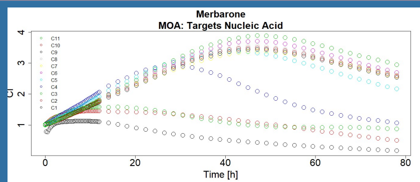
Clustering of Time Series Cytotoxicity Data

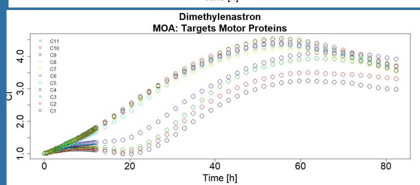
MacEwan University

Dan Richard

Introduction

- It is important to distinguish how different chemicals can harm cells.
- Mode of action (MOA) describes a cellular level functional change, which is a result of exposure of a living organism to a chemical [1]. Various chemicals, each with 11 concentrations, were grouped in 10 clusters according to MOAs and were tested on the HepG2 cell line over a 3 day period
- The goal is to reproduce the MOA clusters using statistical methods.



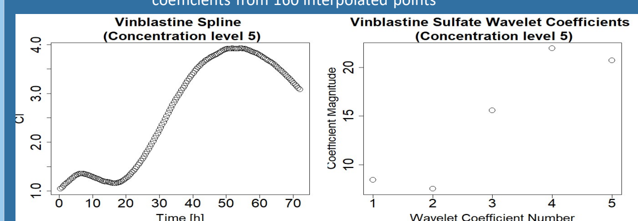


Cubic Spline Interpolation

- Raw data are sampled at uneven timesteps, so we fitted cubic splines and sampled at a fixed rate
- Data are smooth with little noise so there is negligible loss from interpolation.

Wavelet Transformation

- Wavelet analysis is similar to Fourier analysis in that it provides us with information from the frequency domain
- Wavelets differ from Fourier analysis by providing information from the time domain as well
- Wavelet analysis was used to reduce the dimension of the data set:
  - Each chemical concentration curve was reduced to 5 wavelet coefficients from 160 interpolated points



Acknowledgements

This project was supported by a USRI project grant, under the supervision of Prof. Cristina Anton, Department of Mathematics and Statistics, MacEwan University.

Self Organizing Maps

- Self organizing maps (SOMs) are useful for visualizing high dimensional data
- SOMs were used to cluster the wavelet coefficients

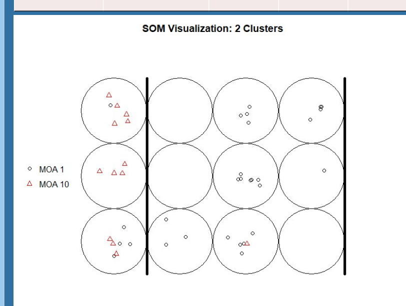
Steps for fitting a SOM [2]:

1. Provide a grid/graph where each node corresponds with a vector in the data space
2. Iteratively update the position of the node vectors based on their distances from the data points in a way similar to K-means clustering
3. Data points are assigned to their closest node

- There are many parameters and settings to consider prior to fitting a SOM in R:
  - Neighborhood Function: Determines the percentage of distance to adjust neighbouring nodes (during step 2 above)
  - Grid Size: how many nodes?
  - Grid Topology: Hexagonal versus Rectangular; Toroidal or Planar
- We first consider data from MOA clusters 1 and 10, and then from MOA clusters 1, 3 and 10 [1].

