Extreme Events Characterization on Time Series

Marcos Wander Rodrigues, Luis Enrique Zárate

Pontifical Catholic University of Minas Gerais, Brazil
marcoswanderrodrigues@gmail.com, zarate@pucminas.br

Abstract. The use of sensors in environments where they require constant monitoring has been increasing in recent years. The main goal is to guarantee the effectiveness, safety, and smooth functioning of the system. To identify the occurrence of abnormal events, we propose a methodology that aims to detect patterns that can lead to abrupt changes in the behavior of the sensor signals. To achieve this objective, we provide a strategy to characterize the time series, and we use a clustering technique to analyze the temporal evolution of the sensor system. To validate our methodology, we propose the clusters' stability index by windowing. Also, we have developed a parameterizable time series generator, which allows us to represent different operational scenarios for a sensor system where extreme anomalies may arise.

CCS Concepts: • **Applied computing**;

Keywords: Anomaly Detection, Characterization of Time Series, Cluster Analysis, Extreme Events, Time Series Analysis

1. INTRODUÇÃO

Ambientes industriais críticos como barragens, sistemas de alto forno, sistemas de fornecimento de energia, gasodutos, e outros, possuem monitoramento contínuo baseado em sensores para garantir a integridade e a segurança, evitando desastres com vítimas, prejuízos econômicos e/ou ambientais.

De acordo com os dados temporais provenientes de sensores, o monitoramento pode apontar três possíveis cenários: a) predomínio da normalidade com algum nível de tendência; b) Variações abruptas que devem ser observadas e analisadas; e c) ocorrência de eventos críticos anômalos ou extremos, os quais devem receber maior atenção.

Para sistemas de monitoramento por sensores, a análise por séries temporais é a modelagem ideal para compreender os padrões de comportamento, prever anomalias e níveis de criticidade. Detectar e caracterizar esses níveis, ainda que não sejam eventos raros (com mínima probabilidade de ocorrer e com maior impacto) [Taleb 2007], permitiria a construção de sistemas de alerta mais eficientes.

A partir dos registros dos sistemas de sensores é possível caracterizar as séries temporais extraindo atributos como valor máximo, mínimo, média, tendência, desvio padrão, além de outras características que podem representar uma série temporal [Lee et al. 2014]. Após caracterização por janelas temporais, é possível aplicar técnicas de agrupamento para reconhecer e prever padrões que melhor descrevam a evolução temporal de determinados eventos.

Em sistemas reais de ambientes críticos, dados de monitoramento, provenientes dos sensores, não são normalmente disponibilizados. Por esse motivo, e para este trabalho optamos por construir um gerador de sinais (séries temporais), o qual permite a geração de bases de dados temporais sintéticas para representar diferentes condições de operação de sensores.

Neste artigo, são extraídas características do conjunto de séries temporais (sinais de sensores) gera-

À CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Código 001, CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico e FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais.

dos sinteticamente, para posterior aplicação de técnicas de agrupamento, a fim de caracterizar anomalias que possam ocorrer de modo a relacioná-las ao surgimento de eventos extremos. Este trabalho, faz parte de um projeto maior que procura observar os movimentos por janelamento, nas respostas dos sensores, extraindo padrões dessas mudanças operacionais.

A contribuição principal deste trabalho é a proposta de um procedimento para analisar a evolução temporal de um sistema de sensores envolvendo condições normais sujeitas à variações abruptas e à variações anômalas de operação.

O artigo possui a seguinte organização: A Seção 2 descreve brevemente os trabalhos relacionados. Na Seção 3 a fundamentação teórica que sustenta nosso trabalho é descrita. Na Seção 4 a metodologia adotada e uma breve descrição do gerador de séries temporais é descrito. Na Seção 5 os experimentos e análise dos resultados são apresentados. Finalmente, a Seção 6 conclui e expõe os trabalhos futuros.

