
Análise e Previsão de Índices de Segurança Operacional

Paulo M. Razaboni^{1,2}

1 Embraer SA

2 paulo.razaboni@embraer.com.br

BIOGRAFIA:

Atua como especialista em Segurança Operacional na Embraer em São José dos Campos, nas áreas de Inovação e Ciência de Dados. Engenheiro Eletricista pela USP, cursou MBA em Gestão da Produção pela Universidade Federal de São Carlos e Gestão Empresarial Competitiva pelo INPG. É membro credenciado Fator Material do SIPAER (Sistema de Investigação e Prevenção de Acidentes Aeronáuticos), com especialização em Segurança Operacional pela Embraer e *Black-belt* em *Six-sigma*. Ministra cursos na área de Segurança Operacional e projeto. Representa a empresa junto a entidades internacionais, como IATA, ICAO, FAA.

RESUMO: É impossível gerenciar o que não se consegue medir. Dessa forma, cada vez mais nos deparamos com a necessidade da determinação de índices representativos para as ocorrências relevantes, dentro do processo estabelecido em que se constitui o Sistema de Gerenciamento da Segurança Operacional. Dentro de um dos pilares deste sistema, chamado de Gerenciamento de Risco, existe o componente de identificação do risco e, dentro dele, o método preditivo, que busca melhorar a Segurança Operacional através do monitoramento dos dados de voo de operações normais. Estamos gerando uma quantidade progressivamente maior de dados, esperando extrair deles as informações para tomadas de decisão quanto à melhor alocação de recursos, a fim de atingir os objetivos estratégicos. A metodologia aqui descrita apresenta o progresso de práticas que se demonstraram válidas para séries temporais, podendo o método ser aplicado a conjuntos de tamanhos variados, que apresentem variações em nível, tendência e sazonalidade. Uma característica chave é a eliminação da subjetividade na análise, ao utilizar-se de critérios estatísticos objetivos, sendo o método apto à inclusão em sistemas computadorizados. Os resultados permitem identificar também o momento em que o comportamento dos dados sofre alteração significativa, além de prover meios para projetar valores futuros a partir do seu histórico. Este conjunto de informações visa prover ao analista os argumentos necessários à tomada de decisão para o elenco de ações no sentido de elevar o nível de segurança para metas estabelecidas, além de se constituir em uma solução adequada à comprovação da eficácia das ações tomadas.

Palavras Chave: Análise, Índices, Previsão.

Analysis and prediction of Operational Safety indexes

ABSTRACT: It is impossible for one to manage what one cannot measure. Therefore, we are increasingly faced with the need to determine representative indexes for relevant occurrences, within the established process in which the Operational Safety Management System is constituted. Within one of the pillars of this system, namely Risk Management, there is the component of risk identification, and within it the predictive method, which seeks to improve Operational Safety by monitoring flight data of normal operations. We are generating a progressively larger amount of data, hoping to extract from these data the pieces of information for the decision-making process aimed at the best allocation of resources to achieve the strategic objectives. The methodology described here presents the progress of practices that have been demonstrated to be valid for time series, and the method can be applied to assorted sizes, varying in level, trend and seasonality. A key feature of the method is the elimination of subjectivity in the analysis, by using objective statistical criteria, being the method suitable for inclusion in computerized systems. The results also allow us to identify the moment when the behavior of the data undergoes significant change, and provide us with means to project future values from their history. This set of pieces of information aims to provide the analyst with the arguments necessary for decision-making for the list of actions, in order to raise the level of safety for established goals, in addition to being a suitable solution for verifying the effectiveness of the actions taken.

Keywords: Analysis. Indexes. Prediction.

Citação: Razaboni, PM. (2018) Análise e Previsão de Índices de Segurança Operacional. *Revista Conexão Sipaer*, Vol. 9, No. 2, pp. 95-105

1 INTRODUÇÃO

Uma necessidade básica existe para todos os sistemas que exigem monitoramento e controle: a adequada interpretação da informação obtida. Em um universo em que oceanos de dados encontram-se progressivamente mais disponíveis, deve-se aprimorar a técnica de extrair informações úteis a partir deles. Termos como "*big data*" são comuns, e muitas técnicas de tratamento e consolidação dos dados já foram desenvolvidas. A intenção deste trabalho não é descrever o tratamento de grandes volumes de dados, mas sim ilustrar técnicas que se mostraram úteis na prática para interpretar os resultados extremamente condensados obtidos a partir destes.

A análise realizada sobre uma tabela de valores numéricos pode produzir noções sobre o comportamento de determinada grandeza. Uma representação gráfica fornecerá melhores recursos visuais. Porém, para tomada de decisão, por muitas vezes,

acaba-se por confiar em interpretações visuais que podem induzir a erros ou então a soluções pouco otimizadas. Uma avaliação de desempenho também fica comprometida se não existe uma base para comparação das situações, projetada e obtida. Nesse sentido, algumas soluções foram selecionadas e combinadas, tendo como alvo principal a projeção de indicadores de desempenho de Segurança Operacional. O método numérico aqui descrito, se resguardadas as suas características, pode ser aplicado a várias outras situações.

2 METODOLOGIA

Os índices de desempenho utilizados como base apresentam as características a seguir, necessárias para a aplicação do método aqui descrito:

- São séries temporais, tomados a intervalos regulares;
- Definem taxas ou contagens (valores positivos);
- Podem apresentar alterações de nível, tendência ou mesmo cíclicas ao longo do tempo (sazonalidade);
- Por consolidarem informações de uma grande massa de dados, não mudam de comportamento de forma abrupta.

