

# TESTES ESTATÍSTICOS APLICADOS À DETERMINAÇÃO DE ALEATORIEDADE PARA GERAÇÃO DE CURVAS POD SIMULADAS\*

Mariana Burrowes M Guimarães<sup>1</sup>  
Gabriela Ribeiro Pereira<sup>2</sup>

## Resumo

Quantificação de confiabilidade de ensaios não destrutivos pode ser realizada através de curvas de probabilidade de detecção (POD). Estudos de confiabilidade através de curvas de probabilidade de detecção dependem, como toda análise estocástica, de parte determinística e parte probabilística. A parte probabilística da geração de curvas POD são representadas basicamente por de parâmetros de incerteza. Tais parâmetros, no caso de curvas de probabilidade de detecção PoD simuladas, são determinados pelo usuário na própria configuração da simulação. Para a geração e propagação dos parâmetros de incerteza, métodos de Monte Carlo são utilizados de modo ser possível gerar uma PoD com o mínimo de aleatoriedade. O intuito do presente trabalho é determinar se existe a aleatoriedade requerida na geração e propagação de parâmetros de incerteza no que tange a construção de curvas PoD simuladas à luz da escolha do tipo de distribuição de probabilidade (PDF) de tais parâmetros. Para tanto, o software CIVA será utilizado para a inspeção virtual e para a simulação das curvas POD. Após realizados testes estatísticos, foi possível concluir que, tanto para distribuição de probabilidade do tipo normal quanto para distribuição do tipo uniforme, os parâmetros de incerteza gerados são satisfatoriamente aleatórios. **Palavras-chave:** Curvas POD Simuladas; Confiabilidade; Inspeções Ultrassônicas.

## STATISTICAL TESTS APLIED TO DETERMINATION OF RANDOMNESS GENERATION FOR SIMULATED POD CURVES

### Abstract

Reliability studies through probability of detection (POD) curves rely, as every statistical analysis, on uncertain parameters (UP). Those UP, regarding simulated POD curves, are set up by the software user during its configuration. Methods called Monte Carlo are used to make sure that the UP are generated and propagated with the randomness that any statistical process requires. The main goal of this paper is to evaluate if the parameters generated by the software CIVA are in fact a random set of values in order to build a POD curve shading light on the probability distribution function (PDF) impact on the results. The present study was able to conclude that CIVA generates random sets of parameters for both Gaussian and Uniform PDF according to pertinent statistical tests.

**Keywords:** Simulated POD curves; Reliability; Ultrasonic Inspections.

<sup>1</sup> Engenheira Civil, MSc. Pesquisadora no LNDC/COPPE/UFRJ, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.

<sup>2</sup> Física, DSc. Professora no PEMM/COPPE/UFRJ, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.

<sup>3</sup>

## 1 INTRODUÇÃO

O uso de curvas PoD como método de quantificação da confiabilidade de ensaios não destrutivos, por exemplo, já é bastante difundido em estudos de análise de risco. De maneira geral, a curva PoD tem a função de estimar a capacidade de inspeção de determinado ensaio quando o mesmo é relativizado à dimensão da descontinuidade que se inspeciona. Sendo assim, se faz imperativa a quantificação do quão confiáveis são as inspeções realizadas para monitorar estruturas e equipamentos. Um método amplamente utilizado para quantificar a confiabilidade de ensaios não destrutivos é através de curvas de probabilidade de detecção (1). Porém, é tarefa nada simples e extremamente cara realizar estudos de confiabilidade através de curvas PoD pois existe uma quantidade mínima de inspeções a serem realizadas bem como há requisitos básicos em relação a quantidade e variabilidade de defeitos projetados e inseridos em corpos de prova.

