

大規模灌漑システムの最適制御に向けた VAR-LiNGAM による因果解析

Causal Analysis for Large-scale Irrigation Systems via VAR-LiNGAM

松本 健宏 (Takehiro Matsumoto)¹, 大野 正夫 (Masao Ono)², 長谷川 文夫 (Fumio Hasegawa)², 関野 智 (Satoshi Sekino)², 吉田 公亮 (Kimiaki Yoshida)², 水藤 寛 (Hiroshi Suito)³

¹ 東北大学理学研究科数学専攻 (Mathematical Institute, Tohoku University),

² 株式会社 IHI (IHI Corporation),

³ 東北大学材料科学高等研究所 (WPI-AIMR, Tohoku University)

e-mail : takehiro.matsumoto.q7@dc.tohoku.ac.jp

1 はじめに

現在, 多くの農業用灌漑システムではゲート操作などの水管理が経験則に基づいて行われているが, 気候変動の影響などによりデータに基づく挙動把握・分類の重要性が高まっている. こうした背景から, G-RIPS Sendai 2025 プログラムの株式会社 IHI 提供課題では, 農業用灌漑システムの効率的な水管理をテーマに, 問題の定式化と時系列データの処理及びその解析に取り組んだ. その過程で, 平滑化フィルタの工夫や, 各ゲート操作の間の因果解析などが必要であることが示された [1].

本研究では, 大規模農業用灌漑システムの挙動把握のため, 実測した水位および流量の時系列データに対して因果探索手法 VAR-LiNGAM [2] を適用する. さらに, 得られた因果グラフをクラスタリングすることで, データ駆動型の挙動分類手法を提案する. 本手法により, 水位および流量の時間変化の原因と紐づけて分類できるため, 原因に応じた効果的な水管理の実現が期待される.

2 VAR-LiNGAM

VAR-LiNGAM は, Hyvärinen ら (2010) によって提案された多変量時系列データに対する因果探索手法である. 多変量時系列データ $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_N(t))^T$ に対し, 即時的因果効果を表す B_0 と, 過去の時刻 $t - \tau$ ($\tau = 1, \dots, p$) からの遅延因果効果を表す B_τ を用いて $\mathbf{x}(t) = B_0 \mathbf{x}(t) + \sum_{\tau=1}^p B_\tau \mathbf{x}(t - \tau) + \boldsymbol{\varepsilon}(t)$ と表す. ここで $\boldsymbol{\varepsilon}(t)$ は誤差項であり, 非ガウス分布に従うと仮定する.

3 因果グラフの作成とクラスタリング

使用データは, ある灌漑システム内の 4 計測地点 c1, d1, d2, e1 で約半年間にわたって 10 分間隔で計測された水位および流量の時系列データである. c1, d1, d2 は上流から下流に並び, e1 は c1 から分岐した流路上に位置する.

まず, 欠損値に対して線形補間を適用した後, 平滑化を行った. また, 弱定常性を確保する目的で隣接する時刻の差分を求めた. 次に, 時系列データを 1 週間 ($N = 1008$) ごとに 1 点ずつずらして分割した. その後, 分割したデータごとに平均 0・分散 1 となるよう標準化を行い, 係数行列 B_0, B_τ を推定した. 次数は $p = 7$ とした. 最後に, 得られた係数行列の全要素を連結して 128 次元のベクトルとし, これらに対して k -平均法によるクラスタリングを行った.

4 結果

クラスタ数 3 でクラスタリングを行った結果を図 1 に示す. ここで, i から $i + 1007$ 番目までに区切られたデータが Cluster k に属する場合, i 番目を Cluster k の色で示している. クラスタリングの

結果として、Cluster 0 は変動が小さい期間、Cluster 1 は 1,2 回ほど大きな変動がある期間、Cluster 2 は計画的にゲート操作が行われることで大きな変動が連続する期間に対応していることがわかる。