Tomemos por base um conjunto de valores, como na **Tabela 1** a seguir. Fontes típicas são índices de desempenho calculados a partir da análise de eventos ocorridos em voos de aeronaves, sendo consideradas de milhares a milhões de operações para a consolidação destes em um único valor mensal.

Tabela 1: Dados de entrada

| Mês \ Ano | 2014 | 2015 | 2016 |
|-----------|------|------|------|
| Janeiro | 377 | 337 | 297 |
| Fevereiro | 343 | 273 | 246 |
| Março | 266 | 226 | 216 |
| Abril | 238 | 204 | 180 |
| Maio | 256 | 209 | 203 |
| Junho | 256 | 237 | 227 |
| Julho | 292 | 244 | 235 |
| Agosto | 320 | 280 | 269 |
| Setembro | 343 | 300 | 301 |
| Outubro | 387 | 332 | 312 |
| Novembro | 413 | 338 | 342 |
| Dezembro | 385 | 313 | 314 |

Os valores da **Tabela 1** representam taxas de eventos em que foram excedidos os limites estabelecidos para desvios de aproximação, velocidade, taxa de descida, atitude etc. É clara a dificuldade de se tomar decisões assertivas com base apenas em uma análise visual dos números.

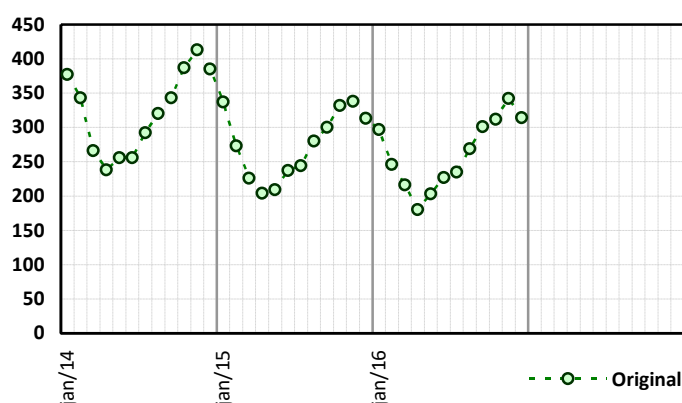


Gráfico 1: Dados de entrada originais

Utilizando-se a representação do **Gráfico 1**, pode-se notar alguns aspectos, como a variação periódica (sazonalidade) e uma tendência de queda no longo prazo. A avaliação, ainda subjetiva, sugere comportamento similar para o próximo ciclo. É possível identificar um período típico, que não raro ocorre neste tipo de indicador (no caso, anual). Valores em torno de abril de cada ano apresentam-se consistentemente menores, e em torno de novembro, mais elevados.

Tais indicadores, por refletirem grandezas de diferentes naturezas, podem apresentar faixas de valores muito amplas, resultando na dificuldade de comparação quando dispostos em um mesmo painel ("dashboard"). Assim, decidiu-se normalizar

o conjunto de dados, tomando-se por base o valor médio de todo o intervalo de entrada, o que equaliza a representação visual de todos, estabelecendo porcentagens de variação. O valor médio do intervalo será:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (1)$$

\bar{y} : valor médio do intervalo

y_i : leituras individuais

n : linhas da tabela (36, no caso da Tabela 1)

Assim, cada valor normalizado pode ser calculado por:

$$y'_i = \frac{y_i}{\bar{y}} \quad (2)$$

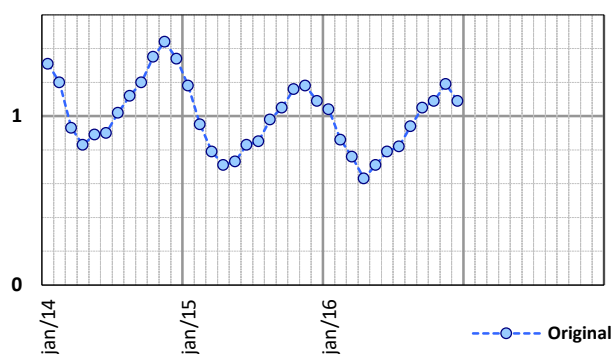


Gráfico 2: Dados de entrada normalizados

Ao dividir-se cada valor pela média do intervalo, estes passam a se distribuir em torno do valor unitário, como no **Gráfico 2**. Para uma previsão sobre o próximo período, utilizou-se a suavização exponencial tripla de Holt-Winters (HOLT, 1957) (WINTERS, 1960). Esta técnica de regressão decompõe a curva original em três componentes: nível, tendência e sazonalidade. Assim, cada ponto futuro será um reflexo dos pontos anteriores, exponencialmente suavizados por coeficientes independentes para cada um desses componentes. O método de Holt-Winters aditivo, aqui utilizado, é aplicável quando a amplitude do ciclo sazonal independe do nível local da série, levando-se em conta que os dados já se encontram normalizados. Seja um modelo cuja série sazonal, de período p , tem seu valor (Z) formado pela soma de nível (L), tendência (T), um fator sazonal (S) e um erro aleatório (ε), dados por:

$$Z_t = L_t + T_t + S_t + \varepsilon_t \quad (3)$$

para $t = 1, 2, \dots$

As projeções dos valores futuros da série são efetuadas através da função de previsão representada por:

$$\hat{Z}_{t+n} = L_t + nT_t + S_{t-p+n} \quad (4)$$

para $n = 1, 2, \dots$

Onde \hat{Z}_{t+n} é a previsão para n períodos à frente ($t+n$), e p é o tamanho do período típico da série temporal.

