it is assumed that quantification of *relative* prominences are.

EVALUATION OF THE RELATIVE PROMINENCE MODEL

RPM will here be contrasted with mainly two other special cases of the additive similarity and bias model. The first is the additive similarity and bias model when F in Eq. (4) is the identity function. This model will be referred to as ASM. The second is a multiplicative variant of the additive similarity and bias model (Eq. (7)), referred to as AMM.

$$p(i, j) = r(i) \cdot c(j) \cdot s(i, j) \tag{7}$$

Also, a symmetric model, referred to as the average model (AVG), will be used. The errors for such a model could be seen as a measure of the magnitude of asymmetry, in that the error increases with it. When data are completely symmetric, the error for AVG will be zero. In AVG, the predicted proximity from i to j, $\tilde{p}(i, j)$, is given by

$$\tilde{p}(i, j) = \tilde{p}(j, i) = \frac{p(i, j) + p(j, i)}{2} \tag{8}$$

In some cases (when predictions are supplied in the literature source) RPM, ASM, AMM and AVG will also be contrasted with the similarity choice model (SCM), reviewed in (Nosofsky, 1991). According to SCM, the probability that stimulus i is identified as stimulus j is given by

$$P(R_j | S_i) = \frac{b_j \eta_{ij}}{\sum_k b_k \eta_{ik}} \tag{9}$$

where b_j $(0 \le b_j)$ is the bias associated with item j, and η_{ij} $(0 \le \eta_{ij}, \eta_{ij} = \eta_{ji})$ is the similarity between items i and j.

The parameters of RPM, ASM and AMM will be estimated using iterative procedures. These procedures, summarised in Appendix A, operate directly upon proximities rather than on distances in a spatial representation, meaning that no specific assumptions regarding the relationship similarity - distance need to be made. The iterative procedures try to minimise the global squared relative error (GSRE) given by

$$GSRE = \sum_{\substack{i,\, j \\ i \neq j}} \left(\frac{p(i,\, j) - \tilde{p}(i,\, j)}{p(i,\, j)} \right)^2 \qquad (10)$$

where $\tilde{p}(i,\, j)$ is the predicted proximity from i to j.

The motivation for using the observed value in the denominator of Eq. (10) is that a given difference between an observed and a predicted value should be weighted more for a small observed value compared to a high. The motivation for raising the expression to the power of two in Eq. (10) is that the error will be more distributed over the whole matrix.

In Nosofsky (1991) different models ability to fit matrices of confusion data are evaluated using a likelihood statistic. However, neither row nor column totals for a matrix of predictions given by RPM, ASM or AMM need to agree with the observed proximity matrix, making evaluation using such a statistic inappropriate. Further, similarity using direct rated data are not represented as frequencies. Here the different models will rather be compared using the average relative error $\overline{RE}$ given by

$$\overline{RE} = \frac{GSRE}{N} \qquad (11)$$

where N is the number of off-diagonal cells.

Prominences according to RPM will be presented in some cases for illustrative purposes, but the focus will be upon relative errors for the models compared.

The proximity data sets analysed for evaluating the models will be both direct rated and indirect proximity data sets (e.g. confusions).

EMPIRICAL EVALUATION OF RPM - DIRECT DATA

Direct Rating of Pairs of Countries

This study deliberately bears some resemblance to the study with pairs of countries conducted by Tversky (Tversky, 1977).

Method

Subjects: Subjects were 5 colleagues (computer scientists) and 1 undergraduate cognitive science student, all at University College of Skövde, Sweden. No subject were paid for participation.

Stimuli: Stimuli were 9 countries (i.e. name of countries): France, Russia, Great Britain, Cuba, Paraguay, Germany, USA, Brazil and China.

Procedure: The stimuli were presented pairwise in left to right order on a computer screen using PsyScope (Cohen, MacWhinney, Flatt & Provost, 1993). All ordered pairs of stimuli $\{(i,j) | i \neq j\}$ (i.e. 72 pairs) were presented in a randomised order that were the same for all subjects. For each pair, subjects were asked to rate how similar the left country was to the right on a 20-graded scale. Each subject was tested individually.