Nesse cenário acima descrito, surge um campo fértil para o uso de ferramentas como modelagem computacional e simulação. Seria extremamente desejável poder não só planejar o experimento de inspeção através de simulação computacional, mas também poder obter indicações de quais e como parâmetros de inspeção afetariam a curva de probabilidade de detecção, sempre com o intuito de otimizar o processo de quantificação de confiabilidade. Entretanto, para tanto, é necessário o processo de validação dessas curvas simuladas e, como primeiro passo, determinar o grau de aleatoriedade dos dados gerados.

Curvas de POD são, como toda análise probabilística, fortemente dependentes do número de observações bem como de parâmetros de incerteza que conferem ao estudo o grau de aleatoriedade que a análise de confiabilidade precisa ter. No caso de dados experimentais, o próprio procedimento prático já confere tais parâmetros de incerteza, uma vez que há parâmetros os quais não se pode afirmar, ou sequer estimar, seus exatos valores. Quando as curvas PoD são baseadas em dados simulados, o próprio usuário do software em questão insere no ambiente computacional os parâmetros que julga serem incertos. Uma vez definidos tais parâmetros, bem como o número de experimentos o usuário deseja que sejam realizados, cabe ao software realizar os experimentos. Porém, parte desses cálculos, como veremos a seguir, são probabilísticos. Sendo assim, é necessário que sejam gerados números pseudo-aleatórios para que essa análise probabilística seja realizada. O conjunto de números gerados são chamados de pseudo-aleatórios pois não são aleatórios puros. São pseudo-aleatórios os números gerados a partir de uma semente e, de posse dessa semente, cálculos matemáticos são realizados para que essa semente seja transformada num conjunto de números não correlacionados. Ou seja, o processo de geração de números pseudo-aleatórios é determinístico. Ora, se o que se pretende é realizar uma análise probabilística de dados simulados, nada mais essencial do que perguntar se esse conjunto de dados pseudo-aleatórios gerados de maneira determinística são, de fato, não correlacionados o suficiente.

O software utilizado neste trabalho, o CIVA, é um software de licença comercial, distribuído pela empresa Extende, com o objetivo original de planejar e otimizar os processos de inspeção em componentes nucleares (2). Ele utiliza métodos de cálculos semi-analíticos e apresenta diferentes módulos de simulação de Ensaios Não Destrutivos (END) tais como ultrassom, correntes parasitas e radiografia. No que diz

respeito à simulação de inspeções, o CIVA já está bastante consolidado no meio acadêmico e industrial. Há validações de suas simulações em relação a todos os tipos de END. O processo de simulação de inspeções é basicamente determinístico uma vez que todos os métodos de END podem ser modelados e seus fenômenos físicos são bastante conhecidos. Contudo, quando curvas PoD são geradas a partir desses resultados de inspeções obtidas por simulações, a parcela probabilística do software se faz presente.

O CIVA utiliza método de Monte-Carlo para geração de seus dados pseudo-aleatórios. Método de Monte-Carlo (MC) pode ser descrito, de uma maneira muito geral, como sendo um método matemático destinado a resolver problemas matemáticos a partir de amostragem (3). Aplicando esse conceito geral ao caso de geração de números pseudo-aleatórios, pode-se dizer que o método de MC é um algoritmo usado para gerar números não correlacionados a partir de uma semente randômica. Desta forma, o que será realizado neste trabalho é, a partir de prévias simulações de inspeções por ultrassom no software CIVA e geração de curvas PoD, também no ambiente CIVA, analisar através de testes estatísticos a aleatoriedade dos conjuntos de parâmetros de incerteza utilizados para construção das curvas. Isso fornece uma primeira análise do quão confiáveis são estudos de confiabilidade realizados com inspeções simuladas. Uma vez determinado se os conjuntos de dados ditos pseudo-aleatórios são de fato não correlacionados, é possível seguir com toda a validação seguinte de curvas PoD geradas por inspeções simuladas.

