

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM SANEAMENTO,**  
**MEIO AMBIENTE E RECURSOS HÍDRICOS**

**ASSIMILAÇÃO DO PADRÃO DE  
VARIABILIDADE DAS VARIÁVEIS DE ESTADO  
DE UM MODELO CHUVA-VAZÃO EM  
ESQUEMAS DE SIMULAÇÃO / PREVISÃO  
HIDROLÓGICA**

**Francisco Eustáquio Oliveira e Silva**

**Belo Horizonte**

**2015**

**ASSIMILAÇÃO DO PADRÃO DE VARIABILIDADE  
DAS VARIÁVEIS DE ESTADO DE UM MODELO  
CHUVA-VAZÃO EM ESQUEMAS DE SIMULAÇÃO /  
PREVISÃO HIDROLÓGICA**

**Francisco Eustáquio Oliveira e Silva**



**Francisco Eustáquio Oliveira e Silva**

**ASSIMILAÇÃO DO PADRÃO DE VARIABILIDADE  
DAS VARIÁVEIS DE ESTADO DE UM MODELO  
CHUVA-VAZÃO EM ESQUEMAS DE SIMULAÇÃO /  
PREVISÃO HIDROLÓGICA**

Trabalho apresentado ao Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da Universidade Federal de Minas Gerais, como requisito à obtenção do título de Doutor em Saneamento, Meio Ambiente e Recursos Hídricos.

Área de concentração: Recursos Hídricos

Linha de pesquisa: Modelagem de processos hidrológicos

Orientador: Mauro da Cunha Naghettini

Belo Horizonte

Escola de Engenharia da UFMG

2015

S586a

Silva, Francisco Eustáquio Oliveira e.

Assimilação do padrão de variabilidade das variáveis de estado de um modelo chuva-vazão em esquemas de simulação / previsão hidrológica [manuscrito] / Francisco Eustáquio Oliveira e Silva.- 2015. xi, 211 f.: il.

Orientador: Mauro da Cunha Naghettini.

Tese (doutorado) – Universidade Federal de Minas Gerais, Escola de Engenharia.

Anexos: f. 131-211.

Bibliografia: f. 121-130.

1. Engenharia sanitária- Teses. 2. Recursos hídricos - Desenvolvimento - Teses. I. Naghettini, Mauro. II. Universidade Federal de Minas Gerais. Escola de Engenharia. III. Título.

CDU: 628(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS

Escola de Engenharia

*Programa de Pós-Graduação em Saneamento, Meio Ambiente e Recursos Hídricos*

*Avenida Antônio Carlos, 6627 - 4º andar - 31270-901 - Belo Horizonte – BRASIL*

*Telefax: 55 (31) 3409-1882 - posgrad@desa.ufmg.br*

*http://www.smarh.eng.ufmg.br*

## FOLHA DE APROVAÇÃO

Assimilação do padrão de variabilidade das variáveis de estado de um modelo chuva vazão, em esquemas de simulação/previsão hidrológica

**FRANCISCO EUSTÁQUIO OLIVEIRA E SILVA**

Tese defendida e aprovada pela banca examinadora constituída pelos Senhores:

Prof. MAURO DA CUNHA NAGHETTINI - Orientador

Prof. LUIZ RAFAEL PALMIER

Prof. NILO DE OLIVEIRA NASCIMENTO

Prof. DIRCEU SILVEIRA REIS JUNIOR

Prof. ALEXANDRE KOLODYNSKIE GUETTER

Aprovada pelo Colegiado do PG SMARH

Prof. Nilo de Oliveira Nascimento  
Coordenador

Versão Final aprovada por

Prof. Mauro da Cunha Naghettini  
Orientador

Belo Horizonte, 31 de agosto de 2015.

## AGRADECIMENTOS

Ao terminar esta tese de doutorado resta-me registrar os meus sinceros agradecimentos às pessoas que de várias formas contribuíram para que se tornasse uma realidade.

À minha esposa Maria Helena e minha filha Alice, pela paciência e compreensão nas noites e finais de semana em que me dediquei à elaboração deste trabalho, sendo sempre meu porto seguro nas horas incertas e minha fonte de alegria e sossego nos momentos difíceis.

Aos meus pais e a minha irmã, pelo constante incentivo ao meu crescimento profissional. Pela compreensão da minha crescente ausência em “nossa” casa.

Aos pais de minha esposa, Antônio e Célia, aos seus respectivos companheiros, Cássia e Nelson, e aos meus cunhados José Eduardo, Fernando e Isadora pelo constante incentivo e imensuráveis ajudas ao longo desse período.

Ao meu orientador, Professor Mauro Naghettini, por toda a dedicação, compreensão e amizade. Pelos desafios que me impôs ao longo do trabalho, pelo estímulo e exigência de qualidade.

Aos amigos do Departamento de Engenharia Hidráulica e Recursos Hídricos: Jussara; Profs: Wilson, Palmier, Nilo, Julian, Éber, Priscilla, Talita, Márcio, Márcia, Veber; e alunos do SMARH: Fernando, Sara, Aline, Hildelano, Eric, Carol e Daniel.

Aos colegas e amigos da POTAMOS Engenharia: Mário, Rodrigo, Márcio, Rossana, Fernando, Alessandra, Frederico, Rodney, Givanildo, Bruno, Tainá, Thiago, Adolfo, Marlon's, Stéphanie, Isabela, Geórgia, Ana Maria, Viviane, Manoela, Filipe e Denis.

Aos colegas e amigos de outras empresas: Marcelo, Flávia, Márcio, Elizabeth, Alice, Letícia, Lucas, Osmar, Gustavo, João e Carlos Augusto.

Aos membros da banca examinadora.

## RESUMO

Modelos chuva-vazão conceituais se propõem a captar a dinâmica do ciclo hidrológico a partir da representação dos principais processos físicos que ocorrem na escala da bacia hidrográfica. A complexidade do processo natural de transformação da chuva em vazão impõe a esses modelos, independentemente de suas estruturas, que a representação do fenômeno ocorra de forma simplificada. Essas simplificações, associadas à complexa representação espaço-temporal do fenômeno, corroboram para o crescente consenso da comunidade técnica acerca da necessidade de se avaliar as incertezas associadas aos prognósticos desses modelos.

O presente estudo orientou-se por essa percepção de que, dada a importância e necessidade da utilização desses modelos, é necessário o desenvolvimento de técnicas que permitam a identificação e redução das incertezas associadas aos seus prognósticos. Para tal, a exemplo de diversos estudos desenvolvidos na última década, foram utilizadas técnicas de inferência Bayesiana, em conjunto com métodos de simulação de Monte Carlo, para determinação das incertezas associadas às estimativas de parâmetros e variáveis de estado presentes na estrutura de um modelo chuva-vazão conceitual.

Analisou-se, na sequência, o padrão de variabilidade das variáveis de estado desse modelo. A hipótese desta tese é que o comportamento das variáveis de estado é capaz de caracterizar os erros na série modelada, já que elas assimilam, a cada passo de tempo, todas as fontes de incertezas associadas à modelagem. A análise do comportamento *input-state-output* foi desenvolvida a partir da formulação de dois procedimentos para a atualização das variáveis de estado do modelo chuva-vazão considerado.

Procedeu-se uma avaliação detalhada do impacto da atualização das variáveis de estado sobre períodos específicos da série. Assim, foi possível contrapor algumas das abordagens metodológicas e conclusões descritas nas principais referências atuais sobre o tema. Acredita-se que as conclusões aqui descritas possam ser empregadas para fundamentar o desenvolvimento de técnicas que permitam reduzir a incerteza global associada ao emprego de modelos chuva-vazão, tanto em esquemas de simulação quanto de previsão hidrológica.

## ABSTRACT

Conceptual rainfall-runoff models are meant to capture the dynamics of the hydrological cycle through the representation of the main physical processes that occur at the river basin scale. The complexity of the natural process of transformation of rain into flow enforces these models, regardless of their structures, given that the representation of phenomenon is made in a simplified manner. These simplifications associated with the complex spatiotemporal representation of the phenomenon, corroborate to the growing consensus of the technical community about the need to assess the uncertainties associated with the predictions of these models.

This study was guided by the perception that given the importance and necessity of using these models, the development of techniques to enable the identification and reduction of the uncertainty associated with its predictions is necessary. For this, as in various studies undertaken in the last decade, we used Bayesian inference techniques, along with Monte Carlo simulation methods for determining the uncertainty associated with estimates of parameters and state variables present in the structure of a conceptual rainfall-runoff model.

In the sequence, the pattern of variability of the state variables of this model was analyzed. The hypothesis of this thesis is that the behavior of state variables is able to characterize the errors in the modeled series, since they assimilate all sources of uncertainties associated with the modeling in each time step. The analysis of *input-state-output* behavior has been developed from the formulation of two procedures for updating the state variables of the rainfall-runoff model evaluated.

A detailed impact assessment of state variables update was developed, on specific periods of the series. It was possible to confront some of the methodological approaches and conclusions described in the main current references on the subject. It is believed that the conclusions described herein can be used to support the development of techniques to reduce the overall uncertainty associated with the use of rainfall-runoff models, in both hydrological simulation and forecasting schemes.

# SUMÁRIO

<b>LISTA DE FIGURAS .....</b>	<b>V</b>
<b>LISTA DE TABELAS .....</b>	<b>VII</b>
<b>LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS .....</b>	<b>VIII</b>
<b>1 INTRODUÇÃO .....</b>	<b>1</b>
<b>2 OBJETIVOS .....</b>	<b>6</b>
2.1 OBJETIVO GERAL .....	6
2.2 OBJETIVOS ESPECÍFICOS .....	6
<b>3 REVISÃO DA LITERATURA .....</b>	<b>7</b>
3.1 MODELOS CHUVA-VAZÃO .....	7
3.1.1 <i>Classificação de modelos chuva-vazão</i> .....	8
3.1.2 <i>Modelagem de processos hidrológicos</i> .....	12
3.1.3 <i>Calibração de parâmetros</i> .....	15
3.2 INCERTEZAS NA MODELAGEM CHUVA-VAZÃO .....	19
3.3 ASSIMILAÇÃO DE DADOS .....	31
3.3.1 <i>Filtro de Kalman e suas variações</i> .....	35
3.3.2 <i>Filtros bayesianos</i> .....	38
3.4 COMENTÁRIO GERAL .....	40
<b>4 METODOLOGIA .....</b>	<b>42</b>
4.1 BACIA HIDROGRÁFICA E INFORMAÇÕES HIDROLÓGICAS .....	44
4.2 MODELO CHUVA-VAZÃO .....	45
4.3 INFERÊNCIA BAYESIANA POR MEIO DE SIMULAÇÕES DE MONTE CARLO .....	59
4.3.1 <i>O método DREAM</i> .....	59
4.3.2 <i>Modelos probabilísticos para os resíduos</i> .....	65
4.4 IDENTIFICAÇÃO DE PERÍODOS DE DISSIMILARIDADE ENTRE VAZÕES MODELADAS E OBSERVADAS .....	70
4.5 ESTIMAÇÃO DAS VARIÁVEIS DE ESTADO E ANÁLISE DAS ALTERAÇÕES EM SUAS TRAJETÓRIAS .....	73
<b>5 RESULTADOS E DISCUSSÃO .....</b>	<b>77</b>
5.1 DADOS DE ENTRADA E PARTICULARIDADES DA SIMULAÇÃO .....	77
5.2 AVALIAÇÃO DOS MODELOS PROBABILÍSTICOS ADMITIDOS PARA OS RESÍDUOS .....	81
5.3 INCERTEZAS NAS ESTIMATIVAS DOS PARÂMETROS .....	85
5.4 PREDIÇÃO DA INCERTEZA NA VAZÃO MODELADA .....	89
5.5 ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO .....	90
5.5.1 <i>Identificação dos períodos de dissimilaridade persistente</i> .....	92
5.5.2 <i>Estimação das variáveis de estado e análise das alterações em suas trajetórias</i> .....	94
5.6 PREDIÇÃO DE INCERTEZA APLICADA À MODELAGEM CHUVA-VAZÃO: SINOPSE DAS ABORDAGENS ATUAIS .....	110
5.6.1 <i>Abordagens que indicam as incertezas associadas aos dados de entrada/saída</i> .....	110
5.6.2 <i>Abordagens que preconizam a estimação conjunta de parâmetros e variáveis de estado</i> .....	112
5.6.3 <i>Abordagens que apregoam o desenvolvimento de algoritmos cada vez mais eficientes/robustos</i> ...	114
5.6.4 <i>Abordagem “propositiva”</i> .....	115
5.7 DISCUSSÃO GERAL .....	116
<b>6 CONCLUSÕES E RECOMENDAÇÕES .....</b>	<b>119</b>
<b>REFERÊNCIAS .....</b>	<b>121</b>
<b>ANEXO I – MAGNITUDE DA VARIAÇÃO DAS VARIÁVEIS DE ESTADO ATUALIZADAS .....</b>	<b>131</b>
<b>ANEXO II – ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO: RESULTADOS GRÁFICOS .....</b>	<b>136</b>

