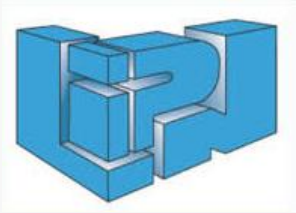


CO-CLUSTERING DE SÉRIES TEMPORELLES POUR LA VALIDATION DE SYSTÈMES D'AIDE À LA CONDUITE PAR SIMULATION MASSIVE

SOCIÉTÉ FRANCOPHONE DE CLASSIFICATION

- Directeur de thèse : Mustapha Lebbah
- Codirecteur de thèse : Hanane Azzag
- Superviseur Renault : Loïc Giraldi

AGENDA

01 **CONTEXT**
ADAS Validation

02 **FUNCTIONAL LATENT BLOCK MODEL**
Time Series Co-clustering

03 **FUNCTIONAL CONDITIONAL LATENT BLOCK MODEL**
Multi-view extension

01

CONTEXT

ADAS - HUGE INCREASE OF THE NUMBER OF SYSTEMS

ADAS SYSTEMS

VEHICLE APPLICATIONS

PLANTS & MARKETS

ADAS – SIMULATION SYSTEM

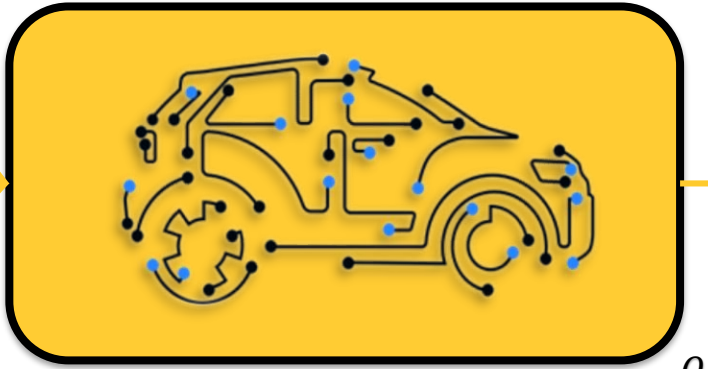
SCENARIO

OBSERVED SIGNALS

ENVIR. DATA (Traffic, driver , ..)

CAR DYNAMICS

CONTROL LOGICS



Simulation system

Data : Multivariate time series

$O(10^6)$ simulations
$O(10^3)$ points per time series
$O(10^2)$ variables

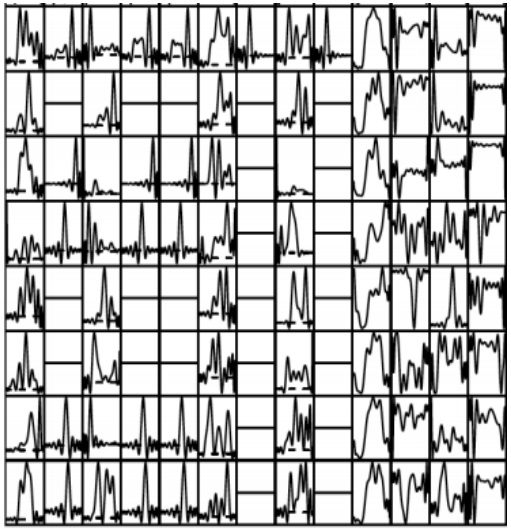
OBJECTIVES

Multivariate Time series

Simulation 1

Simulation 2

…



Picture: Slimen, Y. B., Allio, S., & Jacques, J. (2018). Model-based co-clustering for functional data. *Neurocomputing*, *291*, 97-108.

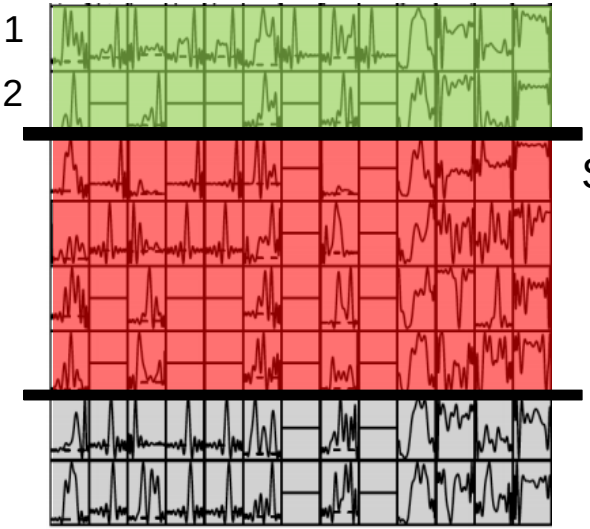
OBJECTIVES

Multivariate Time series

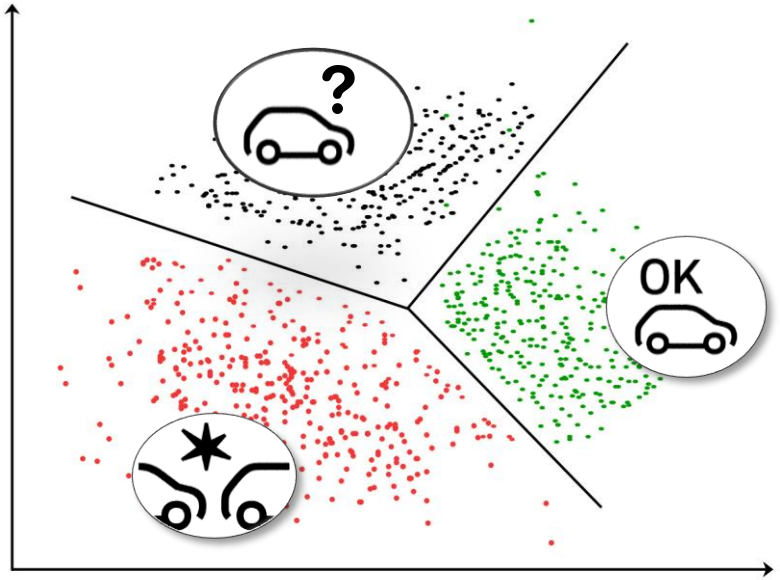
Simulation 1

Simulation 2

…



Simulations Clustering



Picture: Slimen, Y. B., Allio, S., & Jacques, J. (2018). Model-based co-clustering for functional data. *Neurocomputing*, *291*, 97-108.

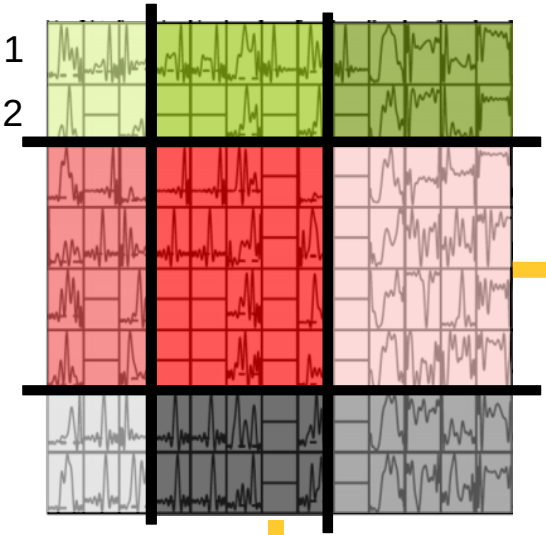
OBJECTIVES

Multivariate Time series

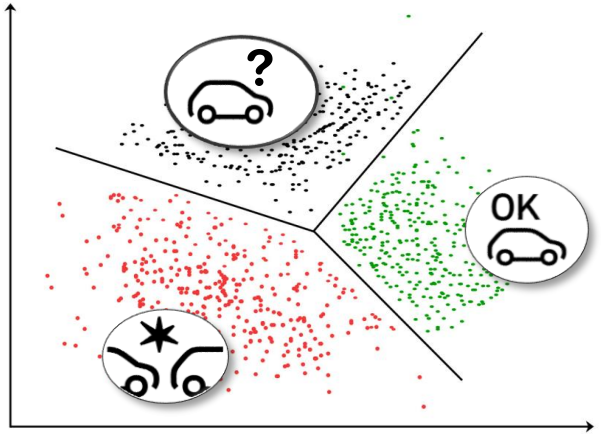
Simulation 1

Simulation 2

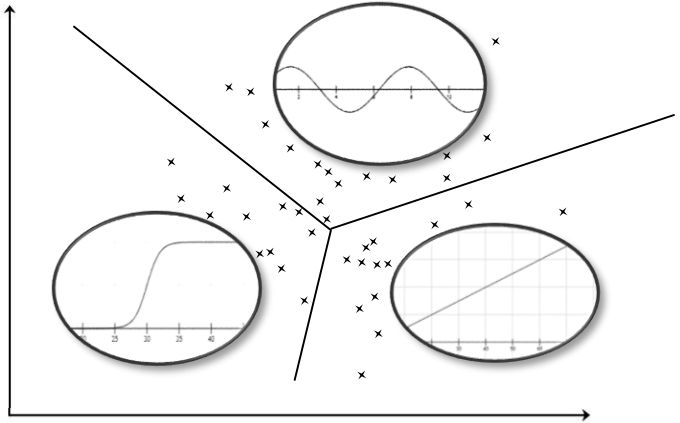
…



Simulations Clustering



Signals Clustering



02

FUNCTIONAL LATENT BLOCK MODEL

BIBLIOGRAPHY

- Mixture Model selection with ICL

Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence*, *22*(7), 719-725.