## 2 DESENVOLVIMENTO

Ao inserir o sólido virtual no software de simulação, as características geométricas e acústicas do corpo de prova real foram respeitadas. O fator geométrico dos defeitos escolhido para gerar as curvas PoD foi a altura. As alturas variavam de 0,35mm a 2,1mm. Após realizar as inspeções no software de simulação, o passo seguinte foi gerar as curvas PoD. O que o software retorna como resultado da geração das curvas, além de uma representação gráfica, é uma tabela com os valores de altura e amplitude correspondentes, conforme pode ser observado na Figura 1. A tabela traz os valores dos parâmetros de incerteza determinados com seus valores e são esses valores de resposta do software relativos aos parâmetros de incerteza que serão analisados quanto à aleatoriedade. Para o presente estudo, foram determinados como parâmetros de incerteza as inclinações do defeito em relação aos eixos x, y e z, chamados respectivamente de *disorientation*, *tilt* e *skew*.

Na prática, a rotação de defeitos em relação aos seus eixos é algo que é extremamente difícil de ser quantificado e ao mesmo tempo é determinante, em casos específicos, para a detectabilidade do defeito. Portanto, definir como parâmetros de incerteza a *disorientation*, *tilt* e o *skew* é bastante representativo da realidade.

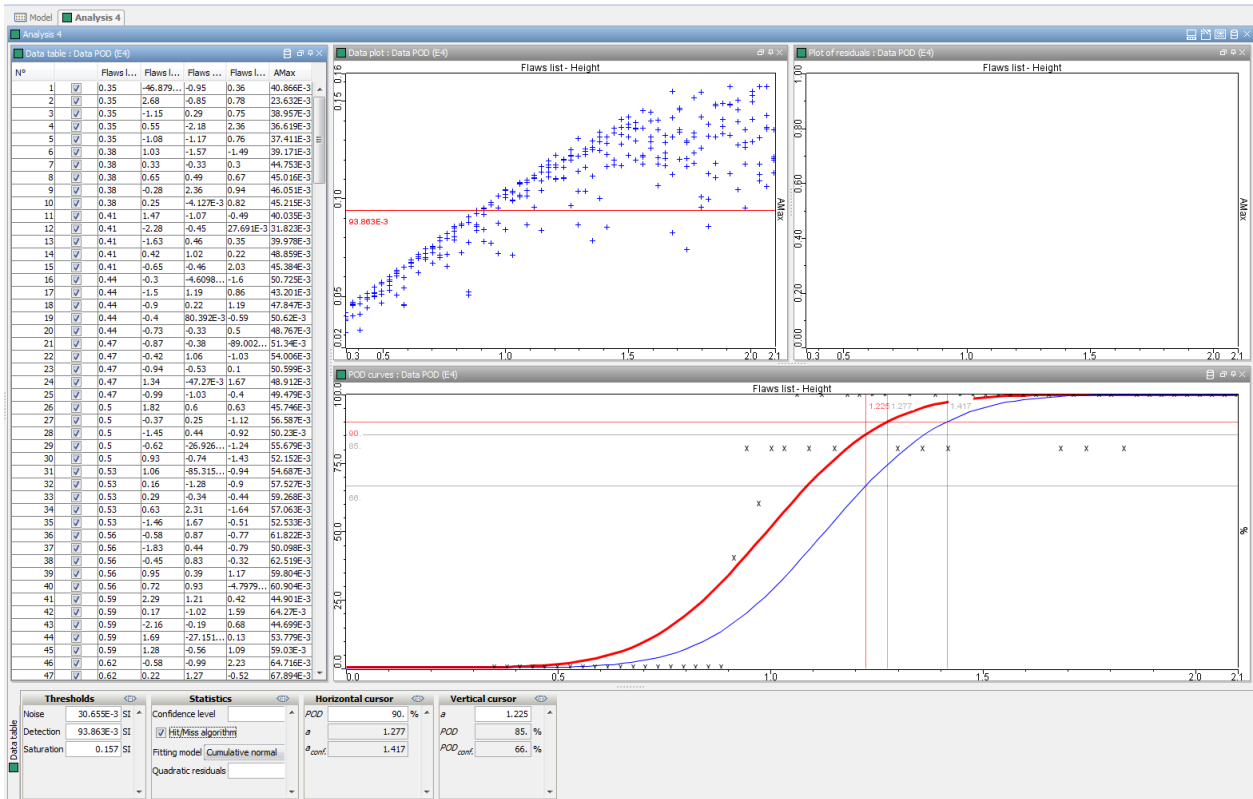