Results: 2 Clusters

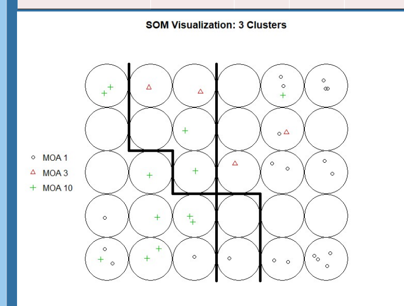
Neighbourhood Function	Grid Topology	Toroidal	Grid Size	Overall Rate	MOA Group	Cluster 1	Cluster 2
GAUSSIAN	HEXAGONAL	TRUE	4 X 3	0.8333	MOA 1	0.7931	0.2069
					MOA 10	0.0769	0.9231
		FALSE	6 X 5	0.8333	MOA 1	0.7931	0.2069
					MOA 10	0.0769	0.9231
	RECTANGULAR	**TRUE**	**4 X 3**	**0.8571**	**MOA 1**	**0.8276**	**0.1724**
					MOA 10	**0.0769**	**0.9231**
		FALSE	4 X 3	0.7857	MOA 1	0.7241	0.2759
					MOA 10	0.0769	0.9231
BUBBLE	HEXAGONAL	**TRUE**	**4 X 3**	**0.8571**	**MOA 1**	**0.8276**	**0.1724**
					MOA 10	**0.0769**	**0.9231**
		FALSE	4 X 3	0.5714	MOA 1	0.4138	0.5862
					MOA 10	0.0769	0.9231
	RECTANGULAR	**TRUE**	**6 X 5**	**0.8571**	**MOA 1**	**0.8276**	**0.1724**
					MOA 10	**0.0769**	**0.9231**
		FALSE	6 X 5	0.7857	MOA 1	0.8276	0.1724
					MOA 10	0.3077	0.6923



- Figure to the left is a SOM mapping plot with a Gaussian neighbourhood function, and Rectangular, toroidal grid
- Toroidal seems to out perform plane grids
- Result is less impacted by the choice of neighbourhood function and grid topology

Results: 3 Clusters

Neighbourhood Function	Grid Topology	Toroidal	Grid Size	Overall Rate	MOA Group	Cluster 1	Cluster 2	Cluster 3
GAUSSIAN	HEXAGONAL	TRUE	6 x 5	0.6579	MOA 1	**0.4762**	0.3333	0.1905
					MOA 3	0.2500	**0.7500**	0.0000
					MOA 10	0.0769	0.0000	**0.9231**
		FALSE	4 x 3	0.6842	MOA 1	**0.5714**	0.2381	0.1905
					MOA 3	0.2500	**0.7500**	0.0000
					MOA 10	0.0769	0.0000	**0.8462**
	RECTANGULAR	TRUE	4 x 3	0.6842	MOA 1	**0.5714**	0.2381	0.1905
					MOA 3	0.2500	**0.5000**	0.2500
					MOA 10	0.0769	0.0000	**0.9231**
		FALSE	**6 x 5**	**0.7105**	**MOA 1**	**0.7619**	**0.0476**	**0.1905**
					MOA 3	**0.5000**	**0.5000**	**0.0000**
					MOA 10	**0.0769**	**0.2308**	**0.6923**
BUBBLE	HEXAGONAL	TRUE	4 x 3	0.6316	MOA 1	**0.4762**	0.3333	0.1905
					MOA 3	0.2500	**0.7500**	0.0000
					MOA 10	0.0769	0.0769	**0.8462**
		FALSE	6 x 5	0.6053	MOA 1	**0.6667**	0.1429	0.1905
					MOA 3	0.2500	**0.0000**	0.7500
					MOA 10	0.0769	0.2308	**0.6923**
	RECTANGULAR	TRUE	6 x 5	0.5526	MOA 1	**0.4762**	0.3333	0.1905
					MOA 3	0.2500	**0.7500**	0.0000
					MOA 10	0.0769	0.3077	**0.6154**
		FALSE	6 x 5	0.6316	MOA 1	**0.4286**	0.3810	0.1905
					MOA 3	0.0000	**1.0000**	0.0000
					MOA 10	0.0769	0.0769	**0.8462**



- Figure to the left is a SOM mapping plot with a Gaussian neighbourhood function, and Rectangular, planar grid
- Gaussian separates the MOA better than the bubble neighbourhood function
- Other parameters appear to have little impact on the overall rate

Conclusions

- Chemical sets of concentration curves tend to group by their MOA clusters, but there is room for improvement.
- The performance is probably affected by the range of the concentrations used in the in-vitro experiments for each toxicant.
- SOMs are a useful tool for visualizing high dimensional data, but require a lot of parameter tuning.

References

[1] Zhang et al., Machine learning algorithms for mode-of-action classification in toxicity assessment, BioData Mining , 9:19, (2016).

[2] Kohonen, Teuvo, Essentials of the self-organizing map. Neural Networks 37, pp. 52-65 (2013).