- Latent Block Model

Govaert, G. and M. Nadif (2013). Co-Clustering. Wiley-ISTE

- Functional Latent Block Model (FLBM) based on functional PCA (FPCA) representation

Slimen, Y. B., Allio, S., & Jacques, J. (2018). Model-based co-clustering for functional data. *Neurocomputing*, *291*, 97-108.

- FLBM based on FPCA representation and subspace clustering

Bouveyron, C., Bozzi, L., Jacques, J., & Jollois, F. X. (2017). The functional latent block model for the co-clustering of electricity consumption curves.

- Multivariate FLBM based on piecewise regression representation

Chamroukhi, F., & Biernacki, C. (2017, July). Model-Based Co-Clustering of Multivariate Functional Data.

- ICL criterion for Latent Block Model

Lomet, A., Govaert, G., & Grandvalet, Y. (2018). Model selection for Gaussian latent block clustering with the integrated classification likelihood. *Advances in Data Analysis and Classification*, *12*(3), 489-508.

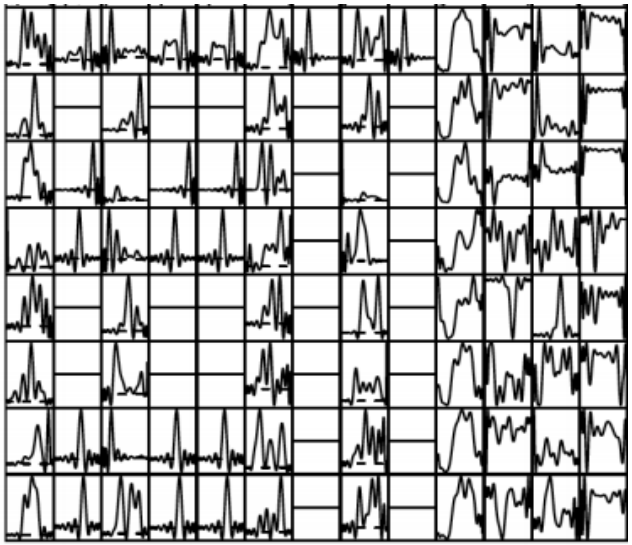
- Multivariate FLBM based on piecewise regression representation

Schmutz, A., Jacques, J., Bouveyron, C., Chèze, L., & Martin, P. (2019, June). Co-clustering de courbes fonctionnelles multivariées.

ILLUSTRATION

Functional Variables

Observations / Simulations



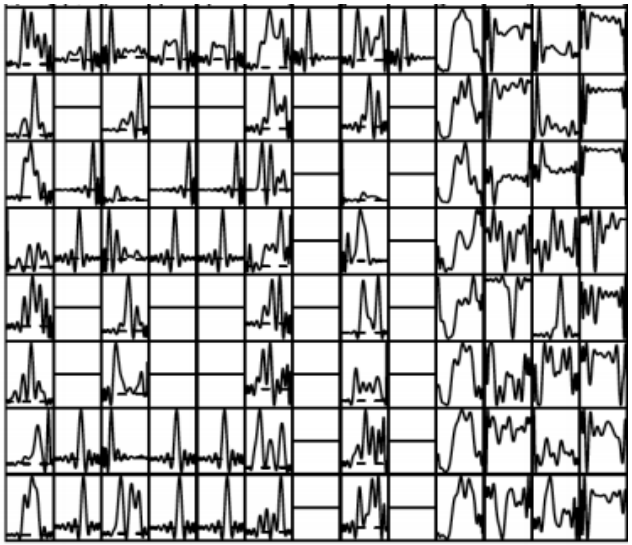
X

ILLUSTRATION

Functional Variables

(speed, acceleration, radius, pitch, system activation, ..)

Observations / Simulations

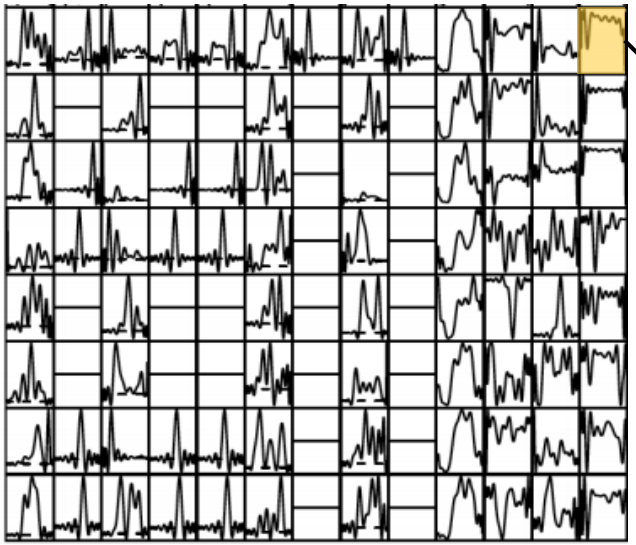


X

ILLUSTRATION

Functional Variables

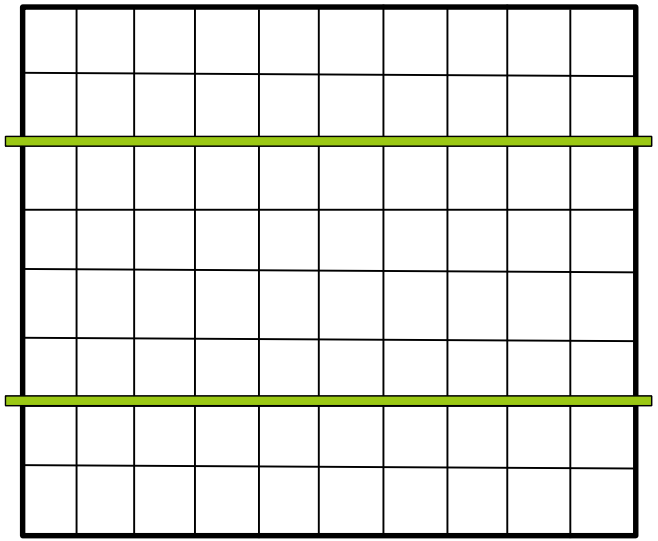
Observations / Simulations



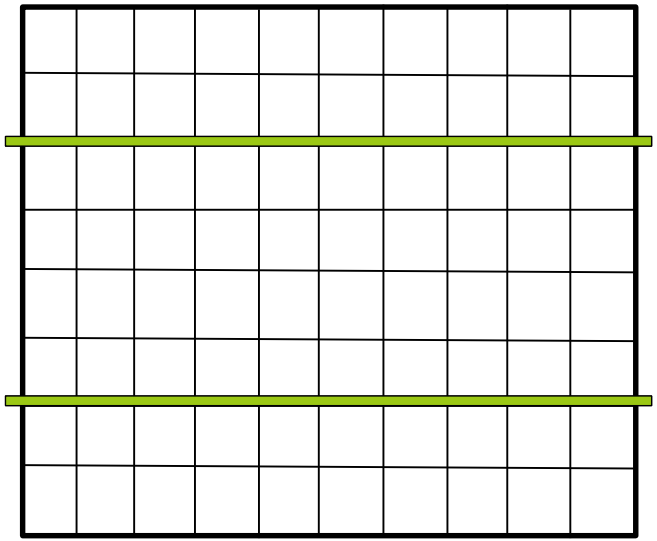
X

1) Expressed in a common Fourier Basis
2) Projected in a lower dimensional space

ILLUSTRATION



ILLUSTRATION



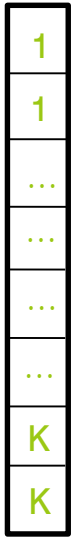
Cluster 1

...

