

UM MÉTODO PARA DETECÇÃO DE CAUSALIDADE DE GRANGER COM SELEÇÃO DE REGRESSORES

RICARDO E. V. VARGAS*, CELSO J. MUNARO*, PATRICK M. CIARELLI*

**Departamento de Engenharia Elétrica, Universidade Federal do Espírito Santo
Av. Fernando Ferrari, 514, Goiabeiras, Vitória - ES - Brasil, CEP: 29060-370, +55 27 4009-2687
E-mails: ricardovvargas@gmail.com, munaro@ele.ufes.br,
patrick.ciarelli@ufes.br*

Abstract— Causal detection methods have been applied in industrial processes data in order to diagnose disturbances that propagate plant-wide. A parametric method based on the Granger causality is proposed in this paper, adding improvements to the estimation of model residuals and procedures for the selection of regressors used to estimate these models. The method is compared to the classic method of Granger using data of an industrial process with known causality relationships, using nine different metrics. The results allow us to define the most appropriate causality detection methodology for the different metrics, and they indicate new perspectives for improvement of these methods.

Keywords— Granger causality, GCCA Toolbox, LASSO, regressor selection, cross-validation, Friedman test.

Resumo— Métodos de detecção de causalidade têm sido aplicados a dados de processos industriais com objetivo de diagnosticar distúrbios que se propagam por eles. Um método paramétrico baseado em causalidade de Granger é proposto neste artigo, agregando melhorias no método de estimação dos resíduos dos modelos e nos procedimentos para seleção dos regressores usados para estimar esses modelos. O método proposto é comparado ao método clássico de Granger utilizando dados de um processo industrial real com relações de causalidade conhecidas, usando para isso nove diferentes métricas. Os resultados permitem definir a metodologia de detecção de causalidade mais adequada para as diferentes métricas, além de indicar novas perspectivas para melhorias desses métodos.

Palavras-chave— causalidade de Granger, GCCA Toolbox, LASSO, seleção de regressores, validação cruzada, teste de Friedman.

1 Introdução

Muitas contribuições têm sido propostas para métodos baseados em dados para monitoramento de processos industriais, pois há grande disponibilidade de dados e existe uma dificuldade em modelar processos cada vez mais complexos e interconectados (Aldrich e Auret, 2013).

A detecção de relações de dependência entre variáveis de um processo contribui para diagnosticar problemas relacionados a busca da causa raiz de distúrbios que se propagam por ele, bem como o diagnóstico de falhas. Essas relações podem ser obtidas manualmente, a partir do conhecimento das equipes de operação, ou por meio dos dados. A segunda tem sido preferida, por ser automática, permitir a verificação periódica das relações e eliminar o esforço das equipes multidisciplinares que devem interagir para gerar essa informação. Métodos de detecção de causalidade têm sido utilizados com a finalidade de automatizar o processo e de reduzir a interferência humana (Bauer et al., 2007).

O requisito de estacionariedade dos dados utilizados pode representar uma restrição na presença de distúrbios. A diferenciação dos dados permite atender esse requisito, porém seu efeito de um filtro passa altas pode dificultar a obtenção de bons modelos para os métodos paramétricos de detecção de causalidade. Esse é o caso do método de detecção de causalidade de Granger (Barnett e Seth, 2014), utilizado neste trabalho. A estacionariedade é necessária para os testes de significância estatística da causalidade, que

também são afetados pelas escolhas das ordens dos modelos e pelos métodos para estimar os resíduos dos modelos.

Apesar do método de Granger ser muito utilizado para detecção de causalidades, ele impõe algumas restrições que podem ser prejudiciais para a identificação de certas causalidades, principalmente em dados do mundo real. Um exemplo dessas restrições é uso desnecessário de regressores associados a tempos mortos.

Este artigo tem como objetivo avaliar o impacto dessas imposições do método de Granger e, a partir das análises, propor um método que agrega as melhorias observadas e foi batizado como *Granger Causality with Regressor Selection* (GCRS) ou Causalidade de Granger com Seleção de Regressores.