Results

The individual similarity matrices were used for estimating the parameters of RPM, ASM and AMM for each subject with the procedures described in Appendix A. The average of errors according to Eq. (11) for each subject were slightly lower for RPM compared to ASM and AMM, which both were lower than AVG (see Table 1.).

RPM	ASM	AMM	AVG
.0429	.0488	.0486	.1541

Table 1. Average Model Errors for Describing Country Proximity Data.

The average estimated relative prominences (Figure 1.) according to RPM appear to be relatively intuitive although these stimuli may be susceptible to contextual effects. For example, it might be that the attention payed by subjects on certain dimensions differs with the pair of countries under consideration.

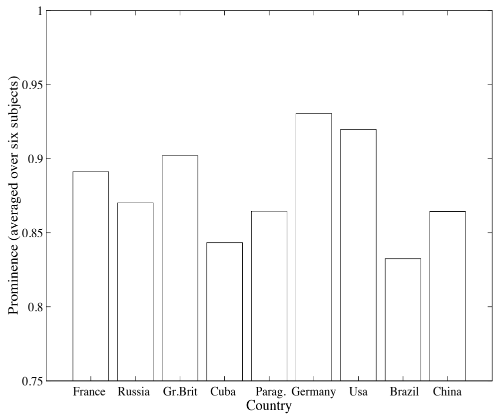

Figure 1. Estimated Relative Prominences for Countries

The "predictive" ability for each of the models was tested by estimating the parameters using subsets of the full off-diagonal matrix, and then compare the resulting "predictions" with the full off-diagonal matrix. The average error curves for about 1200 subsets (evenly distributed from 37 to 71 of 72 cell values[2]) are visualised in Figure 2.

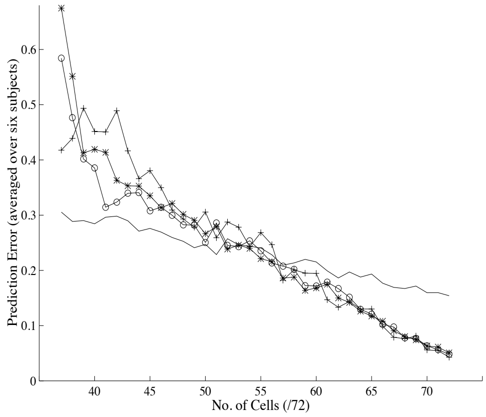

Figure 2. Average Prediction Errors for RPM (+), ASM (*), AMM (o) and AVG (solid) - Countries.

For this particular data set, the asymmetric models had a similar or lower error, compared to the lowest error for AVG, for about 15 or less missing cell values. There appeared to be no major differences between the curves for RPM, ASM and AMM.

Direct Rating of Pairs of Colours

Method

Subjects: Subjects were 5 colleagues (computer and cognitive scientists) and 3 undergraduates (computer and cognitive science students), at University College of Skövde and University of Lund, Sweden. No subject were paid for participation.

Stimuli: Stimuli were 9 quadratic pictures of colours varying in the G and B dimensions of the RGB-system. The colours were: 1: (255,0,255), 2: (255,0,192), 3: (255,0,130), 4: (255,0,64), 5: (255,0,0), 6: (255,33,0), 7: (255,67,0), 8: (255,98,0) and 9: (255,127,0).

Procedure: The stimuli were presented pairwise in left to right order on a computer screen using a html-browser (Netscape). All 72 ordered pairs of stimuli were presented in a block randomised order that were the same for all subjects. Within each block, the left stimuli was held constant whereas the right varied between the remaining 8 stimuli. For each pair, subjects were asked to rate how similar the right colour was to the left on a 20-graded scale. Each subject was tested individually.

2. 37 is the minimum number of cells needed in order to predict asymmetry.

Results

The same procedure as above was used for analysing the data. Again, the error where lower for the asymmetric models compared to the symmetric model (Table 2.). RPM had a slightly lower error than ASM and AMM.