2. TRABALHOS RELACIONADOS

A análise de séries temporais tem um papel fundamental na predição de novos eventos dado os eventos que ocorreram no passado. Seja no contexto uni-variado ou multivariado, é possível agrupar e classificar as séries temporais a fim de realizar predições em cenários dinâmicos devido ao fator temporal.

Os diversos contextos que utilizam as séries temporais demandam diferentes estratégias de análises. O trabalho de [Yanfei Kang 2018] propõe o uso de modelos *mixture autoregressive* para identificar a similaridade entre séries temporais. Em [Abdullah 2014], os autores usam as representações como a transformada discreta de *Fourier*, medidas de similaridade por correlação, e a transformada discreta de *wavelets* para caracterizar a segmentação de séries temporais em diversas aplicações de *Motifs*.

Considerando a ocorrência de valores extremos, o trabalho de [Chavez-Demoulin and Davison 2012] utiliza da análise uni-variada em séries temporais para identificar o surgimento de anomalias nos valores da série. Enquanto que na análise multivariada, o trabalho de [Ranjan et al. 2018] busca identificar um conjunto específico de variáveis que podem ser responsáveis na identificação de um possível evento extremo. O objetivo dos autores é prever o surgimento desses eventos por meio de modelos supervisionados de *machine-learning* (ML).

Algoritmos de agrupamento são técnicas de ML utilizadas para investigar o comportamento de uma série temporal específica dentro dos *clusters* [Nakkeeran et al. 2012]. Assim, os autores buscam extrair informação relevante destes clusters a fim de melhorar a eficácia de sistemas decisórios. O trabalho de [Serra and Zárate 2015] também utiliza as técnicas de agrupamento para tentar descrever o comportamento de séries temporais em um conjunto de *clusters*, bem como o surgimento de novos *clusters*, a partir de características como a componente de nível e tendência extraídas da série.

A provável ocorrência de eventos com valores anômalos na série temporal, torna o seu tratamento e análise inadequados por meio da distribuição gaussiana. Sendo, assim, necessário o uso de métodos estatísticos como a teoria de valores extremos (TVE) para lidar com tais anomalias [Caires 2011].

3. FUNDAMENTAÇÃO TEÓRICA

3.1 Séries Temporais

Uma série temporal (ST) é uma sequência temporal de dados discretos, e um conjunto de STs compõe uma base de dados temporais (TDB). As STs regulares são geradas em intervalos de tempo regularmente espaçados, já as STs irregulares são geradas em intervalos de tempo variados. A TDB gerada para este trabalho é multivariada e formada por sinais de sensores (STs regulares) expressos pela

matriz da Equação 1.

$$[Z] = \begin{bmatrix} Z_{t^{11}} & Z_{t^{12}} & \cdots & Z_{t^{1M}} \\ Z_{t^{21}} & Z_{t^{22}} & \cdots & Z_{t^{2M}} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{t^{N1}} & Z_{t^{N2}} & \cdots & Z_{t^{NM}} \end{bmatrix}_{N \times M,} \tag{1}$$

onde cada elemento $Z_{t^{ij}} = \{Z_{1^{ij}}, Z_{2^{ij}}, \cdots, Z_{T^{ij}}\}$, representa uma série temporal gerada pelo sensor (i,j) para $i = \{1, \cdots, N\}$ e $j = \{1, \cdots, M\}$, onde N representa o número de *arrays* de sensores, e M o número de sensores por *array*. O elemento $Z_{t^{ij}}$ corresponde a uma observação para o sensor j no *array* i, e cada valor de $t = \{1, \cdots, T\}$ corresponde aos períodos de observação na série. O número total de observações de $|Z|$ na TDB é dado por $|Z| = N \times M \times T$.