Cluster K

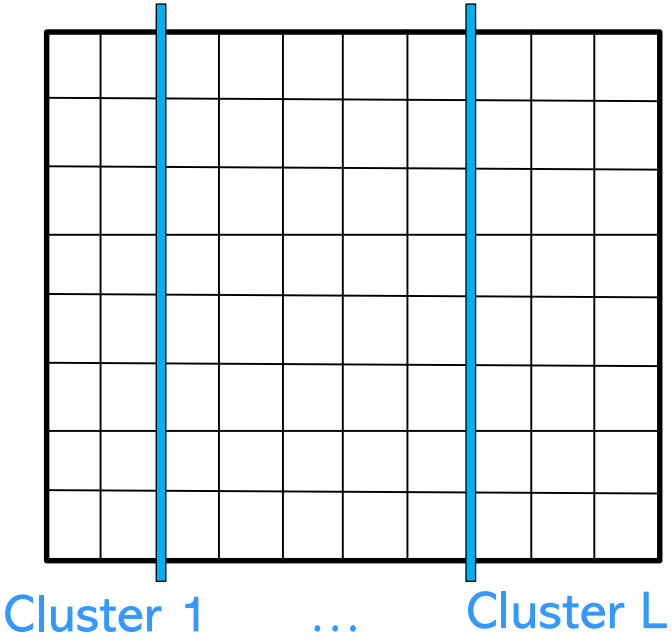
X

Row Cluster Membership

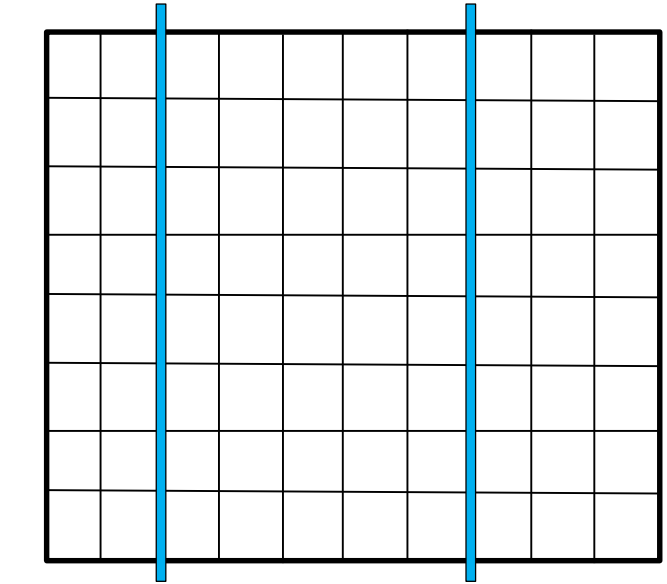


$$Z = (z_i)_{1..n}$$

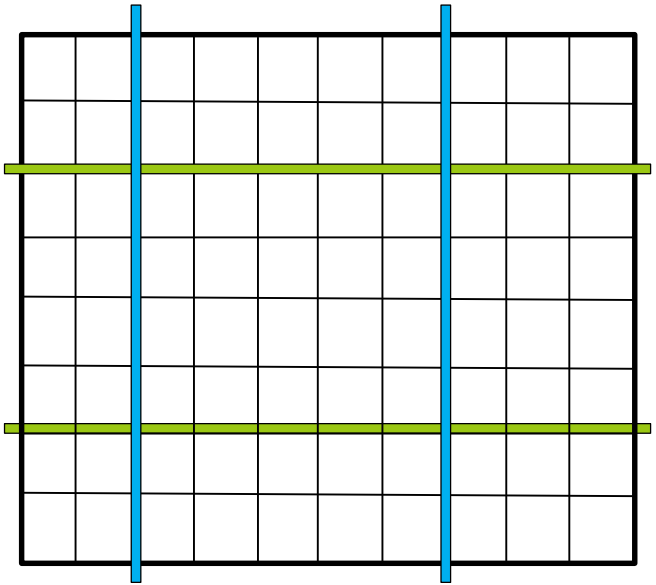
ILLUSTRATION



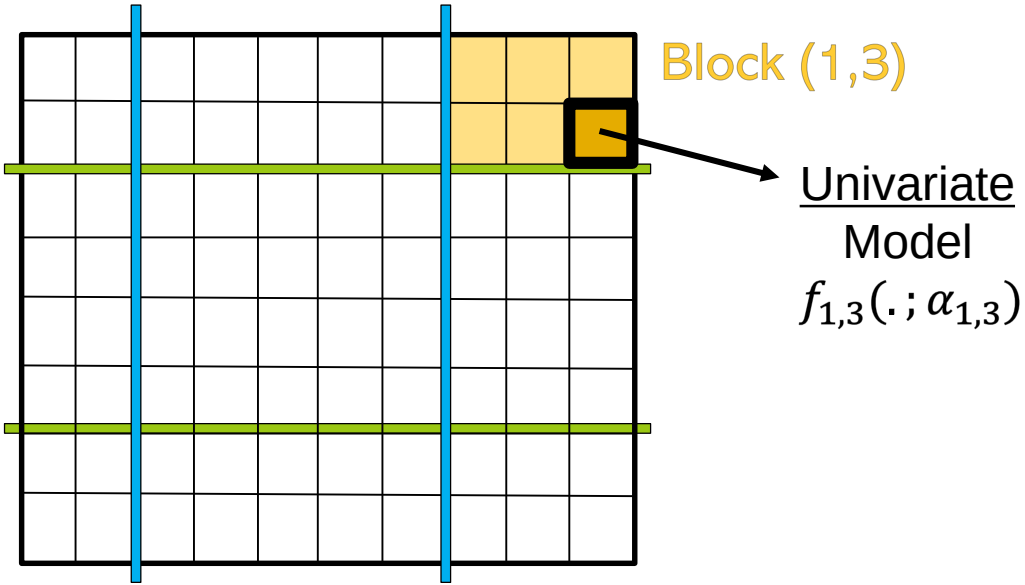
ILLUSTRATION



ILLUSTRATION



ILLUSTRATION

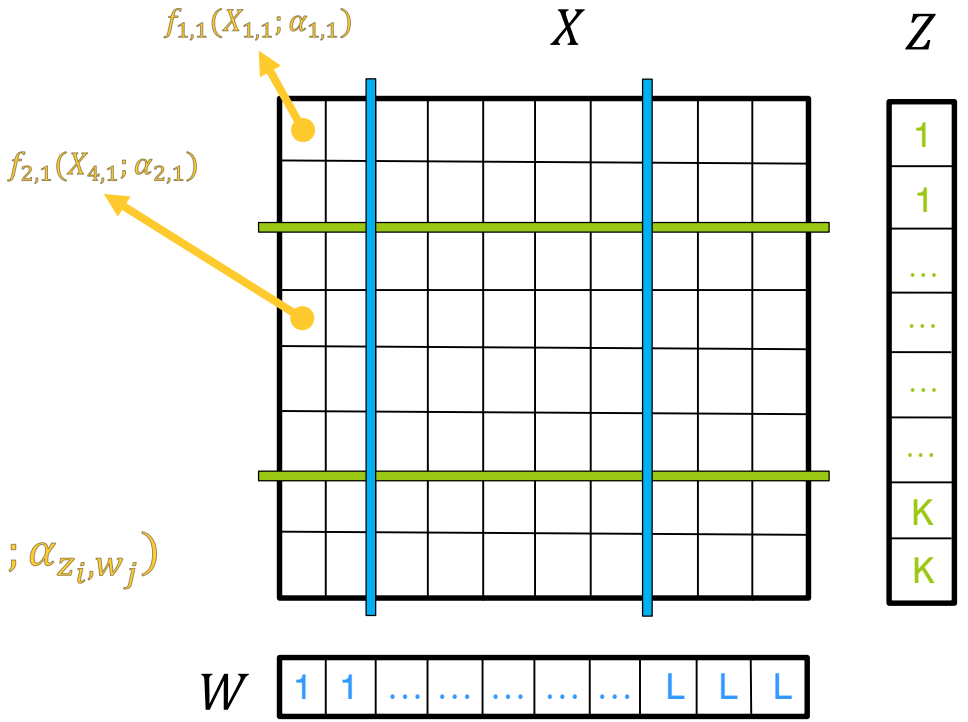


COMPLETED LIKELIHOOD

$$p(X, W, Z; \theta)$$

$$= p(W, Z)\ p(X \mid W, Z)$$
$$= p(W)\, p(Z)\, p(X \mid W, Z)$$

$$= \prod_{j=1}^{p} p(w_j) \prod_{i=1}^{n} p(z_i) \prod_{i,j} f_{z_i, w_j}(X_{i,j}; \alpha_{z_i, w_j})$$



