Identification of causalities in spatio-temporal data

J. Raimbault\textsuperscript{1,2}
juste.raimbault@parisgeo.cnrs.fr

\textsuperscript{1}UMR CNRS 8504 Géographie-cités
\textsuperscript{2}UMR-T IFSTTAR 9403 LVMT

Sageo 2017 - Rouen
November 7th 2017
Spatio-temporal complexity

Source: [Arnaud et al., 2013]
Fieldwork in Pearl River Delta unveils local manifestations of the co-evolution of transportation networks and territories. Source: Author.
From [Bonnafoús and Plassard, 1974] to [Offner, 1993]: do transportation infrastructures structure territories?

→ Existence of co-evolutive processes [Bretagnolle, 2009]
→ At large scale, existence of structural urban system dynamics [Offner et al., 2014]
→ The question of circular causalities arises at all scales (e.g. metropolitan scale and mobility [Cerqueira, 2017]) and in various fields (knowledge spillovers and innovation [Audretsch and Feldman, 1996])
→ Classical geography already investigated causal links in space [Loi, 1985]

→ [Claval, 1985] : beyond reductionist causality in systemic analysis

→ Systemogenesis introduced by [Durand-Dastes, 2003] focuses on dynamics and path-dependency

→ Towards a complex approach to causality ? [Morin, 1976]
Existing approaches in spatio-temporal causality

Transportation Networks and Territories

- Lagged correlations: [Levinson, 2008] London population and network connectivity; [Gargi Chaudhuri and Keith C Clarke, 2015] historical data in North Italy

Spatio-temporal correlations

- Matching method for traffic flows [Liu et al., 2011]
- Generalized granger causality in neuroscience [Ke et al., 2007]
- Spatio-temporal correlations in Computer Vision [Ke et al., 2007]
Research objective:

→ Genericity and operationality of existing approaches?
→ Grasp complexity in a simple way?
→ At the interface of knowledge domains (methodology, modeling and empirical) [Raimbault, 2017]

Research objective:
Explore a generic method based on patterns of spatio-temporal lagged correlations: notion of Granger causality; validation on synthetic data and application to a case study.
Method: Rationale

Spatio-temporal stochastic field

\[ X_1(\vec{x}, t) \]
\[ \vdots \]
\[ X_N(\vec{x}, t) \]

\begin{itemize}
  \item Realizations and Aggregation
\end{itemize}

Trajectories of spatial units

\[ x_{i,1,t}^{(k)} \]
\[ \vdots \]
\[ x_{i,N,t}^{(k)} \]

Lagged Correlations

\[ \rho_{j_1,j_2}(\tau) \]

Datamining: Causality Regimes
Method: Formalization

Correlation estimator $\hat{\rho}$ applying in time, space and repetitions, i.e. covariance is estimated by $\hat{\text{Cov}}[X, Y] = \hat{E}_{i,t,k} [XY] - \hat{E}_{i,t,k} [X] \hat{E}_{i,t,k} [Y]$

Lagged Correlation defined by

$$\rho_{\tau}[X_{j_1}, X_{j_2}] = \hat{\rho} \left[ X_{i,j_1,t-\tau}^{(k)}, X_{i,j_2,t}^{(k)} \right] \quad (1)$$

Patterns of $\text{argmax}_{\tau} \rho_{\tau}[X_{j_1}, X_{j_2}]$ or $\text{argmin}_{\tau} \rho_{\tau}[X_{j_1}, X_{j_2}]$ (assumed clearly defined: e.g. statistical significance, minimal value) capture the sense of causality between $j_1$ and $j_2$

→ Datamining on $\rho_{\tau}$ (possibly parametrized values as $\rho_{\tau}^{(\omega)}$) to understand causality patterns.
Synthetic urban configurations generated by an hybrid morphogenesis model from [Raimbault et al., 2014]
Values of $\rho_\tau$ for all couples of three explicative variables (density, distance to center, distance to roads), for 8 extreme parameter points.
Unveiling Endogenous causality regimes

Intensive exploration of model parameter space (1000 parameters points \( \times \) 100 repetitions) with OpenMole software [Reuillon et al., 2013]