Uma ST é representada por modelos que podem variar de acordo com os objetivos da análise, podendo ser modelos não-paramétricos ou paramétricos. A representação deste último considera a forma geral de uma ST dada pela Equação $Z_t = f_t + a_t$, onde f_t é uma função do tempo somada ao ruído a_t. Neste artigo, usamos o modelo de suavização de *Holt-Winters* [Winters 1960], o qual utiliza as componentes de nível (μ_t), tendência (T_t), sazonalidade (F_t), e o ruído (a_t) associados a cada ST. Em STs geradas por um sistemas de sensores, a componente sazonalidade (F_t) pode não ser considerada, uma vez que o processo de caracterização das séries ocorrem em curtos períodos de tempo, ou seja, são definidas pelas amplitudes das janelas w_h. Assim, o modelo utilizado neste trabalho é definido pela Equação $Z_t = \mu_t + T_t + a_t$.

3.2 Janelamento e Caracterização de Séries Temporais

Neste trabalho, consideramos que cada ST, $(Z_{t^{ij}})$, é dividida em H janelas definidas como $W = \{w_1, w_2, \cdots, w_H\}$. Alguns autores têm discutido diversas estratégias para o janelamento de STs, como pode ser visto em [Huang et al. 2012].

O nosso objetivo é definir janelamentos apropriados para análise, ou seja, onde mudanças extremas podem ocorrer. Assim, o conhecimento acerca do sistema onde as STs são coletadas, nos permite conjecturar a periodicidade das variações, e consequentemente, a amplitude ideal do janelamento. Como critério de escolha da amplitude da janela w_h, devem ser observadas as séries que possuem menor variabilidade no tempo [Serra and Zárate 2015]. Para este trabalho consideramos janelas com tamanho fixo de 6 pontos discretos. As duas primeiras janelas w_1 e w_2 são utilizadas para geração de modelos de referência para efeito de comparação das possíveis movimentações dos sensores.

Como os dados gerados por uma ST variam continuamente no tempo, é relevante utilizar novos atributos que representem melhor todo o conjunto de STs. Como mencionado, o modelo clássico de análise de ST considera a média, tendência e sazonalidade. No entanto, outras técnicas de extração de características podem ser aplicadas para extração de características [Trovero and Leonard 2018]. Neste trabalho, consideramos cinco (5) componentes para cada série $Z_{t^{ij}}$:

1) Mínimo: $\bar{M}_{ijh}$ é o valor mínimo da ST (i,j) na janela h de tempo;
2) Máximo: $\bar{X}_{ijh}$ é o valor máximo da ST (i,j) na janela h de tempo;
3) Média: $\bar{Z}_{ijh}$ é a média de nível da ST (i,j) na janela h de tempo;
4) Inclinação: $\hat{T}_{ijh}$ é a tendência da ST (i,j) na janela h de tempo;
5) Desvio padrão: $\bar{S}_{ijh}$ é a medida de dispersão da ST (i,j) na janela h de tempo.

A matriz Z^* (2) contem os vetores de características Z^*_{ijh} composto pelas componentes citadas anteriormente, $Z^*_{ijh} = \left[\bar{M}_{ijh}, \bar{X}_{ijh}, \bar{Z}_{ijh}, \hat{T}_{ijh}, \bar{S}_{ijh}\right]$. A cardinalidade da matriz $[Z^*]$ corresponde à

$[Z^*] = N \times M \times H \times 5.$

$$[Z^*] = \begin{bmatrix} Z^*_{111} & Z^*_{121} & \cdots & Z^*_{1M1} \\ Z^*_{211} & Z^*_{221} & \cdots & Z^*_{2M1} \\ \vdots & \vdots & \ddots & \vdots \\ Z^*_{N11} & Z^*_{N21} & \cdots & Z^*_{NM1} \\ \cdots & \cdots & \cdots & \cdots \\ Z^*_{112} & Z^*_{122} & \cdots & Z^*_{1M2} \\ Z^*_{212} & Z^*_{222} & \cdots & Z^*_{2M2} \\ \vdots & \vdots & \ddots & \vdots \\ Z^*_{N12} & Z^*_{N22} & \cdots & Z^*_{NM2} \\ \cdots & \cdots & \cdots & \cdots \\ \ddots & \ddots & \ddots & \ddots \\ \cdots & \cdots & \cdots & \cdots \\ Z^*_{11H} & Z^*_{12H} & \cdots & Z^*_{1MH} \\ Z^*_{21H} & Z^*_{22H} & \cdots & Z^*_{2MH} \\ \vdots & \vdots & \ddots & \vdots \\ Z^*_{N1H} & Z^*_{N2H} & \cdots & Z^*_{NMH} \end{bmatrix}_{N \times M \times H} \tag{2}$$