Nesse método, além da função (4) que calcula a previsão, três outras funções são utilizadas para estimar o nível, a tendência da série no período atual e os valores do fator sazonal correspondente ao último período de sazonalidade, conforme eq. (5), (6) e (7) respectivamente,

$$L_t = \alpha(Z_t - S_{t-p}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (5)$$

$$T_t = \beta(L_t - L_{t-p}) + (1 - \beta)T_{t-1} \quad (6)$$

$$S_t = \gamma(Z_t - L_t) + (1 - \gamma)S_{t-p} \quad (7)$$

com: $0 < \alpha < 1$, $0 < \beta < 1$, $0 < \gamma < 1$

Aqui, α , β e γ são constantes de suavização que controlam o peso relativo ao nível (L_t), à tendência (T_t) e à sazonalidade (S_t), respectivamente. Quanto mais próximo de zero forem essas constantes, menor o peso da observação atual e maior o peso

dos valores anteriores na determinação da estimativa. As eq. (5) e (7) assumem que, no período de tempo t , existe uma estimativa do índice sazonal no período de tempo $t-p$ ou que existe um valor S_{t-p} . É possível utilizar-se estimativas a partir das leituras do primeiro período, para assim calcular os valores do segundo período em diante.

Um método numérico de aproximação pode ser utilizado para determinar a combinação das constantes de suavização (α , β e γ) que produz uma curva de previsão com o menor erro em relação à curva original. Neste exemplo, utilizou-se o recurso "Solver" do Excel®. A previsão é calculável a partir do segundo período, com o primeiro servindo de base. Assim, no exemplo em que temos três períodos sucessivos de um ano cada, utiliza-se os valores do período 1 para estimativa e calcula-se os períodos 2 e 3 ajustando os coeficientes para que os valores previstos apresentem o menor erro possível em relação aos reais. A seguir, projeta-se um quarto período. Os valores obtidos, sobrepostos aos dados originais, encontram-se no gráfico a seguir.

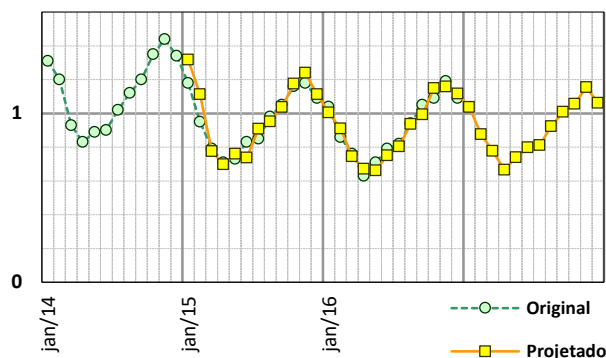


Gráfico 3: Dados de entrada e valores projetados

O aspecto visual corrobora a expectativa de tendência dos valores nos três aspectos: nível, tendência e sazonalidade. O passo a seguir é a translação dos dados para que a média do intervalo inicial seja igual ao valor zero. As variações em torno do zero podem ser representadas como porcentagem.

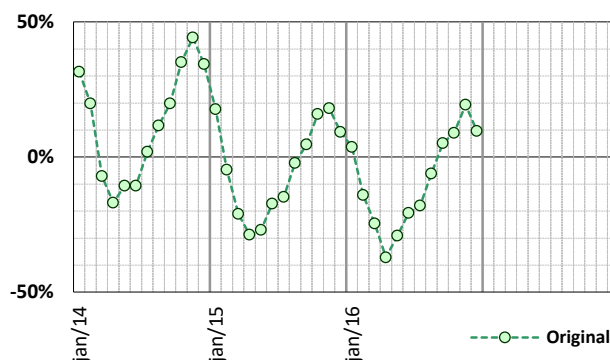


Gráfico 4: Dados de entrada ajustados para média zero

Para a caracterização numérica da sazonalidade, optou-se por buscar a melhor curva senoide que se aproxima destes pontos, excluindo-se a tendência de longo prazo. Para isso, na prática, é realizada a média de todos os valores que apresentam mesma posição relativa no período (no caso, média dos índices nos "janeiros", "fevereiros" etc.).

$$y_m = \frac{\sum_{i=1}^k y'_{(i-1)*p+m}}{k} \quad (8)$$

$$\text{com: } k = \text{Int}(n/p) \quad (9)$$

y' : valor mensal do índice (normalizado)

p : tamanho do período (no caso, 12)

n : linhas da tabela (no caso, 36)

k : períodos no intervalo (parte inteira de $36/12 = 3$)

m : índice do mês (no caso, 1..12)

Com os valores obtidos, determinou-se os coeficientes para:

$$w_m = a \times \sin(x_m) + b \times \cos(x_m) \quad (10)$$

w_m : valor aproximado para cada mês

a, b : coeficientes da função seno, cosseno

x_m : valor para cada fração do período (de zero até $^{11}/_{12}$ frações do período total da senoide, que é 2π)