Figura 1 – Exemplo de resultado de PoD gerado pelo CIVA: à esquerda uma tabela com todos os valores de altura, amplitude e os valores gerados para os parâmetros de incerteza. À direita, as representações gráficas da curva PoD.

Uma vez que os parâmetros de incerteza são definidos, o usuário também deve definir qual tipo de função de distribuição de probabilidade (PDF) eles seguirão. Foram realizados dois testes: o primeiro definindo as PDFs dos três parâmetros de incerteza como sendo normal, ou Gaussiana e o segundo, definindo as PDFs uniformes. O objetivo desses testes foi avaliar o grau de aleatoriedade para outros casos de geração de dados com o intuito de aumentar a variabilidade dos resultados. Sendo assim, são obtidos seis conjuntos de dados a serem testados:

- Disorientation com PDF normal
- Disorientation com PDF uniforme
- Tilt com PDF normal
- Tilt com PDF uniforme
- Skew com PDF normal
- Skew com PDF uniforme

Uma primeira abordagem gráfica para analisar qualquer tendência óbvia em um conjunto de dados é o gráfico de dispersão. Foram plotados os gráficos de dispersão para os três parâmetros de incerteza em função da altura dos defeitos. Os resultados para a PDF normal são mostrados nas Figuras 2, 3 e 4. Pode ser observado que não existe tendência clara no conjunto de dados analisados.

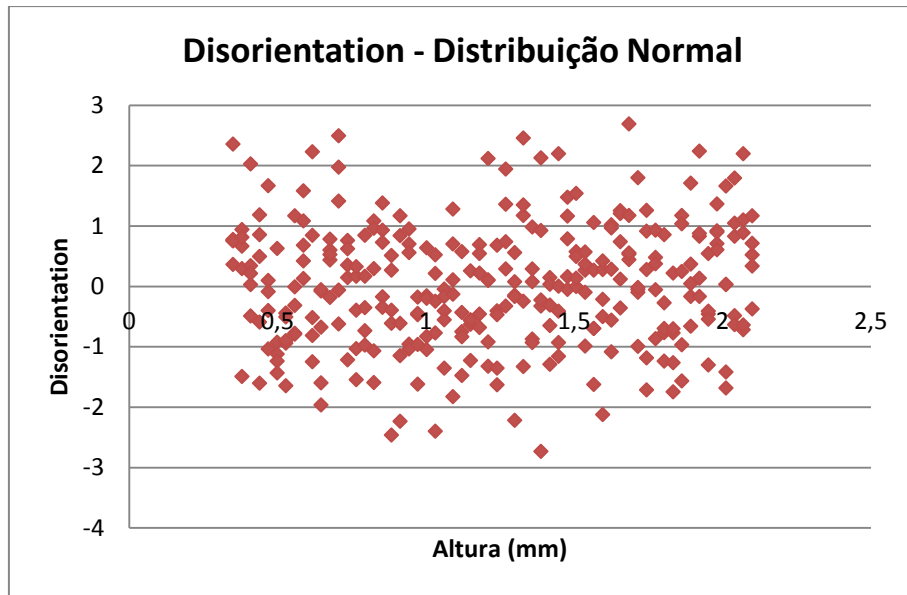


Figura 2 – Gráfico de dispersão Disorientation vs altura, sugerindo que não há tendência clara que relaciona os pontos gerados.