2 Causalidade de Granger

O método de detecção de causalidade de Granger (Granger, 1969) se baseia fundamentalmente em análise de erro de modelagem linear. Originalmente ele foi utilizado para estimar causalidades entre pares de séries temporais, sendo posteriormente adaptado para análises com múltiplas séries.

A existência ou não de causalidade se baseia na análise da variância do erro do modelo completo que considera todas as séries comparada à variância do erro do modelo restrito sem a série cuja causalidade se deseja analisar. Uma maior variância no erro do modelo restrito permite concluir que a variável excluída contribui para explicar a série estimada pelo

modelo completo. Ou seja, nesse caso conclui-se que existe uma relação de causalidade.

Uma condição para utilização desse método é a necessidade das séries serem estacionárias. É possível verificar significância estatística apenas quando essa condição é respeitada.

No método proposto são feitas adaptações em basicamente três características do método de Granger. Tais características são descritas nas subseções a seguir.

2.1 Estimação do Erro de Modelagem com Resubstituição

Na abordagem clássica, o erro de modelagem é estimado utilizando-se as mesmas observações que são utilizadas para a obtenção do modelo, seja ele completo ou restrito. Esse método, chamado de resubstituição, se destaca por ser um tipo de estimativa tendenciosa, otimista (menor que o erro verdadeiro) e que piora com o aumento da ordem (grau de liberdade) do modelo. Destaca-se também que para cada relação são obtidas estimativas únicas de variâncias de erros, tanto para o modelo completo quanto para o modelo restrito. Assim, os valores de causalidade sofrem com baixa confiabilidade estatística (James et al., 2013).

Para ilustrar esse efeito, foi utilizado um modelo de três variáveis x , y , z da literatura (Ding et al., 2006). É conhecido que há apenas duas relações de causalidade nesse modelo: z causa x e y causa z . Para inferir a causalidade entre as variáveis z e x foram obtidos modelos e calculadas as variâncias dos erros com resubstituição e com validação cruzada (Friedman et al., 2009). Essa demonstração foi obtida com implementações próprias do método clássico de Granger (com erro de resubstituição) e do método de Granger ajustado para estimar erro de modelagem com validação cruzada. Na Figura 1, os asteriscos em vermelho representam os resultados com resubstituição e as barras em azul são os *boxplots* do resultado com validação cruzada. Na Figura 2 são mostrados valores da causalidade de z em x (existente) estimando o erro pelos dois métodos. O mesmo é feito na Figura 3 para a causalidade de x em z (inexistente). Os limites estatísticos apresentados nas Figuras 2 e 3 são de F -test com distribuição teórica assintótica e 5% de significância. Em relação a essas figuras é válido destacar:

- Para cada ordem, o erro de validação cruzada foi obtido com 10-fold, sendo que a estimativa final (pontos destacados no *boxplot*) é a média aritmética simples entre os resultados obtidos nas 10 estimativas;
- O erro de resubstituição e a sua variância são monótonas decrescente (ver Figura 1). Por isso o método clássico utiliza algum critério baseado em informação para escolher a ordem dos modelos, por exemplo: BIC (*Bayesian information criterion*) ou AIC (*Akaike information criterion*). Esses critérios são mais indicados quando a

quantidade de observações é muito restrita a ponto de não ser interessante subdividi-las para aplicar validação cruzada (James et al., 2013);

- A dispersão da estimativa, tanto da variância de erro quanto da causalidade, é crescente com o aumento da ordem, como pode ser observada nos *boxplots* das Figuras 1 a 3;
- Espera-se que valores de causalidade diminuam com elevação da ordem do modelo, a ponto de se tornarem não-significativos. Isso acontece porque ao se elevar a ordem do modelo a série causa passa a contribuir relativamente cada vez menos no modelo completo da série efeito. Isso não é observado quando se utiliza erro de resubstituição;
- Uma consequência da observação anterior é que ao se utilizar erro de resubstituição, valores de causalidades inexistentes ficam mais próximos do limite estatístico (ver Figura 3).