RPM	ASM	AMM	AVG
.0329	.0472	.0453	.2157

Table 2. Average Model Errors for Describing Colour Proximity Data.

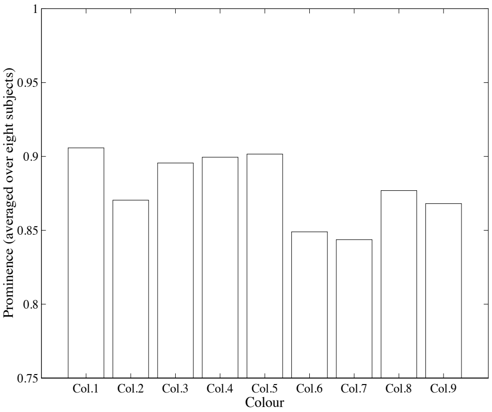

Figure 3. Estimated Relative Prominences for Colours.

The pattern of prominences according to RPM for the colour stimuli (Figure 3.) where not as diversed as for the country stimuli in the previous study. Rather, the colour stimuli appeared to be characterised with basically three different relative magnitudes of prominences.

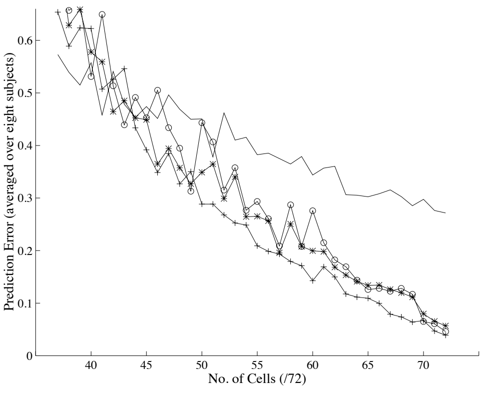

Figure 4. Prediction Errors for RPM (+), ASM (*), AMM (o) and AVG (solid) - Colours.

Regarding the "predictive" ability for the models (Figure 4.), there generally appeared to be a wider gap between the asymmetric models on one hand and the symmetric model on the other, compared to the country stimuli (Figure 2.). This may be caused by differences in contextual effects between the two stimulus sets when rating the similarity of stimuli. Finally, RPM generally had a slightly lower error compared to ASM and AMM.

EMPIRICAL EVALUATION OF RPM - INDIRECT DATA

All indirect proximity data referred to below come from studies reviewed in Nosofsky (1991).

Identification Confusions for Colours

In a study by Nosofsky (1987), identification confusion data was collected as subjects learned to identify 12 Munsell colours varying in brightness and saturation. A version of the MDS model, referred to as the dynamic MDS-choice model (MDS), was used for fitting the data.

The same data was analysed here, using the confusion frequencies as a measure of proximity with the direction from stimuli to response. The errors for describing the original data (Table 3.) were again slightly lower for RPM compared to ASM and AMM. In Table 3. also the error, according to Eq. (11), for the predictions given by using the dynamic MDS-choice model together with SCM (Table 1, cumulated, Nosofsky, 1987) is presented. In this case, SCM got a substantially worse error compared to the other models, including AVG. Note, however, that the error for SCM and the dynamic MDS-choice model should be interpreted with some care since the predictions are derived via the use of a spatial representation (see Nosofsky, 1987), whereas the remaining predictions in Table 3. are derived directly from the identification confusion data.

RPM	ASM	AMM	AVG	SCM
.0609	.0926	.0780	.3097	.6971

Table 3. Average Model Errors for Describing Identification Confusion Data (Nosofsky, 1987).

"Same - Different" Confusions for Morse Codes

Nosofsky (1991) analysed the Morse code "same - different" data reported by Rothkopf (1957), presented in (Kruskal & Wish, 1978), using Eq. (7), i.e. the special case of the additive similarity and bias model here referred to as AMM.