INFERENCE WITH SEM-GIBBS

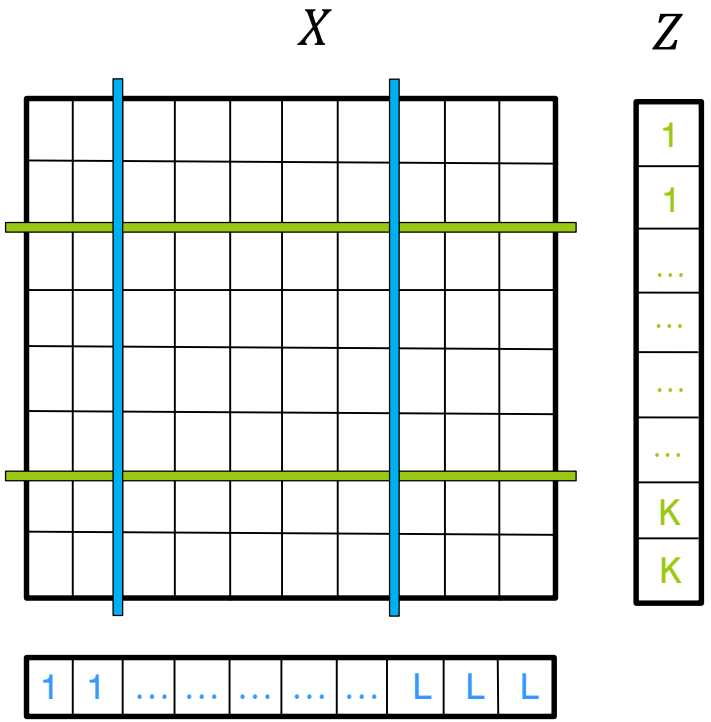
For loop :

- (E step) Gibbs Sampling

→ estimate p(z,w) by alternating the estimation of p(z | w) and p(w | z).

- (M step) Block parameters estimation, given $(\hat{z}, \hat{w})$

End For



MODEL SELECTION - CRITERION

$$Integrated\ Completed\ Likelihood\ (ICL) =$$

$$\log p(X, W, Z) - \frac{C}{2}\log(np) - \frac{K-1}{2}\log n - \frac{L-1}{2}\log p$$

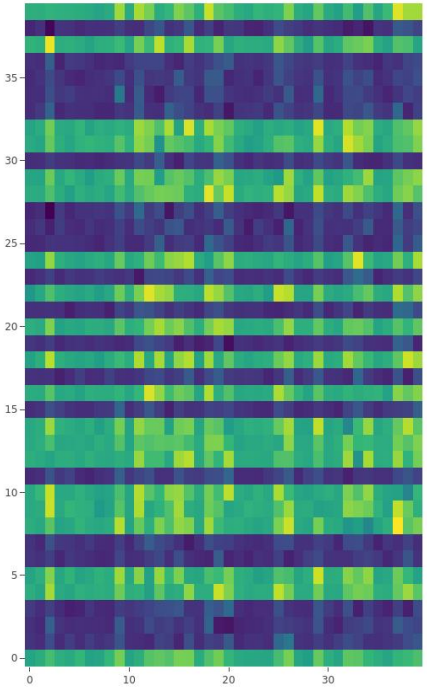
$$\approx \int_{\Theta_{K,L}} L(x, z\,; \theta)\, p(\theta | K, L)\, d\theta$$