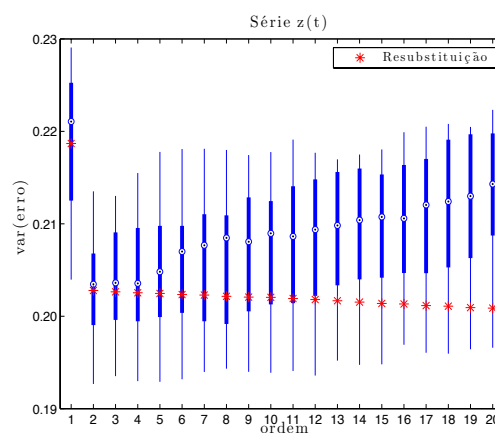


Figura 1. Variância do erro do modelo completo versus ordem.

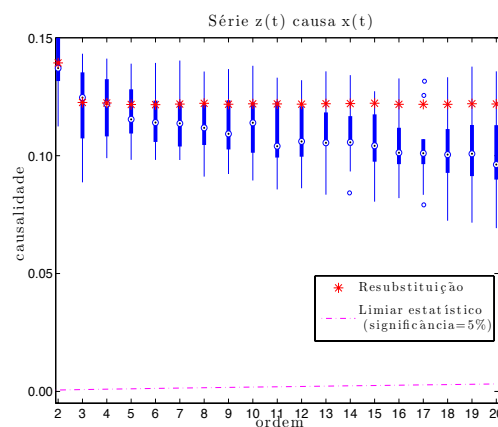


Figura 2. Causalidade existente versus ordem.

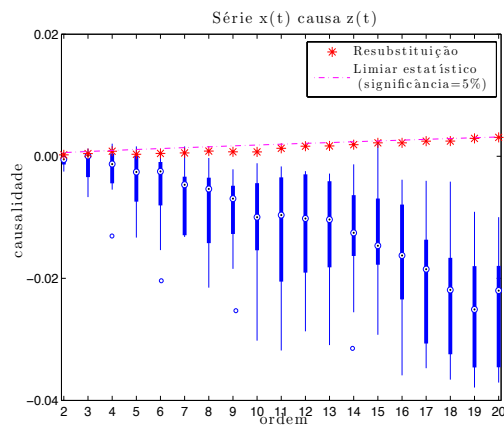


Figura 3. Causalidade inexistente versus ordem.

2.2 Utiliza  o de Ordem  nica para Todas as S ries

No m todo de Granger, utiliza-se uma  nica ordem para todos os modelos. Assim, quando certa s rie   melhor modelada em termos de vari ncia de erro com ordens menores que   escolhida, a abordagem cl ssica utiliza graus de liberdade desnecess rios nos modelos. Em (Thornton e Batten, 1984),   mostrado como causalidade de Granger   sens vel em rela  o   escolha da ordem dos modelos.

2.3 Utiliza  o de Todos os Regressores at  a Ordem Estimada

Quando uma ordem p   escolhida, a abordagem cl ssica inclui em cada modelo completo e restrito todos os regressores de todas as s ries, sem considerar se eles contribuem com a redu  o da vari ncia do erro de modelagem.

3 Granger Causality with Regressor Selection (GCRS)

O principal pilar das propostas contidas no m todo GCRS   o *Bias-Variance Tradeoff* (Friedman et al., 2009). O objetivo   respeitar a complexidade (quantidade de graus de liberdade) de cada s rie, mas utilizar no seu modelo completo apenas os regressores que minimizam a vari ncia do erro de modelagem.

Essas propostas foram elaboradas visualizando-se causalidade de Granger sob a  tica de aprendizado estat stico: regress o linear e sele  o de regressores. O m todo GCRS estima as rela  es de causalidade de cada s rie, uma de cada vez, em at  tr s fases. A 3  fase   obrigat ria e uma das duas primeiras fases pode ser suprimida a crit rio do usu rio. Ou seja, pode-se utilizar as tr s fases, a 1  e a 3  fases ou a 2  e a 3  fases. Cada fase   descrita em subse  o pr pria a seguir.