Ajusta-se a e b para se obter a menor soma dos quadrados das diferenças entre os valores calculados w_m e os originais, y_m . A correlação ρ entre os valores mensais e os aproximados pela senoide pode ser calculada como segue (MUKAKA, 2012):

$$\rho = \frac{\sum_{m=1}^p (w_m - \bar{w})(y_m - \bar{y})}{\sqrt{\sum_{m=1}^p (w_m - \bar{w})^2} \cdot \sqrt{\sum_{m=1}^p (y_m - \bar{y})^2}} \quad (11)$$

w_m : valor aproximados para cada mês

\bar{w} : média geral dos valores aproximados

y_m : valor médio lido para cada mês

\bar{y} : média geral dos valores lidos

p : tamanho do período (no caso, 12)

m : índice do mês (no caso, 1..12)

Tabela 2: Interpretação para correlação (ρ)

| Faixa | Interpretação |
|-------------|---------------|
| 0,00 a 0,19 | Muito fraca |
| 0,20 a 0,39 | Fraca |
| 0,40 a 0,69 | Moderada |
| 0,70 a 0,89 | Forte |
| 0,90 a 1,00 | Muito forte |

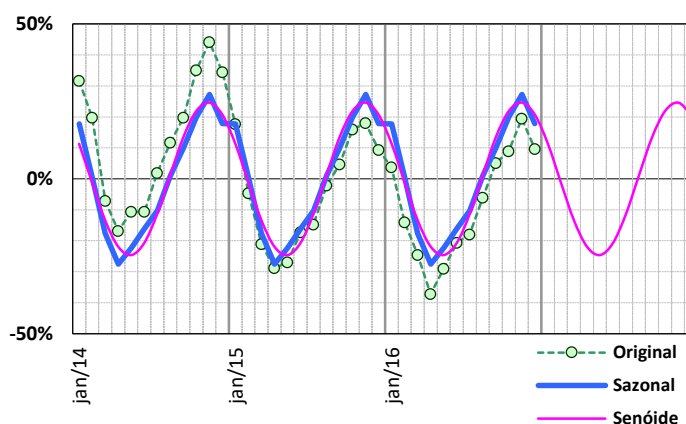


Gráfico 5: Dados de entrada e sazonalidade (senoide correspondente à Eq. 10, após o cálculo de a e b)

Assim, caracterizada a sazonalidade (**Gráfico 5**), o próximo passo é efetuar a sua compensação sobre os dados originais, subtraindo-se a contribuição sazonal (valor de w_m) destes, mês a mês, a fim de isolar e analisar a tendência geral.

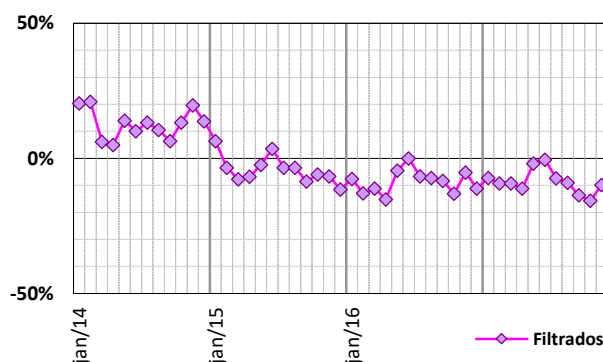


Gráfico 6: Tendência para dados sem sazonalidade

A partir do conjunto filtrado de dados (**Gráfico 6**), aplica-se o algoritmo denominado "Ponto de Mudança" ("Change Point" - vide APÊNDICE), segundo o qual mudanças com significância estatística podem ser identificadas.

Este algoritmo percorre os pontos da curva buscando instantes em que a colaboração para o valor médio sofre inflexão. Ou seja, caso leituras anteriores a um determinado ponto intermediário produzam contribuições à média do período total em um sentido oposto às leituras posteriores, este ponto é identificado como candidato.

A sua validação é feita produzindo-se uma série de combinações aleatórias das leituras do intervalo sob análise. Caso a disposição original dos pontos produza um efeito maior que as combinações aleatórias geradas, em determinado grau (no caso, considerou-se pelo menos 95% dos casos), o ponto pode ser caracterizado como "de mudança" (já que seu efeito é maior do que combinações aleatórias). No caso, cada candidato a ponto de mudança é testado contra 10.000 combinações de leituras combinadas aleatoriamente. Caso seu efeito seja mais apreciável que pelo menos 9.500 combinações (95% de 10.000), é identificado como ponto de mudança.

A seguir, calcula-se a média das leituras no sub-intervalo anterior e no posterior. Então, cada subintervalo é pesquisado recursivamente em busca de pontos de mudança de nível seguinte. Este procedimento é aplicado até que exista uma quantidade de pontos no subintervalo acima de um valor mínimo, ou que determinada profundidade de níveis já tenha sido alcançada. (RAZABONI, 2015).

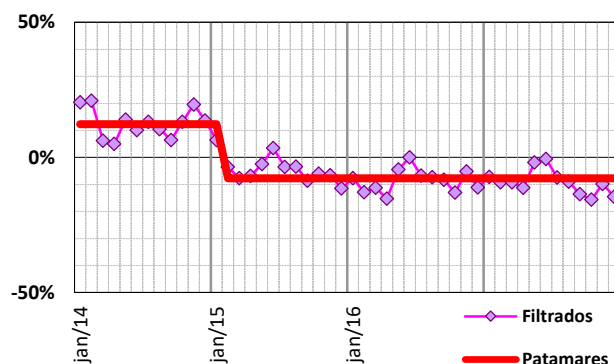


Gráfico 7: Representação gráfica do ponto de mudança

No caso deste trabalho específico, como valor prático, definiu-se um intervalo mínimo de 10 leituras e até 5 níveis de busca. Aqui, a mudança posicionou-se de janeiro para fevereiro de 2015 (com grau de confiança $\geq 95\%$). A média dos valores anteriores e dos posteriores está representada visualmente no Gráfico 7.