4. METODOLOGIA

A metodologia adotada é ilustrada na Fig. 1. Esta tem início com a parametrização do gerador de séries temporais *SyTSBox*, desenvolvido para este trabalho, e que permitiu gerar uma TDB para representar a operação de S sensores em ambientes industrias ou de IoT. A segunda fase consiste no janelamento temporal (w_h) da TDB. Na terceira fase da metodologia, é realizada a extração de características a fim de representar o conjunto de sensores $s_j \in S$. Na quarta fase, aplicamos algoritmo de clusterização Aglomerativa (AGNES) para analisar a movimentação dos sensores. Por fim, a quinta fase, analisamos o surgimento de eventos considerados anômalos ou extremos.

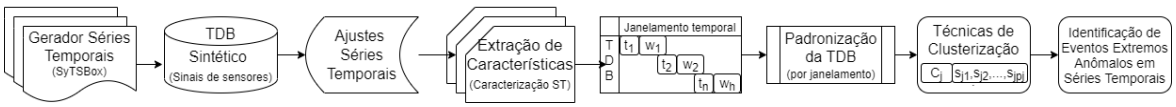

Fig. 1. Esboço da metodologia proposta

4.1 Base de dados Temporal

Devido à indisponibilidade de dados de STs em ambientes críticos, desenvolvemos um gerador de séries temporais capaz de emular diversos tipos de STs, gerando assim uma TDB sintética. O gerador, denominado de SyTSBox (*Synthetic Time Series Box*), é capaz de representar diferentes cenários de *streaming* devido às diversas configurações possíveis, dada por sua capacidade parametrizável.

Para este trabalho, a TDB gerada pelo SyTSBox é composta por 4 tipos de STs: 1) STs com indicador de tendência; 2) STs com ocorrência de picos; 3) STs com deslocamento abrupto de nível; e 4) STs com deslocamento gradual de nível. Exemplos dos tipos de STs são ilustradas nas Fig. 2, 3, 4 e 5. As STs que compõem a TDB são formadas por 40 sensores operando individualmente, onde cada sensor possui 240 pontos/instâncias de série. Os sensores foram definidos com os seguintes parâmetros operacionais: a) média (μ) aleatória entre 1 e 10 por sensor; b) desvio padrão (σ) igual a 1; c) desvio padrão (σ) entre 1 e 6; e d) cada sensor possui 5% de pontos com valores extremos.

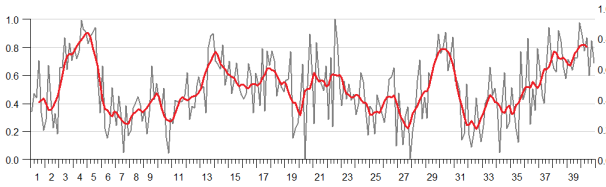

Fig. 2. ST com tendências

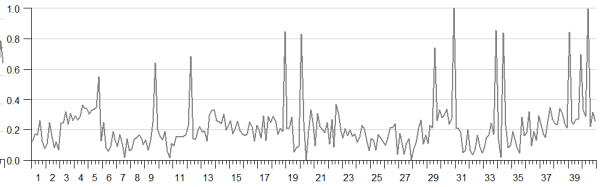

Fig. 3. ST com ocorrência de picos

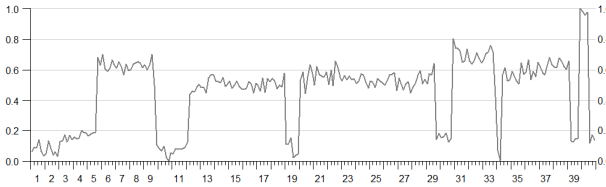

Fig. 4. ST com deslocamentos abruptos de média

Fig. 5. ST com deslocamentos gradual de média

A Fig. 6 ilustra a variação dos sinais que compõem a TDB, tal como a presença de *outliers* nas STs com ocorrência de picos (VP1∼VP10), e nas STs com deslocamento gradual de nível (VSG1∼VSG10). No entanto, existe menor distância dos máximos em relação ao 3^o quartil nas STs com tendência (VT1∼VT10), e maior distância nas STs com deslocamento abrupto de nível (VSA1∼VSA10).