Unsupervised classification (robust k-means) on \( \tau_{\text{min}}, \tau_{\text{max}} \) features: (Left) Derivative of clustering coefficient for number of clusters \( k \); (Right) PCA visualisation of classification for “optimal” \( k \)
Consistence and interpretation of regimes

Values of cluster centers in terms of $\rho_\tau$
Consistency and interpretation of regimes

Position of clusters in the parameter space $w_i$.
Successive projects for the Grand Paris new transportation infrastructure
Values of $\rho_\tau$ for the different projects (columns) and different variables (rows), with accessibility differentials
Implications
→ Lagged correlation patterns on real data to investigate “structuring effects” in complex systems
→ The operational concept of Causality Regimes introduces a novel way to look at co-evolution in models of simulation

Developments
→ Characterisation of spatio-temporal diffusion: testing the spatial diffusion of innovation and the evolutive urban theory [Pumain, 2010]
→ Optimal spatial scales for stationarity: link with GWR? [Brunsdon et al., 1998]
→ A method validated on synthetic data and showed operational on a real system

→ At the interface of knowledge domains: theory, modeling, empirical, methodological

→ At the interface of disciplines: spatial analysis, statistics, datamining

- Code, data and results available at https://github.com/JusteRaimbault/CityNetwork
- Acknowledgments: We thank the European Grid Infrastructure and its National Grid Initiatives (France-Grilles in particular) to give the technical support and the infrastructure.
Reserve Slides
Granger causality

Granger causality test based on VAR processes:

\[ X(t) = \sum_{0 \leq \tau \leq \tau_Y} b_\tau Y(t - \tau) \]

If there exists \( b_\tau \) such that \( |b_\tau| > 0 \) significantly, then \( Y \) Granger-causes \( X \).

We have then \( \rho_\tau(Y, X) > 0 \).
Morphogenesis (Oxford dictionary)

1. **Biology**: The origin and development of morphological characteristics
2. **Geology**: The formation of landforms or other structures.

**History of the notion**

→ Started significantly with embryology around 1930 [Abercrombie, 1977]
→ Turing’s 1952 paper [Turing, 1952], linked to the development of Cybernetics
→ first use in 1871, large peak in usage between 1907-1909, increase until 1990, decrease until today. *Scientific fashion?*
Defining Morphogenesis

**Meta-epistemological framework of imbricated notions:**
Self-organization ⊇ Morphogenesis ⊇ Autopoiesis ⊇ Life

**Properties:**
- Architecture links form and function
- Emergence strength [Bedau, 2002] increases with notion depth, as bifurcations [Thom, 1974]

**Definition of Morphogenesis:** *Emergence of the form and the function in a strongly coupled manner, producing an emergent architecture* [Doursat et al., 2012]
Causality Regimes in a model of co-evolution

Unsupervised learning on lagged correlations between local variables unveils a diversity of causality regimes in a model co-evolving urban form and network topology

→ Link between co-evolution regime and morphogenetic properties of the urban system

(Left) Lagged correlation profiles of cluster centers; (Right) Distribution of regimes
Application to South Africa: Stationarity scales

**Optimal estimation time window and spatial range for accessibility**

[Graphs showing the relationship between Tw and meanabcorr/signcorr for different values of d0]
Application to South Africa: Causality Patterns

Clear inversion of the sense of Granger causality suggests a structural segregation effect of the apartheid laws.
Application to France: Stationarity scales

Optimal estimation time window and spatial range for accessibility
Application to France: No significant correlation
Macroscopic co-evolution: Correlation Patterns

```
gridWeight=5e−04; nwThreshold=4.5
```

Graph showing the correlation patterns with different parameters:

- `var`: ClosenessAccessibility, PopAccessibility, PopCloseness
- `tau`: -6, -3, 0, 3, 6
- `rho`: 0.00, 0.25, 0.50, 0.75

The graph illustrates the correlation patterns with varying `tau` and `rho` values.
References I


Les effets structurants des infrastructures de transport. 

Une théorie géographique des villes. 

An applied knowledge framework to study complex systems. 
*ArXiv e-prints.*