Em resumo, os dados (excluída a sazonalidade) apresentam um comportamento médio de 12,3% acima da média geral do intervalo de amostragem até janeiro de 2015, passando a 7,7% abaixo a partir de fevereiro de 2015.

Como os dados projetados para o quarto período (ano de 2017) foram considerados e integrados ao universo de cálculo, a análise aponta que não é esperada uma mudança significativa até o horizonte do estudo, ou seja, durante o ano seguinte, 2017.

Para uma melhor visibilidade, nem todos esses resultados intermediários precisam ser exibidos. Uma visualização sucinta pode ser apresentada, como no **Gráfico 8** a seguir:

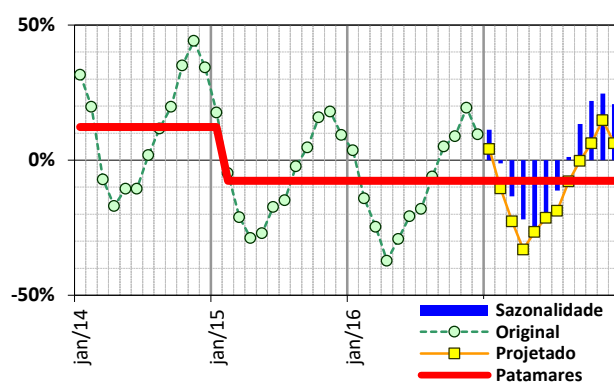


Gráfico 8: Dados de entrada e valores projetados

O valor obtido para correlação (**Eq. 11**) foi de 0,98. A partir de todas essas informações, é possível afirmar:

"O conjunto de dados dos últimos três anos, incluindo-se uma projeção para o próximo ano, apresenta uma mudança de patamar entre janeiro e fevereiro de 2015, de 12,3% acima para 7,7% abaixo do valor médio para o intervalo de amostragem. Considerando-se periodicidade de doze meses, os valores apresentam uma alta correlação sazonal (0,98), de amplitude média 25%, com máximos em torno de novembro e mínimos em torno de maio. As projeções para o próximo ano (2017), com resolução mensal (apresentadas no **Gráfico 8**), refletem uma redução no valor de pico esperado. Como ações possíveis em busca da redução deste índice, pode-se investigar a influência de fatores com variação anual no comportamento dos dados, bem como a ocorrência de eventos significativos (com efeito positivo, ou seja, de redução do valor médio) cuja influência foi observada a partir de fevereiro de 2015. O resultado de ações tomadas no momento presente, a serem refletidas no próximo período, pode ser comparado a esta previsão para verificação de sua efetividade."

Em função destas informações, pode-se realizar a alocação de recursos para programar as ações de mitigação no tempo adequado, bem como buscar identificar colaborações eficazes, que geraram resultados favoráveis e se constituem em lições positivas aprendidas.

3 RESULTADOS

A exemplo do conjunto de dados utilizado para ilustrar o procedimento (taxa de eventos onde foram excedidos os limites estabelecidos para desvios de aproximação, velocidade, taxa de descida, atitude etc., típico de índices obtidos por consolidação de dados de voo de aeronaves), vários outros conjuntos foram testados como forma de confirmar a aplicabilidade do método. Ressalva-se a necessidade de utilizar sempre um universo amostral suficientemente extenso para gerar dados consistentes, o que não se constitui em um inconveniente em se tratando de *big data*.

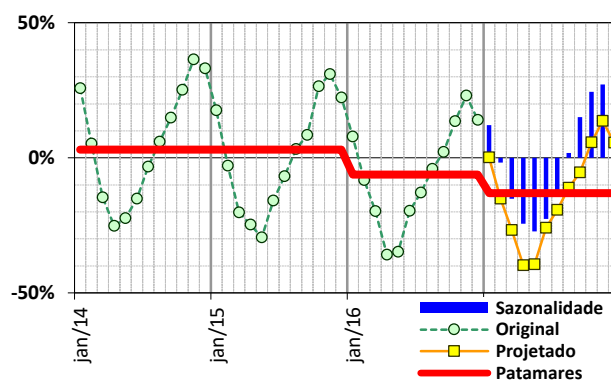


Gráfico 9: Exemplo adicional

No Gráfico 9, a exemplo do modelo desenvolvido a partir da Tabela 1, existe uma forte sazonalidade aliada à queda do nível. Um ponto de mudança em janeiro de 2016 e outro, em janeiro de 2017, foram identificados.

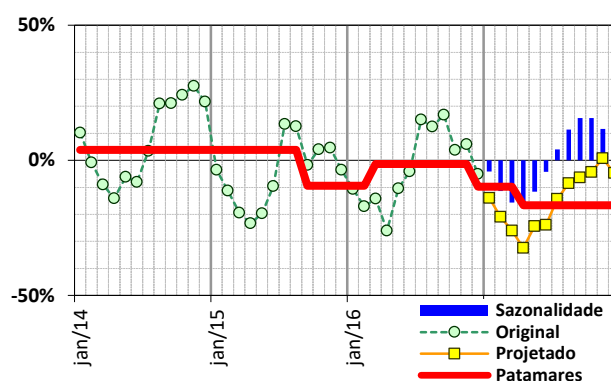


Gráfico 10: Exemplo adicional

No Gráfico 10 observa-se certa sazonalidade. E, conforme ilustra a linha vermelha contínua, quatro pontos de mudança: descida em setembro de 2015, subida em março de 2016, descidas em dezembro de 2016 e abril de 2017.