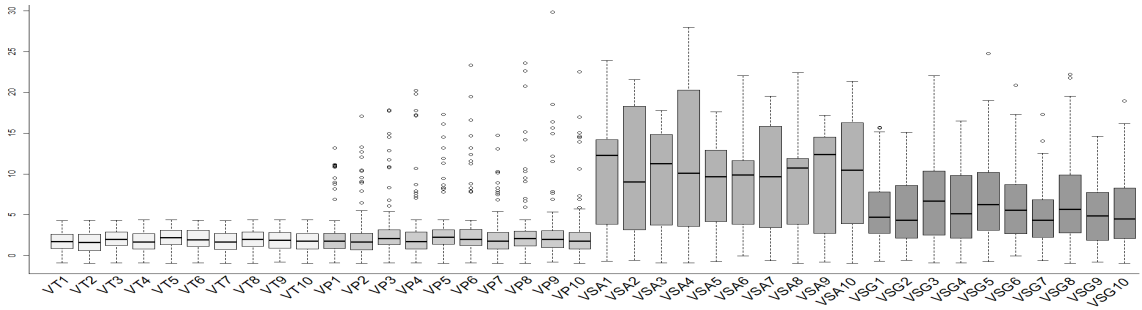

Fig. 6. Distribuição e valores dos sinais do sistema de sensores

4.2 Tipos de Cenários

Seja s_j um sensor do conjunto S de sensores e K o número de *clusters* obtidos sob condições normais de operação. C_j representa um determinado *cluster* definido por $C_j = \{s_{j1}, \cdots, s_{jp_j}\}$, $\forall\ s_j \subseteq S$, $j = \{1, \cdots, K\}$, onde, p_j é o número de sensores no *cluster* C_j. De acordo com a evolução temporal nos sinais dos sensores, consideramos dois possíveis cenários, são eles:

Cenário I - Condições normais de operação:
O número de *clusters* é constante ao longo do tempo:

$$|C_j^{Referencia}| = |C_j^t| = K \tag{3}$$

Os sensores podem ou não mudar de agrupamentos:

$$\begin{cases} C_j^{Referencia} = C_j^t \\ C_j^{Referencia} \neq C_j^t \end{cases} \forall\ j = \{1, \cdots, K\}, t = \{1, \cdots, T\} \tag{4}$$

Cenário II - Existe variação significativa (eventos anômalos) nos sinais dos sensores: Número de *clusters* pode variar:

$$|C_j^{Referencia}| \neq |C_j^t| \tag{5}$$

4.3 Clusterização das Séries Temporais

A clusterização é feita sobre o conjunto de características Z_{ijh}^* extraídas da janela $w_h \in W$ pertencentes aos sensores $s_i \in S$ da TDB. O número ótimo de *clusters* C_j é obtido pela medida do *Silhouette index*. As estratégias para a clusterização das STs que atendem aos cenários I e II são:

- ***Estratégia 1 (Cenário I)*** - O *Silhouette index* permite determinar o número ótimo de *clusters* para as janelas de referência w_1 e w_2, usando esse mesmo número para as janelas seguintes. Sendo o número de *clusters* constante (Equação 3), é possível verificar o índice de Estabilidade (Equação 6) dos sensores, o qual indica o grau de movimentação dos sensores entre os *clusters*.
- ***Estratégia 2 (Cenário II)*** - Por meio do *Silhouette index* é determinado o número ótimo de *clusters* para cada janela $w_h \in W$, onde $h = \{3, \cdots, 40\}$. Como o número de *clusters* pode ser variável (Equações 5), verifica-se o surgimento, ou o desaparecimento de *clusters*, o que indica alterações nas condições normais de operação de um ou mais sensores no sistema.