With C the number of free parameters of the Gaussian distribution

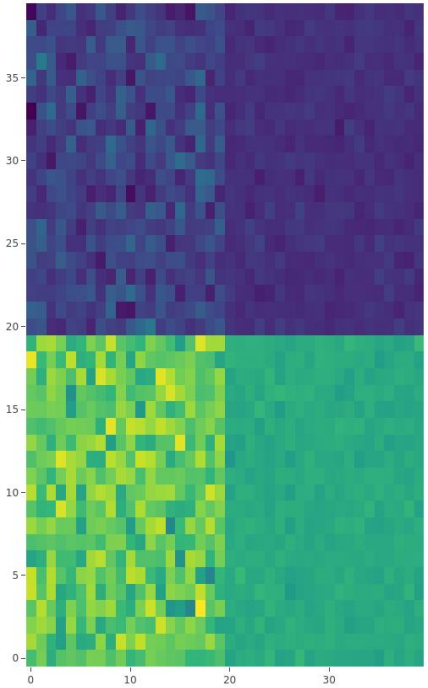
LIMITS OF THE LATENT BLOCK MODEL

LBM LIMITS – APPROPRIATE LATENT DATASET STRUCTURE

Raw Data

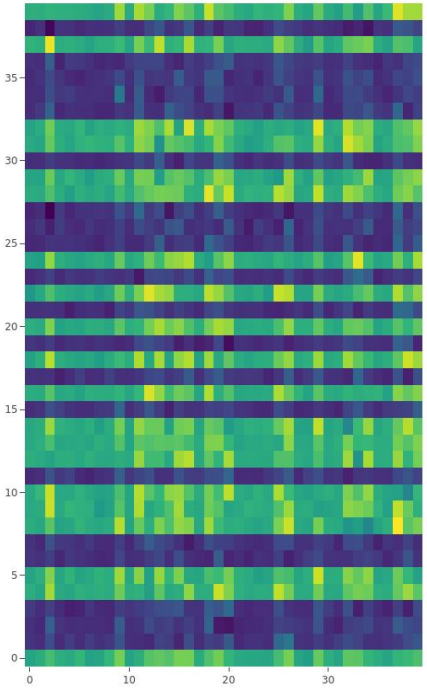


« True » Latent Structure

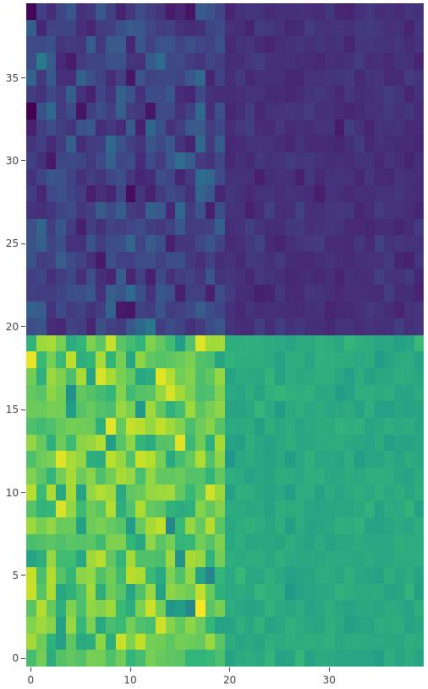


LBM LIMITS – APPROPRIATE LATENT DATASET STRUCTURE

Raw Data



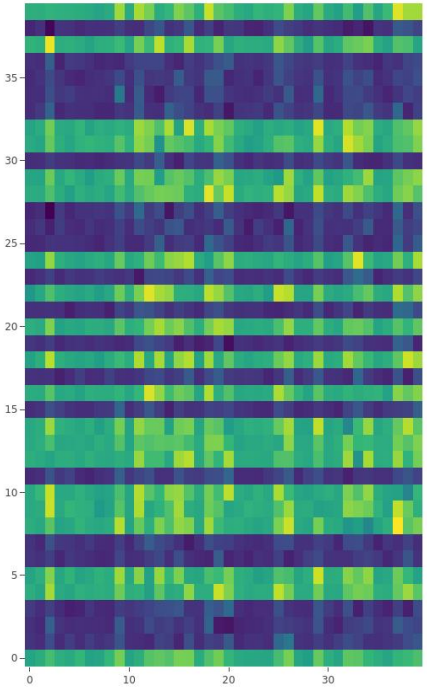
« True » Latent Structure



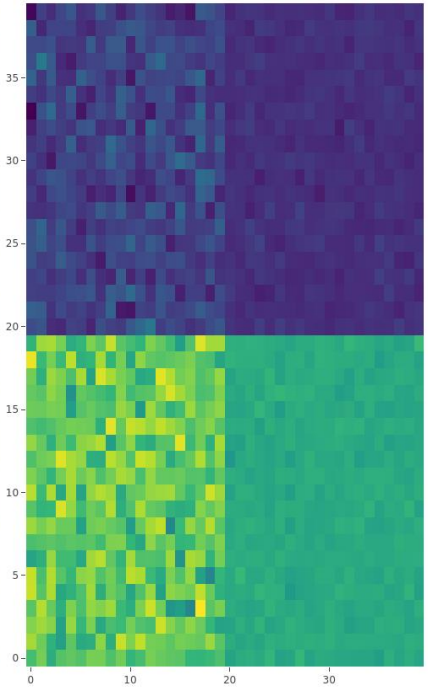
LBM Model Inference

LBM LIMITS – APPROPRIATE LATENT DATASET STRUCTURE

Raw Data

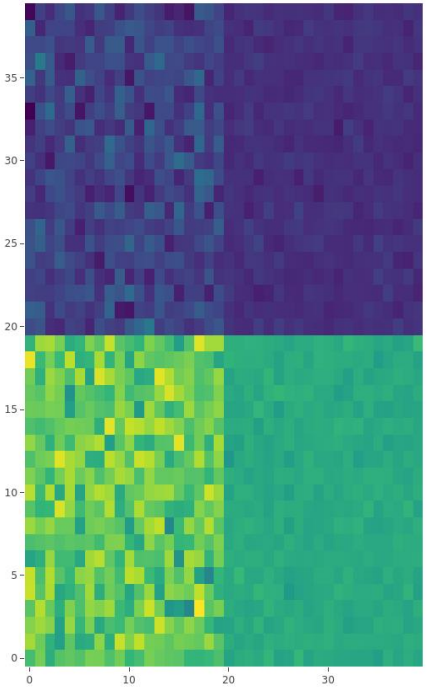


« True » Latent Structure



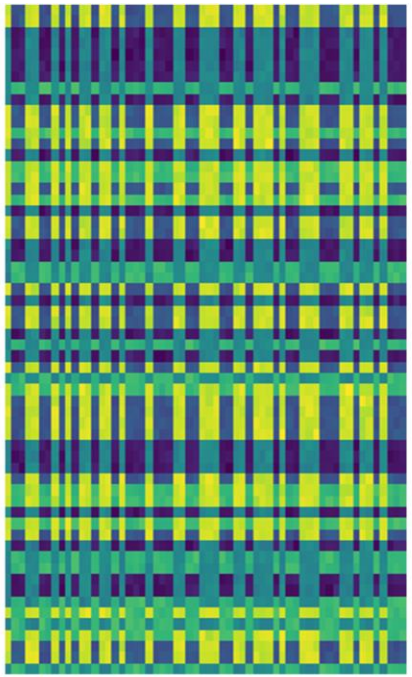
LBM Model Inference

Inferred Structure

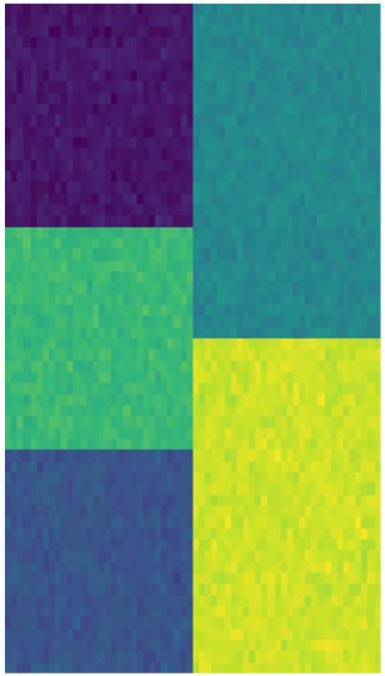


LBM LIMITS – INADAPTED DATASET STRUCTURE

Raw Data

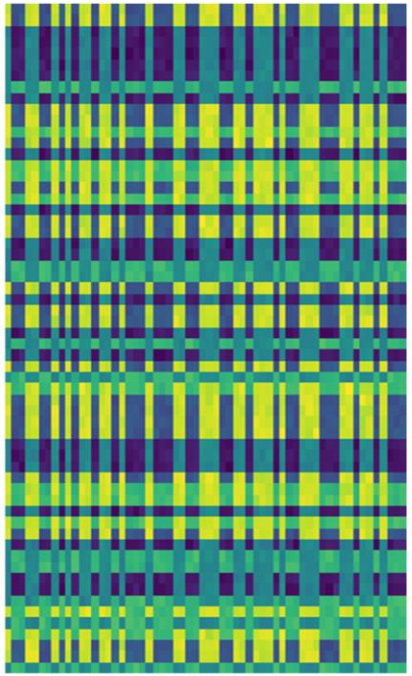


« True » Latent Structure

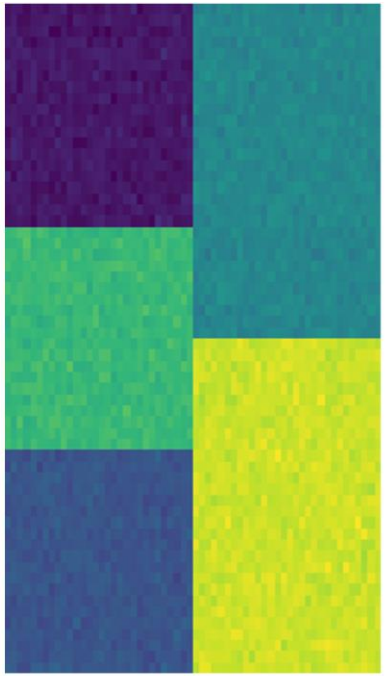


LBM LIMITS – INADAPTED DATASET STRUCTURE

Raw Data

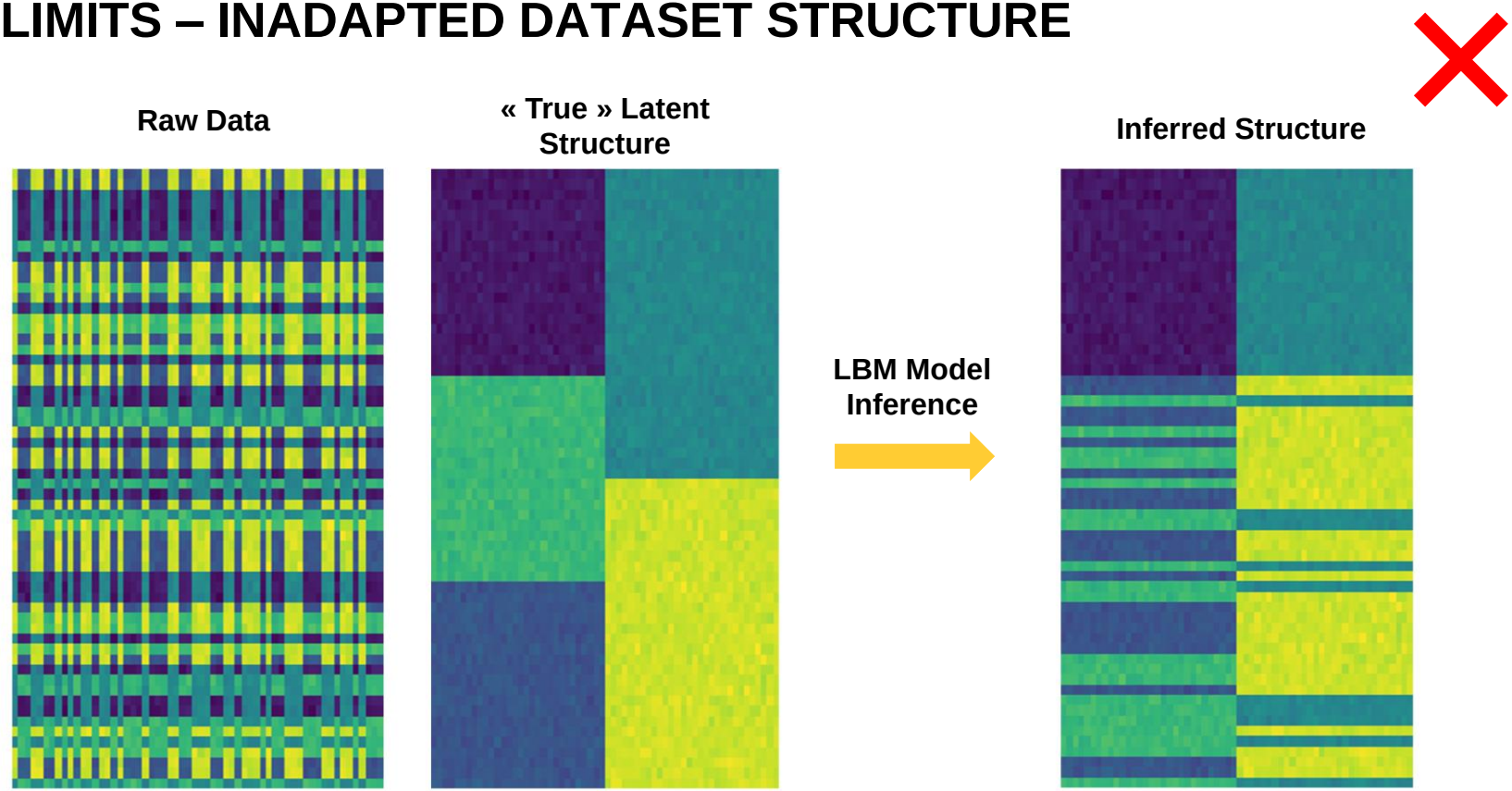


« True » Latent Structure



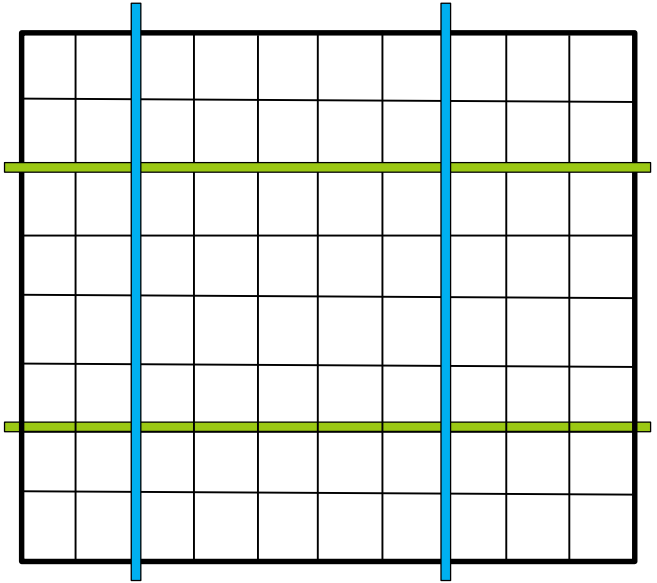
LBM Model Inference

LBM LIMITS – INADAPTED DATASET STRUCTURE



MODEL EXTENSION:
CONDITIONAL LATENT BLOCK MODEL

LBM



Row Cluster Membership

1
1
…
…
…
…
K
K

$Z = (z_i)_{1..n}$

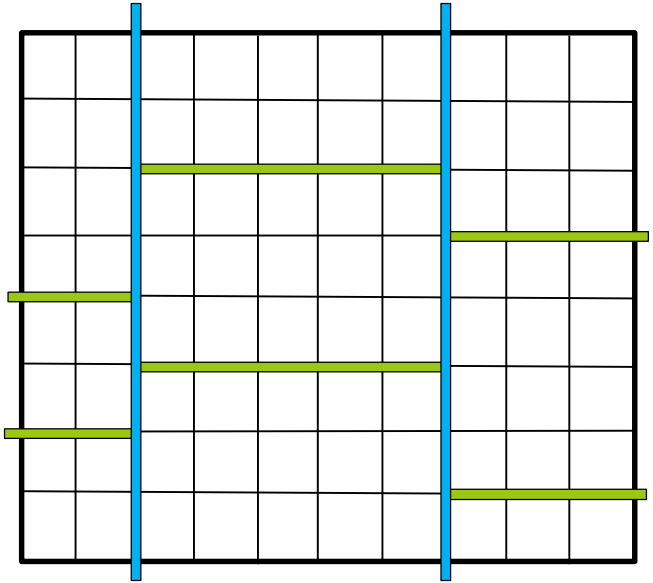
Cluster 1 … Cluster L

$W = (w_i)_{1..p}$

1	1	…	…	…	…	…	L	L	L

Column Cluster Membership

CONDITIONAL LBM



Conditional Row Cluster Membership

1,1	1,2	1,3
1,1	1,2	1,3
1,1	2,2	1,3
1,1	2,2	2,3
2,1	2,2	2,3
2,1	3,2	2,3
3,1	3,2	2,3
3,1	3,2	3,3

$$Z = (z_{i,l})_{\{1..n\}\times\{1,..L\}}$$

Cluster 1 … Cluster L

$$W = (w_i)_{1..p}$$

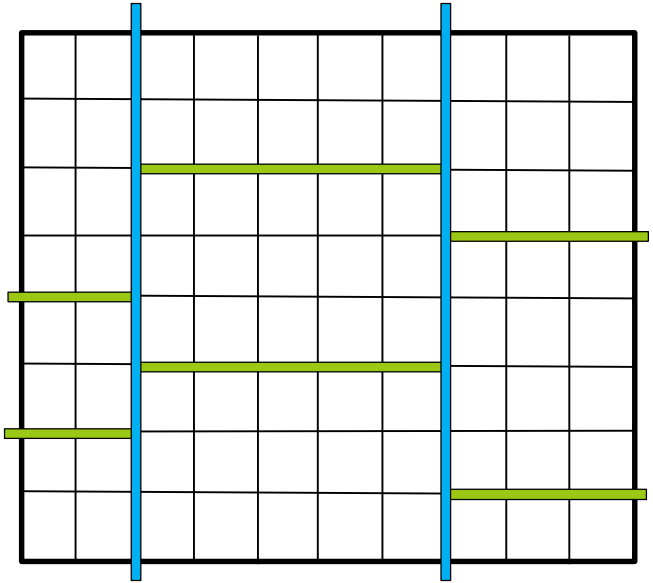
1	1	…	…	…	…	…	L	L	L

