Comparing internal and ideal point preference mapping solutions

Wakeling, Hasted and Macfie

Overview

- Motivation
- Simulation conditions
- Results
- Application to real data
- Conclusion

Motivation

- Mackay (2001) and Rousseau (2009) have proposed that internal preference mapping using PCA will give wrong configurations when a high proportion of ideal point individuals are present in the data

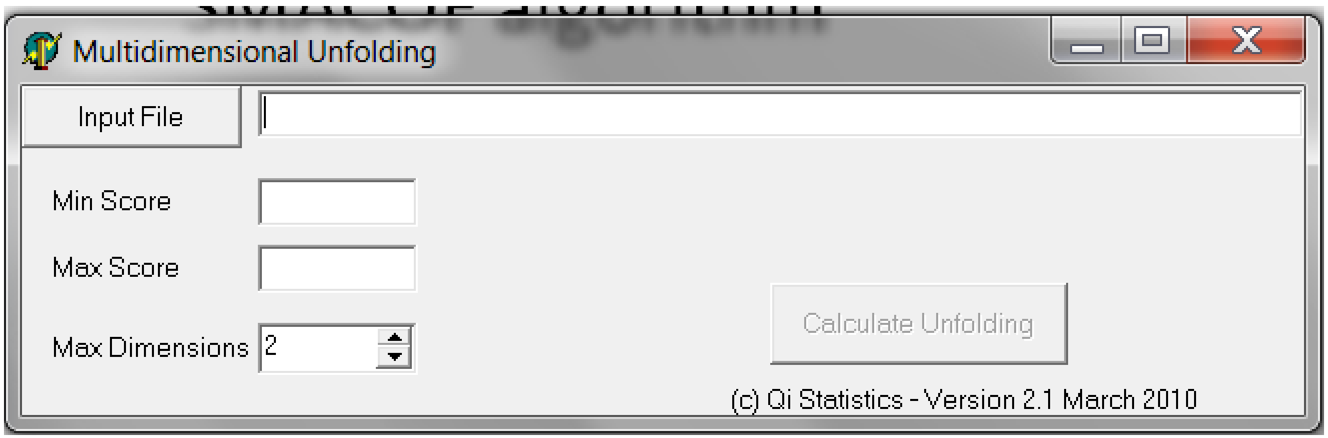
- Our experience and others (Meullenet 2009) has been that the PCA solution gives stable configurations that correlate well with sensory data

- Smacof algorithm in R gives us a chance to compare solutions to see how well this works out in practice.

- Simulation conditions varying the proportion of ideal point individuals should show when the PCA solution starts to break down

UNFOLDER

- A QI Statistics program to run the SMACOF algorithm



- Free download from QI web site contact Ian
- Outputs coordinates and Stress values

10 point fixed solution to form basis of simulated data

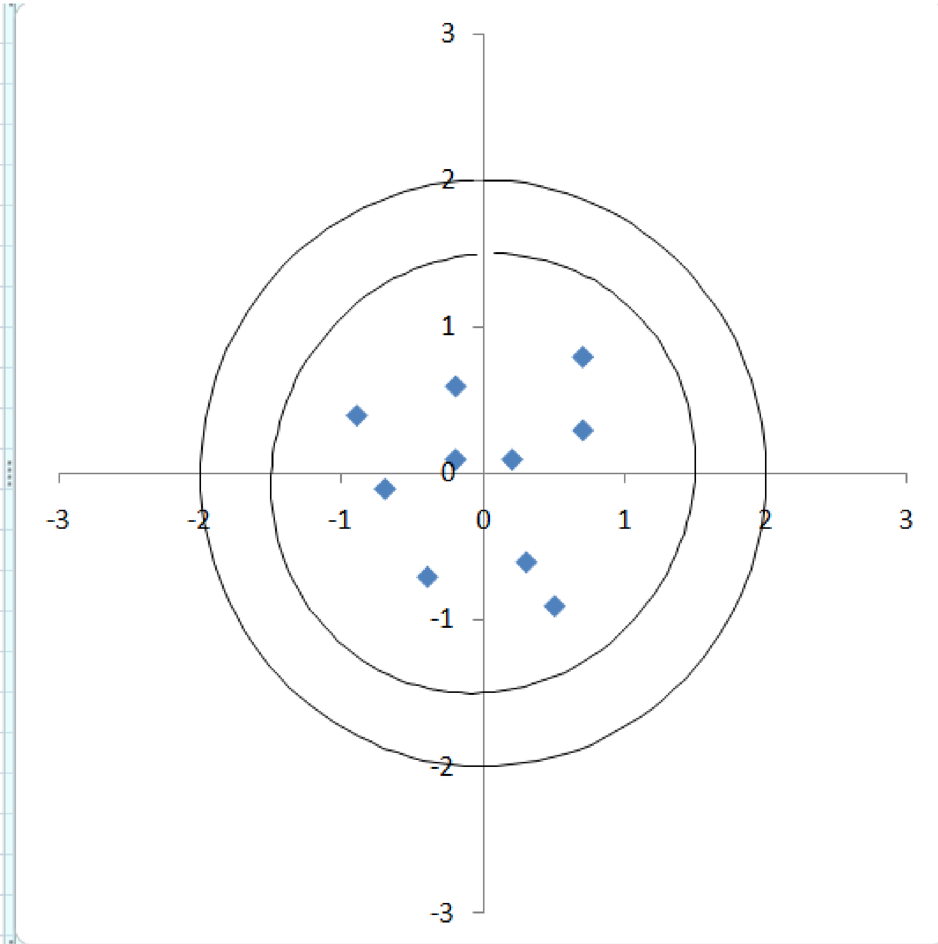
Fixed Config for Simulations

0.7	0.8
0.7	0.3
0.3	-0.6
-0.4	-0.7
-0.7	-0.1
-0.9	0.4
-0.2	0.6
-0.2	0.1
0.2	0.1
0.5	-0.9
0	0