# LISTA DE FIGURAS

<b>FIGURA 1.1</b> – IMPACTO DE UM PROCEDIMENTO DE ATUALIZAÇÃO DOS ESTADOS DE UM MODELO CHUVA-VAZÃO (FONTE: ADAPTADO DE BERTHET <i>ET AL.</i> , 2009).....	3
<b>FIGURA 3.1</b> – REPRESENTAÇÃO DOS PROCESSOS ENVOLVIDOS EM UM MODELO DE PERCEPÇÃO DOS FENÔMENOS HIDROLÓGICOS EM UMA ENCOSTA (FONTE: ADAPTADO DE BEVEN, 2012). ....	9
<b>FIGURA 3.2</b> – CLASSIFICAÇÃO DOS MODELOS HIDROLÓGICOS BASEADA NA ESTRUTURA UTILIZADA PARA DESCRIÇÃO DOS PROCESSOS FÍSICOS (FONTE: ADAPTADO DE REFSGAARD, 1996 <i>APUD</i> SHRESTHA, 2009). ....	11
<b>FIGURA 3.3</b> – DISCRIMINAÇÃO DOS PROCESSOS HIDROLÓGICOS REPRESENTADOS EM UM MODELO CHUVA-VAZÃO ATRAVÉS DE FASES: BACIA E CANAL (FONTE: ADAPTADO DE NETO <i>ET AL.</i> , 2007). ....	12
<b>FIGURA 3.4</b> – REPRESENTAÇÃO ESQUEMÁTICA DOS PROCESSOS HIDROLÓGICOS EM UMA BACIA HIDROGRÁFICA (FONTE: ADAPTADO DE BECKER, 1992 <i>APUD</i> WAGENER <i>ET AL.</i> , 2004). ....	13
<b>FIGURA 3.5</b> – ESCOAMENTO SUPERFICIAL EM UMA ENCOSTA PRODUZIDO PELO EXCESSO DE CHUVAS SOBRE A CAPACIDADE DE INFILTRAÇÃO (FONTE: ADAPTADO DE CHOW <i>ET AL.</i> , 1988).....	15
<b>FIGURA 3.6</b> – REPRESENTAÇÃO ESQUEMÁTICA DE UM PROBLEMA DIRETO E DE UM PROBLEMA INVERSO (FONTE: ADAPTADO DE MORADKHANI E SOROOSHIAN, 2008). ....	18
<b>FIGURA 3.7</b> – COMPONENTES E FONTES DE INCERTEZA DE UM MODELO CHUVA-VAZÃO (ADAPTADO DE LIU E GUPTA, 2007). ....	20
<b>FIGURA 3.8</b> – REPRESENTAÇÃO ESQUEMÁTICA DA SIMULAÇÃO DE MONTE CARLO PARA ANÁLISE DE INCERTEZAS (ADAPTADO DE SHRESTHA, 2009). ....	27
<b>FIGURA 3.9</b> – REPRESENTAÇÃO GRÁFICA DO TEOREMA DE BAYES (ADAPTADO DE KAVETSKI <i>ET AL.</i> , DATA DESCONHECIDA). ....	28
<b>FIGURA 3.10</b> – AVALIAÇÃO DAS INCERTEZAS SOBRE A PERSPECTIVA DA CALIBRAÇÃO DE MODELOS CHUVA-VAZÃO (ADAPTADO DE VRUGT <i>ET AL.</i> , 2008A).....	29
<b>FIGURA 3.11</b> – DIAGRAMA ESQUEMÁTICO DO PROCEDIMENTO DE ASSIMILAÇÃO DE DADOS (ADAPTADO DE SERBAN E ASKEW, 1991 <i>APUD</i> WMO, 2011; E PERRIN, 2003 <i>APUD</i> WMO, 2011). ....	33
<b>FIGURA 3.12</b> – TÉCNICAS DE ASSIMILAÇÃO: (A) SEQUENCIAL; (B) DINÂMICA (ADAPTADO DE HOUSER <i>ET AL.</i> , 2012). ....	34
<b>FIGURA 4.1</b> – LOCALIZAÇÃO DA BACIA E ESTAÇÕES DE MONITORAMENTO ANALISADAS ( <i>DATUM</i> : WGS84). ....	44
<b>FIGURA 4.2</b> – VAZÕES E PRECIPITAÇÕES MÉDIAS NO PERÍODO NA BACIA ANALISADA: (I) MÉDIA ANUAL; (II) MÉDIA MENSAL; (III) VARIABILIDADE MENSAL DA PRECIPITAÇÃO; (IV) VARIABILIDADE MENSAL DA VAZÃO. ....	45
<b>FIGURA 4.3</b> – FLUXOGRAMA DO MÓDULO DE PRODUÇÃO DO MODELO RIO GRANDE. ....	47
<b>FIGURA 4.4</b> – DISTRIBUIÇÃO ESPACIAL DA CAPACIDADE DE RETENÇÃO DE ÁGUA DO SOLO. ....	48
<b>FIGURA 4.5</b> – DISTRIBUIÇÃO ESPACIAL DA CAPACIDADE DE ÁGUA LIVRE.....	50
<b>FIGURA 4.6</b> – INFLUÊNCIA DA DISTRIBUIÇÃO PROPOSTA SOBRE AS AMOSTRAS OBTIDAS NO ALGORITMO DE <i>METROPOLIS</i> : PAINEL A ESQUERDA: $N(x, \sigma = 1)$ ; (II) PAINEL CENTRAL: DISTRIBUIÇÃO MUITO RESTRITA – $N(x, \sigma = 0,025)$ ; E (III) PAINEL A DIREITA: DISTRIBUIÇÃO MUITO AMPLA – $N(x, \sigma = 50)$ (ADAPATADO DE: NAVARRO E PERFOR, DATA DESCONHECIDA). ....	61
<b>FIGURA 5.1</b> – HIDROGRAMA UNITÁRIO DE DURAÇÃO IGUAL A 1H CALCULADO PARA O CASO DE ESTUDO.....	78
<b>FIGURA 5.2</b> – SÉRIE TEMPORAL DE VAZÕES OBSERVADAS E SIMULADAS ADMITINDO-SE A NORMALIDADE DOS RESÍDUOS.....	81
<b>FIGURA 5.3</b> – COMPORTAMENTO DOS RESÍDUOS ADMITIDOS NORMAIS DURANTE A CALIBRAÇÃO.....	82
<b>FIGURA 5.4</b> – SÉRIE TEMPORAL DE VAZÕES OBSERVADAS E SIMULADAS ADMITINDO-SE O MODELO GL.....	83
<b>FIGURA 5.5</b> – COMPORTAMENTO DOS RESÍDUOS ADMITIDOS HETEROCEDÁSTICOS, AUTOCORRELACIONADOS E MODELADOS PROBABILISTICAMENTE SEGUNDO A FDP DE LAPLACE ASSIMÉTRICA DURANTE A CALIBRAÇÃO.....	84
<b>FIGURA 5.6</b> – DISTRIBUIÇÕES A <i>POSTERIORI</i> DOS PARÂMETROS DO MODELO RIO GRANDE. ....	86
<b>FIGURA 5.7</b> – COMPORTAMENTO DA FUNÇÃO DE VEROSSIMILHANÇA PARA O PARÂMETRO <i>KG</i> : CALIBRAÇÃO (ACIMA) E VALIDAÇÃO (ABAIXO). ....	88



<b>FIGURA 5.8</b> – COMPORTAMENTO DA FUNÇÃO DE VEROSSIMILHANÇA PARA O PARÂMETRO <i>C</i> : CALIBRAÇÃO (ACIMA) E VALIDAÇÃO (ABAIXO). .....	88
<b>FIGURA 5.9</b> – INCERTEZA NA PREDIÇÃO DA VAZÃO MODELADA DURANTE O PERÍODO DE VALIDAÇÃO UTILIZANDO-SE O MODELO DE RESÍDUOS GL – SÉRIE TEMPORAL OBSERVADA (LARANJA) E INTERVALO COM 95% DE INCERTEZA: (I) DECORRENTE DAS ESTIMATIVAS NOS PARÂMETROS (CINZA CLARO); E (II) TOTAL (CINZA ESCURO). .....	89
<b>FIGURA 5.10</b> – DIFERENÇA ENTRE OS LIMITES SUPERIOR (97,5%) E INFERIOR (2,5%) DOS INTERVALOS DE CREDIBILIDADE ( <i>DIC</i> ): DA VARIÁVEL DE ESTADO <i>Q<sub>i</sub></i> VERSUS A DO ERRO NA VAZÃO SIMULADA. ....	91
<b>FIGURA 5.11</b> – DISPERSÃO DOS RESÍDUOS DE UM MODELO DE REGRESSÃO LINEAR MÚLTIPLA CONSTRUÍDO PARA A ESTIMAÇÃO DA MEDIANA DA SÉRIE DE ERROS. ....	91
<b>FIGURA 5.12</b> – EVENTOS DE DISSIMILARIDADE SELECIONADOS PARA ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO.....	94
<b>FIGURA 5.13</b> – COMPORTAMENTO PROBABILÍSTICO EXIBIDO PELOS RESÍDUOS DA SIMULAÇÃO APÓS A ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO: <i>LAG</i> = 0 DIAS E JANELA ADMITIDA = 1X. ....	97
<b>FIGURA 5.14</b> – COMPARAÇÃO DO COMPORTAMENTO EXIBIDO PELA RAIZ DO ERRO MÉDIO QUADRÁTICO, NOS INTRAPERÍODOS SELECIONADOS, QUANDO ADMITIDO O PROCEDIMENTO PARA ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO BASEADO EM PERCENTUAIS. ....	99
<b>FIGURA 5.15</b> – COMPARAÇÃO DO COMPORTAMENTO EXIBIDO PELA RAIZ DO ERRO MÉDIO QUADRÁTICO, NOS INTRAPERÍODOS SELECIONADOS, QUANDO ADMITIDO O PROCEDIMENTO PARA ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO BASEADO EM MULTIPLICADORES.....	100
<b>FIGURA 5.16</b> – PERSISTÊNCIA DA ALTERAÇÃO PONTUAL DA VARIÁVEL DE ESTADO NO TEMPO. ....	102
<b>FIGURA 5.17</b> – AVALIAÇÃO DA QUALIDADE DO AJUSTE, EM JANELAS DE 10 DIAS, DURANTE O PERÍODO NO QUAL A ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO IMPÕE AS VAZÕES PREVISTAS E ATUALIZADAS DIFERENÇAS SUPERIORES A 5%: EVENTO 04, <i>LAG</i> = 0 DIAS, JANELA ADMITIDA = 1X, PROCEDIMENTO DE ATUALIZAÇÃO BASEADO EM MULTIPLICADORES.....	103
<b>FIGURA 5.18</b> – COMPARAÇÃO DO COMPORTAMENTO EXIBIDO PELA RAIZ DO ERRO MÉDIO QUADRÁTICO, CALCULADO EM JANELAS DE 10 DIAS, PARA OS PERÍODOS EM QUE OCORREU UMA MELHORA NA PERFORMANCE DO MODELO DEVIDO AO EMPREGO DE UM PROCEDIMENTO PARA ATUALIZAÇÃO DOS ESTADOS X.....	105
<b>FIGURA 5.18</b> – MAGNITUDE DA ALTERAÇÃO DOS AFLUXOS SUBSUPERFICIAIS À REDE DE DRENAGEM OBTIDOS A PARTIR DOS PROCEDIMENTOS DE ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO ANALISADOS. ....	107
<b>FIGURA 5.19</b> – MAGNITUDE DA ALTERAÇÃO DOS AFLUXOS SUBTERRÂNEOS À REDE DE DRENAGEM OBTIDOS A PARTIR DOS PROCEDIMENTOS DE ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO ANALISADAS. ....	108
<b>FIGURA 5.20</b> – EVENTO 03: VARIAÇÃO DAS VAZÕES PREVISTAS E ATUALIZADAS POR MEIO DE UM PROCEDIMENTO BASEADO EM PERCENTUAIS. ....	109