Column Cluster Membership

CONDITIONAL LBM

Each **column** cluster produces a specific **row** clustering

→ Create **multi-views** clustering. The expert choose among the most relevant one w.r.t variable cluster of interest.



Conditional Row Cluster Membership

1,1	1,2	1,3
1,1	1,2	1,3
1,1	2,2	1,3
1,1	2,2	2,3
2,1	2,2	2,3
2,1	3,2	2,3
3,1	3,2	2,3
3,1	3,2	3,3

$$Z = (z_{i,l})_{\{1..n\}\times\{1,..L\}}$$

$$W = (w_i)_{1..p}$$

Cluster 1		…					Cluster L		
1	1	…	…	…	…	…	L	L	L

Column Cluster Membership

CONDITIONAL LBM – MODEL SELECTION

For given L_{max} and K_{max}, the number of possible models is:

$$\sum_{l=1}^{L_{max}} \binom{K_{max} + l - 1}{l},$$

the number of combination with repetition
and without order

CONDITIONAL LBM – MODEL SELECTION

For given L_{max} and K_{max}, the number of possible models is:

$$\sum_{l=1}^{L_{max}} \binom{K_{max} + l - 1}{l},$$

the number of combination with repetition
and without order

→ With $L_{max} = K_{max} = 10$, 184755 combinations (vs 100 in the LBM case)
→ Impossible to perfom a Grid Search

CONDITIONAL LBM – MODEL SELECTION

For given L_{max} and K_{max}, the number of possible models is:

$$\sum_{l=1}^{L_{max}} \binom{K_{max} + l - 1}{l},$$

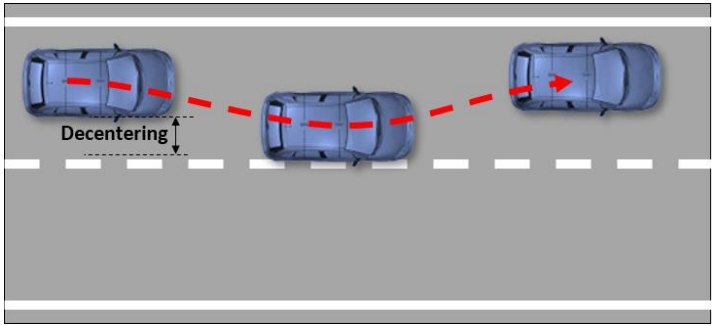
the number of combination with repetition
and without order

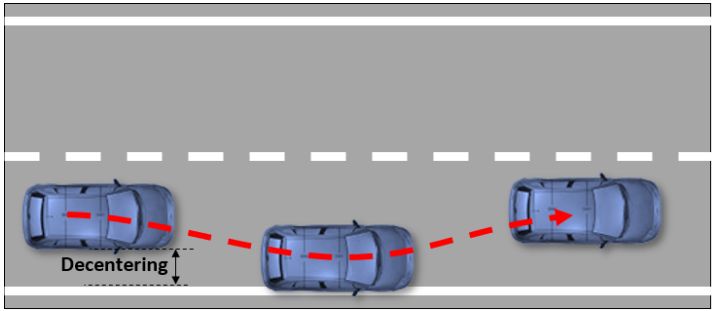
- → With $L_{max} = K_{max} = 10$, 184755 combinations (vs 100 in the LBM case)
- → Impossible to perfom a Grid Search
- → Strategies in development..

APPLICATION: LANE KEEPING ASSIST

APPLICATION: LANE KEEPING ASSIST (LKA)

The car “of interest” moves in a straight line and then drifts to one side of the road.
The LKA system puts the vehicle back to the lane center.