Correlation Coeff

0.003424738

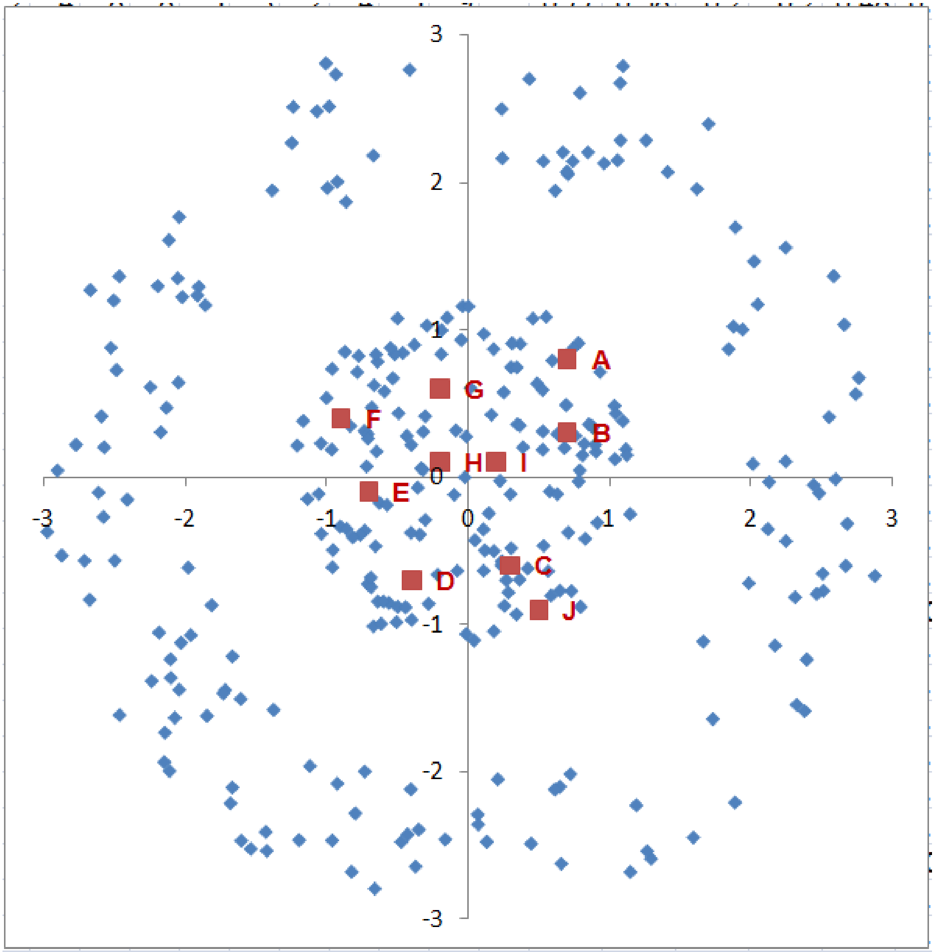
0.000012



Simulation Procedure

- The fixed orthogonal configuration is chosen.
- If p is the proportion of consumer created ideal. Then 300*(1-p) random vector points are chosen, products are projected onto these vectors and the projections are scaled and rounded to become a score on a nine point scale.
- 300p ideal points are placed uniformly randomly in and around the product configuration - the distance from all the ideal points to all the product points are rescaled and rounded to a nine point scale.
- Optional normal noise rounded to the nearest integer is added.

Example of 50% ideal 50% vector data set

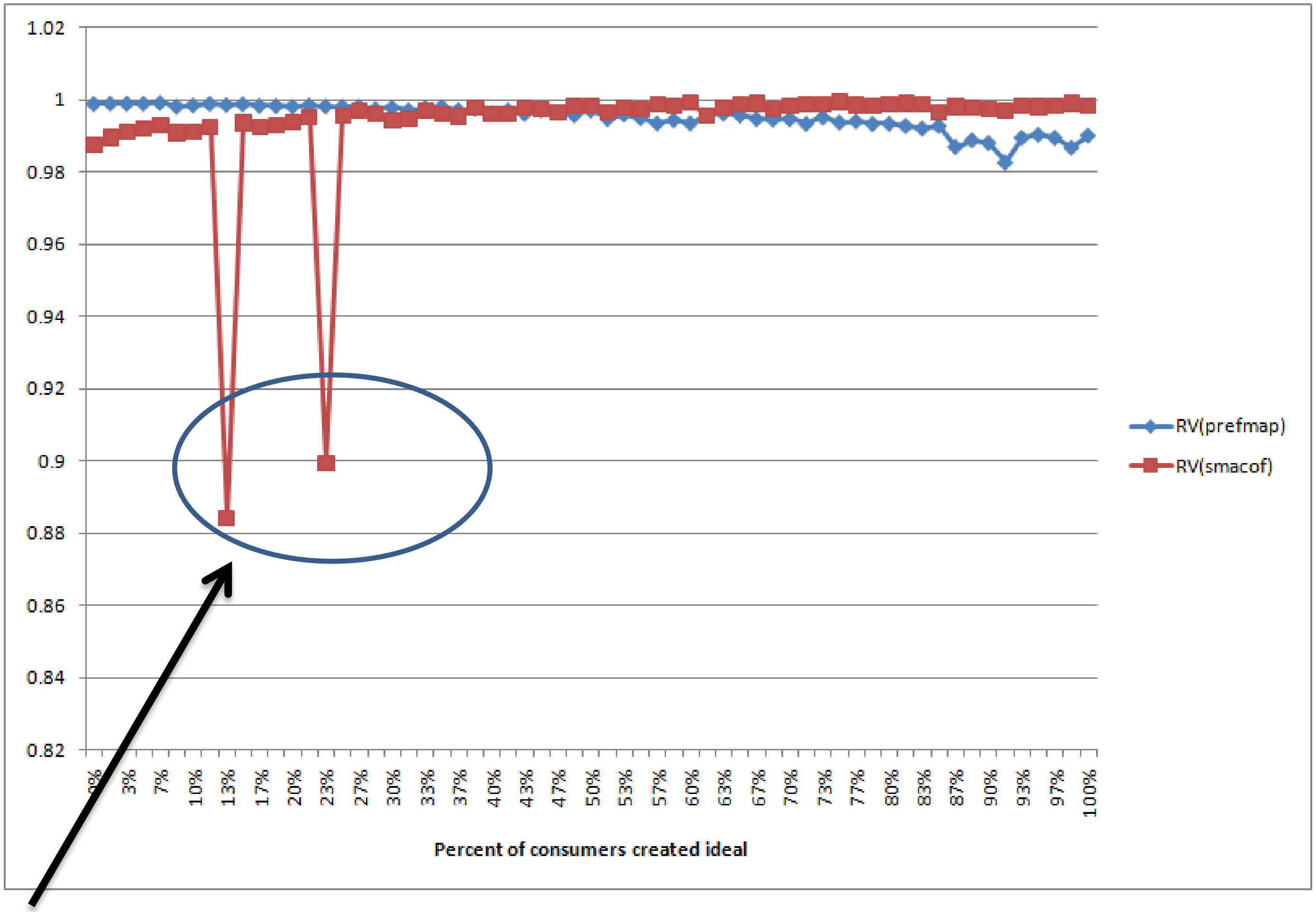


Only the direction of the vectors has a significance in the simulation, since the samples are projected on to the vectors to obtain simulated hedonic scores for the consumers.

Comparing PCA and UNFOLDER Solutions

- The two product configurations are adjusted to the same sum of squares as the original configuration and rigidly rotated towards the original configuration.
- For unfolding the distances from each consumer point to the products are calculated and then correlated with the original scores.
- The average squared correlations across all consumers becomes the variance explained measure for the unfolding solutions.

Comparison of RV Coefficients between initial configuration and recovered configurations using PCA and SMACOF (no noise added)



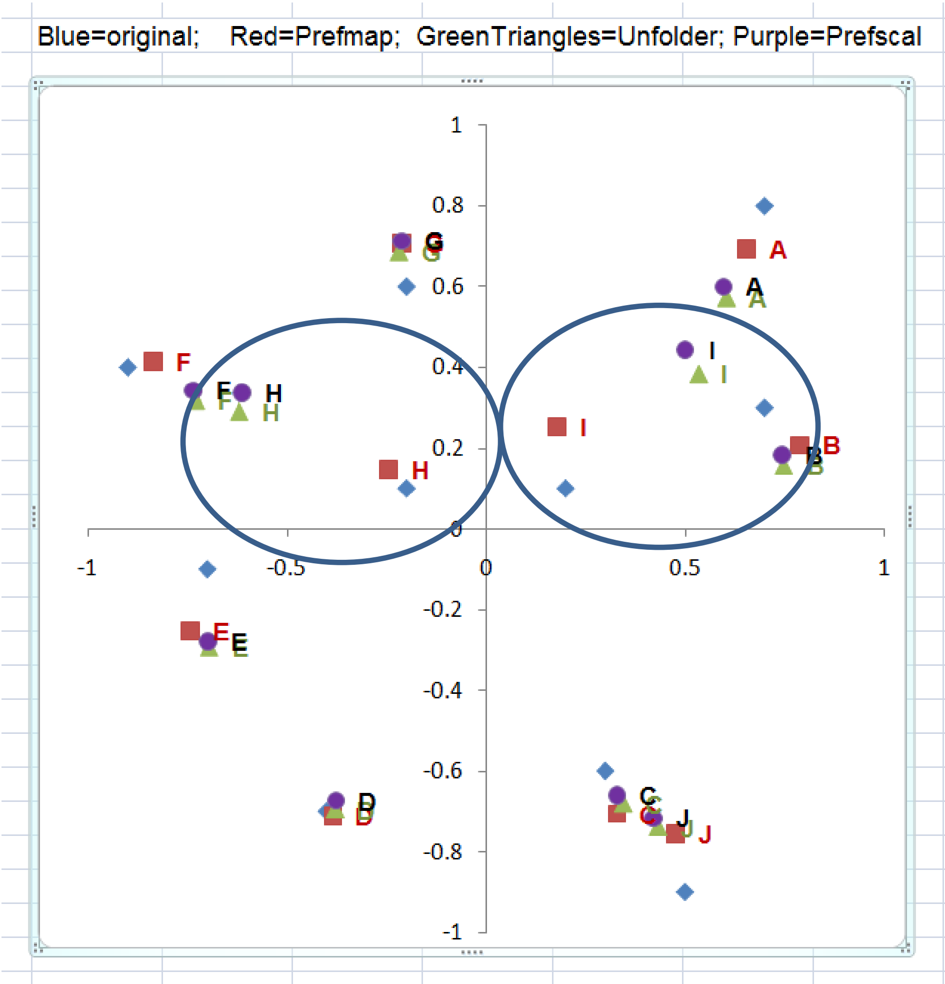
These are two odd solutions with high stress