## LISTA DE TABELAS

<b>TABELA 3.1</b> – TÉCNICAS UTILIZADAS NA AVALIAÇÃO DA INCERTEZA NA MODELAGEM HIDROLÓGICA (ADAPTADO DE SHRESTHA, 2009 E MONTANARI, 2011). .....	23
<b>TABELA 4.1</b> – INFORMAÇÕES HIDROLÓGICAS EMPREGADAS NO ESTUDO.....	45
<b>TABELA 4.2</b> – PARÂMETROS DO MODELO RIO GRANDE. ....	58
<b>TABELA 4.3</b> – VARIÁVEIS DE ESTADO DO MODELO RIO GRANDE. ....	58
<b>TABELA 4.4</b> – INFORMAÇÕES NECESSÁRIAS À UTILIZAÇÃO DO MÉTODO DREAM. ....	65
<b>TABELA 4.5</b> – PARÂMETROS DAS FUNÇÕES DE VEROSSIMILHANÇA QUE DEVEM SER ESTIMADOS. ....	70
<b>TABELA 5.1</b> – PARÂMETROS DO MODELO RIO GRANDE ADMITIDOS A <i>PRIORI</i> . ....	77
<b>TABELA 5.2</b> – INTERVALOS DE INCERTEZAS ADMITIDOS A <i>PRIORI</i> PARA OS PARÂMETROS DAS FUNÇÕES DE VEROSSIMILHANÇA AVALIADAS.....	79
<b>TABELA 5.3</b> – VALORES DOS PARÂMETROS EMPREGADOS NA UTILIZAÇÃO DO MÉTODO DREAM. ....	80
<b>TABELA 5.4</b> – ESTATÍSTICAS DESCRITIVAS DOS RESULTADOS ADMITINDO-SE A NORMALIDADE DOS RESÍDUOS. ....	81
<b>TABELA 5.5</b> – ESTATÍSTICAS DESCRITIVAS DOS RESULTADOS ADMITINDO-SE O MODELO GL. ....	83
<b>TABELA 5.6</b> – PERÍODOS DE DISSIMILARIDADE SELECIONADOS PARA ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO. ....	93
<b>TABELA 5.7</b> – PERCENTUAIS DE MELHORA DA PERFORMANCE DO MODELO NO TEMPO PARA PERÍODOS SUBSEQUENTES À ATUALIZAÇÃO DOS ESTADOS $x_t$ , NO QUAL A DIFERENÇA ENTRE VAZÕES PREVISTAS E ATUALIZADAS SÃO SUPERIORES À 5%. ....	104

# LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

$au$	– ORDENADA QUE REPRESENTA O ESTADO INICIAL DO ARMAZENAMENTO DA ÁGUA LIVRE
$AI$	– FRAÇÃO DA ÁREA TOTAL
$AR(p)$	– MODELO AUTOREGRESSIVO
$AT$	– ÁREA TOTAL
$B$	– CONDIÇÕES DE CONTORNO DO MODELO HIDROLÓGICO
$b$	– UNIFORMIDADE DA CURVA QUE REPRESENTA A DISTRIBUIÇÃO ESPACIAL DA CAPACIDADE DE RETENÇÃO DE ÁGUA NO SOLO
$burn-in$	– PERÍODO DE AQUECIMENTO DA CADEIA
$c$	– CONSTANTE DE NORMALIZAÇÃO
$c$	– COEFICIENTE DE EVAPOTRANSPIRAÇÃO DA ZONA AERADA PROFUNDA
$C$	– CÓPULA
$cg$	– COEFICIENTE DE RECESSÃO DO FLUXO SUBTERRÂNEO DIÁRIO
$ci$	– COEFICIENTE DE RECESSÃO DO FLUXO SUB-SUPERFICIAL DIÁRIO
$cin$	– PULSOS DISCRETOS DE LÂMINAS DE ESCOAMENTO TOTAL
$cov(u,x)$	– COVARIÂNCIA ENTRE AS VARIÁVEIS $u$ E $x$
$C_t$	– COVARIÂNCIA DA FUNÇÃO GAUSSIANA PROPOSTA
$d$	– DIMENSÃO
$DE-MC$	– <i>DIFFERENTIAL EVOLUTION MARKOV CHAIN</i>
$DREAM$	– <i>DIFFERENTIAL EVOLUTION ADAPTIVE METROPOLIS</i>
$\epsilon$	– VALOR OBTIDIDO DE UMA DISTRIBUIÇÃO D-DIMENSIONAL SIMÉTRICA COM VARIÂNCIA PEQUENA
$e$	– ERRO
$e$	– EVAPOTRANSPIRAÇÃO EFETIVA TOTAL
$E_t$	– VAZÕES MÉDIAS CORRESPONDENTES ÀS OBSERVADAS AO LONGO DO TEMPO
$ed$	– EVAPORAÇÃO REAL NA ZONA PROFUNDA
$el$	– EVAPORAÇÃO REAL NA ZONA INFERIOR
$em$	– EVAPOTRANSPIRAÇÃO POTENCIAL
$eu$	– EVAPOTRANSPIRAÇÃO REAL OU EFETIVA NA ZONA SUPERIOR
$ev$	– TANQUE EVAPORIMÉTRICO
$E(f(\theta))$	– ESPERANÇA MATEMÁTICA DE $f(\theta)$ EM RELAÇÃO À DISTRIBUIÇÃO $p(\theta)$
$EKF$	– EXTEND KALMAN FILTER
$EnKF$	– ENSEMBLE KALMAN FILTER
$f$	– CAPACIDADE DE INFILTRAÇÃO (MM/H)
$F$	– DISTRIBUIÇÃO CONJUNTA DAS VARIÁVEIS ALEATÓRIAS $u$ E $x$
$FDA$	– FUNÇÃO DENSIDADE DE PROBABILIDADE ACUMULADA MULTIVARIADA
$FDP$	– FUNÇÃO DENSIDADE DE PROBABILIDADE
$f/F$	– FRAÇÃO DE ÁREA PERMEÁVEL ( $f$ ) EM RELAÇÃO A ÁREA TOTAL ( $F$ )
$func.type$	– ESPECIFICAÇÃO DA FUNÇÃO DE VEROSSIMILHANÇA A SER CONSIDERADA
$fr$	– FRAÇÃO DE ÁREA PERMEÁVEL GERADORA DO ESCOAMENTO
$FP$	– FORWARD PROBLEM
$fr$	– FRAÇÃO DE ÁREA PERMEÁVEL DA BACIA
$F_{ux}(u,x)$	– DISTRIBUIÇÃO CONJUNTA DE $u$ E $x$
$G$	– DISTRIBUIÇÃO CONJUNTA DAS VARIÁVEIS $S_u$ E $S_x$
$GL$	– MODELO GENERALIZADO
$H$	– OPERADOR DE OBSERVAÇÕES
$HU$	– HIDROGRAMA UNITÁRIO
$HTA$	– HISTOGRAMA TEMPO-ÁREA

$i$	– INTENSIDADE DE PRECIPITAÇÃO
$IQR$	– INTER-QUARTILE RANGE
$imp$	– FRAÇÃO DE ÁREA IMPERMEÁVEL
$IP$	– INVERSE PROBLEM
$J$	– FUNÇÃO CUSTO
$k$	– COEFICIENTE DO TANQUE EVAPORIMÉTRICO
$K$	– FILTRO DE KALMAN
$kg$	– COEFICIENTES DO FLUXO DIÁRIO DO ESCOAMENTO SUBTERRÂNEO
$kss$	– COEFICIENTES DO FLUXO DIÁRIO DO ESCOAMENTO SUB-SUPERFICIAL
$L(\theta \tilde{y})$	– FUNÇÃO DE VEROSSIMILHANÇA
$M$	– MODELO HIDROLÓGICO
$M_{\theta}(x_t x_{t-1})$	– FUNÇÃO DENSIDADE DE PROBABILIDADE DE TRANSIÇÃO DO MODELO DE ESTADOS
$MAE$	– ERRO ABSOLUTO MÉDIO
$MCMC$	– SIMULAÇÃO DE MONTE CARLO VIA CADEIAS DE MARKOV
$MH$	– <i>METROPOLIS-HASTINGS</i>
$n$	– PARÂMETRO QUE DEFINE A FORMA DOMINANTE DA BACIA
$NA$	– NÍVEL DE ÁGUA
$NA_t$	– NÍVEL DE ÁGUA NO INSTANTE ATUAL
$ndraw$	– NÚMERO MÁXIMO DE ITERAÇÕES
$NL$	– MODELO NORMAL
$NS$	– NASH-SUTCLIFFE
$nseq$	– NÚMERO DE CADEIAS AVALIADAS EM PARALELO
$p$	– PRECIPITAÇÃO
$P$	– MATRIZ DE COVARIÂNCIA
$pars$	– LISTA O NOME E RESPECTIVOS INTERVALOS DE INCERTEZA DAS VARIÁVEIS ALEATÓRIAS
$p(\theta)$	– DISTRIBUIÇÃO <i>A PRIORI</i>
$p(\theta \tilde{y})$	– DISTRIBUIÇÃO <i>A POSTERIORI</i>
$p_{\theta}(x_{1:t} \tilde{y}_{1:t})$	– EVOLUÇÃO DA DISTRIBUIÇÃO <i>A POSTERIORI</i>
$q(. .)$	– DISTRIBUIÇÃO PROPOSTA PARA O MÉTODO <i>METROPOLIS</i>
$qg$	– AFLUXOS SUBTERRÂNEOS À REDE DE DRENAGEM
$qi$	– AFLUXOS SUB-SUPERFICIAIS À REDE DE DRENAGEM
$qs$	– AFLUXOS SUPERFICIAIS À REDE DE DRENAGEM
$r$	– ESCOAMENTO TOTAL
$R$	– RESÍDUOS DA SIMULAÇÃO
$\hat{R}_j$	– ESTATÍSTICA DE GELMAN E RUBIN
$rb$	– ESCOAMENTO SUPERFICIAIS GERADOS PELAS ÁREAS IMPERMEÁVEIS
$rg$	– ESCOAMENTO SUBTERRÂNEO
$ri$	– ESCOAMENTO SUB-SUPERFICIAL
$rs$	– ESCOAMENTO SUPERFICIAL GERADOS PELAS ÁREAS PERMEÁVEIS
$r[p(r \theta)]$	– MODELO PROBABILÍSTICO PARA OS RESÍDUOS
$R$	– COEFICIENTE DE ARMAZENAMENTO DA SUB-BACIA
$R^2$	– COEFICIENTE DE DETERMINAÇÃO
$RMSE$	– RAIZ DO ERRO MÉDIO QUADRÁTICO
$RV$	– RELAÇÃO DE VOLUMES
$RWA$	– RANDOM WALK METROPOLIS
$s$	– ARMAZENAMENTO DA ÁGUA LIVRE
$S$	– ESTADOS (ASSIMILAÇÃO DE DADOS)
$SMC$	– <i>SEQUENTIAL MONTE CARLO</i>
$SCEM-UA$	– <i>EVOLUTION METROPOLIS ALGORITHM</i>