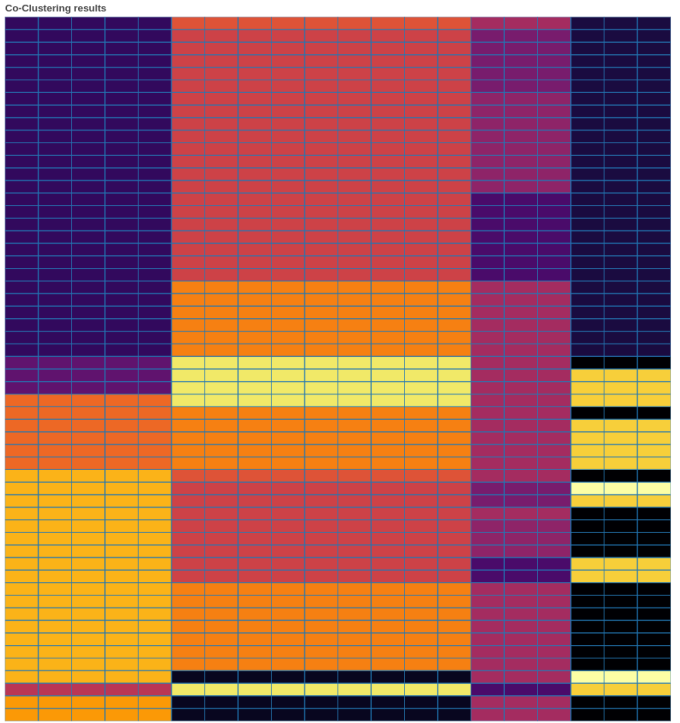




APPLICATION: LANE KEEPING ASSIST

20 signals

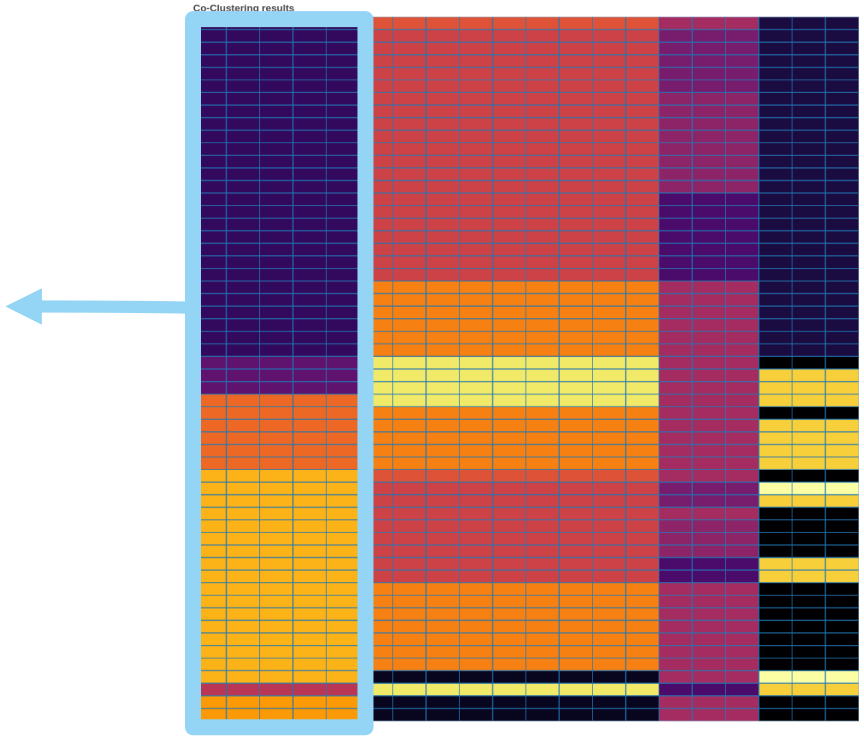
56 simulations



APPLICATION: LANE KEEPING ASSIST

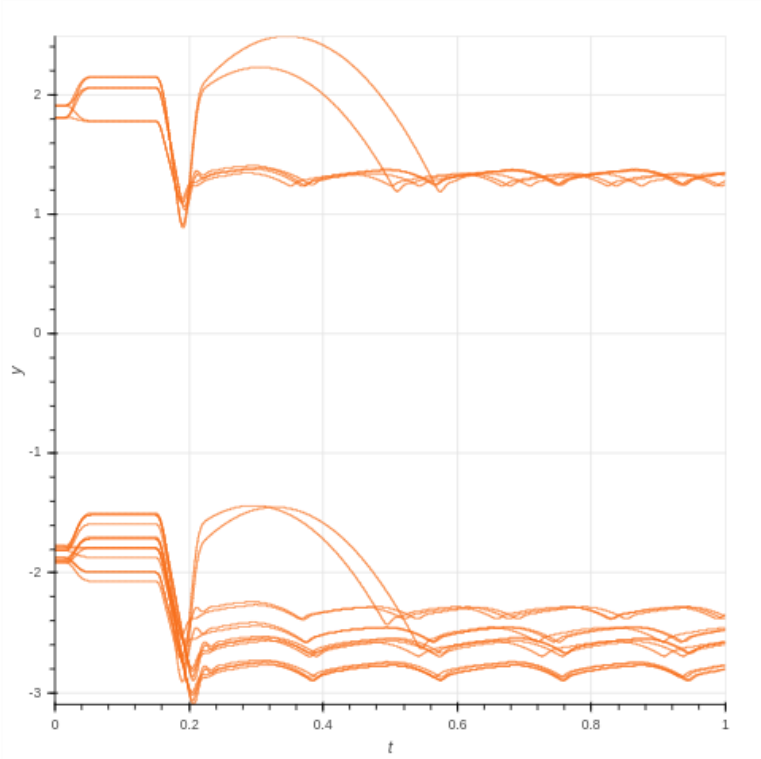
→ **Position signals**

- Position in lane (continuous - duplicated)
- Index of current lane (discrete)
- Marking line type on Ego's right (discrete)
- Marking line type on Ego's left (discrete)



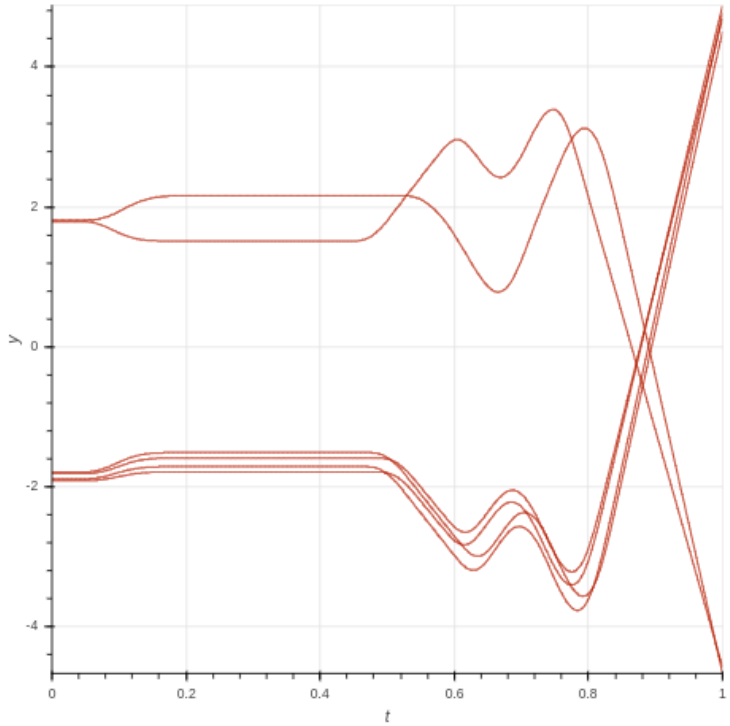
APPLICATION: LANE KEEPING ASSIST

Lateral position in Lane: block (1, 1)