3.1 Primeira Fase do GCRS

Seu principal prop sito   selecionar as quantidades iniciais de regressores de cada s rie causa (entradas do modelo) mais adequadas para o modelo completo inicial de cada s rie efeito (sa da do modelo). S o quantidades iniciais porque se a 2  fase for aplicada, elas podem ser reduzidas.

Nesta fase as regress es lineares s o feitas com M nimos Quadrados Ordin rios (MQO), por conta do seu baixo custo computacional. Para cada s rie efeito, s o realizadas regress es lineares com todas as poss veis combina  es de n mero de regressores associados a cada s rie causa, variando entre zero e um valor m ximo escolhido pelo usu rio. Nesse sentido, o caminho de busca   praticamente completo. Apenas o modelo com zero regressores n o   considerado. Ao final desta fase,   escolhido como modelo completo inicial aquele composto pelo n mero de regressores de cada s rie causa que gera a menor vari ncia do erro de modelagem. Convencionou-se assumir que a ordem de cada modelo completo   igual ao maior n mero selecionado de regressores.

  importante destacar que n o   utilizado um caminho de busca completo em termos dos regressores em si. Isso porque para qualquer n mero p de regressores associados a qualquer s rie causa, considera-se sempre o conjunto de regressores entre 1 e p . Outros poss veis conjuntos, por exemplo entre 2 e $p+1$, n o s o avaliados. Essa abordagem   similar   adotada no m todo cl ssico e diminui o tempo de processamento durante a busca.

Mesmo no caso espec fico de s ries temporais, existem v rias formas de se realizar valida  o cruzada e ainda n o h  consenso sobre qual   o melhor m todo (Barrow, 2013; Arlot e Celisse, 2010). Em fun  o da experi ncia dos autores deste artigo, adotou-se a seguinte abordagem. S o formadas 10 amostras (grupos) de observa  es cont guas, com tamanhos iguais, disjuntas e em sequ ncia. Dessa forma, as poss veis autocorrela  es s o preservadas. Durante a i - sima valida  o, utiliza-se a i - sima amostra no teste (avalia  o) e uma  nica outra amostra aleat ria diferente de i no treinamento (modelagem). Assim, cada observa  o   utilizada apenas uma  nica vez em treinamento e uma  nica vez em teste. Considera-se como vari ncia do erro de modelagem a m dia aritm tica simples entre as suas 10 estimativas.

Portanto, logo na 1  fase do m todo GCRS j  s o propostas alternativas  s tr s caracter sticas descritas nas Subse  es 2.1, 2.2 e 2.3.

3.2 Segunda Fase do GCRS

O prop sito principal da 2  fase   reduzir o grau de liberdade dos modelos completos por meio de sele  o de regressores.

Nesta fase   utilizado o m todo LASSO (*Least Absolute Shrinkage and Selection Operator*), por ser um eficiente seletor de regressores (Friedman et al., 2009; Hsu et al., 2008; Chiuso e Pilonetto, 2012). Para cada s rie efeito, s o selecionados os regresso-

res que geram ou o menor MSE (*mean squared error* – erro quadr tico m dio) ou menor 1SE (*mean squared error plus its standard error* – erro quadr tico m dio mais seu erro padr o), a crit rio do usu rio. A primeira op  o tende a zerar menos coeficientes, gerando um maior n mero de regressores.

  importante destacar que o m todo LASSO pode selecionar qualquer subconjunto de regressores, inclusive com regressores n o-sequenciais e/ou que n o se inicia a partir do primeiro regressor. Com essa caracter stica, certos regressores n o s o selecionados para modelos completos associados a s ries que apresentam certo tempo morto.

A mesma estrat gia de valida  o cruzada descrita anteriormente   utilizada para estimar as vari ncias de erros de modelagem nesta fase.

  importante observar que a 2  fase do m todo GCRS tamb m modifica as tr s caracter sticas descritas nas Subse  es 2.1, 2.2 e 2.3.

3.3 Terceira Fase GCRS

Esta fase visa estimar os valores de causalidade, sendo assim a  nica fase obrigat ria do m todo GCRS.

Os c lculos s o da mesma forma que o m todo de Granger. Portanto, qualquer teste estat stico aplic vel   causalidade de Granger tamb m pode ser utilizado com o m todo GCRS.