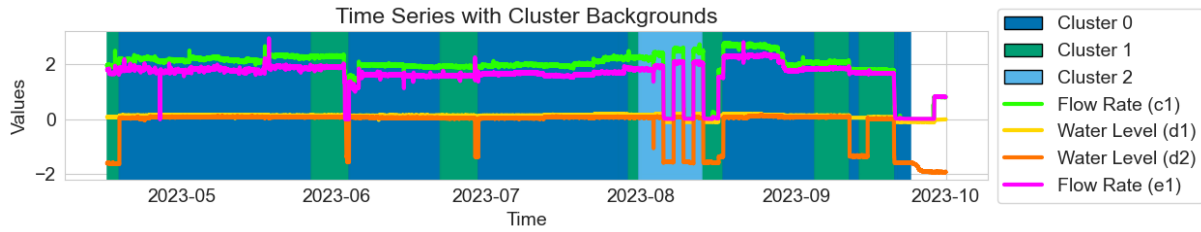


図 1: クラスタリング結果（クラスタ数 3）

次に、各クラスタの中心に対応する因果グラフを図 2 に示す。矢印の太さはエッジの重みに対応しており、時系列データの特徴と整合する因果構造が推定された。Cluster 0 は変動が少ない期間であり、各観測地点における時系列内の因果効果が支配的である。Cluster 1 は変動が大きい期間であり、Cluster 0 と比べて各観測地点における時系列内の因果効果は弱まっている。また、Cluster 2 は大きな変動が繰り返される期間であり、観測地点間の因果効果が最も強く現れている。

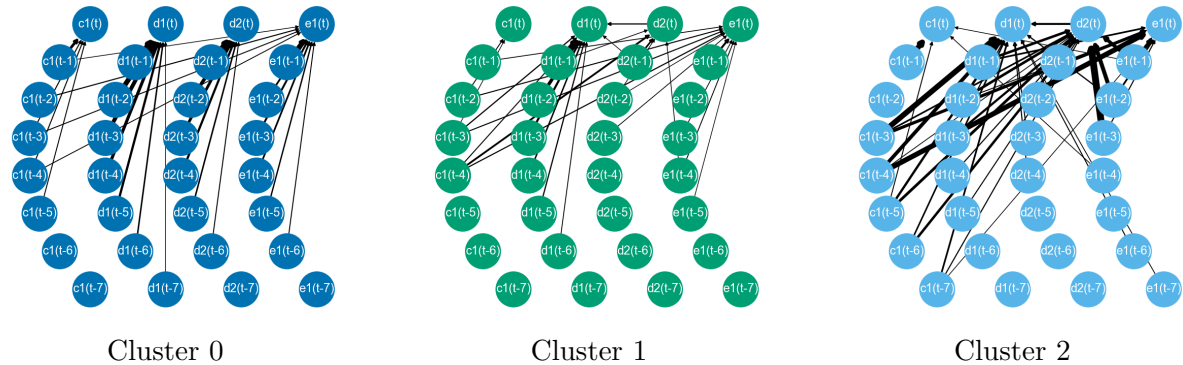


図 2: 各クラスタの中心に対応する因果グラフ

これらの結果から、因果構造を用いたクラスタリングにより灌漑システムの挙動を捉え、その制御のための有用な情報が得られることが示唆される。今後の課題は、分析に用いる期間やクラスタリング手法の最適化を図るとともに、得られた因果構造の分類方法およびその精度を向上させ、現場への実装に向けた検討を進めていくことである。

参考文献

- [1] Ishikawa, M., Matsumoto, T., Moen, K., Stockdill, A. and Tokonami, W., Resilient water management against global warming and for sustainable food supply, Final report of the IHI project, G-RIPS Sendai 2024, https://www.mccs.tohoku.ac.jp/g-rips/report/2024/pdf/ihi_final_report.pdf (2024) .
- [2] Hyvärinen, A., Zhang, K., Shimizu, S. and Hoyer, P. O., Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity, Journal of Machine Learning Research, 11 (2010), 1709–1731.