Comparing sample recovery for the 50/50 simulated no noise added

Note that we have also applied PREFSCAL (the SPSS iterative majorization algorithm)

The solution matches up well with SMACOF results

Note that only PCA satisfactorily places samples H and I close to the solution

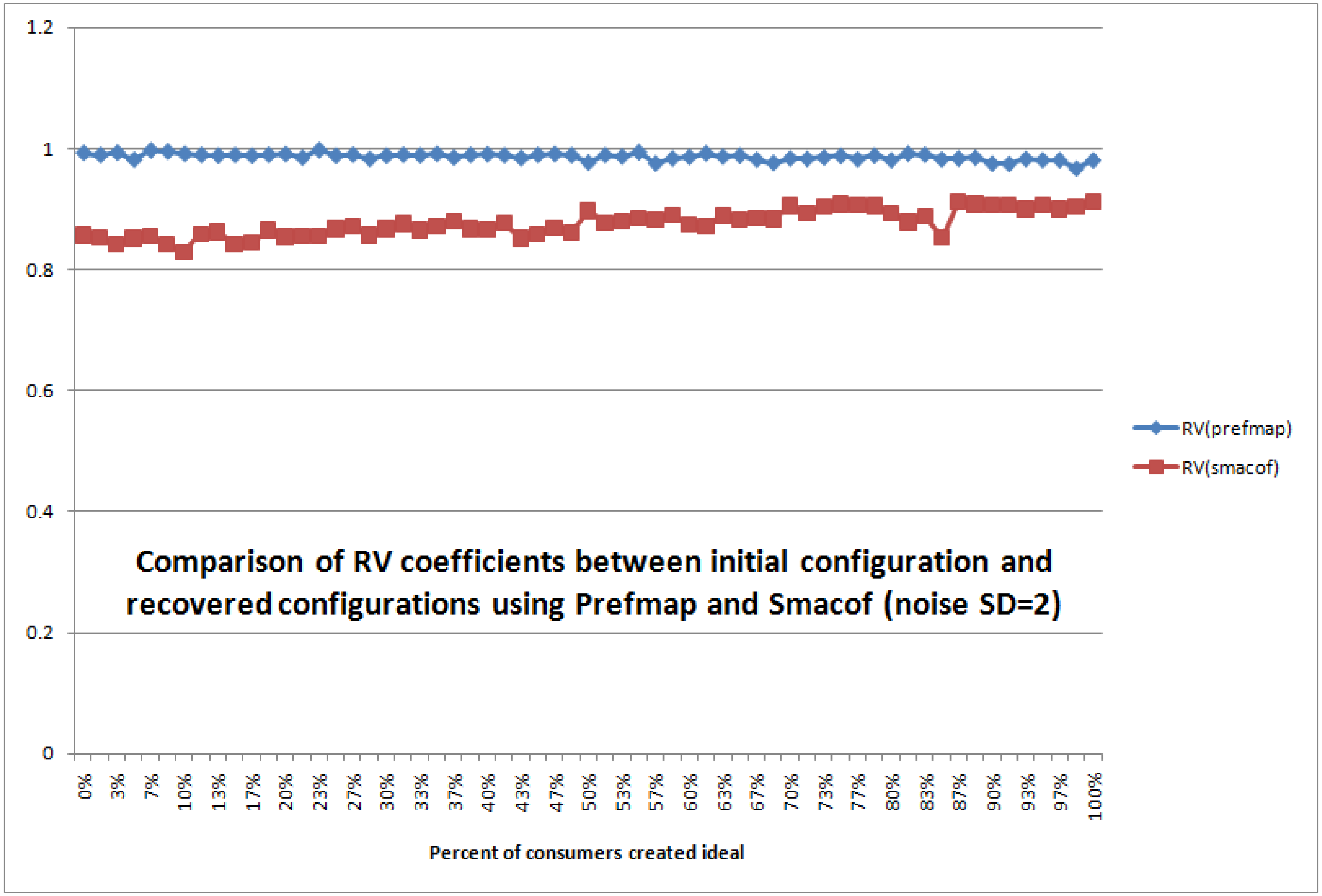


S

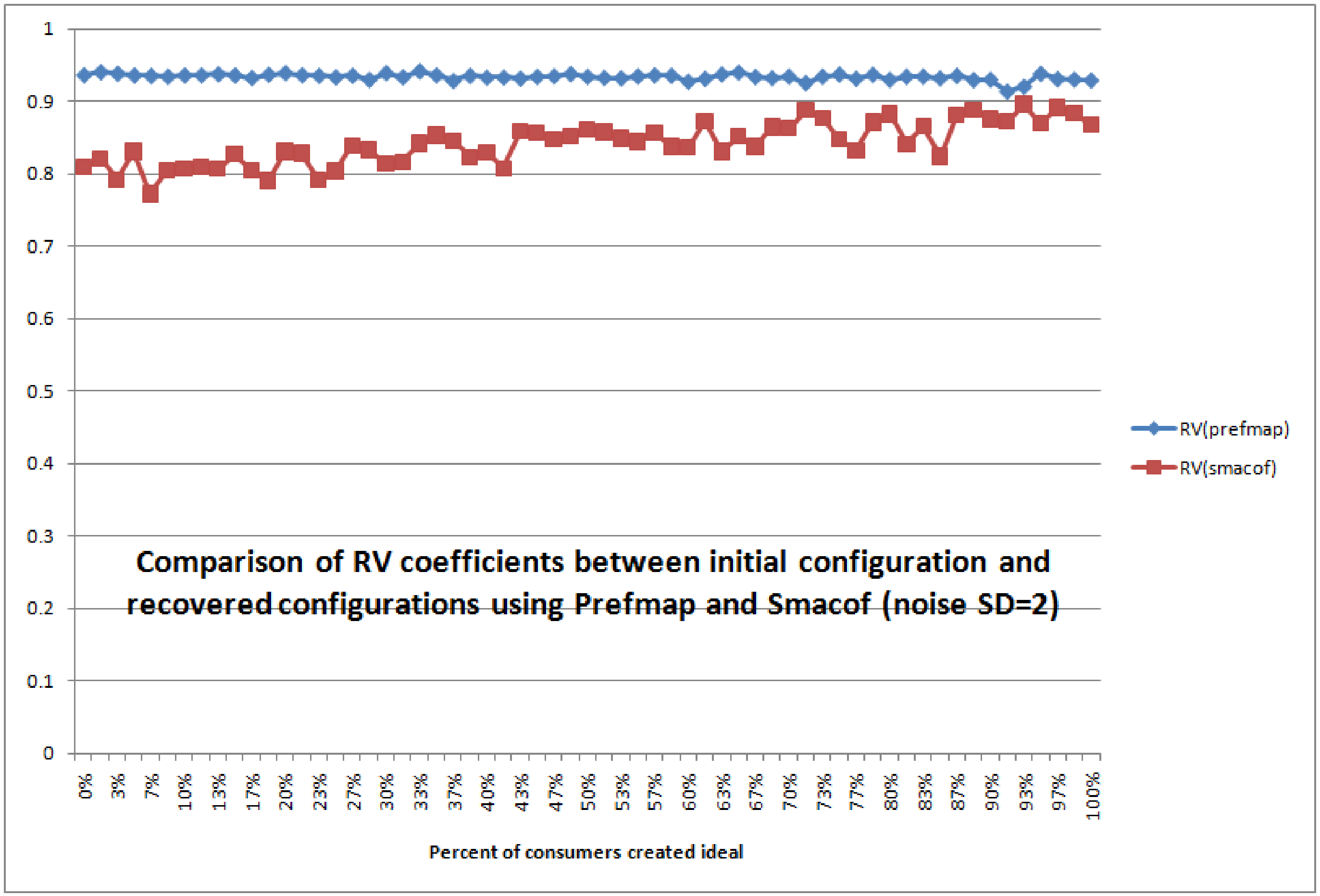
Adding noise to the configuration

- Hough et al indicated that the basic noise of a respondent scoring on a liking scale in a CLT context to be 25% of the scale length.
- Repeated the simulation adding this noise to the configuration