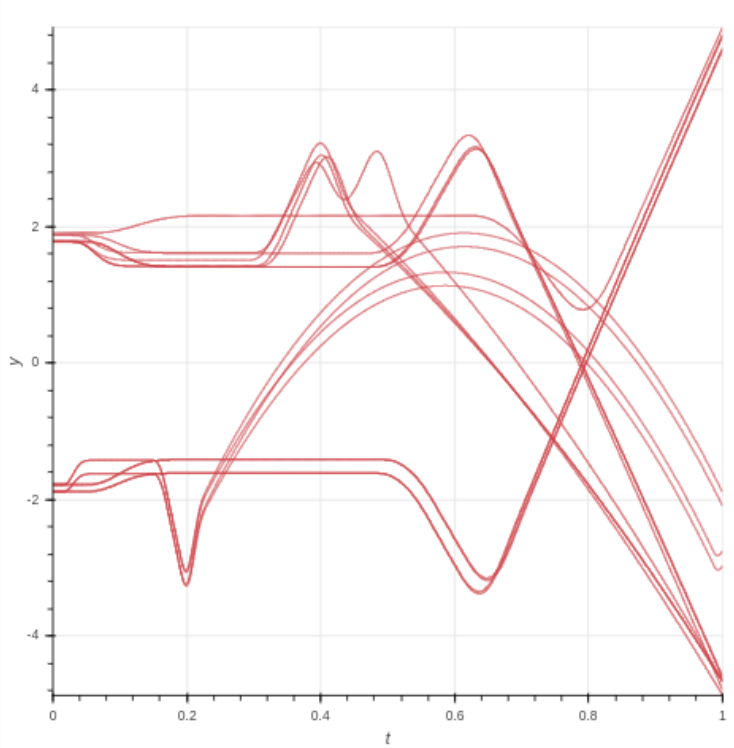
APPLICATION: LANE KEEPING ASSIST

Lateral position in Lane: block (2, 1)



APPLICATION: LANE KEEPING ASSIST

Lateral position in Lane: block (3, 1)

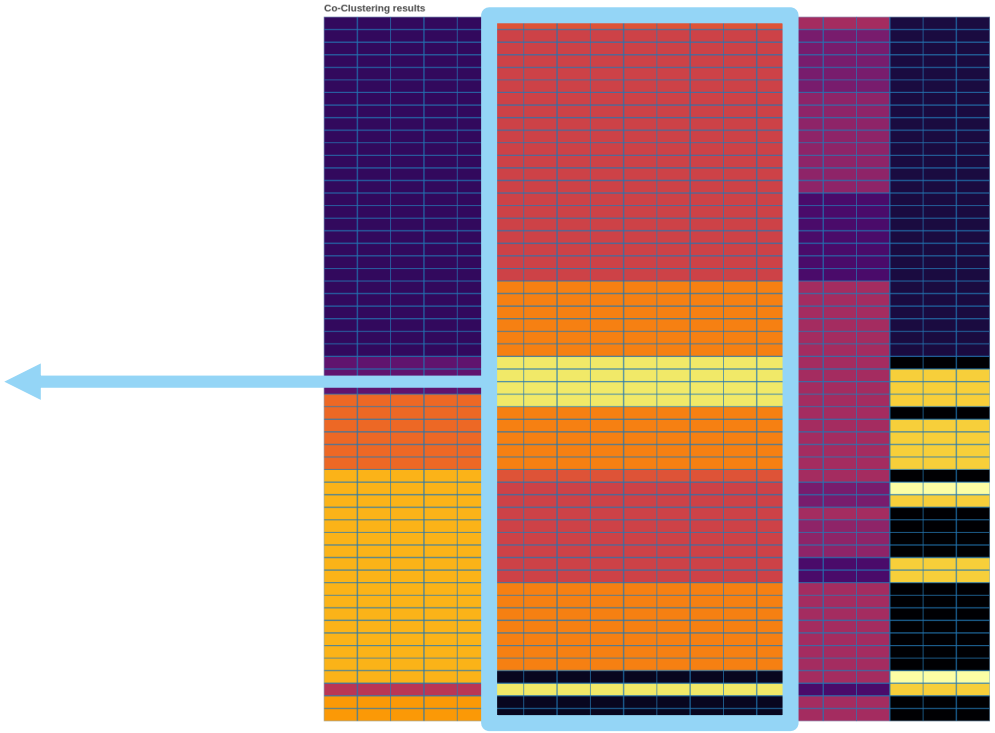


APPLICATION: LANE KEEPING ASSIST

→ **Bag of uninformative/ irrelevant signals**

- Constant values (vehicle length, width, distance between wheels, road bend radius)

- Linearly increasing (distance to origin)

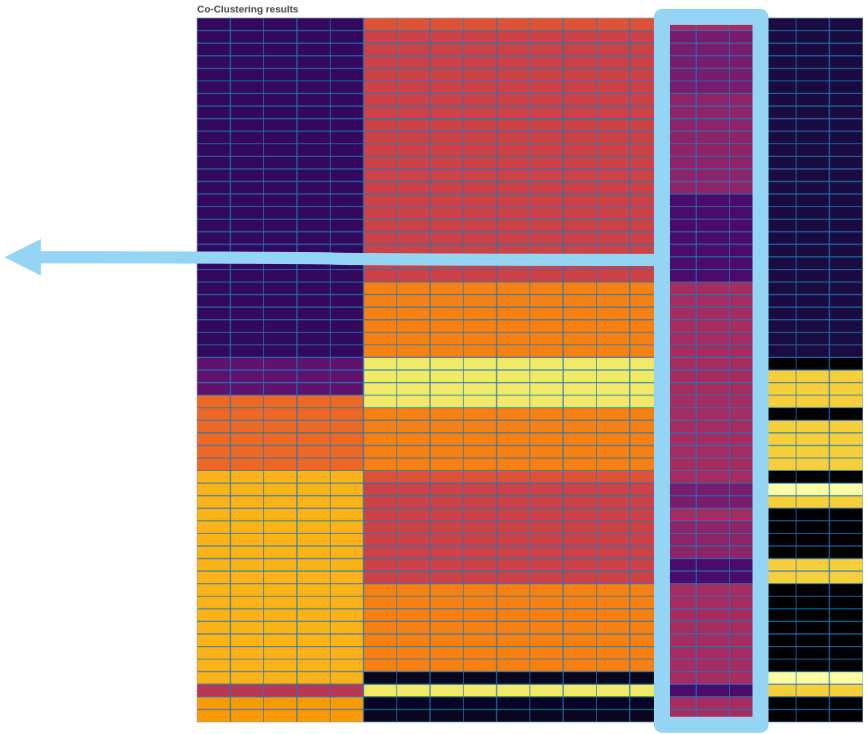


APPLICATION: LANE KEEPING ASSIST

→ **Leverage signals**

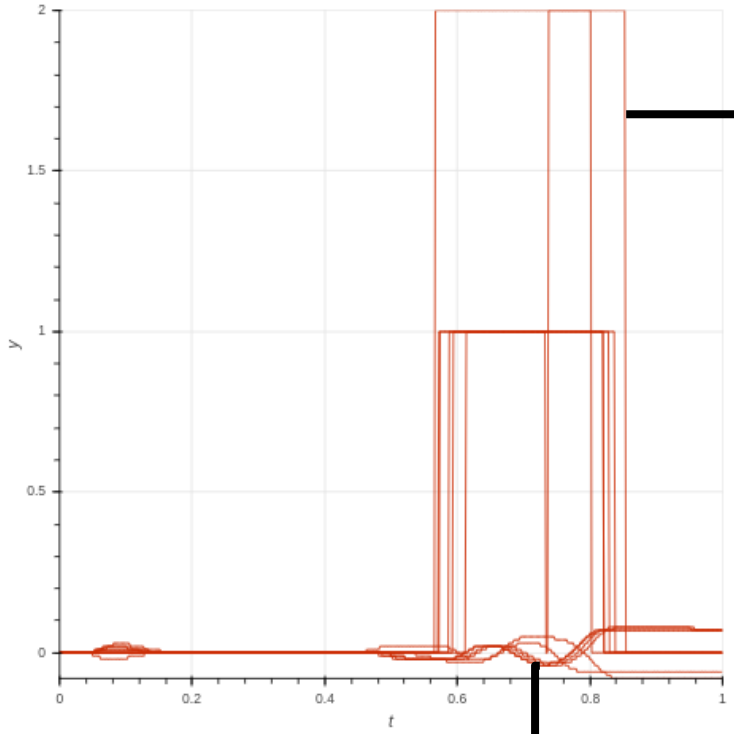
- Control logic activation

- Changes in Ego's heading



APPLICATION: LANE KEEPING ASSIST

Block (1,3)



Control Logic Active

Ego's heading change

THANK YOU

APPENDIX

MODEL SELECTION

How to choose K and L ?

→ For small values of K and L:

Model Selection Criterion + Grid Search

NUMBER OF POSSIBLE MODELS

$$L_{max} = K_{max} = 3$$

Number of combination with replacement & without order

$$\sum_{L=1}^{L_{max}} \binom{K_{max} + L - 1}{L}$$

$$= \binom{3}{1} + \binom{4}{2} + \binom{5}{3}$$

$$= 3 + 6 + 10 = 19$$

3	6	10
1	1,1	1,1,1
2	1,2	1,1,2
3	1,3	1,1,3
	2,2	1,2,2
	2,3	1,2,3
	3,3	1,3,3
		2,2,2
		2,2,3
		2,3,3
		3,3,3