Penalized GSTAR Models for Indonesian Rainfall Forecasting

Arif Nurwahid¹

¹Graduate School of Science and Engineering, Ehime University
e-mail : m819004b@mails.cc.ehime-u.ac.jp

1 Introduction

Rainfall forecasting in Indonesia, an archipelagic nation of 38 provinces, is a critical challenge for agricultural planning and disaster mitigation. The complexity of spatial interactions and temporal dependencies makes this a difficult modeling task. The Generalized Space-Time Autoregressive (GSTAR) model is a common approach for such geospatial time series data. However, it can suffer from over-parameterization when applied to high-dimensional data, which reduces interpretability and increases the risk of overfitting.

To address these limitations, this study proposes a penalized regression approach within the GSTAR framework. Specifically, we integrate the Least Absolute Shrinkage and Selection Operator (LASSO) and Fused LASSO penalties to perform automatic spatial variable selection and identify regional clusters. The objective is to produce a model that is not only accurate in forecasting but also more parsimonious and interpretable.

2 Methodology

The Generalized Space-Time Autoregressive (GSTAR) model, a specialized form of the Space-Time Autoregressive (STAR) model, is used for this analysis. The model describes the observation at all N locations simultaneously, defined in its vectorized form as:

$$Z(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \Phi_{kl} W^{(l)} Z(t-k) + \varepsilon(t)$$

where $Z(t)$ is the $N \times 1$ vector of observations at time t , Φ_{kl} is the diagonal square matrix spatio-temporal parameter, $W^{(l)}$ is the $N \times N$ spatial weight matrix for lag l , and $\varepsilon(t)$ is the error term. For estimation, this GSTAR structure is rewritten in the standard linear model form suitable for Ordinary Least Squares (OLS), $Y = X\beta + \varepsilon$. Here, Y is a vector of all $Z(t)$ observations stacked over time, X is the design matrix constructed from the lagged spatio-temporal variables $W^{(l)} Z(t-k)$, and β is the vector of all parameters Φ_{kl} . Instead of standard OLS, we estimate β by minimizing a penalized objective function that incorporates both LASSO and Fused LASSO penalties:

$$\underset{\beta}{\operatorname{argmin}} \left(\|Y - X\beta\|_2^2 + \alpha_1 \|\beta\|_1 + \alpha_2 \sum_{(i,j) \in E} |\beta_i - \beta_j| \right)$$

The first term is the OLS sum of squared errors, while the penalty terms perform variable selection and encourage regional parameter similarity. This convex optimization problem is then reformulated and solved efficiently using Quadratic Programming (QP).

3 Expected Results and Discussion

Model evaluation will be based on forecasting accuracy (e.g., RMSE), comparing the penalized models against the standard GSTAR.

- **Forecasting Performance:** The penalized models are expected to show improved or comparable accuracy with a more parsimonious structure.
- **Model Interpretability:** We expect the LASSO penalty to yield a sparse parameter vector. This result can be visualized on a map of Indonesia to highlight the most significant inter-provincial dependencies. The Fused LASSO penalty is expected to group provinces into distinct zones with similar parameter values, which can also be visualized on a map.

4 Conclusion

We propose a penalized GSTAR framework for modeling Indonesian rainfall. This approach is expected to improve forecasting accuracy and provide a more interpretable model for understanding complex regional climate dynamics.

References

- [1] B. N. Ruchjana, The Generalized SpaceTime Autoregressive Order One Model and Its Application to Oil production Data, Unpublished Dissertation, Department of Mathematics, Institut Teknologi Bandung, (2002).
- [2] B. N. Ruchjana, S. A. Borovkova, H. P. Lopuhaä, Least Squares Estimation of Generalized Space Time AutoRegressive (GSTAR) Model and Its Properties, Journal of the Indonesian Mathematical Society, 13 (2007), 1–14.
- [3] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society: Series B (Methodological), 58 (1996), 267–288.
- [4] Y. Yan, H. Huang, M. G. Genton, Vector Autoregressive Models with Spatially Structured Coefficients for Time Series on a Spatial Grid, Journal of the American Statistical Association, 117 (2022), 860–873.