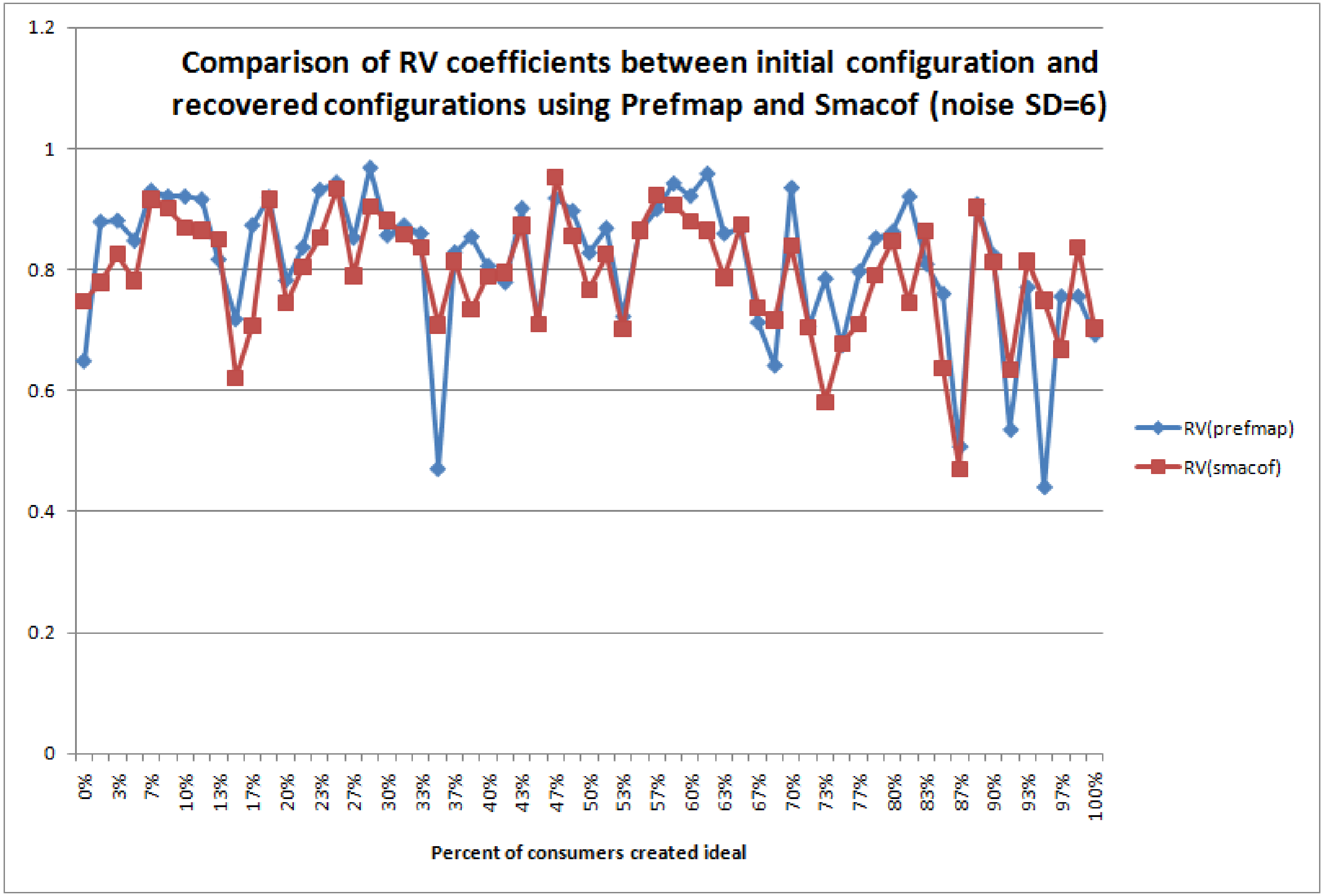
Comparison of RV Coefficients between initial configuration and recovered configurations using PCA and SMACOF UNFOLDER (Noise 25% of scale length) – Configuration 1



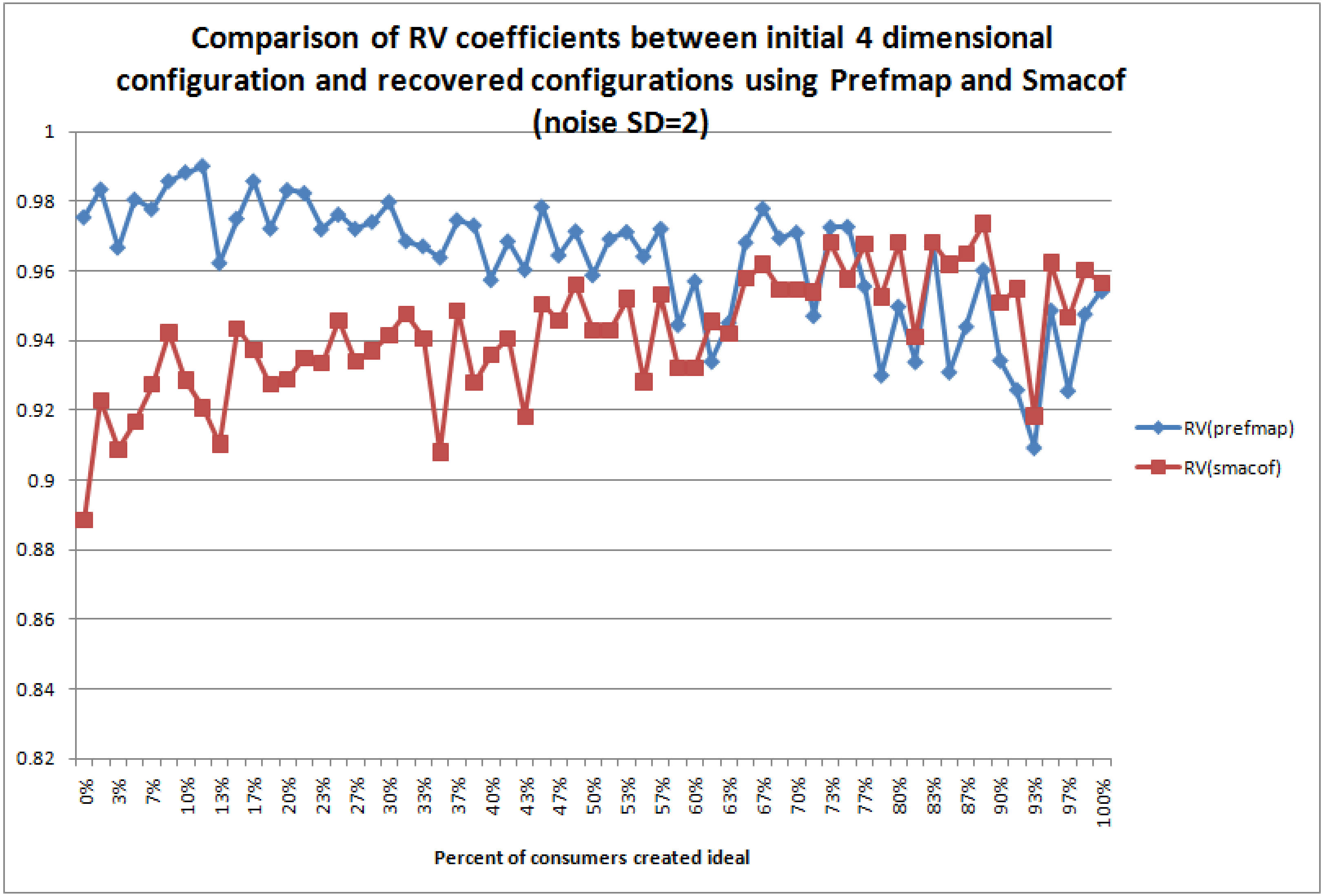
Comparison of RV Coefficients between initial configuration and recovered configurations using PCA and SMACOF UNFOLDER (Noise 25% of scale length) – Configuration 2