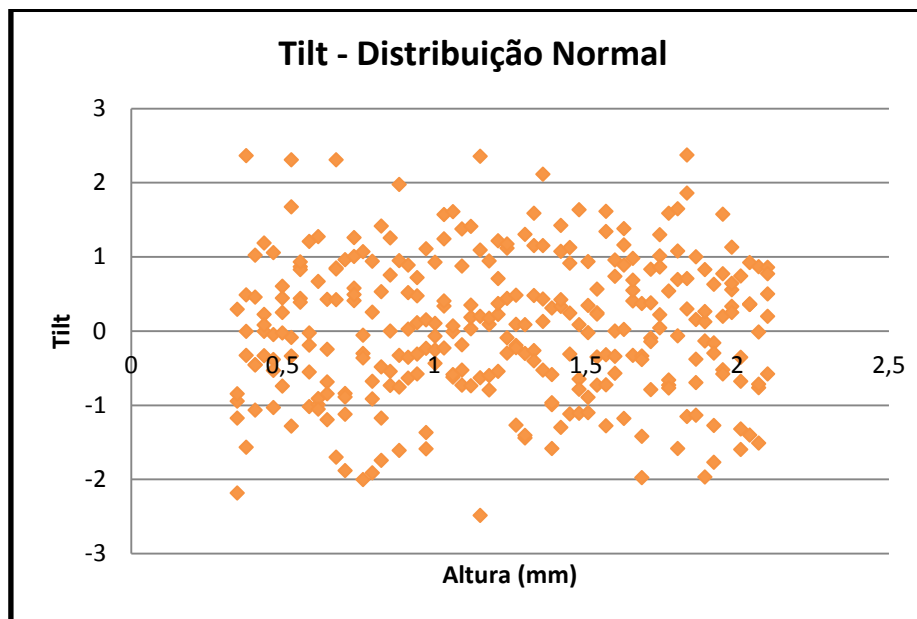


Figura 3 – Gráfico de dispersão Tilt vs altura, sugerindo que não há tendência clara que relaciona os pontos gerados.

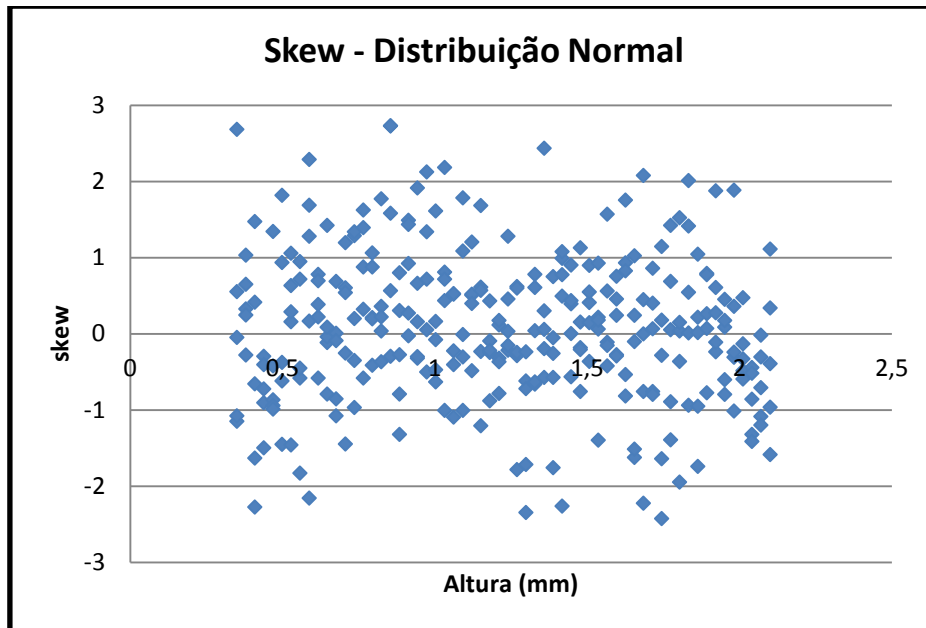


Figura 4 – Gráfico de dispersão Skew vs altura, sugerindo que não há tendência clara que relaciona os pontos gerados.