Kalman Filter Algorithm's role in SARIMAX model's MLE Optimization

Egide NTAKIRUTIMANA

Graduate School of Science and Engineering, Ehime University

e-mail : ntakireg@gmail.com

1 Introduction

Time series analysis has significantly contributed to the Industry and applied mathematics. By blending probability theory, statistics, and algebra, it enables and enhances the modeling and optimization efficiency. Industries utilize the field of mathematics to draw conclusions and provide reliable information to decision-makers. Kalman Filter (KF) Seasonal Autoregressive Integrated Moving Average with exogenous variables (KF-SARIMAX) model is utilized to forecast future events. Hence, the Kalman Filter algorithm optimizes SARIMAX models via Maximum Likelihood Estimation (MLE) regardless of the stationarity condition.

2 SARIMAX Model

It is well known that a stationary process Z_t with a continuous spectrum may be expressed in a linear form as:

$$z_t = \mu + \sum_{j=1}^{\infty} \psi_j z_{t-j} + a_t \quad (1)$$

where $a_t \sim WN(0,1)$ is a sequence of zero mean, homoskedastic, and uncorrelated random variables(r.v) and $\sum_{j=1}^{\infty} |\psi_j| < \infty$ [1].

Approximating the spectrum by rational functions leads to a parsimonious representation of Z_t as a finite order linear model in the class $ARMA(p, q)$:

$$\phi_p(B)z_t = \mu + \theta_q(B)a_t \quad (2)$$

where the autoregressive and moving average parts of the model satisfy, respectively, the stationarity and invertibility conditions [2].

Supporting nonstationary time series, ARMA(p,q) extends to ARIMA(p,d,q) given by:

$$\phi_p(B)(1-B)^d z_t = \mu + \theta_q(B)a_t \quad (3)$$

Along with seasonal component, three hyper-parameters are added to generate ARIMA(p, d, q) \times (P, D, Q)_s defined by:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D z_t = \mu + \theta_q(B)\Theta_Q(B^s)a_t \quad (4)$$

[3] Considering outer factors'impact on a time series,we need to import eXogenous or input variables.SARIMAX models stand for describing the functionality of input variables and are given by:

$$\begin{cases} z_t = \beta_t \gamma_t + w_t \\ \phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D w_t = \mu + \theta_q(B)\Theta_Q(B^s)a_t \end{cases} \quad (5)$$

[4].

The (5) can be seen as a special form of ARMAX(p+sP,q+sQ) given by:

$$\phi_{p+sP}^*(B)\tilde{z}_t = \beta_t\gamma_t + \theta_{q+sQ}^*(B)a_t \quad (6)$$

Where $\mu=0$ for easing calculations [5].

The state space model for Kalman Filter Algorithm components identification is:

$$\begin{cases} x_{t+1} = Fx_t + H\gamma_{t+1} + Ga_t \\ z_t = Ax_t + B\gamma_t + a_t \end{cases} \quad (7)$$

The likelihood function is:

$$f_{z|\gamma_t, z_{t-1}}(z_t|\gamma_t, z_{t-1}) = \prod_{t=1}^n \sqrt{\frac{1}{2\pi|\Sigma_t|}} \exp\left\{-\frac{1}{2} \sum_{t=1}^n (\epsilon_t^T) \Sigma_t^{-1} \epsilon_t\right\} \quad (8)$$

Assuming the Innovations in functions of parameter Θ , and ignoring the constant, the log-likelihood function is:

$$-logf_{z|\gamma_t, z_{t-1}}(z_t|\gamma_t, z_{t-1}) = \frac{1}{2} \sum_{t=1}^n \log|\Sigma_t| + \frac{1}{2} \sum_{t=1}^n \epsilon_t^T(\Theta) \Sigma_t^{-1}(\Theta) \epsilon_t(\Theta) \quad (9)$$

Thus, (9) gives the exact log-likelihood function regardless whether the moving average representation is invertible, merely, non-stationary process .

.....

References

- [1] Tsay et al., Nonlinearity tests for time series, Biometrika, 73, 1986, **--**.
- [2] R.H.Shumway and D.S.Stoffer, Time Series Analysis and Its Applications, Springer, 2006.
- [3] Nontapa et al., A new hybrid forecasting using decomposition method with SARIMAX model and artificial neural network, International Journal of Mathematics and Computer Science, 16. (2021), **--**.
- [4] Huang et al., Applied time series analysis and forecasting with Python, Springer, 2022.
- [5] Box et al., Time series analysis: forecasting and control, John Wiley & Sons, Vol. **, pp.379 **--**, 2015.