4.4 Métricas de Avaliação

4.4.1 *Índice de Estabilidade.* O *índice Estabilidade* $B(w_h)$ mede o quanto os sensores apresentam-se estáveis nos *clusters* ao longo das janelas. $B(w_h)$ é a taxa de permanência do sensor no *cluster* representativo durante as janelas w_h, em relação aos modelos de referência, ver Equação 6.

$$B(w_j) = \frac{|clusters : w_{[1,2]} \cap clusters : w_h|}{|clusters : w_{[1,2]}| + |clusters : w_h|} \tag{6}$$

onde, obtemos uma taxa de acerto ao comparar cada janela w_h com os modelos de referência $w_{[1,2]}$.

4.4.2 *Índice de Instabilidade.* O *índice Instabilidade* $L(s_i)$ mede o quanto os sensores movimentam-se entre os *clusters*. Assim, definimos a instabilidade (entropia) como, o quanto que cada sensor s_i movimenta-se entre diferentes *clusters* nas w_j janelas, ver Equação 7.

$$L(s_i) = -\sum_{s \in S} P(s_i) Log_2(P(s_i)) \tag{7}$$

O cálculo da instabilidade não utiliza os modelos de referência ($w_{[1,2]}$), apenas a variação de s_i em w_j.

5. EXPERIMENTOS E ANÁLISE DOS RESULTADOS

Considerando a Estratégia 1 (Seção 4.3), a medida *Silhouette index* indicou o número ideal de 2 *clusters* para ambos os modelos de referência, w_1 e w_2. Os *clusters* foram avaliados pelo teste de hipóteses multivariado T^2 de *Hotelling* [Hotelling 1992].

A divisão dos sensores nos *clusters* evidenciou a separação do conjunto de STs com tendência (em condições normais de operação) no *cluster* 1 (VT1∼VT10); e em relação ao *cluster* 2 composto pelas séries com picos (VP1∼VP10), séries com deslocamentos abruptos (VSA1∼VSA10) e deslocamentos graduais (VSG1∼VSG10), os quais apresentam valores anômalos, discrepantes das séries com condições normais de operação.

Pela Estratégia 1, foi aplicada o algoritmo de clusterização a cada janela $w_{1..40} \subseteq W$, individualmente (considerando a Equação 3). Assim, os sensores localizados nos *clusters* de cada janelamento

($w_{[3..40]}$) são comparados com os *clusters* do modelo de referência ($w_{[1,2]}$). Para avaliar o equilíbrio dos sensores nos *clusters* foi calculado o índice de Estabilidade (Equação 6) ilustrado na Tabela I.

Table I. Índice de Estabilidade por janelamento

Wnd	Clst	Estabilidade Ref[1,2]	Wnd	Clst	Estabilidade Ref[1,2]	Wnd	Clst	Estabilidade Ref[1,2]	Wnd	Clst	Estabilidade Ref[1,2]
W1	2	1.000	W11	2	1.000	W21	2	0.650	W31	2	0.325
W2	2	1.000	W12	2	0.625	W22	2	0.725	W32	2	0.475
W3	2	0.975	W13	2	0.725	W23	2	0.275	W33	2	0.400
W4	2	1.000	W14	2	0.700	W24	2	0.525	W34	2	0.275
W5	2	0.625	W15	2	0.500	W25	2	0.475	W35	2	0.725
W6	2	0.225	W16	2	0.500	W26	2	0.500	W36	2	0.400
W7	2	0.275	W17	2	0.525	W27	2	0.575	W37	2	0.500
W8	2	0.500	W18	2	0.450	W28	2	0.525	W38	2	0.525
W9	2	0.600	W19	2	1.000	W29	2	0.675	W39	2	0.675
W10	2	0.025	W20	2	0.600	W30	2	0.925	W40	2	0.650

Considerando a Estratégia 2 (Seção 4.3), o número ideal de *clusters* foi obtido para cada janelamento $w^6_{1..40} \subseteq W$ Por este motivo, obtivemos um número diferente de *clusters* a cada janelamento w_j.