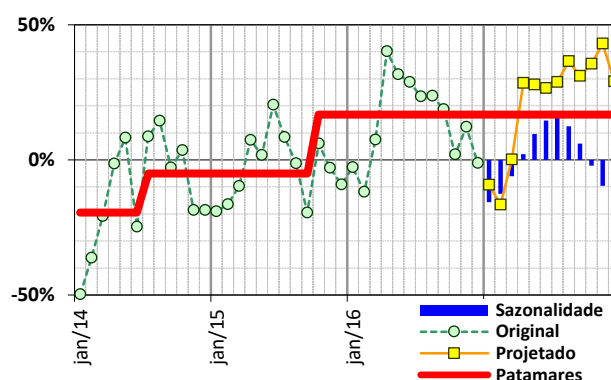


Gráfico 11: Exemplo adicional

Já no caso do Gráfico 11, torna-se evidente a elevação do nível ainda dentro do universo amostral (julho de 2014 e outubro de 2015), e a previsão para 2017 de um pico de amplitude comparável ao já ocorrido no período anterior, 2016. Os perfis crescentes dentro de cada início de período são dignos de nota. A sazonalidade identificada, embora seja a melhor

correspondência possível para o conjunto de dados fornecido, parece não traduzir bem seu comportamento. Na verdade, não existe um período claramente definido e, nesse caso, os valores previstos podem sofrer grande influência. Métodos de previsão diferentes da proposta deste trabalho podem gerar resultados mais coerentes.

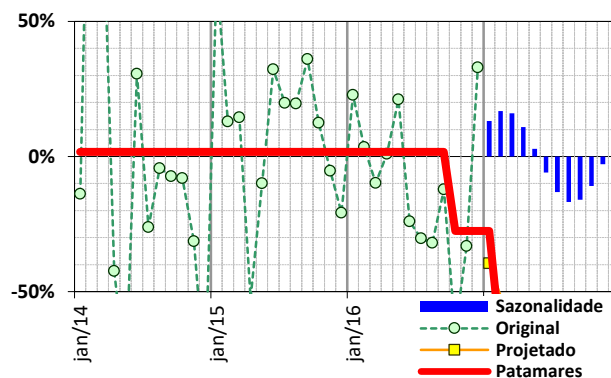


Gráfico 12: Exemplo adicional

O **Gráfico 12**, por sua vez, apresenta previsão com comportamento aparentemente incoerente em relação ao conjunto de dados inicial (curva vermelha contínua fora da escala). Nesse caso específico, não foi detectado um período de variação nos dados coincidente com a premissa do cálculo (12 meses), o que gerou este distúrbio. Esse resultado foi mantido assim para enfatizar ao leitor a necessidade de se realizar um teste inicial para determinar a correspondência entre os períodos utilizados no método e o período típico dos dados.

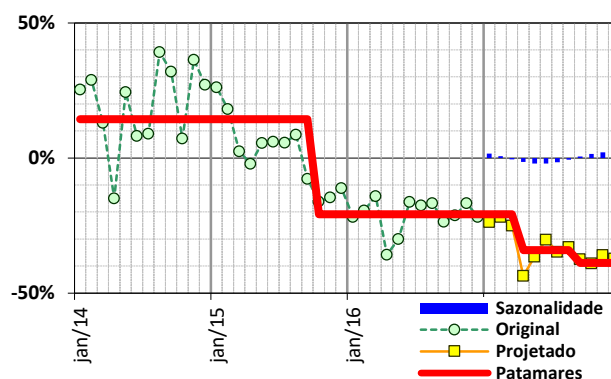


Gráfico 13: Exemplo adicional

Já no **Gráfico 13**, devido à sazonalidade calculada com base no período de 12 meses se encontrar a um nível extremamente baixo, ela pouco influenciará no cálculo da projeção para o período seguinte. A verificação do período típico dos dados continua a ser necessária para uma adequada aplicação do método, uma vez que poderia haver um período tal que sua influência fosse mais marcante na projeção final, e que deixou de figurar ao se assumir um período inadequado.

Todos os dados utilizados neste estudo são reais e objeto constante de análise para redução dos índices de eventos, apenas descaracterizados quanto às fontes por necessidade de manutenção de sigilo, porém sem prejuízo do método.

4 CONCLUSÃO

O método se apresentou adequado às necessidades iniciais, ou seja, a partir da análise de taxas de eventos relacionados à segurança operacional, fornecer suporte à tomada de decisão, planejamento e avaliação de eficácia de medidas tomadas no sentido de reduzir tais índices. A solução aqui apresentada visa apresentar ao leitor uma ferramenta adicional, que se espera seja útil dentro da busca contínua da excelência.