<i>SEP</i>	– <i>SKEW EXPONENTIAL POWER</i>
<i>si</i>	– ORDEM DAS VARIÁVEIS
<i>SLS</i>	– STANDARD LEAST SQUARES
<i>sm</i>	– CAPACIDADE MÉDIA ESPACIAL DE ARMAZENAMENTO DE ÁGUA LIVRE
<i>smm</i>	– MÁXIMA CAPACIDADE DE ARMAZENAMENTO PONTUAL
<i>s'm</i>	– CAPACIDADE DE ARMAZENAMENTO DE ÁGUA LIVRE
<i>t</i>	– INSTANTE DE TEMPO
<i>T</i>	– FRAÇÃO DO TEMPO DE CONCENTRAÇÃO
<i>tc</i>	– TEMPO DE CONCENTRAÇÃO DA SUB-BACIA
<i>ti</i>	– INSTANTE DE TEMPO INICIAL
<i>tf</i>	– INSTANTE DE TEMPO FINAL
<i>thin.t</i>	– INTERVALOS ENTRE AVALIAÇÕES DA FUNÇÃO DE VEROSSIMILHANÇA UTILIZADA PARA COMPOSIÇÃO DA AMOSTRA
<i>U</i>	– INPUTS
<i>USACE</i>	– CORPO DE ENGENHEIROS DO EXÉRCITO DOS ESTADOS UNIDOS
$u_t^T, x$	– DESVIOS QUADRÁTICOS DO MODELO DE PREDIÇÃO
<i>w</i>	– TEOR DE UMIDADE DO SOLO OU TENSÃO MÉDIA ESPACIAL
<i>W</i>	– MATRIZ DE PESOS DE INOVAÇÃO
<i>wd</i>	– QUANTIDADE DE ÁGUA RETIDA NA ZONA PROFUNDA
<i>wl</i>	– QUANTIDADE DE ÁGUA RETIDA NA ZONA INFERIOR
<i>wu</i>	– QUANTIDADE DE ÁGUA RETIDA NA ZONA SUPERIOR
<i>wm</i>	– CAPACIDADE MÉDIA ESPACIAL DE RETENÇÃO DE UMIDADE NO SOLO
<i>w'm</i>	– CAPACIDADE DE RETENÇÃO PONTUAL
<i>wmm</i>	– MÁXIMA CAPACIDADE DE RETENÇÃO PONTUAL DE UMIDADE DO SOLO NA BACIA
<i>wdm</i>	– ZONA PROFUNDA DE WM
<i>wlm</i>	– ZONA INFERIOR DE WM
<i>wum</i>	– ZONA SUPERIOR DE WM
$w_t$ e $v_{t+1}$	– ERROS ALEATÓRIOS
<i>x</i>	– VARIÁVEIS DE ESTADO
$x_0$	– VARIÁVEL DE ESTADO INICIAL
<i>y</i>	– OUTPUT, VAZÃO MODELADA
$y^*$	– OUTPUT ATUALIZADO
$\tilde{y}$	– VAZÃO OBSERVADA
<i>Z</i>	– VALOR CANDIDATO NO MÉTODO <i>METROPOLIS</i>
$\Gamma$	– ESCALAR DEFINIDO A PARTIR DO MÉTODO RWM
$\Delta t$	– INTERVALO DE TEMPO EM HORAS
$\Delta s$	– ESCOAMENTO DADO PELA DIFERENÇA DE R - RS
$\alpha(x_{t-1}, z)$	– TAXA DE ACEITAÇÃO NO ALGORITMO <i>METROPOLIS</i>
$\beta$	– CURTOSE
$\theta$	– PARÂMETROS
$\xi$	– ASSIMETRIA
$\pi$	– FREQUÊNCIA ESTÁVEL E ESTACIONÁRIA DA DISTRIBUIÇÃO DE PROBABILIDADE
$\pi(.)$	– DENSIDADE DA DISTRIBUIÇÃO ALVO
$\rho(u, x)$	– CORRELAÇÃO DE PEARSON
$\phi$	– PARÂMETROS DE MEDIÇÃO
$\phi_i$	– COEFICIENTES DE AUTOCORRELAÇÃO
$\Sigma$	– DESVIO PADRÃO
$\sigma_0$	– HETEROCEDASTICIDADE:INTERCEPÇÃO
$\sigma_1$	– HETEROCEDASTICIDADE:INCLINAÇÃO
$\sigma_u(\sigma_x)$	– DESVIO PADRÃO DA VARIÁVEL $u(x)$

- $\mu_M$  — VIÉS SOBRE A VAZÃO MODELADA
- $\mu_t$  — VIÉS INTRODUZIDO A VAZÃO MODELADA DEVIDO AOS ERROS NOS DADOS OBSERVADOS

# 1 INTRODUÇÃO

O desenvolvimento e a avaliação de modelos que descrevem o processo de transformação da chuva em vazão têm sido o foco de diversas pesquisas realizadas nas últimas décadas (BURNASH *et al.*, 1973 *apud* SINGH e WOOLHISER, 2002; ZHAO, 1992; LINDSTROM *et al.*, 1997; MOORE, 2007; JONES *et al.*, 2008). As razões para o desenvolvimento e aplicação desses modelos são diversas, conquanto podem ser sumariadas pelas limitações das técnicas de medição das variáveis hidrológicas no espaço e no tempo (BEVEN, 2012). Essas limitações, em geral, são impostas por fatores tecnológicos, de caráter econômico ou físico. A utilização de modelos permite aos hidrólogos extrapolar o conhecimento adquirido por meio de observações, tanto em termos espaciais, particularmente úteis no caso de bacias não monitoradas, quanto temporais, quando o objetivo é realizar previsões dos estados futuros de variáveis associadas ao fenômeno modelado (vazão, volume, nível do curso d'água etc.).

Os modelos chuva-vazão possuem uma vasta gama de aplicações práticas, entre as quais o dimensionamento e operação de sistemas hídricos para diversos fins (abastecimento público, geração hidroenergética, irrigação etc.), a previsão do impacto da alteração da paisagem (urbanização, desertificação etc.) e de fenômenos climáticos (*El Niño*, *La Niña*) sobre a resposta hidrológica, o prognóstico e alerta de eventos hidrológicos extremos (cheias ou secas), entre outros. De fato, o objetivo primordial da modelagem é prover informações acerca do comportamento das variáveis hidrológicas que permitam a tomada ou aprimoramento do processo de decisão.

Modelos chuva-vazão conceituais se propõem a captar a dinâmica do ciclo hidrológico, a partir da representação dos principais processos físicos que ocorrem na escala da bacia hidrográfica. Eles se caracterizam por possuir uma estrutura definida *a priori*, formulada deterministicamente por meio de relações entre estados e eventos (MONTANARI e KOUTSOYIANNIS, 2012). Beven (2012) observa ser inevitável algum grau de aproximação nesses modelos. A representação do sistema hidrológico exige, em geral, simplificações dos processos físicos, representações espaciais médias e, por vezes, parametrizações estatísticas.

Essas simplificações associadas à complexa representação espaço-temporal do fenômeno de transformação da chuva em vazão a partir das variáveis hidrológicas observadas e empregadas na modelagem, corroboram para o crescente consenso da comunidade técnica acerca da necessidade de se avaliar as incertezas associadas aos prognósticos desses modelos.

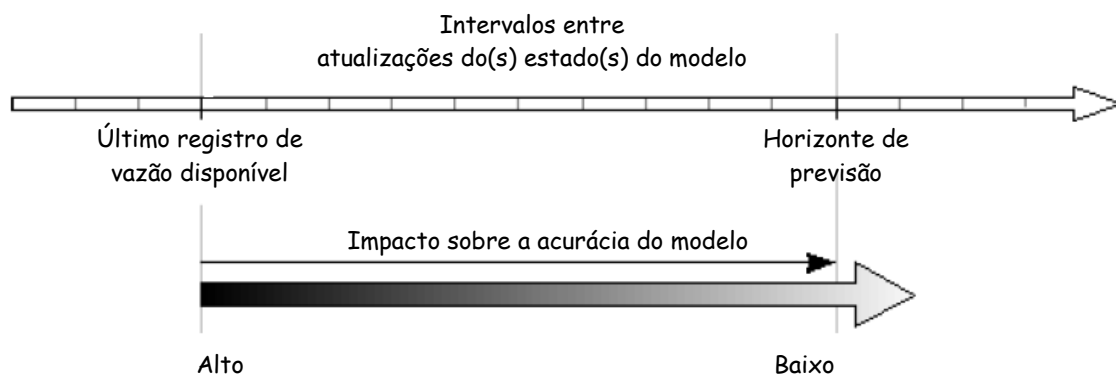
Segundo Sivapalan *et al.* (2003), alguns hidrólogos expressaram sua convicção de que a incerteza na hidrologia é de natureza epistêmica e, portanto, pode ser, em princípio, eliminada ou reduzida por meio de uma representação mais precisa dos processos físicos. Entretanto, há hidrólogos que sugerem que ela é inevitável, proveniente da variabilidade (aleatoriedade) natural do fenômeno hidrológico (MONTANARI *et al.*, 2009). Independentemente de sua natureza, verifica-se na literatura hidrológica produzida na última década a prevalência do emprego de técnicas de inferência Bayesiana associadas a métodos de amostragem baseados em simulações de Monte Carlo para estimação das incertezas na modelagem hidrológica.

Não obstante a importância da definição quantitativa dessas incertezas, é necessário o desenvolvimento de métodos que permitam a redução das mesmas. Liu e Gupta (2007) apontam o desenvolvimento de técnicas que possibilitem a extração e assimilação das informações contidas nos dados observados como uma possível ação para o alcance desse objetivo.

Métodos de assimilação de dados lidam com as incertezas dos distintos estados (*inputs*, parâmetros, variáveis de estado e/ ou *output*) associadas à modelagem por meio da análise da interação entre as diferenças entre os valores preditos e os dados observados (VRUGT *et al.*, 2005). Dentre as técnicas aplicadas à modelagem chuva-vazão, o filtro de Kalman e suas variações são as mais comumente verificadas na literatura. Kavetski *et al.* (2003), entretanto, são céticos quanto à aplicação da referida técnica em decorrência: (i) da não linearidade dos modelos hidrológicos; (ii) da suposição de normalidade dos erros na modelagem (prevista na formulação original do filtro); e (iii) da complexidade na adequação dos modelos à representação matricial requerida pela técnica.

Apesar dessas limitações, diversos estudos demonstraram que a atualização dos estados de um modelo pode aprimorar significativamente a acurácia de suas previsões, ainda que se observe, em geral, um aumento da complexidade operacional do sistema (BERTHET *et al.*, 2009). Os efeitos da atualização, no entanto, estão limitados a um horizonte de tempo, a partir do qual a assimilação de informações adicionais não mais produz impacto sobre as previsões do modelo (Figura 1.1). Em geral, quanto maior o tempo decorrido entre o início (reinício) do modelo, menor a efetividade do procedimento de atualização.





**Figura 1.1** – Impacto de um procedimento de atualização dos estados de um modelo chuva-vazão (Fonte: adaptado de BERTHET *et al.*, 2009).

A maioria dos estudos dedicados à análise de métodos de assimilação de dados aplicados à modelagem chuva-vazão orienta-se pela verificação da “efetividade” dessas técnicas. A ênfase dá-se, em geral, à maior similaridade entre vazões observadas e simuladas obtidas a partir da aplicação de determinada técnica. A análise do comportamento *input-state-output* do modelo é, em geral, preterida em favor da verificação do impacto na acurácia.

Há que se ponderar, entretanto, que a compreensão da interação entre as variáveis de estado de um modelo, as quais são função da estrutura conceitual admitida *a priori*, e os eventos hidrológicos que as produziram permitiria, em princípio, o alcance de resultados semelhantes (redução do viés na predição).

As incertezas associadas à evolução temporal das variáveis de estado podem ser caracterizadas de duas formas distintas: (i) uma intrínseca, imposta pela estrutura conceitual; e (ii) outra explícita, caracterizada pelos erros (sistemáticos ou não) nas observações empregadas nas etapas de calibração/ validação. Alguns trabalhos se propuseram a analisar a primeira (NEUMAN, 2002 *apud* WAGENER e GUPTA, 2005; BULYGINA, 2007; BULYGINA e GUPTA, 2009). No entanto, é a segunda que tem recebido expressiva atenção da comunidade hidrológica (KAVETSKI *et al.*, 2002, 2006a, 2006b; WAGENER *et al.*, 2003; AJAMI *et al.*, 2007; VRUGT, 2008a, 2009).

Partindo-se da premissa de que a representação conceitual do fenômeno é suficientemente adequada, é plausível admitir que eventuais vieses em períodos específicos seriam impostos ora por erros e/ou baixa representatividade (espaço-temporal) dos dados observados

empregados, ora por simplificações admitidas na representação do fenômeno modelado, ou, ainda, por ação de ambos os fatores.

O presente estudo orienta-se pela percepção da comunidade técnica de que, dada a importância e necessidade da utilização de modelos chuva-vazão, é necessário o desenvolvimento de técnicas que permitam a identificação e redução das incertezas associadas aos prognósticos desses modelos. Para tal, a exemplo de diversos estudos desenvolvidos na última década, utilizar-se-ão técnicas de inferência Bayesiana, em conjunto com métodos de simulação de Monte Carlo, para determinação das incertezas associadas às estimativas de parâmetros e variáveis de estado presentes na estrutura de um modelo chuva-vazão conceitual.