The same set[3] of data have here been reanalysed assuming that the direction of the proximities was in the direction from column to row[4]. The pattern so far (see Table 4.), that the three asymmetric models describe empirical data roughly equally well, became broken by ASM. The morse code data were slightly better described by RPM compared to AMM.

RPM	ASM	AMM	AVG
.0532	.0879	.0580	.1739

Table 4. Average Model Errors for Describing Rothkopf's (1957) Morse Code Data.

Identification Confusions for "Feature sets"

Nosofsky (1991) re-presents two sets of data reported by Garner and Haun (1978) and the corresponding predictions according to SCM (Eq. (9)). The data sets where identification confusions for four symbols under a state-limited condition and a process-limited condition.

In the present paper, as in the above case of Nosofsky's (1987) identification confusions for colours, the direction of the proximities were assumed to be in the direction from stimuli to response. The average errors according to Eq. (11) are presented in Table 5.. It is clear that RPM and AMM once again give about the same average error, but with RPM having a slightly lower error. Regarding SCM, it had a slightly higher error than RPM and AMM for the state-limited condition, whereas it was substantially worse for the process limited condition.

Set	RPM	ASM	AMM	AVG	SCM
SL	.0106	.0365	.0106	.6409	.0143
PL	.0130	.0136	.0135	.2797	.0322

Table 5. Average Model Errors for Garner and Haun's "Feature-Set" in the State-Limited (SL) and Process-Limited (PL) Conditions.

Conditionally Biased Identification Confusions

Nosofsky (1991) presents a subset of data and predictions according to SCM (Eq. (9)) from a payoff-biased experiment reported by Kornbrot (1978). In this experiment, subjects were motivated to underestimate magnitudes of loudness stimuli (see Nosofsky, 1991 pp. 132-134 for details).

Again, the direction of the proximities were here assumed to be in the direction from stimuli to response.

3. As in Nosofsky (1991), cell (5,N), which originally were 0, was here set to 1.

4. Note that an assumption regarding the direction does not affect the magnitude of the error. However, the relative magnitude of the prominences will be inverted if the assumption is wrong.

The errors according to Eq. (11) are presented in Table 6.

RPM	ASM	AMM	AVG	SCM
.1696	.6169	.3250	12.2449	.1907

Table 6. Average Model Errors for a Subset of Kornbrot's (1978) Data.

In this particular case, the magnitudes of the errors were high for all models, saying that no one of them described the data set very well. However, although not low, RPM and SCM had substantially lower errors compared to the other, and RPM, in turn, had a slightly lower error than SCM.

DISCUSSION

Evaluation using seven asymmetric data sets, both direct rated data and confusion data, shows that the relative prominence model tend to describe proximity data slightly better than ASM and AMM. However, it is difficult to say how general this result is since it is based upon the use of estimation procedures that lead to suboptimal rather than optimal results. A possibly more important property of the relative prominence model is that it incorporates one bias parameter (prominence) per object and therefore is simpler than ASM and AMM that incorporates two (row- and column - bias).

The relative prominence model as presented here does not predict any differences regarding self similarity between stimuli. This could mean that the model may be more appropriate for stimulus sets where the differences are not too small, compared to stimulus sets with very small pairwise differences (see also (Melara, 1992) for a discussion of differences between confusability and similarity), and also that it is more appropriate for studies of higher level cognition than pure perception.

A gain with using a simple model like RPM is that the vague concept of prominence could be studied and possibly be more well understood. This in turn could have implications for e.g. theories of concept formation. Studies of relative prominence could be performed by searching for relations between relative prominences and properties of the proximity data studied. For example, above, the relative prominences for direct rated colours were presented (Figure 3.). If the relative prominences for the colours would mainly have been determined by their focality, something that sounds quite intuitive, the prominence would have been the largest for Colour 5 and decrease with "higher" and "lower" numbers in the figure. Since this was not the case, it means that other properties of the stimuli have to be considered.

Finally, it is here suggested that studies of prominence should be performed also on the individual level (i.e. on individual data), not only for summary data as in this paper.