Results with noise increased to 75% of scale length



Using a 4 dimensional solution with 25% noise



Summary Simulation

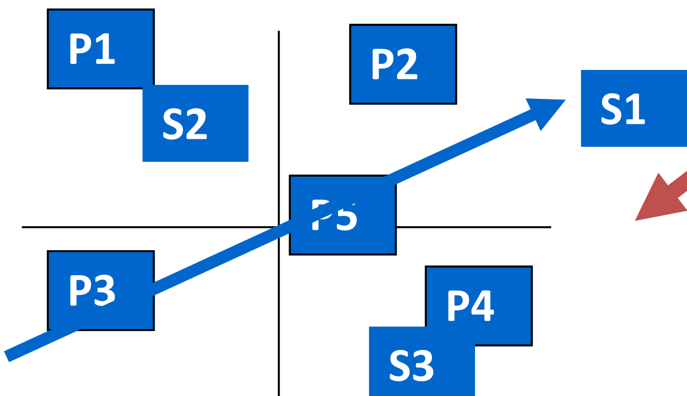
- With 25% noise added, PCA appears to give a numerically close recovery of the original configuration
- The SMACOF algorithm appears to have a problem placing points in the centre of the plot.

Real data sets

Preference Mapping (Cross Over)

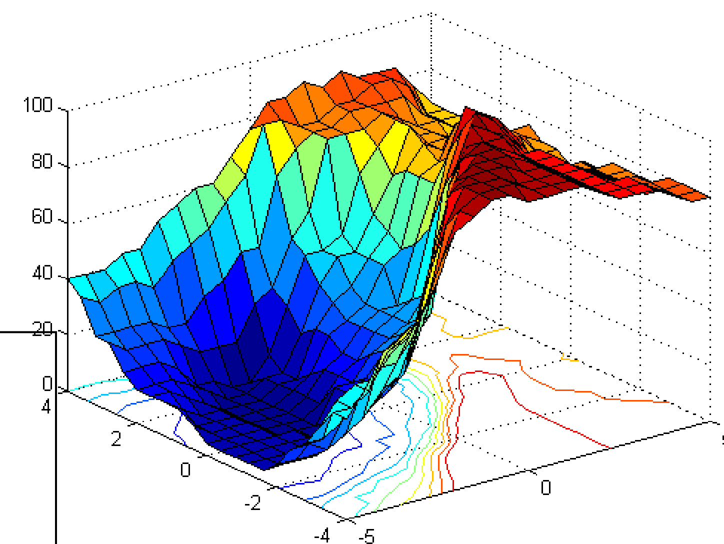
Consumer liking data

PCA
MDS
Unfolding



Consumer liking data

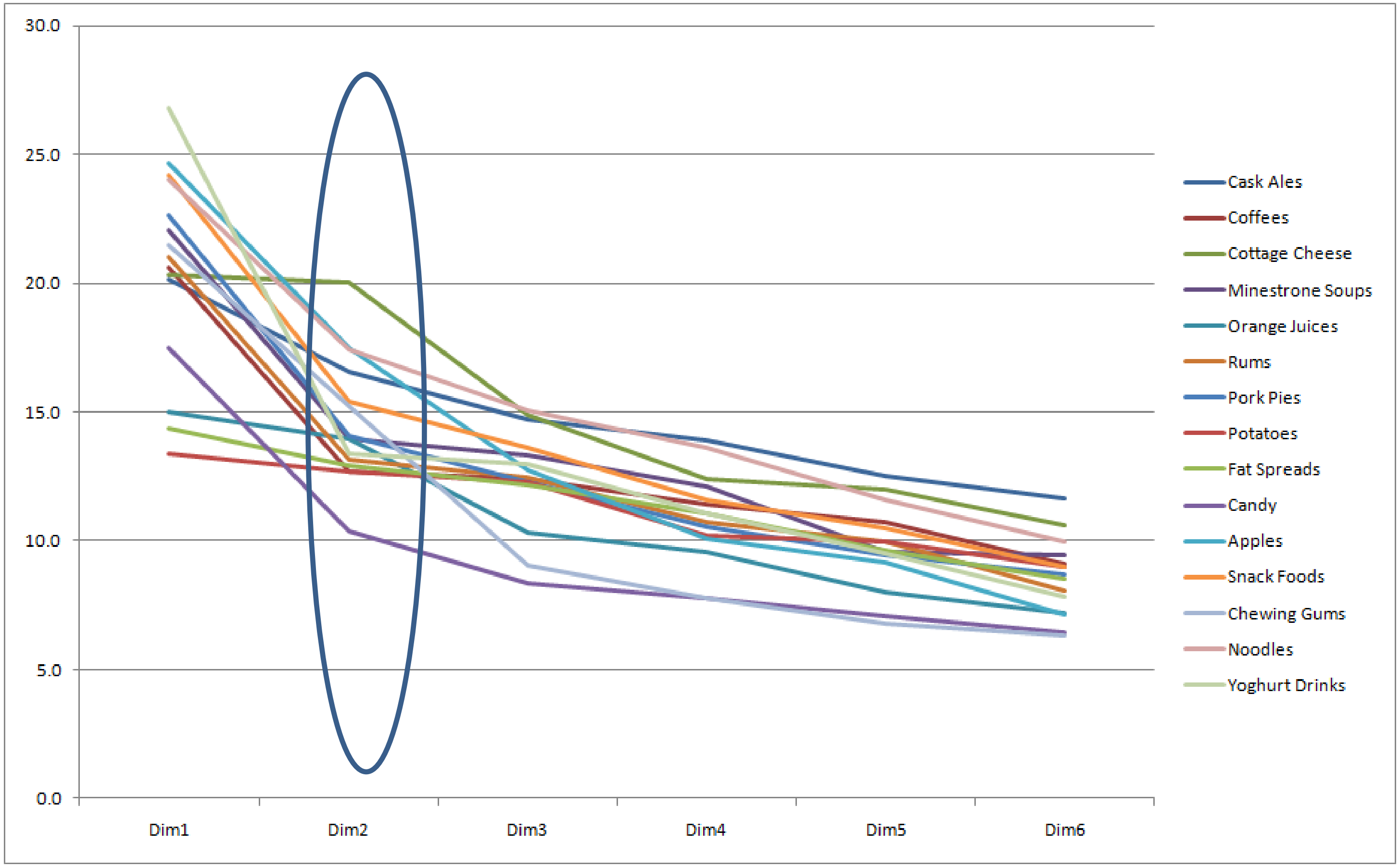
Individual vector and ideal point modelling



So the map is obtained from the consumers using UNFOLDER and we use conventional regression techniques to fit consumers to the map.