Analisar-se-á, na sequência, o padrão de variabilidade das variáveis de estado do modelo. A hipótese do presente trabalho é a de que o comportamento (trajetória temporal) das variáveis de estado é capaz de caracterizar erros em períodos específicos da série modelada, já que elas assimilam, a cada passo de tempo, todas as fontes de incertezas associadas à modelagem. A análise do comportamento *input-state-output* foi desenvolvida sob a perspectiva: (i) de caracterizar esses erros; (ii) permitir sua identificação *a priori*; e (iii) avaliar a possibilidade de sua correção por meio da atualização das variáveis de estado.

A inovação introduzida por esta tese é metodológica. Propõem-se aqui a avaliação das incertezas nas trajetórias dos estados como instrumento para identificação de períodos de dissimilaridade crescente entre as séries (de vazão) observada e modelada. Ademais, a partir de uma avaliação detalhada do impacto da atualização das variáveis de estado sobre períodos específicos da série, foi possível contrapor algumas das abordagens metodológicas e conclusões descritas nas principais referências sobre o tema na atualidade. Acredita-se que as conclusões aqui descritas podem ser empregadas para fundamentar o desenvolvimento de técnicas que permitam reduzir a incerteza global associada ao emprego de modelos chuva-vazão, tanto em esquemas de simulação quanto de previsão hidrológica.

O presente documento é um requisito à obtenção do título de Doutor em Saneamento, Meio Ambiente e Recursos Hídricos. Ele foi organizado em 6 capítulos, além deste de natureza introdutória. No segundo capítulo (2) são apresentados os objetivos do estudo, enquanto no capítulo (3) faz-se a revisão da literatura associada aos principais temas abordados. A metodologia sugerida é apresentada no capítulo (4). No capítulo (5) os resultados obtidos são descritos e discutidos, seguido pelo capítulo (6), onde são apresentadas as conclusões acerca

dos resultados obtidos e da metodologia aqui empregada. Por fim, no último capítulo (7) as referências consultadas são listadas.

## **2 OBJETIVOS**

### **2.1 *Objetivo geral***

Formular um procedimento para assimilar (compreender) o padrão de variabilidade das variáveis de estado de um modelo chuva-vazão conceitual, a partir do emprego de técnicas de inferência Bayesiana.

### **2.2 *Objetivos específicos***

- Calibrar um modelo chuva-vazão a partir do emprego de técnicas de inferência Bayesiana por meio de simulações de Monte Carlo;
- Avaliar o impacto da adoção de funções de verossimilhança construídas a partir de diferentes hipóteses sobre: (i) os resíduos do modelo; e (ii) a distribuição *a posteriori* dos parâmetros;
- Predizer as incertezas inerentes às vazões modeladas decorrentes das estimativas incertas dos parâmetros;
- Identificar períodos na série de vazões modeladas nos quais seja possível observar um viés na previsão/simulação hidrológica;
- Atualizar as variáveis de estados ao início desses períodos, valendo-se da técnica empregada para estimação de parâmetros/ incertezas; e
- Analisar a variação imposta às trajetórias das variáveis de estado à luz das incertezas associadas à modelagem hidrológica.

### 3 REVISÃO DA LITERATURA

Este capítulo apresenta uma revisão dos principais tópicos relacionados ao presente estudo. É válido destacar que o intuito não é descrever todos esses tópicos em sua completude. O objetivo primordial é apresentar os elementos centrais dos temas associados à pesquisa, enfatizando-se os principais aspectos relacionados à abordagem metodológica sugerida.

#### 3.1 Modelos chuva-vazão

Os processos hidrológicos que ocorrem em uma bacia hidrográfica, desde a ocorrência da chuva à formação do escoamento em vertentes e cursos de água, são diversos e complexos (WAGENER *et al.*, 2004). A natureza dessa complexidade pode ser notada em diferentes níveis (BEVEN, 2009): (i) na descrição do processo ao longo do tempo; (ii) na definição dos limites do sistema de interesse, ponderando-se que esse é somente parte de um ambiente maior; (iii) na compreensão, seleção e quantificação das principais forças externas que atuam sobre o referido sistema; e (iv) na variação espacial do fenômeno, considerando-se as características locais a ele associadas. Essa complexidade pode ser descrita em termos qualitativos para os diferentes níveis mencionados. No entanto, uma descrição quantitativa é, em geral, necessária para o gerenciamento e tomada de decisões no âmbito do sistema de interesse. A descrição matemática do fenômeno, fundamentada a partir de princípios científicos, e construída com vistas à precisão e à acurácia, é o fundamento dos modelos hidrológicos, entre os quais, os modelos chuva-vazão.

A complexidade do processo natural de transformação da chuva em vazão impõe a esses modelos, independentemente de suas estruturas, que a representação do fenômeno ocorra de forma simplificada. É necessário, entretanto, assegurar que os principais processos hidrológicos associados ao problema de interesse sejam adequadamente representados (WAGENER *et al.*, 2004). De fato, em situações práticas, a utilização de um modelo é condicionada, sobretudo, pela inter-relação dos seguintes fatores:

- os objetivos da modelagem, expressos pela abordagem intrínseca ao modelo e pelos resultados esperados de sua aplicação;

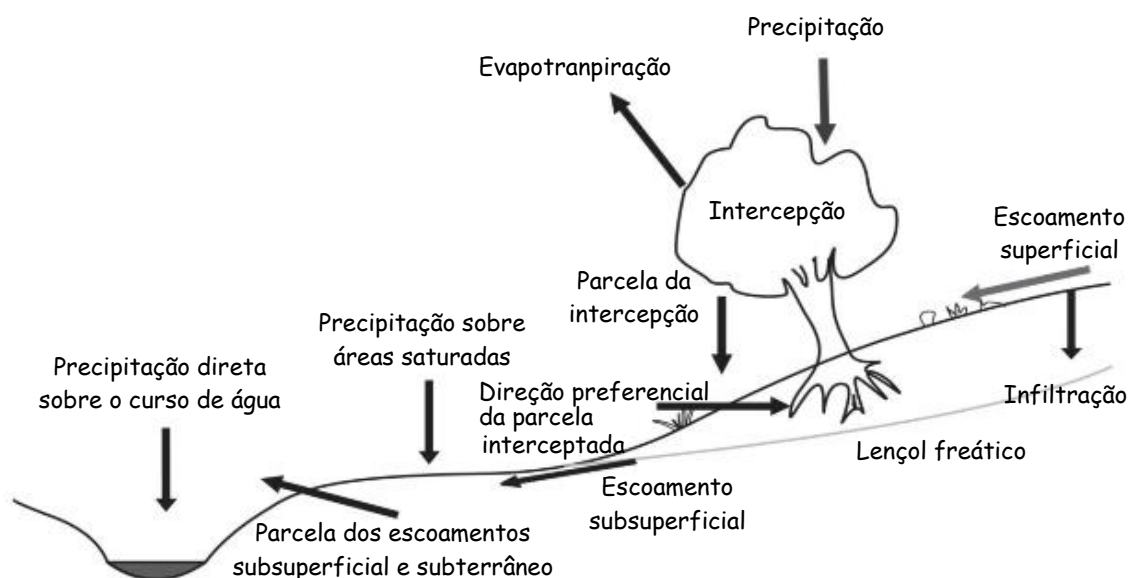
- a disponibilidade de informações hidrológicas, as quais implicam em uma representação adequada (ou não) da escala espaço-temporal do fenômeno em análise; e
- a estrutura do modelo, a qual apoia-se na parametrização dos processos hidrológicos, e cujo equilíbrio reside na parcimônia admitida durante a conceitualização do fenômeno, preservada a correção da representação desses processos.

Não obstante esses fatores, a familiaridade, as ferramentas disponíveis, o tempo a ser despendido na modelagem e o estágio de desenvolvimento norteiam a seleção e aplicação dessas técnicas.

Modelos chuva-vazão auxiliam os hidrólogos na compreensão das interações dinâmicas entre o clima e a superfície terrestre. Ademais, permitem a previsão do comportamento do sistema hidrológico frente a condições distintas das outrora observadas. Por consequência, observa-se o desenvolvimento de um grande número de modelos que se diferenciam em função dos dados utilizados, da discretização, das prioridades da representação dos processos e dos objetivos a serem alcançados.

### **3.1.1 Classificação de modelos chuva-vazão**

Em função da natureza complexa do fenômeno de transformação da chuva em vazão, representada por uma série de processos na atmosfera, no solo e nos corpos de água, altamente interligados e em diferentes escalas espaciais e temporais, os hidrólogos divergem acerca da identificação e representação dos principais processos nos modelos hidrológicos (WAGENER *et al.*, 2004). Beven (2012) observa que essas percepções distintas dos fenômenos do ciclo hidrológico (Figura 3.1) influenciadas, sobretudo, pelas experiências prévias dos hidrólogos acerca dos sistemas a modelar, conduziu-os a elencar uma série de classificações.



**Figura 3.1** – Representação dos processos envolvidos em um modelo de percepção dos fenômenos hidrológicos em uma encosta (Fonte: adaptado de BEVEN, 2012).

Os modelos chuva-vazão podem ser classificados segundo diversos atributos com base: (i) na resposta aos mesmos dados e parâmetros de entrada, em determinístico ou estocástico; (ii) na abordagem utilizada para a representação do fenômeno, em físico, conceitual ou caixa preta (*black-box*); (iii) na linearidade (quando o princípio da superposição é respeitado ou quando, sobre um enfoque estatístico, uma relação linear é admitida para explicar a variação das variáveis hidrológicas), em lineares ou não lineares; (iv) na representação discreta, ou não, no tempo, em contínuo ou por eventos; (v) na amplitude da representação no espaço, em concentrados ou distribuídos; entre outras. Sob um enfoque mais geral, Marshal (2005) observa a existência de duas diferentes abordagens no desenvolvimento desses modelos: a abordagem *top-down*, na qual a interpretação da relação entre as variáveis monitoradas conduz à construção do modelo; e outra denominada *bottom-up*, na qual o conhecimento dos principais processos físicos que governam a transformação da chuva em vazão guia a concepção de sua estrutura.

Uma descrição pormenorizada de todas essas categorias excede os objetivos deste documento. Por conseguinte, ater-se-á, aqui, àquelas diretamente relacionadas à metodologia empregada nesta tese. Uma apresentação detalhada desse tema pode ser vista em Tucci (2005).

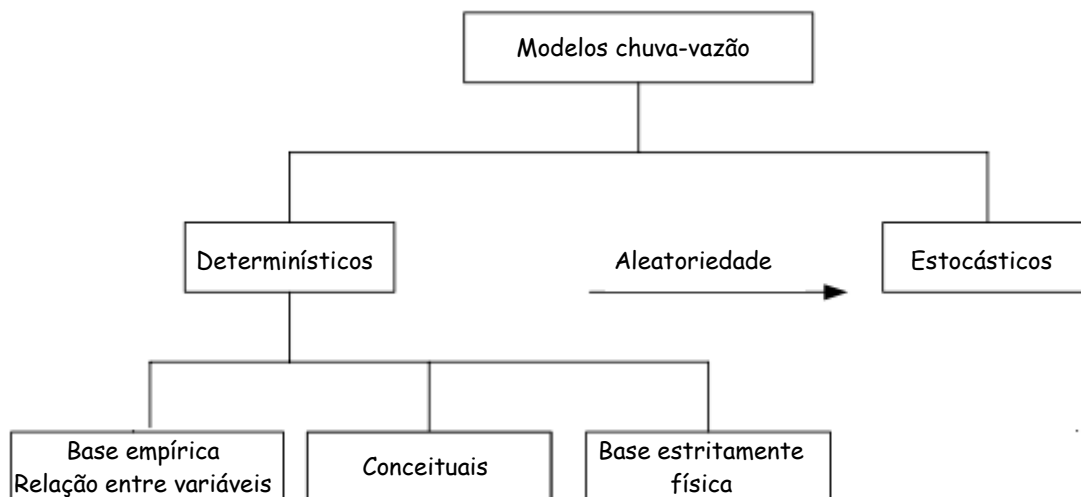
A primeira categoria refere-se aos modelos determinísticos e estocásticos (BEVEN, 2012). Modelos (*M*) são ditos determinísticos quando, valendo-se de um mesmo conjunto de dados e

parâmetros de entrada, um único *output* ( $y$ ) é obtido. Nos modelos estocásticos a aleatoriedade, representada pela variabilidade de  $\tilde{y}$ , decorre das incertezas presentes nas variáveis de entrada ( $u$ ), nas condições de contorno ( $B$ ), nos parâmetros ( $\theta$ ) e nas variáveis de estado ( $x$ ) do modelo. Existem, também, os modelos híbridos, os quais combinam componentes determinísticos e estocásticos (EHR/UFMG, 2001). Por vezes, o simples tratamento dado ao processo de calibração de parâmetros de um modelo inteiramente determinístico pode torná-lo um modelo híbrido. Isso ocorre, por exemplo, na análise das incertezas da modelagem, quando os parâmetros de um modelo determinístico são descritos por distribuições de probabilidade, ou quando se atribui um intervalo de confiança à variável de saída.