A principal inova  o desta fase decorre da possibilidade de todos os regressores associados a uma s rie causa serem descartados do modelo completo de uma s rie efeito, o que resulta na aus ncia de causalidade dessa rela  o. Em outras palavras, se nenhum regressor associado a uma s rie causa for selecionado   porque a rela  o causa-feito n o existe.

4 Estudo de Caso

Para estudo de caso foram utilizadas 6 vari veis de processo (PIC400, LIC434, FIC408, LIC400, LIC430 e TIC400) obtidas em um planta real de automa  o industrial (Marques et al., 2015). As equipes de opera  o dessa planta comprovaram a exist ncia de 7 rela  es de causalidade entre as vari veis.

4.1 Metodologia

O m todo GCRS foi implementado em MATLAB vers o R2013a e comparado neste trabalho com o GCCA *Toolbox*, que implementa causalidade de Granger e   gratuito (<http://www.sussex.ac.uk/sackler/mvge>).

Para o teste estat stico aplicado em valores de causalidade foi implementada e utilizada a op  o padr o do GCCA *Toolbox*: *F-test* com distribui  o te rica assint tica, 5% de signific ncia e corre  o por FDR (Barnett e Seth, 2014). Como no GCCA a ordem   a mesma para todas as s ries e no GCRS isso n o necessariamente acontece, a implementa  o original

desse teste precisou ser adaptada, sem altera  o na teoria envolvida. No GCRS os graus de liberdade do modelo completo e dos modelos restritos s o calculados para cada par de s ries. J  no GCCA tais graus s o calculados uma  nica vez, pois s o fixos em todos os pares de s ries.

Foram avaliados os 6 m todos descritos na Tabela 1, sendo que 5 s o variantes do m todo GCRS. A proposta foi investigar a import ncia e as caracter sticas das duas primeiras fases isoladas ou juntas.

Tabela 1: M todos avaliados.

M�todo	Descri��o
GCCA	M�todo cl�ssico.
GCRS (Fase 1)	Seleciona regressores apenas com a 1� fase.
GCRS (1SE - Fase 2)	Seleciona regressores apenas com a 2� fase com o crit�rio 1SE.
GCRS (1SE - 2 Fases)	Seleciona regressores tanto com a 1� quanto com a 2� fase com o crit�rio 1SE.
GCRS (MSE - Fase 2)	Seleciona regressores apenas com a 2� fase com o crit�rio MSE.
GCRS (MSE - 2 Fases)	Seleciona regressores tanto com a 1� quanto com a 2� fase com o crit�rio MSE.

Ao todo, foram realizadas 20 rodadas de estimativas de causalidades. Em cada rodada foram utilizadas 400 observa  es distintas de cada vari vel de processo e foram obtidas estimativas de causalidade para cada m todo. Cada m todo recebeu como entrada rigorosamente as mesmas observa  es a cada rodada. Para garantir estacionariedade, foram utilizadas apenas observa  es de per odos nos quais a planta estava em regime estacion rio. Al m disso, as observa  es foram diferenciadas uma  nica vez.

Para o GCCA, sempre foi utilizada a maior ordem entre AIC e BIC, estimadas pelo pr prio GCCA *Toolbox*. Para o m todo GCRS, sempre foi utilizada como limite para busca de graus de liberdade a mesma ordem utilizada pelo GCCA.

4.2 M tricas

Em cada rodada, foram calculadas as m tricas da Tabela 2 em rela  o  s estimativas de cada m todo avaliado.

Tabela 2: M tricas cl ssicas utilizadas.

M�trica	Descri��o
VP	Verdadeiro-positivo: percentual de causalidades existentes estimadas como existentes.
FN	Falso-negativo: percentual de causalidades existentes estimadas como inexistentes.
VN	Verdadeiro-negativo: percentual de

	causalidades inexistentes estimadas como inexistentes.
FP	Falso-positivo: percentual de causalidades inexistentes estimadas como existentes.