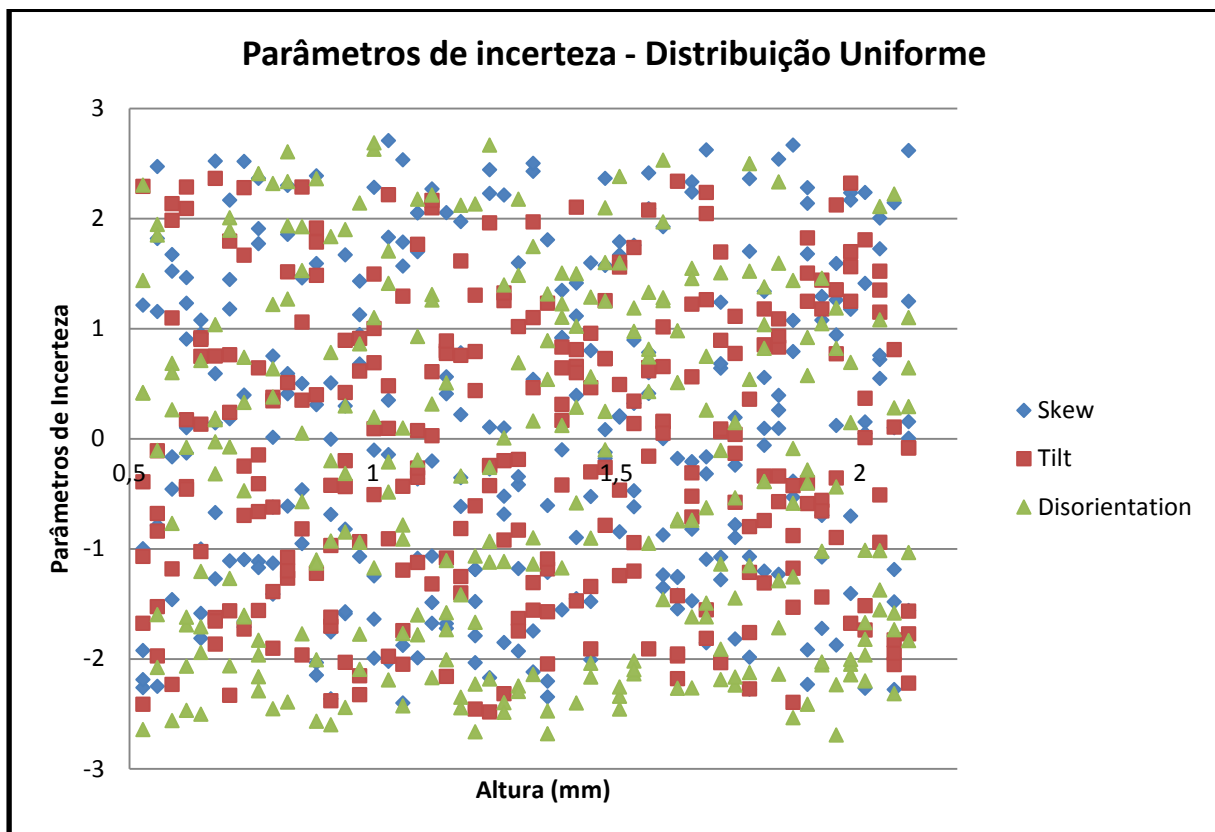


Figura 5 – Gráfico de dispersão dos três parâmetros de incerteza avaliados vs altura, segundo função de probabilidade uniforme, sugerindo que não há tendência clara que relaciona os pontos gerados.

Da mesma forma, ao analisar o gráfico de dispersão dos parâmetros de incerteza em função da altura agora em relação à PDF uniforme, obtém-se resultados equivalentes

aos anteriores. Não é possível, como mostrado na Figura 5, observar tendência entre os dados gerados. No caso da Figura 5, foram plotados os três parâmetros de incerteza juntos com o objetivo de comparar seus comportamentos.