No tocante à abordagem admitida para representação do fenômeno, os denominados modelos conceituais procuram descrever processos complexos por meio de sistemas simplificados que guardam uma relação de analogia, ainda que incompleta, com os processos físicos. Um conjunto significativo desses modelos procura solucionar a equação de balanço hídrico a partir do emprego de reservatórios (lineares, ou não) para a representação das relações de armazenamento e transferência de água no solo. Caso, na formulação do modelo, nem mesmo uma analogia conceitual com o fenômeno seja estabelecida, então fala-se em modelo empírico, *black-box* ou de caixa preta.

A principal característica dos modelos chuva-vazão conceituais encontra-se no objetivo de atribuir algum realismo físico à representação do fenômeno. Eles podem ser interpretados como um intermédio entre os modelos puramente empíricos e os estritamente físicos (Figura 3.2). Eles permitem a representação matemática baseada na percepção ou conhecimento *a priori* do hidrólogo sobre o processo de transformação chuva-vazão, sem, no entanto, ater-se à estrita observância dos princípios físicos de conservação da massa, quantidade de movimento e energia, na conceituação do fenômeno.





**Figura 3.2** – Classificação dos modelos hidrológicos baseada na estrutura utilizada para descrição dos processos físicos (Fonte: adaptado de REFSGAARD, 1996 *apud* SHRESTHA, 2009).

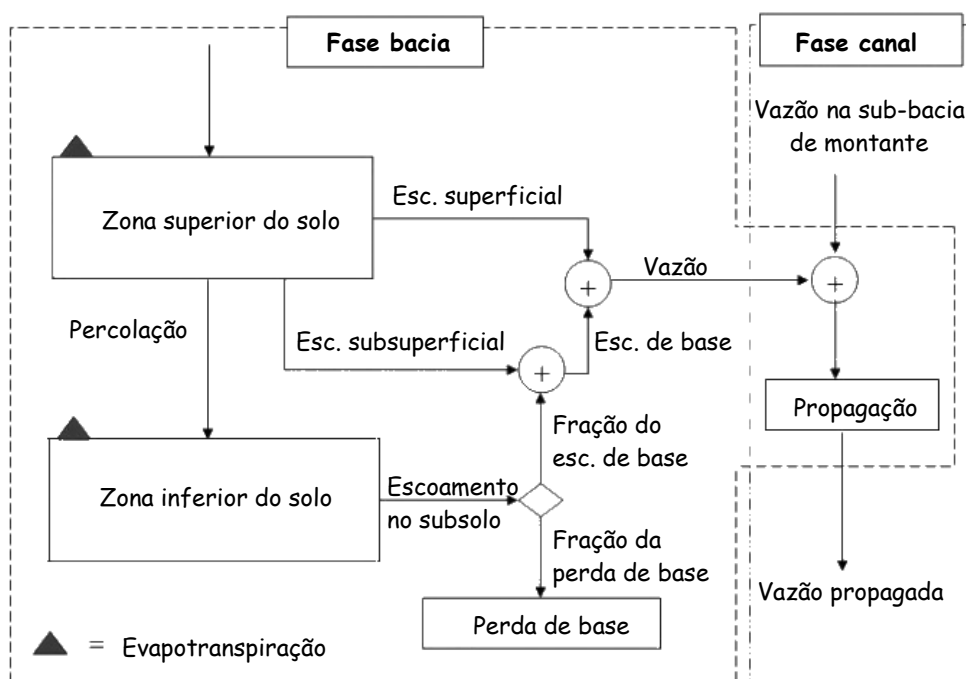
Outra classificação dos modelos chuva-vazão faz-se quanto ao tratamento do espaço e do tempo (QUEIROGA, 2003). No caso do tempo, os modelos podem ser de simulação contínua ou por evento. Os modelos por evento interessam-se apenas em reproduzir uma cheia em um ponto de controle da bacia hidrográfica, a partir de um evento de precipitação de origem. Os modelos de simulação contínua interessam-se pela reprodução do balanço hídrico por longos períodos, contemplando o cálculo de vazões mesmo em ausência de precipitação. Para isso, eles devem integrar em sua estrutura uma representação bem mais detalhada da água no solo e da evapotranspiração, quando comparados aos modelos por evento.

Quanto ao espaço, os modelos são classificados como concentrados ou distribuídos. Nos modelos chuva-vazão concentrados, a bacia hidrográfica é considerada como uma unidade única, na qual as variáveis de entrada ( $u$ ), os parâmetros ( $\theta$ ) e as variáveis de estado ( $x$ ) representam valores médios sobre a área da bacia (MORADKHANI e SOROOSHIAN, 2008; BEVEN, 2012). Nos modelos distribuídos  $u$ ,  $x$  e  $\theta$  (opcionalmente) variam espacialmente por meio da discretização da bacia em subunidades. Quando essas subunidades são sub-bacias (em contraposição a um *grid* regular) diz-se que o modelo é semi-distribuído.

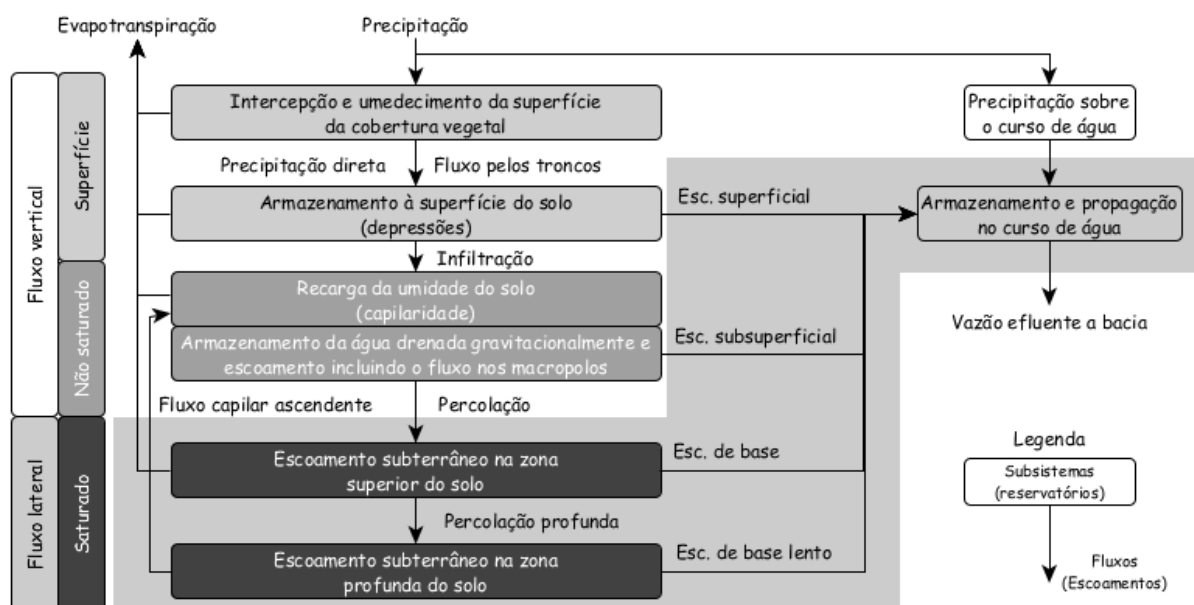
### 3.1.2 Modelagem de processos hidrológicos

Considerando-se a abordagem admitida nesta tese, a discussão a seguir refere-se aos modelos chuva-vazão determinísticos (híbridos), conceituais, concentrados (semi-distribuídos) e de simulação contínua.

Em geral, o tipo de modelo assim particularizado representa os processos hidrológicos em uma bacia hidrográfica considerando duas fases principais: a fase bacia e a fase canal (Figura 3.3). Na fase bacia, encontra-se, com maior ou menor detalhamento, a descrição de processos como interceptção, infiltração, armazenamento de água no solo, escoamentos em meios saturado e não-saturado, armazenamento superficial, evapotranspiração e escoamento superficial. Na fase canal, modela-se o escoamento não permanente em cursos d'água, lagos e reservatórios que compõem a rede de drenagem da bacia. As representações esquemáticas desses processos são ilustradas na Figura 3.4.



**Figura 3.3** – Discriminação dos processos hidrológicos representados em um modelo chuva-vazão através de fases: bacia e canal (Fonte: adaptado de NETO *et al.*, 2007).



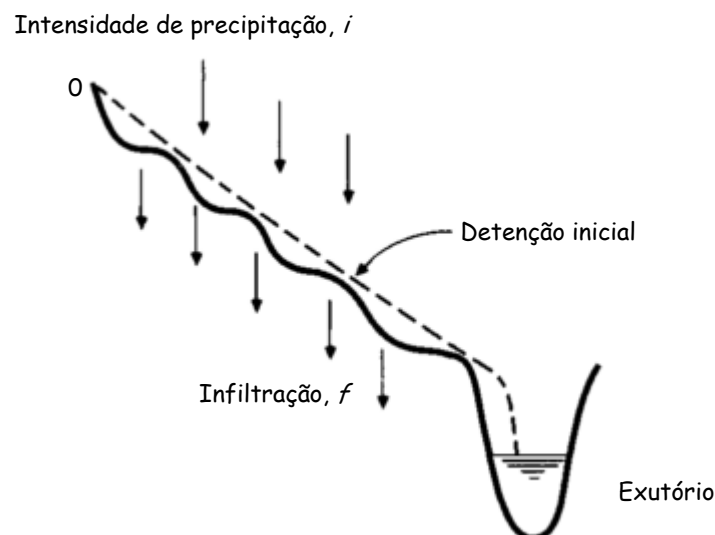
**Figura 3.4** – Representação esquemática dos processos hidrológicos em uma bacia hidrográfica (Fonte: adaptado de BECKER, 1992 *apud* WAGENER *et al.*, 2004).

Em geral, os modelos concentrados não possuem reprodução da fase canal, embora alguns deles separem claramente os processos de transferência do escoamento superficial na bacia e de amortecimento do hidrograma na rede de drenagem. Um exemplo é o modelo do hidrograma unitário de Clark, no qual a transferência de escoamentos na bacia é feita por um hidrograma tempo-área construído a partir da definição de isócronas e o amortecimento é realizado por um reservatório linear (CLARK, 1943 *apud* HEC, 1973).

A fase bacia é, usualmente, dividida em água no solo e água em escoamento superficial. A chamada “água no solo” designa um conjunto de processos que ocorrem sob a cobertura vegetal, à superfície do solo e em seu interior, na zona radicular, usualmente não saturada, e na zona saturada. Modelos do tipo chuva-vazão procuram reproduzir esses processos com o fim de estabelecer a separação dos aportes de precipitação entre a água que será absorvida pelo solo, por infiltração, ou que ficará retida à superfície, por intercepção e armazenamento superficial, e a água destinada ao escoamento à superfície do solo, caso ele ocorra. Essa etapa da modelagem chuva-vazão é, seguramente, a mais importante e a de maior dificuldade de representação, tendo em vista, por um lado, a complexidade dos processos envolvidos e, por outro, a carência de dados sobre esses processos, em escalas espaço-temporais apropriadas, que permitam a adequada formulação dos modelos e a calibração de seus parâmetros (EHR/UFMG, 2001).