クーロン相互作用を有する格子気体モデルにおける 秩序形成と拡散特性

Diffusion and Ordering in a One-Dimensional Coulomb Lattice Gas

木原 開 (Kai Kihara)¹, 西成 活裕 (Katsuhiko Nishinari)¹,
安藤 康伸 (Yasunobu Ando)², 江崎 貴裕 (Takahiro Ezaki)³

¹ 東京大学工学系研究科航空宇宙工学専攻

(Dept. of Aero. Astro., Sch. of Eng., Univ. of Tokyo),

² 東京科学大学総合研究院化学生命科学研究所 (CLS, IIR, Science Tokyo),

³ 東京大学先端科学技術研究センター (RCAST, Univ. of Tokyo)

e-mail : kihara-kai@g.ecc.u-tokyo.ac.jp

1 概要

クーロン相互作用を伴う格子気体モデルは、固体電解質中のイオン拡散を表すモデルとして研究されてきた [1]。クーロン格子気体モデルでは、ポテンシャルに粒子が捕捉されることで拡散が著しく遅くなることがわかっている [2]。しかし、長距離相互作用の効果は解析が困難であることから、拡散抑制機構の理論的な説明は不十分である。そこで、本研究では、Euclidean string 理論 [3] を用いて粒子配置を粗視化し、拡散流量を近似的に導出する方法を提案する。

本研究では、周期境界をもつ一次元クーロン格子気体モデルを調べる。各サイト上の粒子は、隣接するサイトに粒子が存在しないとき、遷移率 $p = \exp[-\beta(E_0 + \Delta U_{\text{int}}/2)]$ で隣接サイトに移動する。ここで、 β は逆温度、 E_0 は活性化エネルギーであり、また ΔU_{int} は粒子の移動前後のポテンシャル変化である。粒子のポテンシャルは

$$U_{\text{int}} = K \sum_{i=1}^L \sum_{\substack{j=1 \\ j \neq i}}^L \left(\frac{1}{d_{ij}} + \frac{1}{d_{ji}} \right) \quad (1)$$

である。ここで、 K はクーロン力の強さである。また、 L はサイト数であり、 d_{ij} は i 番目のサイトから j 番目のサイトまでの距離である。拡散流量 J は、定常状態において、単位時間あたりに 1 サイトに粒子がホップする確率として定義する。

2 クーロンポテンシャル下での粒子の秩序構造

低温極限では、粒子が一様に分布した配置に収束する。これは、均等な配置において、式 (1) で示されるポテンシャル U_{int} が最も低くなるからである。例えば、粒子密度 $\rho = 2/5$ のときは $\circ \circ \bullet \circ \bullet$ (● は粒子、○ は空サイトを表す) となる配置パターンの繰り返しが基底状態となる。

本研究ではこの配置パターンを Euclidean string 理論を用いて表現する方法を提案する。Euclidean string $\mathcal{E}(q, r)$ とは、長さ $q (\geq 2)$ の数字列であり、 $r (\leq q)$ 個の 1 と $q - r$ 個の 0 を均等に分配したものである。また、均等な並べ方のうち 2 進数として最も小さいという制約から、Euclidean string は各 q, r に対し一意に決まる。例えば、 $\mathcal{E}(3, 1) = 001$, $\mathcal{E}(5, 2) = 00101$ である。これは、基底状態の配置パターンの条件と合致しており、密度 $\rho = r/q$ ($\gcd(q, r) = 1$) 配置パターン $\mathcal{U}(q, r)$ は、 $\mathcal{U}(1, 0) = 0$, $\mathcal{U}(1, 1) = 1$, $q \geq 2$ のときは $\mathcal{U}(q, r) = \mathcal{E}(q, r)$ として定義できる。