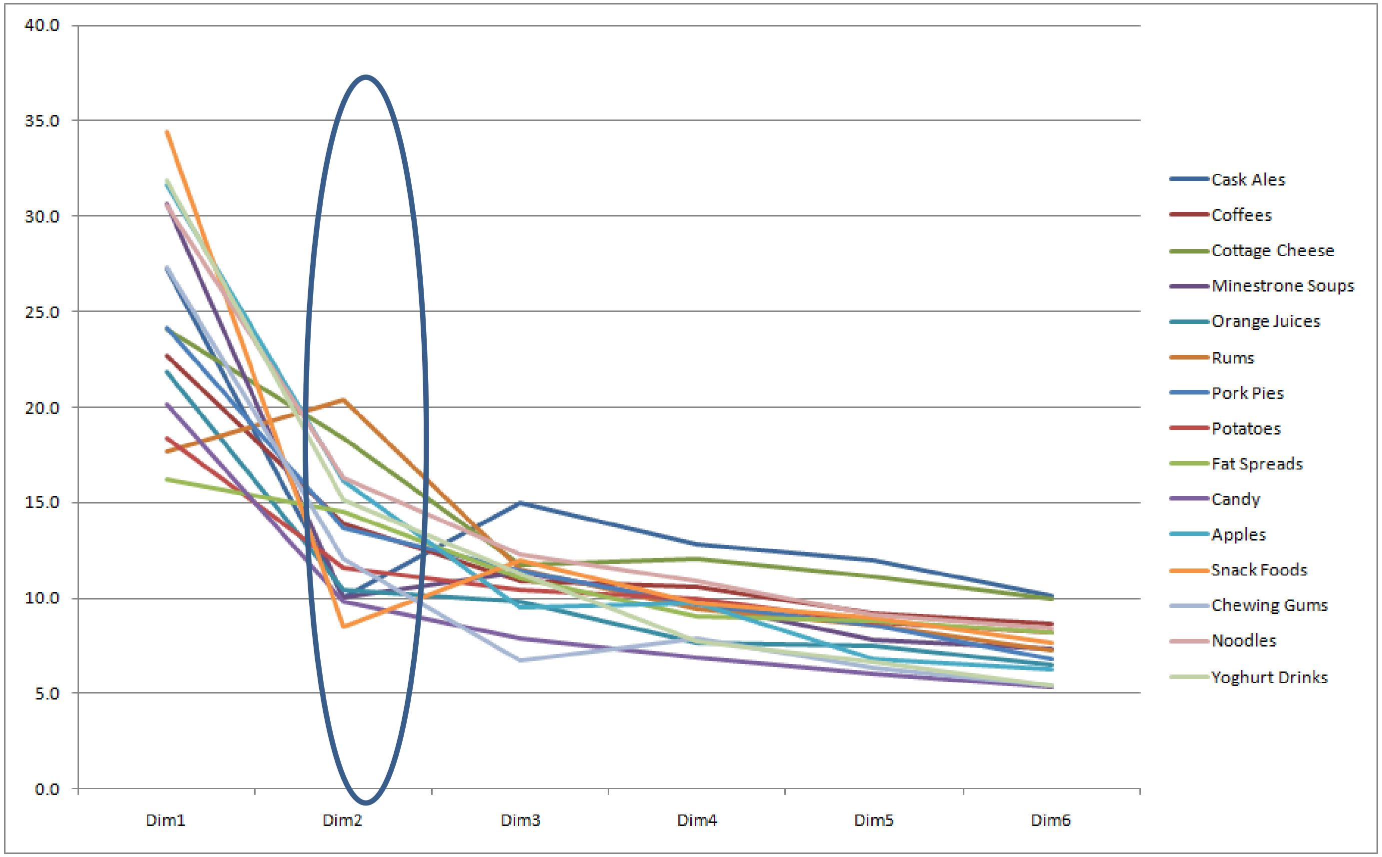
Enables a test of how many individuals are ideal point

Scree Plot of % variance explained by applying PCA to 15 real data sets



Most sets show an elbow at a 2 dimensional fit but a few indicate a 3 dimensional fit

Scree Plot of % variance explained by applying SMACOF to 15 real data sets

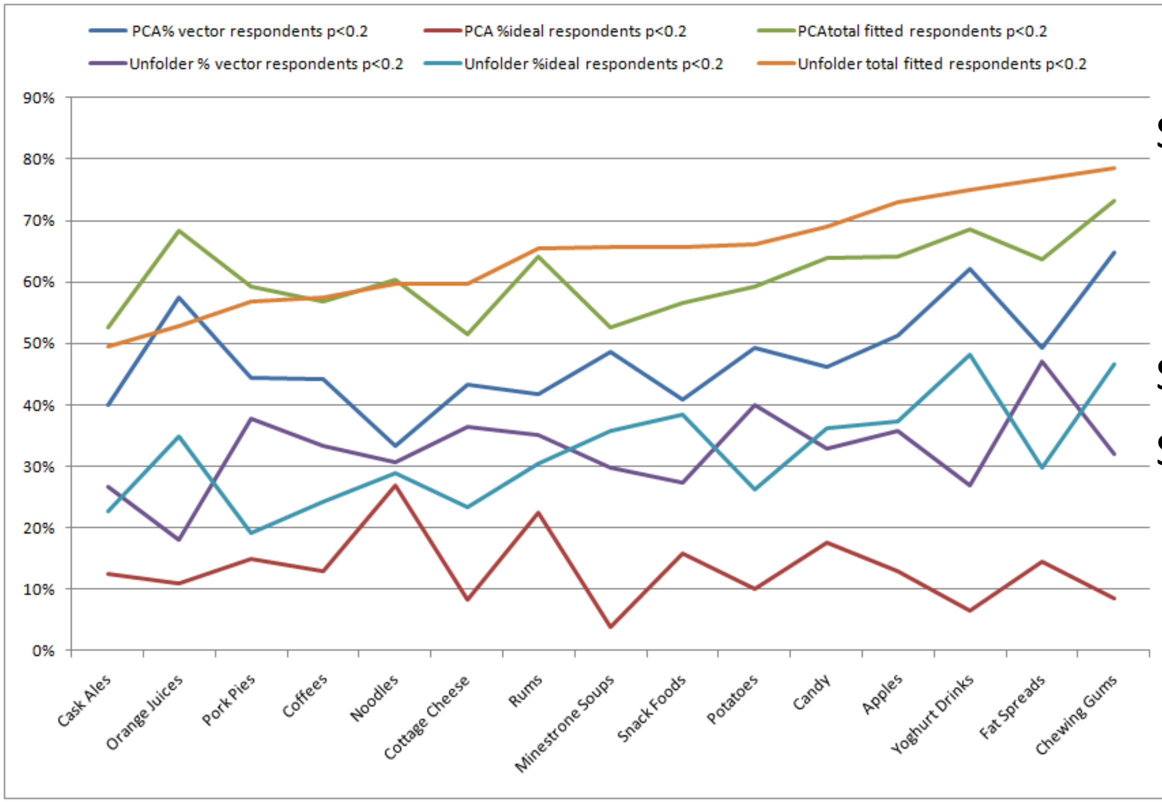


Most sets show an elbow at a 2 dimensional fit but a few indicate a 3 dimensional fit

Applying External modelling to PCA configuration of 15 real data sets

Data Set	Consumers	Products	Scale	RSD	Dimensionality	% vector respondents $p<0.2$	%ideal respondents $p<0.2$	total fitted respondents $p<0.2$
Cask Ales	255	8	1--10	2.16	2	40%	13%	53%
Coffees	304	10	1--9	2.07	2	44%	13%	57%
Cottage Cheese	160	8	1--9	1.68	3	40%	11%	52%
Minestrone Soups	154	10	1--9	1.82	2	49%	4%	53%
Orange Juices	155	14	1--9	1.95	3	55%	9%	64%
Rums	151	11	1--9	2.13	2	42%	23%	64%
Pork Pies	167	10	1--9	1.81	2	44%	15%	59%
Potatoes	130	12	1--10	1.94	3	39%	10%	49%
Fat Spreads	138	15	1--9	1.88	3	46%	8%	54%
Candy	120	18	1--9	1.40	2	46%	18%	64%
Apples	604	10	0--10	2.74	2	51%	13%	64%
Snack Foods	242	9	0--100	19.87	2	41%	16%	57%
Chewing Gums	309	14	1--7	1.04	2	20%	8%	28%
Noodles	158	8	1--9	1.74	2	33%	27%	60%
Yoghurt Drinks	108	10	1--10	2.00	2	62%	6%	69%