A exemplo de qualquer outra ferramenta estatística, deve-se ter em mente que a sua aplicação pura não exime o analista de uma comprovação da sua validade dentro da natureza dos dados que manipula. Uma simulação que gere uma previsão ainda dentro de um intervalo passado conhecido pode fornecer um método de validação. Em outras palavras, pode-se utilizar um conjunto total de dados de quatro anos, aplicar o método com os dados dos três primeiros e verificar a aderência do resultado projetado para o quarto ano contra os valores reais. De qualquer forma, parâmetros fora do universo de controle, ou variações pontuais como alterações de produto, incorporação de novas tecnologias etc., podem afetar tal aderência. Um produto em que não se tenha uniformidade de comportamento a ponto de se realizar uma análise estatística, por si só demanda outras ações práticas que estão além do escopo deste trabalho. O analista ainda poderá buscar períodos mais representativos (semanais, por exemplo). Porém, mesmo se caracterizada a imprevisibilidade dos dados, esta é por si uma informação relevante, que poderá demandar as devidas ações a critério do responsável técnico. Ao final, caberá sempre ao analista a avaliação da aplicabilidade

deste método ou de qualquer outro, e decidir se uma projeção baseada simplesmente na média móvel atende à sua necessidade ou se é necessária uma ferramenta mais elaborada.

Assim, esta solução não se propõe cobrir toda a gama de situações que podem ser encontradas na prática, mas provê, dentro de sua gama de aplicação, soluções de alto valor a um custo de processamento extremamente baixo e rápido. Parafraseando o mercado de investimentos, onde "resultado passado, não é garantia de rendimento futuro", não há como definir o resultado futuro, apenas como estimar, mas isso pode ser feito de várias formas, algumas mais adequadas.

Atualmente, esta metodologia encontra-se em fase de implementação em um sistema de análise de dados da frota de aeronaves Embraer acessível aos operadores, e em fase de proposta para um sistema mundial de coleta e análise de dados para monitoramento da segurança operacional.

Na metodologia utilizada, manteve-se algumas premissas:

- As leituras devem ser igualmente espaçadas;
- Se houver periodicidade, esta deve permanecer constante e ser condizente com o tamanho do período assumido;
- O método se aplica a valores positivos, e os resultados são apresentados normalizados;
- A previsão utiliza o método aditivo de Holt-Winters, com coeficientes α , β e γ ajustados automaticamente - isso pressupõe que projeções exponenciais podem ser utilizadas para a série temporal;
- A sazonalidade é estimada por interpolação senoidal, e seu resultado é apenas uma referência, não sendo utilizada para a projeção;
- Para a caracterização de ponto de mudança, é necessário que os candidatos apresentem um nível de confiança $\geq 95\%$ em 10.000 interações, considerando buscas recursivas até 5º nível, onde cada sub-intervalo utilizado deve possuir ao menos 10 amostras;
- O método de convergência, utilizado para determinação dos coeficientes de ajuste que geram menor erro total, foi o suplemento "Solver" do Excel®.

O autor se compromete a remeter a planilha eletrônica utilizada como base ao longo deste trabalho para os interessados, mediante solicitação.

AGRADECIMENTOS

Agradeço aos profissionais com quem interajo na realização das minhas atividades, pelos desafios que me apresentaram e levaram ao desenvolvimento deste método em particular. Estendo minha gratidão ao revisor desse artigo, pelo esmero na análise, a qual proporcionou a melhoria do mesmo. Também agradeço à minha família pelo apoio e pelo tempo de convívio que cederam. Finalmente, agradeço a Deus pela oportunidade da vida e pelos meios que me concedeu para buscar compreender um pouco melhor a Sua criação.

REFERÊNCIAS

RAZABONI, P.M. **Change-Point Analysis Applied to SMS**, ISASI Forum magazine – Journal of the International Society of Air Safety Investigators, Volume 48, Number 4, October–December 2015, p. 26–30. Disponível em <<http://www.isasi.org/Documents/ForumMagazines/Forum-V48-N4-LowRes%20Oct%20Dec%202015.pdf>>. Acesso em 17 abr. 2018, 10:30:00.

TAYLOR, W.A. **Change-Point Analysis: A Powerful New Tool for Detecting Changes**, 2000. Disponível em <<http://www.variation.com/cpa/tech/changepoint.html>>. Acesso em 17 abr. 2018, 10:00:00.

WINTERS, P.R. **Forecasting sales by exponentially weighted moving averages**, Management Science, 1960. Vol. 6, p. 324–342

....

Apêndice - Ponto de mudança

Para um conjunto de dados particular, constituído por uma série temporal, onde os mesmos se encontram regularmente espaçados, pode-se gerar uma representação gráfica como a que se segue (já incluindo o valor médio do intervalo):

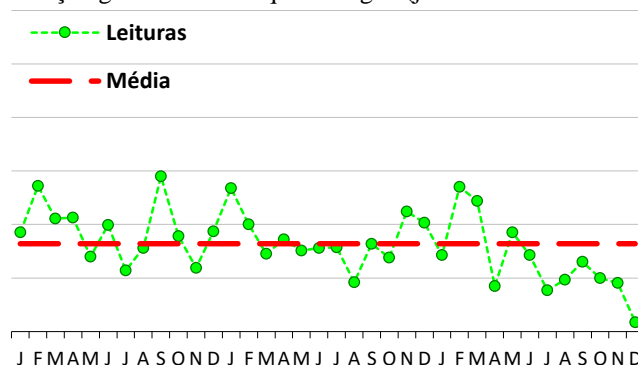
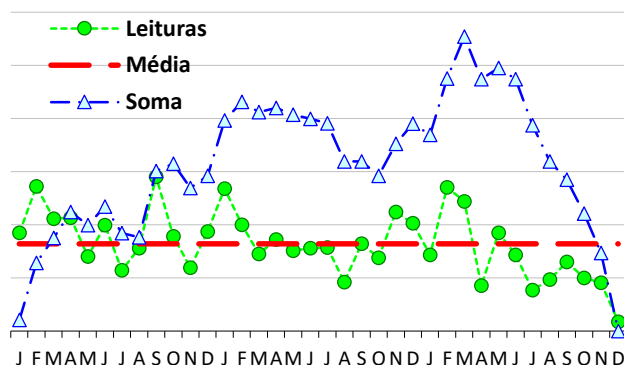


Gráfico 1: Conjunto de dados típico, com média

Determinar um ponto de mudança a partir do conjunto de dados acima pode ser desafiador. Vários métodos podem ser tentados. Por exemplo, um conjunto de pontos sequencialmente acima ou abaixo da média... Embora todas as técnicas possam, em princípio, ser implementadas por lógicas, algumas delas poderão gerar resultados discrepantes entre si, e um valor numérico pode ser difícil de se obter.