Mas apenas atestar que não há tendência clara que relaciona os pontos gerados, não significa dizer que os dados são não correlacionados ou pseudo-aleatórios. É preciso realizar testes adicionais para que, numericamente, se verifique a aleatoriedade dos pontos. Para tanto, foi realizado o Run teste, que é um teste não-paramétrico e tem como objetivo avaliar se uma série ocorre aleatoriamente (4). Basicamente, consiste em determinar os valores acima e abaixo da mediana de uma série de dados ordenada. Admite-se a hipótese nula quando a distribuição dos dados é normal e o conjunto analisado é aleatório testando-se a seguinte estatística:

$$Z = \frac{u - E(u)}{\sqrt{\text{var}(u)}} \quad \text{Equação 1}$$

Os valores obtidos são comparados aos valores de z para distribuição normal com nível de significância de 5%. Caso os valores obtidos sejam superiores aos valores esperados de z (entre -1,96 e 1,96), rejeita-se a hipótese nula e o conjunto é considerado composto de dados não aleatórios. Vale lembrar que os conjuntos de dados originais são: disorientation, skew e tilt com distribuições de probabilidade normal e disorientation, skew e tilt com distribuições de probabilidade uniforme.

Para aplicar o Run teste no segundo grupo de dados, os que foram gerados com PDF uniforme, é necessário normalizar a sequência de dados, o que foi feito. Sendo assim, é possível aplicar o teste de aleatoriedade em questão nos dois blocos de dados dos parâmetros de incerteza simulados.

Os resultados obtidos podem ser observados na Tabela 1. O valores de Z evidenciam que a hipótese nula não pode ser rejeitada, uma vez que todos os valores calculados de Z estão no intervalo tabelado [-1,96; 1,96]. Ou seja, todos os conjuntos são compostos de dados pseudo-aleatórios.

Tabela 1 – Valores obtidos de Z evidenciando o caráter aleatório de todos os grupos de dados testados.

	PDF Normal			PDF Uniforme		
	Skew	Tilt	Disorientation	Skew	Tilt	Disorientation
Normal padronizada (Z)	-0,366	1,5566	0,045782056	-0,504	0,8699	-0,778294951

### 3 CONCLUSÕES

Em resumo, o que se pretendia era avaliar a aleatoriedade, ou pseudo-aleatoriedade, dos números gerados pelo software CIVA que foram usados como parâmetros de incerteza para construção de curvas de probabilidade simuladas. Para tanto, foram considerados como parâmetros de incerteza as orientações de defeito virtual que foram computacionalmente inspecionados por ultrassom: tilt, skew e disorientation. Tais

parâmetros de incerteza apresentavam, por sua vez, PDFs Normal e Uniforme. Uma primeira análise qualitativa através de gráficos de dispersão, sugeriu que todos os grupos de dados não apresentaram correlação clara entre seus valores constituintes. Mas para quantificar a não correlação entre os dados gerados, foi necessária uma análise estatística mais formal. Aplicando o Run Teste, foi possível rejeitar a hipótese de nulidade através dos valores de Z e, pôde-se afirmar que, de fato, os dados gerados pelo software através de MC são pseudo-aleatórios e toda a etapa de validação de curvas PoD simuladas pode ser, desta forma, desenvolvida.

## REFERÊNCIAS

- 1 GEORGIU, G.A. “PoD Curves, their Derivation, Application and Limitations”. In: Insight, vol. 49, n.7, p. 409-413, 2007
- 2 <http://www.extende.com/> - Última visualização em 10/03/2017 às 12:34.
- 3 <http://statweb.stanford.edu/~owen/mc/> - Última visualização em 10/03/2017 às 12:36.
- 4 THOM, H. C. S. “Some methods of climatological analysis”. Genève : World Meteorological Organization, 1966. 54 p. (WMO Technical Note, 81).