Ainda segundo EHR/UFMG (2001), duas razões principais explicam a importância da modelagem da separação dos escoamentos. A primeira está ligada ao fato de que os processos físicos de escoamento superficial e subterrâneo são muito distintos. O escoamento superficial estabelece-se, quase sempre, em regime turbulento e não-permanente e é responsável pela resposta rápida da bacia a precipitações de maior intensidade. O escoamento subterrâneo faz-se em regime laminar, com velocidades muito baixas, e sua correta representação permite uma melhor reprodução do trecho de recessão do hidrograma de cheia. A segunda razão está associada ao fato de que nem toda a água absorvida pelo solo escoar e aparece à saída da bacia. Estimar adequadamente o volume de água absorvida pelo solo, a parcela retida na zona não saturada e a parcela que percola até a zona saturada é importante para reproduzir, nos períodos sem chuva, a evolução do estado de umidade do solo e os processos de evapotranspiração e de escoamento de base. O estado de umidade do solo será, igualmente, importante para a simulação da reação da bacia a novas precipitações, ao final de um período sem chuva.

A concepção dos modelos chuva-vazão conceituais, foco desta pesquisa, é fortemente influenciada pela teoria Hortoniana de formação de escoamentos (CHOW *et al.*, 1988). Para Horton a formação do escoamento superficial na bacia hidrográfica é controlada pela capacidade de infiltração do solo (Figura 3.5). Se a intensidade de precipitação,  $i$  [mm/h], for inferior à capacidade de infiltração,  $f$  [mm/h], toda a precipitação é absorvida pelo solo, não ocorrendo escoamento superficial. Porém, caso  $i > f$ , então há um excedente de precipitação ( $i-f$ ) que não é absorvido pelo solo e que escoar superficialmente pelas encostas da bacia hidrográfica e concentra-se em vertentes, talvegues e cursos d'água.



**Figura 3.5** – Escoamento superficial em uma encosta produzido pelo excesso de chuvas sobre a capacidade de infiltração (Fonte: adaptado de CHOW *et al.*, 1988).

Essa concepção da gênese dos escoamentos, originária de estudos em regiões semi-áridas, é capaz de explicar como bacias hidrográficas cujos solos são pouco espessos e muito expostos, em razão de cobertura vegetal pouco densa, respondem rapidamente a eventos de precipitação de elevada intensidade. Entretanto, hidrólogos com experiência em bacias hidrográficas localizadas em regiões temperadas, com solos profundos e cobertura vegetal densa, observam que o escoamento superficial do tipo Hortoniano, nesses casos, é raro (DUNNE *et al.*, 1975 *apud* EHR/UFMG, 2001). Experimentos que se valem de traçadores naturais para investigação desse tema têm sido realizados há vários anos (SKLASH e FARVOLDEN, 1979; CASSIE *et al.*, 1996; LAUDON e SLAYMAKER, 1997). A partir de resultados de estudos dessa natureza foram propostas novas teorias sobre a formação de escoamentos, as quais, posteriormente, foram incorporadas a alguns modelos chuva-vazão, a exemplo dos modelos Xinanjiang (ZHAO, 1992), TOPMODEL (BEVEN *et al.*, 1984) e ARNO (TODINI, 1996).

### 3.1.3 Calibração de parâmetros

Modelos chuva-vazão baseiam-se na premissa de que os processos hidrológicos “dominantes”, os quais controlam o fenômeno de transformação da chuva (sobre a bacia) em vazão (no exutório dessa), são estruturalmente similares na maioria dos sistemas hidrológicos e, por conseguinte, é possível valer-se de uma mesma estrutura conceitual para obter-se a resposta em diferentes sistemas. Para tal, é necessário apenas ajustar os valores dos

parâmetros admitidos para a conceitualização do fenômeno, confrontando-se as respostas do modelo com as vazões observadas (GUPTA *et al.*, 2003).

Moradkhani e Sorooshian (2008) observam que os parâmetros dos modelos conceituais podem ser interpretados como quantidades que caracterizam o sistema hidrológico. Há parâmetros que podem ser medidos diretamente e independem da resposta da bacia (ex.: área de drenagem, porção de área impermeável na bacia, fração de área vegetada, percentual de corpos de água etc.). Há também aqueles que não possuem uma interpretação física direta, não podem ser medidos (por restrições técnicas, conceituais e/ou econômicas) e cujos valores propõem-se a representar médias indiretas de propriedades relevantes à caracterização do sistema hidrológico (ex.: umidade do solo, condutividade hidráulica, taxa de escoamento em reservatórios hipotéticos utilizados para representação do fluxo da água no solo, fração de área permeável que contribui para geração do escoamento superficial etc.).

A fim de que um modelo seja capaz de simular “consistentemente” o comportamento dinâmico das vazões em uma bacia hidrográfica, é necessário “ajustar” os valores de seus parâmetros, sob a perspectiva de aproximar, o quanto possível, a resposta do modelo (a um *input*) aos registros de vazões na seção de interesse (exutório). Nesse procedimento, denominado calibração de parâmetros, a qualidade do ajuste é avaliada com base em critérios quantitativos, ou ainda por meio de inspeção gráfica e pela análise dos resíduos da simulação.

Em geral, uma primeira aproximação do conjunto “ótimo” de parâmetros é realizada a partir de um processo de tentativa e erro, conhecido como calibração manual. Esse procedimento pode ser categorizado em três níveis (BOYLE *et al.*, 2000 *apud* GUPTA *et al.*, 2003): (i) no nível zero, os dados da bacia são examinados e estimativas *a priori* do intervalo de variação dos parâmetros são estabelecidas; (ii) no segundo nível, procede-se à análise de alguns segmentos específicos dos hidrogramas observados (ex.: vazões de pico, recessão etc.), a fim de se identificar o(s) parâmetro(s) que predominantemente controla(m) esses trechos; e (iii) no terceiro nível, são realizadas simulações diversas (*input-output*) da resposta da bacia, valendo-se de trechos representativos da variação fluviométrica e de limites paramétricos plausíveis, para identificação dos parâmetros que produzem a maior similaridade entre vazões simuladas e observadas.

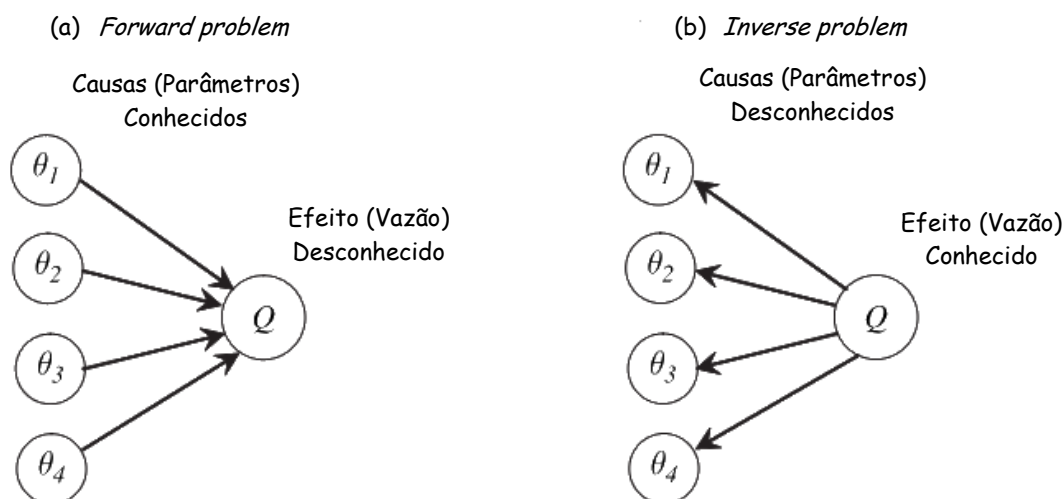
Em aplicações práticas, a implementação das etapas mencionadas revela-se desafiadora, sobretudo, devido (GUPTA *et al.*, 2003; BEVEN, 2009; BEVEN, 2012):

- ao número, em geral, elevado de parâmetros a serem calibrados;
- às interações não-lineares entre os parâmetros, as quais produzem efeitos similares ou compensatórios sobre diferentes segmentos dos hidrogramas modelados;
- à inexistência de uma forma inequívoca de se avaliar a similaridade entre as séries de vazões simuladas e observadas; e
- aos erros e incertezas presentes nos dados observados (*input*  $u$  e *output*  $\tilde{y}$ ) e na estrutura admitida para conceitualização do fenômeno.

A fim de lidar, ao menos em parte, com essas dificuldades, os hidrólogos dedicaram-se ao desenvolvimento de técnicas automáticas para calibração de parâmetros.

De fato, a calibração de parâmetros em modelos hidrológicos, sob um cenário de otimização, pode ser tratada como um problema inverso (*IP: inverse problem*), no qual, à exceção dos parâmetros, as observações ( $u$  e  $\tilde{y}$ ), e, por vezes as variáveis de estado  $x$ , são conhecidos (MORADKHANI e SOROOSHIAN, 2008). Isso contrasta com um problema direto (*FP: forward problem*), no qual as propriedades relevantes do modelo, incluindo as condições de contorno e os parâmetros do sistema, são conhecidas.

Em um *FP*, o objetivo é determinar o efeito (vazão) a partir das causas (parâmetros), enquanto que, em um problema inverso, o propósito é identificar as causas partindo-se de uma realização do fenômeno (Figura 3.6). No caso *IP*, em geral, a solução não é única. Isso equivale a afirmar que o mesmo efeito (vazões) pode ser produzido por diferentes causas (parâmetros).



**Figura 3.6** – Representação esquemática de um problema direto e de um problema inverso (Fonte: adaptado de MORADKHANI e SOROOSHIAN, 2008).

Na calibração automática, o problema é formulado sob uma perspectiva de otimização a partir de duas componentes principais (SOROOSHIAN e GUPTA, 1995):

- função objetivo (por vezes denominada função perda ou custo): é uma medida quantitativa da capacidade do modelo de reproduzir as respostas observadas em um sistema hidrológico. Ex.: coeficiente de *Nash-Sutcliffe*, coeficiente de determinação ( $r^2$ ), erro absoluto médio (*MAE*), raiz do erro médio quadrático (*RMSE*) etc.; e
- algoritmo de calibração: são formulações matemáticas de processos de tentativa e erro, segundo regras formais, que buscam reduzir os passos necessários para se atingir um objetivo pré-estabelecido, por exemplo, o de minimizar o erro quadrático (função objetivo) de um modelo matemático por meio da estimação dos valores de seus parâmetros. Ex.: (i) busca local: *Rosenbrock*, *Downhill Simplex*, *Pattern Search* etc.; e (ii) busca global: *Adaptive Random Search*, Algoritmo Genético, *Simulated Annealing* etc.

A filosofia por trás de qualquer algoritmo de calibração reside na busca da melhoria da função objetivo por meio da avaliação de diferentes regiões do espaço paramétrico. Em um quadro exploratório, essa busca ocorre de forma aleatória, enquanto em um cenário de refinamento, a busca baseia-se no conhecimento adquirido em amostras construídas previamente. Os métodos diferenciam-se na forma como a realizam, sob os argumentos de eficiência,



estabilidade e robustez, ao lidar com a ocorrência de regiões de mínimos (ou máximos) locais e com a multidimensionalidade dos modelos hidrológicos.

Gupta *et al.* (2005) enumeram as condições para que a calibração dos parâmetros de um modelo hidrológico seja considerada satisfatória:

- similaridade entre o *input-state-output* do modelo e o comportamento exibido pelas variáveis hidrológicas monitoradas;
- acurácia (viés desprezível) e precisão (incerteza relativamente baixa) das previsões do modelo; e
- consistência entre a estrutura conceitual admitida pelo modelo e o comportamento exibido pelos processos hidrológicos na bacia analisada.

### **3.2 Incertezas na modelagem chuva-vazão**

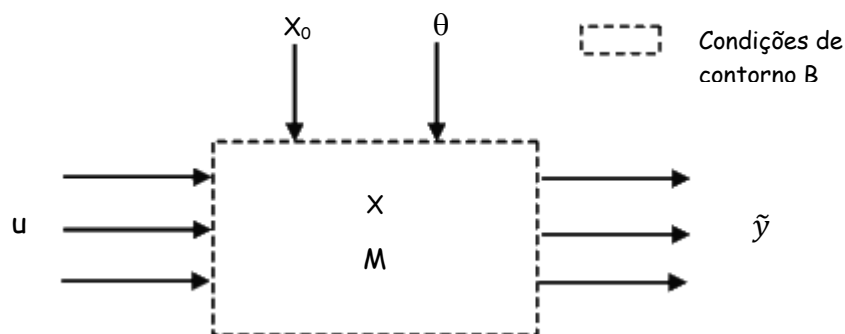
O desenvolvimento dos modelos e o consequente aprimoramento da compreensão da física e dinâmica dos sistemas hidrológicos foram fortemente beneficiados pela crescente capacidade computacional, aliada a um aumento na disponibilidade de informações hidrológicas (LIU e GUPTA, 2007). Paradoxalmente, enquanto esses avanços refletiram em uma maior compreensão do fenômeno, eles também consolidaram a necessidade do desenvolvimento de métodos para avaliação das incertezas associadas aos prognósticos realizados por esses modelos.