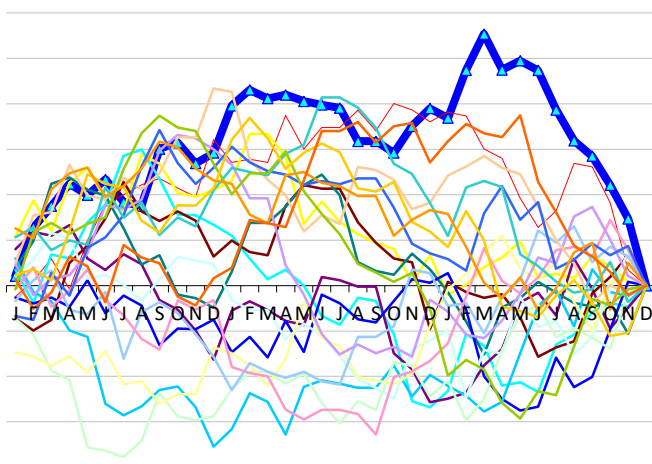
Para a detecção de pontos de mudança, foi utilizado um algoritmo chamado "Soma cumulativa das diferenças" em relação à média (OAKLAND, 2008). O cálculo realizado a partir do conjunto de dados possui o aspecto mostrado no gráfico a seguir:

**Gráfico 2:** Soma cumulativa das diferenças

A curva da soma cumulativa das diferenças inicia com a diferença entre a primeira leitura e a média, e necessariamente termina no valor zero, já que, por definição, a soma de todas as diferenças em relação à média é zero. A coordenada "x" onde o valor "y" encontra-se mais distante da média (acima ou abaixo desta) define o candidato a ponto de mudança. Conceitualmente, a partir desse ponto, os valores passam a contribuir de forma inversa para o valor da média em relação aos predecessores.

Nesse momento, torna-se necessário verificar se a sequência de dados original apresenta um padrão de distribuição suficientemente significativo, não se constituindo meramente em uma sequência aleatória de dados que poderia vir a produzir um valor de pico em algum momento.

A maneira mais direta de fazer isso é forçar essa situação, embaralhando-se os dados e verificando se alguma distribuição aleatória seria capaz de gerar resultados comparáveis (nesse caso, uma amplitude maior na curva de Soma Cumulativa). De acordo com uma das referências (TAYLOR, 2000), proceder desta forma mil vezes deve ser suficiente para classificar o candidato como um ponto de mudança ou não. Por exemplo, se os dados forem aleatoriamente embaralhados e a amplitude da curva cumulativa for calculada 1.000 vezes, e para 950 vezes as amplitudes permanecerem abaixo da original, o candidato pode ser considerado um ponto de mudança com 95% de confiança. De forma análoga, se nenhuma outra distribuição fosse capaz de produzir amplitudes maiores, o candidato seria considerado um ponto de mudança com 100% de confiança. Neste exemplo, um ponto de mudança pode ser atribuído ao processo, da seguinte maneira:

**Gráfico 3:** Soma cumulativa para várias séries

O Gráfico 3 mostra a soma cumulativa das diferenças em relação à média, para vinte conjuntos de dados obtidos a partir da combinação aleatória das leituras originais (estas, indicadas pela linha azul em negrito com marcadores). A amplitude de cada curva (distância em relação à média) é comparada com a original, para determinação do nível de confiança.

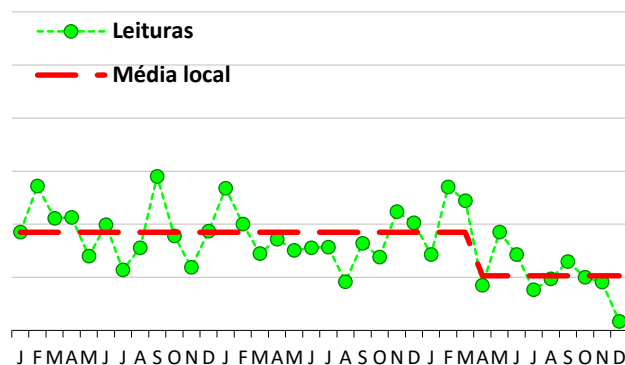


Gráfico 4: Representação das médias locais

Como o candidato foi classificado positivamente como um ponto de mudança, ou seja, revelou-se uma mudança de comportamento, calcula-se a média para ambos os sub-intervalos (antes / depois). Uma análise semelhante pode ser executada novamente para cada segmento, de forma recursiva, enquanto condições previamente estabelecidas permaneçam válidas. Como exemplo, pode-se limitar a busca a intervalos não menores que determinada quantidade de pontos, ou o número de níveis de busca. Obviamente, uma análise que não encontre pontos de mudança no intervalo, dentro do grau de confiança estabelecido, é determinante do término da busca.