Foram calculadas também as seguintes métricas (Siyal et al., 2014):

- Precisão (P): proporção entre causalidades existentes estimadas como existentes e todas as estimadas como existentes. Ou seja, entre todas que foram estimadas como existentes, percentual que realmente são existentes (Equação 1);
- Sensibilidade (S): proporção entre causalidades existentes estimadas como existentes e todas as causalidades existentes. Ou seja, entre todas que realmente são existentes, percentual estimadas como existentes (Equação 2);
- Acurácia (A): percentual de causalidades estimadas corretamente, sejam existentes ou inexistentes (Equação 3);
- Medida F (F1): média harmônica entre precisão e sensibilidade (Equação 4).

$$P = \frac{VP}{VP+FP} \quad (1)$$

$$S = \frac{VP}{VP+FN} \quad (2)$$

$$A = \frac{VP+VN}{VP+VN+FP+FN} \quad (3)$$

$$F1 = \frac{2PS}{P+S} \quad (4)$$

Também foram registrados os tempos de execução (T) em um Intel Core i5 2.7 GHz, 8 GB de RAM DDR3 1867 MHz e OS X 10.11.1.

Tabela 3. Resultados do estudo de caso.

MÉTODO	VP [%]	FN [%]	VN [%]	FP [%]	P [%]	S [%]	A [%]	F1 [%]	T [s]
GCCA	50.0	50.0	84.8	15.2	77.2	50.0	67.4	60.2	0.2
GCRS	Fase 1	45.0	55.0	78.9	21.1	67.6	45.0	62.0	53.4
	Fase 2	28.6	71.4	94.1	5.9	87.0	28.6	61.4	41.9
	2 Fases	30.0	70.0	95.0	5.0	87.0	30.0	62.5	43.4
	Fase 2	42.9	57.1	81.1	18.9	69.9	42.9	62.0	51.8
	2 Fases	38.6	61.4	83.9	16.1	71.1	38.6	61.2	49.4

5 Conclusões

O objetivo do trabalho foi desenvolver uma alternativa ao método causalidade de Granger com maior confiabilidade estatística.

Um método batizado como GCRS foi elaborado pensando-se fundamentalmente no *Bias-Variance Tradeoff*. Nele são adaptadas três características do método clássico: estimação do erro de modelagem com resubstituição, utilização de ordem única para todas

4.3 Teste Estatístico Aplicado nas Métricas

Para cada métrica, os valores associados a cada método foram avaliados com o teste estatístico de comparação múltipla de Friedman (Hollander e Wolfe, 1999). Utilizou-se 5% de significância e correção de Tukey-Kramer (Hochberg e Tamhane, 1987).

4.4 Resultados

As médias de todas as métricas e dos tempos de execução são apresentados na Tabela 3. Em relação a essa tabela, é válido destacar que:

- O símbolo ★ indica métricas estatisticamente diferentes e para melhor em relação ao obtido pelo GCCA. Ou seja, eles indicam por métrica quais variantes do método GCRS venceram o GCCA. Já o símbolo ⊗ indicam por métrica quais variantes do método GCRS perderam para o GCCA. Ausência desses símbolos indica que houve empate com o GCCA;
- O método GCRS tende a obter melhores resultados para as métricas VN e FP, justamente por conta do GCCA utilizar erro de resubstituição conforme descrito na Subseção 2.1. Utilizando-se a opção MSE do LASSO, essas métricas melhoraram em relação à variante Fase 1, mas não o suficiente para vencerem estatisticamente os respectivos resultados do GCCA;
- Uma consequência negativa da execução da 2ª fase do método GCRS (LASSO) foi a piora nos resultados das métricas VP e FN. Como consequência, o GCRS não obteve bons resultados nas métricas S, A e F1;
- Em relação a tempo de execução, o GCCA venceu as três variantes do método GCRS que executam a sua 1ª fase.

as séries e utilização de todos os regressores até a ordem estimada.