Comparing PCA and SMACOF fitting by respondents who are significant at 20%



SMACOF total %fitted

PCA total %fitted

PCA total % vector fitted

SMACOF total % vector fitted

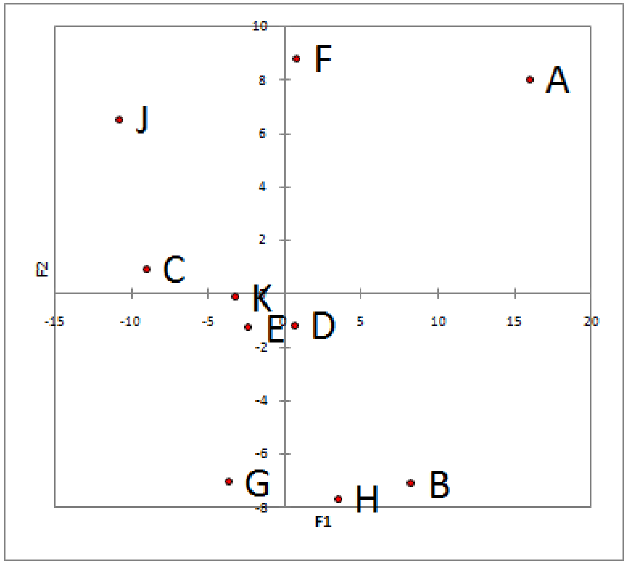
SMACOF total % ideal fitted

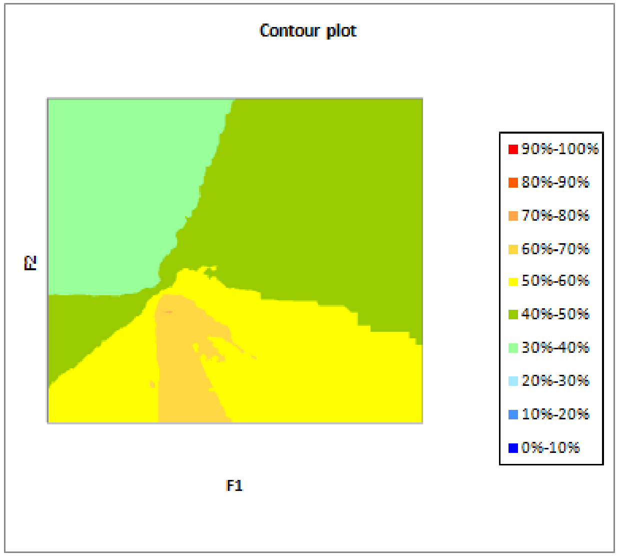
PCA total % ideal fitted

PCA configuration generally fits 50% or more as vector models
SMACOF configuration generally fits slightly more respondents and fits roughly equal numbers as vector or ideal.