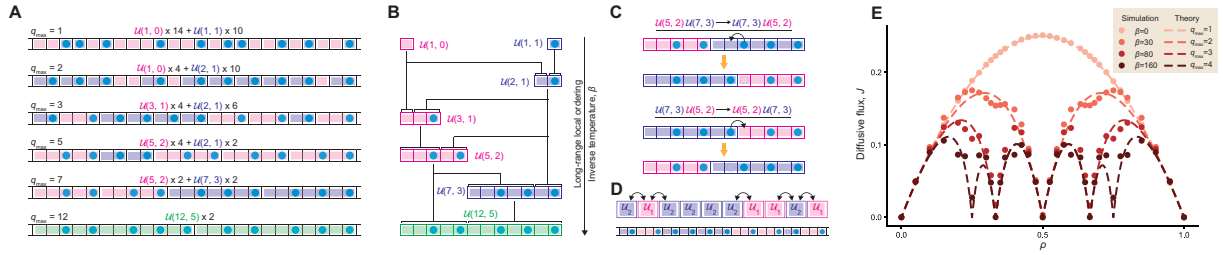


図 1. クーロン力による秩序化と粗視化近似.

3 有限温度における拡散係数の粗視化近似による導出

低温極限では Euclidean string により特徴づけられる配置パターン $\mathcal{U}(q, r)$ の繰り返しとなる秩序構造になると見なせることを説明したが、一方で有限温度では熱運動により秩序構造が崩れ、粒子配置は不規則になる。しかし、有限温度下でもクーロン力は限られた範囲内で十分な影響を及ぼすため、有効距離内の粒子間で部分的に秩序化が起きると考えられる。我々は、逆温度 β に対応した有効距離 q_{\max} 以下の粒子群が秩序構造を形成し、2 種類の配置パターンの組み合わせで書けるという仮説を立てた。配置パターン $\mathcal{U}_1 = \mathcal{U}(q_1, r_1)$, $\mathcal{U}_2 = \mathcal{U}(q_2, r_2)$ は以下の条件より一意的に選ばれる。

- 1) r_1/q_1 , r_2/q_2 が分母 q_{\max} 以下の既約分数列 (ファレイ数列, $F_{q_{\max}}$) で隣り合う。
- 2) 粒子密度 ρ が $r_1/q_1 < \rho < r_2/q_2$ を満たす。

より直観的には、クーロン力によって系の密度ゆらぎは条件 2) の範囲に制限されており、 q_{\max} が大きくなる、すなわち温度が低くなるほど制限が厳しくなる (図 1A)。 \mathcal{U}_1 と \mathcal{U}_2 は、Euclidean string の性質 $\mathcal{U}(q_1 + q_2, r_1 + r_2) = \mathcal{U}(q_1, r_1)\mathcal{U}(q_2, r_2)$ を用いて元の配置パターンを分解することで得られる (図 1B)。

粒子配置が \mathcal{U}_1 と \mathcal{U}_2 の組み合わせに制限されるとき、粒子の移動もその条件を破らない場合のみ許される。この場合、粒子の移動は \mathcal{U}_1 と \mathcal{U}_2 の入れ替えに対応する (図 1C)。この移動前後でのポテンシャル変化は比較的小さいため、移動が起こる確率を 1 と置くと、系はサイトを \mathcal{U}_1 と \mathcal{U}_2 に粗視化した対称単純排他過程 (Symmetric Simple Exclusion Process: SSEP) に近似できる (図 1D)。

以上の近似とモンテカルロシミュレーションの結果を図 1E に示す。近似解の傾向はシミュレーション結果と概ね一致した。また、粒子配置が非常に高い割合で \mathcal{U}_1 および \mathcal{U}_2 の組み合わせで表現できていることも確かめられた。

謝辞 本研究は科研費学術変革領域研究 (A) 「イオン渋滞学の構築に向けた数理・計算シミュレーション解析 (課題番号:24H02203)」の支援を受けました。

参考文献

- [1] K. Funke, Jump relaxation in solid electrolytes, *Progress in Solid State Chemistry*, 22 (1993).
- [2] A. Bunde, D. K. Chaturvedi, and W. Dieterich, Transport in charged lattice-gases, *Solid State Ionics*, 9-10 (1983).
- [3] J. Ellis, et al., Euclidean strings, *Theoretical Computer Science*, 301 (2002).