Para avaliar a movimentação dos sensores entre os *clusters* em cada janelamento, foi calculado o índice de Instabilidade (Equação 7), tanto para Estratégia 1 quanto para a Estratégia 2. Por meio do índice de Instabilidade foi possível identificar os sensores com mais instabilidades, ou seja, os sensores que tiveram maior variação nos valores de seus sinais, ver Tabela II.

Table II. Índice de Instabilidade por sensor com número fixo e variável de clusters

Sensores	Clstr Fixo	Instabilidade Estratégia-1	Clstr Var.	Instabilidade Estratégia-2	Sensores	Clstr Fixo	Instabilidade Estratégia-1	Clstr Var.	Instabilidade Estratégia-2
VT1	2	0.0000	2	0.0000	VSA1	2	0.9710	3	1.7348
VT2	2	0.6690	2	0.9341	VSA2	2	0.9982	6	1.9333
VT3	2	0.7219	2	1.2853	VSA3	2	0.9710	6	2.0380
VT4	2	0.9544	2	1.6102	VSA4	2	0.9982	8	2.0700
VT5	2	0.8485	3	1.5730	VSA5	2	0.9710	2	2.0380
VT6	2	0.7219	2	1.1263	VSA6	2	0.9710	6	1.8848
VT7	2	0.7692	2	1.5832	VSA7	2	0.9097	4	**2.1151**
VT8	2	0.7692	4	1.4202	VSA8	2	0.9710	4	1.8159
VT9	2	0.9341	2	1.6477	VSA9	2	0.9982	2	**2.1504**
VT10	2	0.9837	2	1.6819	VSA10	2	0.9837	3	2.0380
VP1	2	0.9710	2	1.9834	VSG1	2	0.9982	2	1.7666
VP2	2	0.9097	3	1.8376	VSG2	2	**1.0000**	2	1.9802
VP3	2	0.9097	2	1.9322	VSG3	2	**1.0000**	6	2.0229
VP4	2	0.8813	2	1.8421	VSG4	2	0.9928	9	1.8016
VP5	2	0.8813	4	1.8747	VSG5	2	0.9982	3	**2.1290**
VP6	2	0.9341	2	1.9508	VSG6	2	**1.0000**	6	1.9991
VP7	2	0.9097	2	1.8101	VSG7	2	0.9982	2	**2.1290**
VP8	2	0.9341	2	1.9508	VSG8	2	**1.0000**	4	2.0502
VP9	2	0.8813	2	1.8697	VSG9	2	0.9982	2	1.8481
VP10	2	0.9097	4	2.0095	VSG10	2	0.9097	2	2.0595

Note que, por meio do índice de Instabilidade, foi possível identificar os sensores que se mostraram mais instáveis, obtendo os valores de 1.0 e 2.1 para as Estratégias 1 e Estratégia 2, respectivamente.

Como mencionado, os sensores que compõem o *cluster* 1 contêm apenas as séries com indicadores de tendência (Fig. 2), não sendo possível identificar valores anômalos, ou mesmo indícios de algum valor extremo. Assim, os sensores contidos neste *cluster* não estão associados a eventos extremos. Porém, o *cluster* 2 é composto por séries que apresentam, além do comportamento anômalo, como picos (Fig. 3), deslocamento abrupto de médias (Fig. 4), e deslocamento gradual de médias (Fig. 5), também possui sinais com valores extremos, tornando o sistema de sensores que compõem este *cluster*, ideal para serem monitorados pela Teoria de Valores Extremos (*Leonard Tippett* 1902–1985). O índice de estabilidade *1.0* indica que não houve mudanças dos sensores entre os *clusters*. Valor menores no índice, indica movimentação dos sensores entre os *clusters*.