ACKNOWLEDGEMENTS

Research for this paper has been supported by the Department of Computer Science, University College of Skövde. I would like to thank Anders Eklund and Magnus Johansson at University College of Skövde for their extensive programing help, and Magnus Haake for the excellent drawing on the cover. I would also like to thank my supervisor Professor Peter Gärdenfors, Lund University Cognitive Science, and Dr. Lars Niklasson, University College of Skövde, for valuable comments on an earlier draft.

REFERENCES

Cohen, J. D. MacWhinney, B. Flatt, M. & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers*, 25(2), 257–271.

Garner, W. R. & Haun, F. (1978). Letter identification as a function of perceptual limitation and type of attribute. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 199-209.

Holman, E. W. (1979). Monotonic models for asymmetric proximities. *Journal of Mathematical Psychology*, 20, 1-15.

Kornbrot, D. E. (1978). Theoretical and empirical comparison of Luce's choice model and logistic Thurstone model of categorical judgment. *Perception & Psychophysics*, 24, 193-208.

Kruskal, J. B. & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage Publications.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 85, 445-463.

Melara, R. D. (1992). The concept of perceptual similarity: from psychophysics to cognitive psychology. In D. Algom (Ed.), *Psychophysical Approaches to Cognition* (pp. 303-388). Elsevier Science Publishers B.V.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 87-108.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23, 91-140.

Palmer, S. E. (1978). Fundamenthal aspects of cognitive representation. In E. Rosch and B. B. Lloyds

(Eds.) *Cognition and Categorization* (pp. 259-303). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, 53, 93-101.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance. I and II. *Psychometrika*, 27, 125–140, 219–246.

Tversky, A. (1977). Features of Similarity. *Psychological Review*, 84, 327-352.

APPENDIX A

The RPM-, ASM, and AMM - procedures

1. Initiate $s(i, j)$ for all i, j by setting

RPM: $(i, j) = \sqrt{p(i, j) \cdot p(j, i)}$

ASM and AMM: $s(i, j) = \frac{p(i, j) + p(j, i)}{2}$

2. Initiate prominences/biases

RPM: initiate i_p to a random number for all i

ASM and AMM: initiate $r(i)$, $c(i)$ to a random number for all i (note: $r(i)$, $c(i)$ could be negative for ASM)

3. Calculate "proposed" prominences/biases:

RPM: for each i, calculate the averaged "proposed" prominence from all j, *PP(ij),* given j's current prominence

$$PP(ij) = \frac{\frac{s(i, j) \cdot j_p}{p(i, j)} + \frac{p(j, i) \cdot j_p}{s(i, j)}}{2}$$

ASM and AMM: for each i, calculate the "proposed" changes of $r(i)$, $c(i)$ given all other objects current values of r and c

ASM

(row): $PCr(i) = \sum_{j=1}^{N} \frac{p(i, j) - s(i, j) - r(i) - c(j)}{2}$

ASM
(col): $PCc(i) = \sum_{j=1}^{N} \frac{p(j, i) - s(i, j) - r(j) - c(i)}{2}$

AMM
(row): $PCr(i) = \sum_{j=1}^{N} \frac{p(i, j) - (s(i, j) \cdot r(j) \cdot c(i))}{2}$

AMM
(col): $PCc(i) = \sum_{j=1}^{N} \frac{p(i, j) - (s(i, j) \cdot r(j) \cdot c(i))}{2}$

where *PCr(i)* and *PCc(i)* are the "proposed" changes for *r(i)* and *c(i)* respectively.

4. RPM: for each i, adjust the prominence proportional to the *average* proposition from all j.

ASM and AMM: for each i, adjust the prominence proportional to *PCr(i)* and *PCc(i)* respectively.

5. Calculate the error according Eq. (10) and save the fitted parameters if the error is the lowest so far.

6. Repeat steps 3 - 5 N times.

7. Given the best parameter values so far, adjust $s(i, j)$ for all i, j so that the error according to Eq. (10) is minimised.