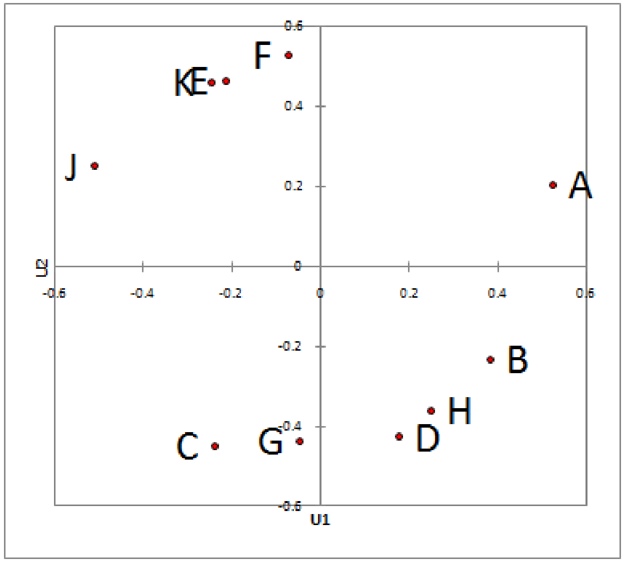
Coffees

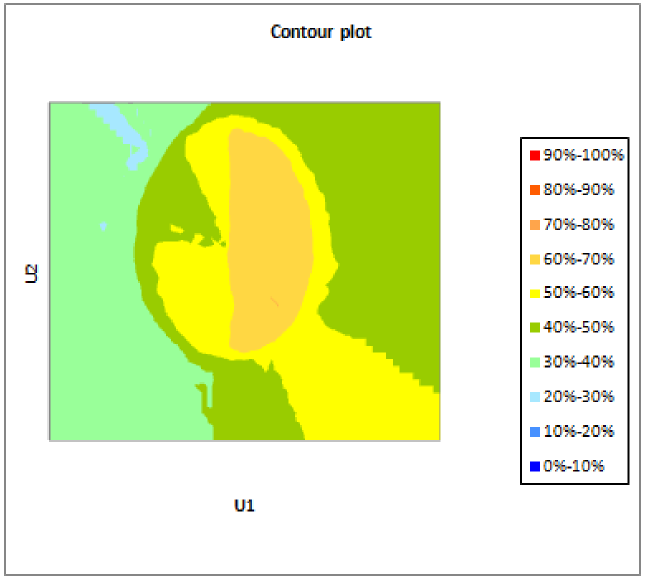
PCA





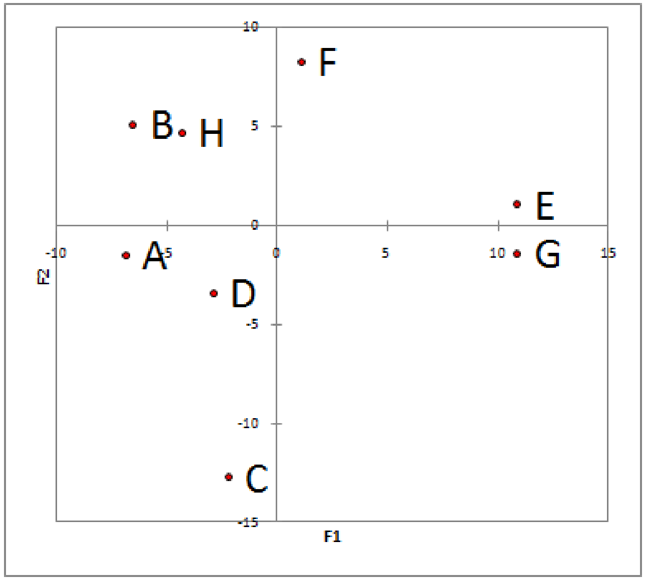
SMACOF

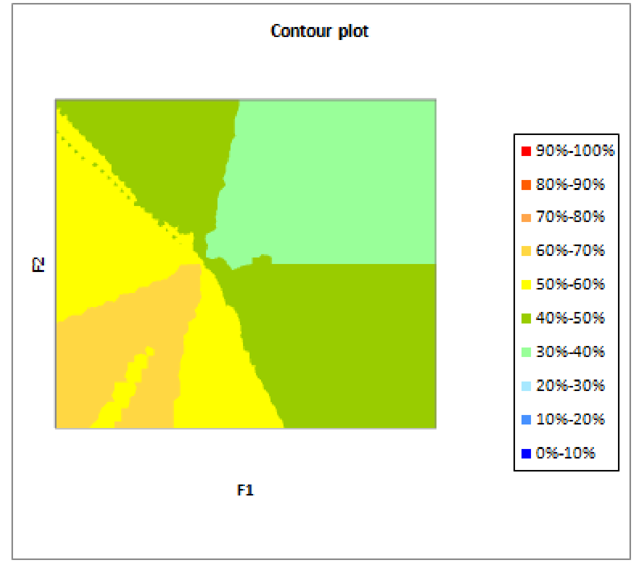




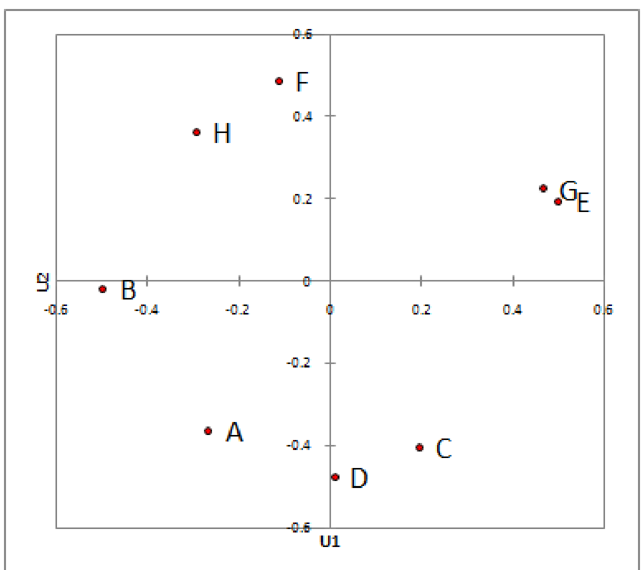
Cask Ales

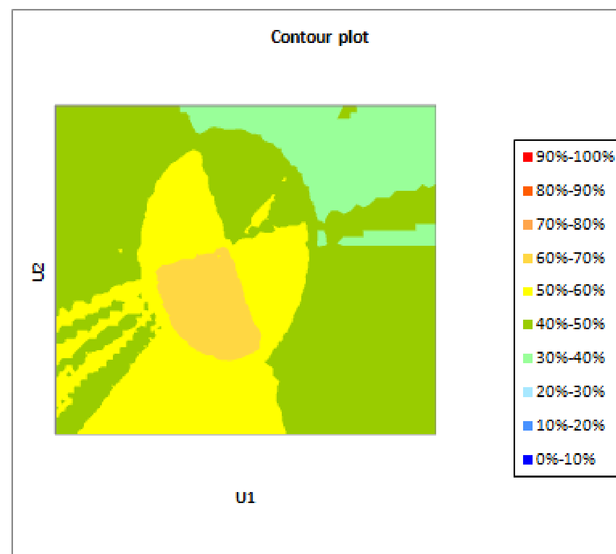
PCA





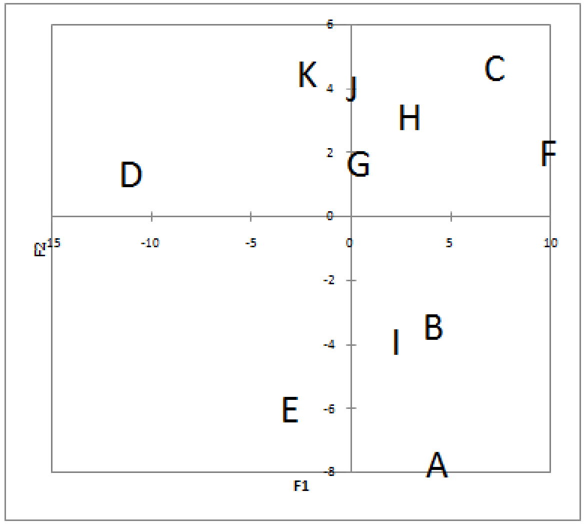
SMACOF

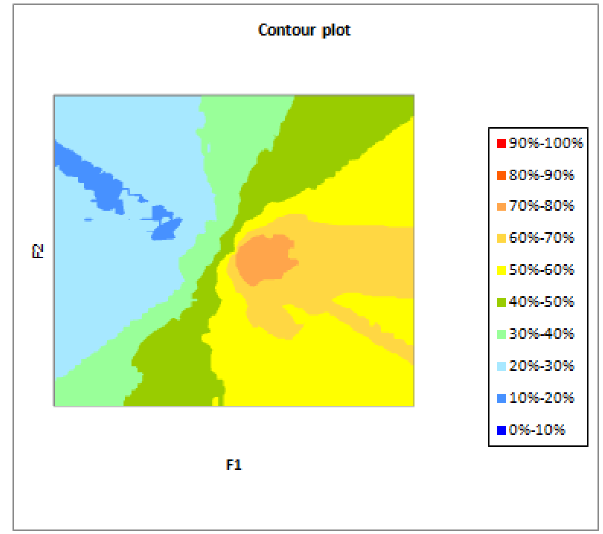




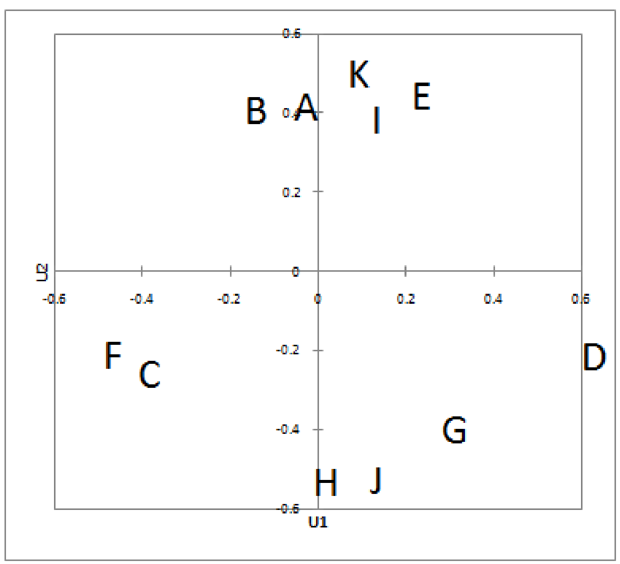
Rums

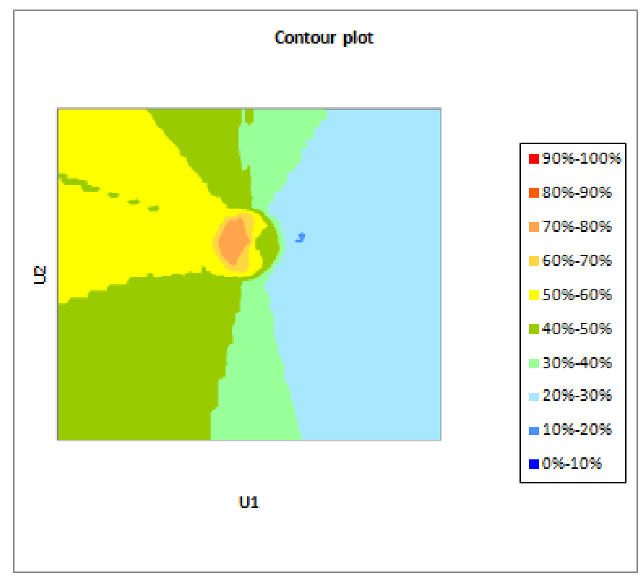
PCA





SMACOF





Conclusions

- In practice it looks like less than 50% of respondents show ideal point behaviour in the sense that the ideal point is well within the sample space

.

- PCA appears to perform better in the sense of sample recovery for the simulated case where less than 50% of respondents show clear ideal point behaviour and there is no noise, and in a simulated noise situation fits better than SMACOF in the sense of recovering the original configuration.
- In real data sets, SMACOF fits slightly more respondents significantly and appears to fit around equal numbers of ideal and vector models
- Comparing contour and sample plots across the two methods indicates a clear trend for SMACOF to not place points in the centre of the plot
- Not happy with SMACOF solution at this point
- PCA and Crossover Pref mapping giving a reasonable recovery.