6. CONCLUSÃO

Neste trabalho, abordamos o problema de análise em séries temporais que representam diferentes sistemas de sensores, onde a criticidade em termos de segurança deve ser monitorada. A estratégia de janelamento temporal foi efetiva, uma vez que o tamanho da janela foi definido pela menor variação da série. Esta efetividade foi demonstrada ao monitorar as demais janelas em relação aos modelos de referência, sendo possível identificar as mudanças de comportamento dos sensores entre *clusters*.

Como pontos a serem melhorados em nossa abordagem, apontamos: a) utilizar outras características para representar as séries temporais; b) construir modelos supervisionados para explicar as mudanças detectadas; c) considerar outros ajustes no tamanho do janelamento; d) utilizar um *threshold* semi-supervisionado para diferenciar anomalias comuns dos grandes desvios; e) analisar o comportamento e a evolução temporal; e f) considerar a evolução dos modelos de referência pelas mudanças nos sensores.

Como trabalhos futuros, considera-se: implementar novas funções ao gerador SyTSBox, considerando outros cenários de séries temporais; definir um *threshold* que diferencie valores anômalos e extremos; guardar estados intermediários para ajustar o modelo de referência; e utilizar a teoria de valores extremos, como gatilho para detecção de eventos anômalos ou extremos em séries temporais.

ACKNOWLEDGMENT

The authors acknowledge the financial support from the CNPq (Brazilian National Council for Scientific and Technological Development), CAPES (Coordination for the Improvement of Higher Education Personnel), FAPEMIG (Foundation for Research Support of the State of Minas Gerais), and Pontifical Catholic University of Minas Gerais, Brazil.

REFERENCES

Abdullah, M. Time series motif discovery: dimensions and applications. *WIREs Data Mining Knowl Discov* vol. 4, pp. 152–159, 2014.

Caires, S. Extreme value analysis: wave data. Tech. Rep. 57, JCOMM Technical Report 57, 2011.

Chavez-Demoulin, V. and Davison, A. C. Modelling the time series extremes. *Revstat Statistical Journal* vol. 10, pp. 109–133, 03, 2012.

Hotelling, H. The generalization of student's ratio. In *Breakthroughs in Statistics: Foundations and Basic Theory*, S. Kotz and N. L. Johnson (Eds.). Springer New York, New York, NY, pp. 54–65, 1992.

Huang, D., Koh, Y. S., and Dobbie, G. Rare pattern mining on data streams. In *Proceedings of the 14th International Conference on Data Warehousing and Knowledge Discovery (DaWaK)*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 303–314, 2012.

Lee, T., Zhang, R., Xiao, Y., and Dean, J. Feature extraction methods for time series data in sas enterprise miner. Tech. rep., SAS Institute, 2014.

Nakkeeran, K., Garla, S., and Chakraborty, G. Application of time series clustering using sas enterprise miner for a retail chain. Tech. rep., SAS Global Forum 2012. 04, 2012.

Ranjan, C., Reddy, M., Mustonen, M., Paynabar, K., and Pourak, K. Dataset: Rare event classification in multivariate time series. *ArXiv* vol. abs/1809.10717, pp. 1–7, 2018.

Serra, A. P. and Zárate, L. E. Characterization of time series for analyzing of the evolution of time series clusters. *Expert Systems with Applications* 42 (1): 596 – 611, 2015.

Taleb, N. N. *The Black Swan: The Impact of the Highly Improbable.* Incerto. Random House Publishing Group, London, 2007.

Trovero, M. A. and Leonard, M. J. Time series feature extraction. Tech. rep., SAS 2020-2018, 2018.

Winters, P. R. Forecasting sales by exponentially weighted moving averages. *Manag. Science* 6 (3): 324–342, 1960.

Yanfei Kang, Rob J Hyndman, F. L. Efficient generation of timeseries with diverse and controllable characteristics. Tech. Rep. 15/18, Monash Business School, 2018.