Um estudo de caso com 6 variáveis de processo obtidas em planta real de automação industrial foi realizado para comparar o método GCRS com o GCCA, uma implementação gratuita do método clássico de causalidade de Granger. Mostrou-se que o GCRS tende a vencer o GCCA nas métricas VN e FP, justamente por conta das três adaptações realizadas. Isso pode significar ganhos relevantes na prática. Por exemplo, em análise de causa raiz de perturbações e de oscilações em malhas industriais de controle, VN

e FP melhores diminuem as chances de falsas causas raiz serem selecionadas. Ou seja, nesse tipo de aplicação, o GCRS tenderia a indicar com mais sucesso as causas raiz de fato, que poderiam então ser tratadas mais rapidamente e com menos custo operacional. Contudo, o GCRS tende a gerar piores resultados em relação ao GCCA para as outras métricas avaliadas.

Investigações adicionais se mostram necessárias. Por exemplos:

- Alternativas ao caminho de busca utilizado na 1ª fase do método para torná-la menos custosa em termos computacionais;
- Alternativas ao MQO e ao LASSO;
- Alternativa intermediária entre MSE e 1SE, que tenda a obter bons resultados tanto em termos de VN e FP quanto em termos de VP e FN;
- Estimar causalidades com algum critério de votação dentre os valores obtidos durante validação cruzada. Ou seja, aproveitar os múltiplos cálculos de variância de erro para serem obtidos também múltiplos valores de causalidade para cada par de séries.

Agradecimentos

Agradecemos à empresa Petróleo Brasileiro S.A. (Petrobras) por ter liberado parcialmente seu empregado (Ricardo E. V. Vargas) para o desenvolvimento deste trabalho.

Referências Bibliográficas

- Aldrich, C. e Auret., L. (2013). *Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods*, Springer.
- Arlot, S. e Celisse, A. (2010). A survey of cross validation procedures for model selection, *Statistics Surveys*, 4(0), pp. 40–79.
- Barnett, L. e Seth, A.K. (2014). The MVGC Multivariate Granger Causality Toolbox: A New Approach to Granger-causal Inference, *Journal of Neuroscience Methods*, 223, pp. 50–68.
- Barrow, D. K. (2013). Crogging (Cross-Validation Aggregation) for Forecasting - a novel algorithm of Neural Network Ensembles on Time Series Subsamples, *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8.
- Bauer, M., Cox, J. W., Caveness, M. H., Downs J. J. e Thornill N. F. (2007). *Finding the direction of disturbance propagation in a chemical process using transfer entropy*, IEEE Trans. Cont. Syst. Tech, vol. 15, no. 1.
- Chiuso, A. e Pilonetto, G. (2012). A Bayesian approach to sparse dynamic network identification, *Automatica*, 48(8), pp. 1553–1565.
- Ding, M; Chen, Y e Bressler, S. L., 2006. *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*. Wiley-VCH, 1, pp. 437–460.
- Friedman, J., Hastie, T. e Tibshirani, R. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition*. New York, NY: Springer-Verlag.
- Granger, C. W. J., August 1969. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, Vol. 37, No. 3, pp. 424–438.
- Hochberg, Y., e A. C. Tamhane. (1987). *Multiple Comparison Procedures*. Hoboken, NJ: John Wiley & Sons, Inc.
- Hollander, M., e D. A. Wolfe. (1999). *Nonparametric Statistical Methods*. Hoboken, NJ: John Wiley & Sons, Inc.
- Hsu, N.-J., Hung, H.-L. e Chang, Y.-M. (2008). Subset selection for vector autoregressive processes using lasso, *Computational Statistics & Data Analysis*, 52(7), pp. 3645–3657.
- James, G., Hastie, T. e Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. United States: Springer-Verlag New York.
- Marques, V. M., Munaro, C. J. e Shah, S. L. (2015). Detection of causal relationships based on residual analysis, *IEEE Transactions on Automation Science and Engineering*, 12(4), pp. 1525–1534.
- Siyal, M. Y., Furqan, M. S. e Monir, S. M. G. (2014). Granger causality: Comparative Analysis of Implementations for Gene Regulatory Networks, *13th International Conference on Control Automation Robotics & Vision (ICARCV)*, pp. 793–798.
- Thornton, D.L. e Batten, D.S. (1984). Lag-length Selection and Granger Causality, *Technical report, Federal Reserve Bank of St. Louis*.