As incertezas, no contexto da modelagem chuva-vazão, decorrem da incompletude da compreensão do processo físico de transformação da chuva em vazão, associada à imperfeição das técnicas de medição utilizadas para obtenção das variáveis hidrológicas. Elas provêm essencialmente de três fontes (SHRESTHA, 2009):

- dados (*input*  $u$ ; e *output*  $\tilde{y}$ ): erros em  $u$  (ex.: precipitação, evaporação), e em  $\tilde{y}$  (ex.: vazão), ou ainda nas variáveis de estado iniciais ( $x_0$ ), quando as mesmas são obtidas a partir de observações de campo (ex.: teor de umidade do solo, nível de água no subsolo). Eles decorrem essencialmente de erros no processo de medição, seja por limitação dos instrumentos ou falha humana, e/ou da inadequada representatividade da amostra, em termos espaciais e temporais;

- estrutura ( $M$ ) e condições de contorno ( $B$ ) do modelo: modelos conceituais são apenas abstrações da realidade, as quais geralmente envolvem certo grau de simplificação e/ou idealização. Aproximações inapropriadas, ou mesmo desconsideradas, dos principais processos físicos podem resultar em erros significativos. Ademais, esses erros podem advir da implementação matemática (discretização espacial e temporal), a qual converte a estrutura conceitual em um modelo numérico; e
- parâmetros ( $\theta$ ) e variáveis de estado ( $x$ ): incertezas nos parâmetros e variáveis de estado derivam da inability de obterem-se medidas precisas dos dados de entrada do modelo. Os parâmetros do modelo nem sempre possuem um significado físico direto. Por conseguinte, eles não podem ser medidos diretamente (ou o custo associado inviabiliza a medição) em campo. Nesse caso, os valores são estimados a partir de métodos indiretos (ex.: baseados na experiência do modelador, por meio da calibração dos parâmetros), os quais se constituem em fontes de incerteza.

A Figura 3.7 apresenta esquematicamente os componentes e fontes de incerteza de um modelo.



**Figura 3.7** – Componentes e fontes de incerteza de um modelo chuva-vazão (adaptado de LIU e GUPTA, 2007).

É necessário ponderar que, independentemente da fonte, existem duas categorias distintas de incerteza (BEVEN, 2009; GONG *et al.*, 2013): (i) uma aleatória (ou intrínseca), cuja natureza reside na diferença observada nos resultados obtidos por meio da repetição de um experimento sob as mesmas condições; e (ii) outra epistêmica, que decorre da insuficiência ou ausência de compreensão e/ou informação sobre todos os processos físicos que governam o fenômeno modelado. A redução da incerteza epistêmica ocorre por meio da obtenção de

informações complementares, aprimoramento dos modelos e/ ou de uma melhor estimativa dos parâmetros e variáveis de estado. A presente pesquisa concentra-se nesta última categoria, mais especificamente na análise das incertezas nos resultados da modelagem hidrológica associadas à evolução das variáveis de estado ao longo do período de simulação.

### **Avaliação da incerteza**

O termo avaliação da incerteza refere-se, no contexto da modelagem chuva-vazão, a uma análise quantitativa da incerteza associada às variáveis ( $u$  e  $\tilde{y}$ ), aos parâmetros ( $\theta$ ) e/ ou ao modelo hidrológico ( $M$ ). Ela fornece uma indicação quantitativa da confiança em uma dada “quantidade” hidrológica, seja essa observada ou inferida a partir de um modelo (MONTANARI, 2011). Essa confiança pode ser expressa pela: (i) estimativa do erro que afeta essa quantidade; ou (ii) pelo padrão de variabilidade de seu valor em um intervalo pré-estabelecido.

O grau de dissimilaridade entre o *output*  $y$  de um modelo e o valor verdadeiro da variável modelada ( $\tilde{y}$ ) pode ser compreendido como uma medida da incerteza global. Essa medida é composta pelas incertezas específicas, as quais são definidas a partir de suas fontes (causas) (ex.: incerteza dos parâmetros, incerteza nos dados de entrada etc.). Incertezas específicas podem compensar-se mutuamente e influenciar a composição da incerteza global. Por exemplo, erros nos parâmetros podem compensar, ao menos em parte, erros nos dados observados ( $u$  e  $\tilde{y}$ ) e na estrutura do modelo (MORADKHANI e SOROOSHIAN, 2008).

De forma geral, os métodos empregados na avaliação da incerteza são dicotomicamente categorizados em probabilísticos e não probabilísticos. Na primeira categoria, a estimação das incertezas ocorre a partir de técnicas “consistentemente” apoiadas na teoria da probabilidade. Nos métodos não probabilísticos são admitidas “generalizações” dessa teoria, o que permite que algumas das premissas admitidas sejam parcialmente violadas (ex.: que a função de verossimilhança seja definida a partir de um modelo não estatístico).

Montanari (2011) observa que, em alguns casos, não é simples classificar uma técnica em probabilística ou não. De fato, alguns métodos baseiam-se na teoria de probabilidades, no entanto, em aplicações reais, valem-se de hipóteses simplificadoras sob a perspectiva de superar problemas práticos, tal qual a insuficiência de dados necessários à aplicação de técnicas de inferência estatística.

Recentemente, classificações mais detalhadas foram propostas. Matott *et al.* (2009) categorizaram essas técnicas em: (i) métodos para análise de dados, as quais incluem procedimentos estatísticos e analíticos para avaliação da acurácia dos dados empregados; (ii) análises de identificabilidade, as quais dedicam-se à detecção da suficiência dos dados empregados; (iii) métodos para estimação de parâmetros, propostos para fins de avaliação da incerteza associada à estimação dos parâmetros presentes na estrutura dos modelos; (iv) técnicas para análise (propagação) da incerteza decorrente de fontes distintas; (v) análise de sensibilidade, as quais dedicam-se à avaliação do impacto da variação dos *inputs* sobre os *outputs*; (vi) análises multimodelos, construídas para fins de comparação de modelos; e (vii) métodos Bayesianos, cuja definição decorre da classificação probabilística e o emprego é observado nas categorias anteriores, em especial, em (iii) e (iv). Destaca-se que essa categorização não é unívoca e que, portanto, determinada técnica pode ser classificada em mais de uma categoria.

A Tabela 3.1 lista algumas das técnicas de avaliação da incerteza aplicadas à modelagem hidrológica mais comumente observadas na literatura hidrológica. Elas estão apresentadas segundo o tipo de incerteza estimada (global ou específica) e classificadas quanto à abordagem (probabilística ou não) e ao propósito de sua utilização.

**Tabela 3.1** – Técnicas utilizadas na avaliação da incerteza na modelagem hidrológica (adaptado de SHRESTHA, 2009 e MONTANARI, 2011).

Método	Autor(es)	Tipo de incerteza estimada	Classificação
AMALGAM <i>Multialgorithm genetically adaptive method for multiobjective optimization</i>	Vrugt e Robinson (2007)	Parâmetros	Não probabilístico, estimação de parâmetros
BATEA <i>Bayesian total error analysis</i>	Kavetski <i>et al.</i> (2002, 2006a, 2006b)	Dados (induzida pela precipitação)	Probabilístico, estimação de parâmetros, avaliação da incerteza, análise de sensibilidade
BFS <i>Bayesian forecasting system</i>	Krzysztofowicz (1999, 2001, 2002)	Global	Probabilístico, bayesiano
BMA <i>Bayesian multimodel analysis</i>	Hoeting <i>et al.</i> (1999)	Global	Probabilístico, multimodelo
DREAM <i>Differential Evolution Adaptive Metropolis</i>	Vrugt (2008a, 2009)	Global, parâmetros e dados	Probabilístico
DYNIA <i>Dynamic identifiability analysis</i>	Wagener <i>et al.</i> (2003)	Parâmetros	Probabilístico, análise de identificabilidade
GLUE <i>Generalized likelihood uncertainty estimation</i>	Beven e Binley (1992) <i>apud</i> Beven (2009)	Global, parâmetros, dados e estrutura	Não probabilístico (quando uma função de verossimilhança “informal” é utilizada), estimação de parâmetros, avaliação da incerteza, análise de sensibilidade

Método	Autor(es)	Tipo de incerteza estimada	Classificação
IBUNE <i>Integrated bayesian uncertainty estimator</i>	Ajami <i>et al.</i> (2007)	Global (induzida pela precipitação e pela estrutura do modelo)	Probabilístico, estimação de parâmetros, avaliação da incerteza, análise de sensibilidade
<i>Machine Learning</i>	Solomatine e Shrestha (2009)	Usualmente global	Não probabilístico
<i>Meta-Gaussian</i>	Montanari e Brath (2004)	Global	Probabilístico, análise de dados
MOSCEM-UA <i>Multiobjective shuffled complex evolution – University of Arizona</i>	Gupta <i>et al.</i> (1998)	Parâmetros	Não probabilístico, estimação de parâmetros, análise de sensibilidade
SCE-UA <i>Shuffled complex evolution – University of Arizona</i>	Duan <i>et al.</i> (1992) Sorooshian <i>et al.</i> (1993)	Parâmetros	Probabilístico, estimação de parâmetros

A adequação de determinada técnica (probabilística ou não) reside no conhecimento prévio do modelador acerca da estrutura dos erros do modelo (MONTANARI, 2011). Em princípio, essas técnicas deveriam conduzir a resultados semelhantes. Há, no entanto, alguns autores que são céticos quanto à aplicabilidade dos métodos probabilísticos devido, sobretudo, às hipóteses simplificadores acerca da natureza dos erros admitidas nesses métodos (ex.: os resíduos seguem uma distribuição Normal  $N(0, \sigma^2)$ ). Não obstante essas dificuldades, há um crescente consenso acerca da capacidade das abordagens probabilísticas de resumir eficientemente a informação contida nos dados.

De fato, o desconhecimento da estrutura dos erros e a escassez de dados requerem, além da experiência do modelador, a integração de diferentes tipos de informações dentro de um quadro que é, inevitavelmente subjetivo, uma vez que a própria informação é, muitas vezes, discutível (ex.: representatividade de um pluviômetro frente a um evento pluvial convectivo, a frequente escassez de medições de descarga líquida nos ramos superiores das curvas de descarga etc.). Diante desse quadro, há hidrólogos que optam pelo emprego de métodos não probabilísticos para avaliação da incertezas. Dentre essas técnicas, destacam-se na literatura hidrológica as baseadas nas teorias da evidência, da possibilidade e dos conjuntos difusos (*fuzzy set theory*) (BEVEN, 2009).

A presente tese orientou-se pelo emprego de um método probabilístico para avaliação da incerteza. Detalhes adicionais quanto a técnicas não probabilísticas podem ser vistos em Beven (2009), Shrestha (2009) e Montanari (2011).

### **Métodos probabilísticos**

Entre os diversos métodos utilizados para avaliação da incerteza, os denominados probabilísticos são, de longe, os mais tradicionais, bem compreendidos e mais testados (SHRESTHA, 2009). Segundo Paulino *et al.* (2003), há duas abordagens acerca do conceito de probabilidade: a frequentista ou clássica e a subjetiva ou bayesiana. Ambas compreendem que as incertezas sobre as variáveis aleatórias devem ser mensuradas via probabilidade. No entanto, no contexto clássico, a medida de probabilidade capta a variabilidade inerente ao processo e, no contexto Bayesiano, tal medida captura o desconhecimento do indivíduo sobre a variável em estudo. Como consequência, surge uma das diferenças entre a abordagem

Bayesiana e a frequentista: a forma como cada uma enxerga o significado dos parâmetros  $\theta$  dos modelos (FERNANDES, 2009).

Segundo Brooks (2003) os frequentistas veem o parâmetro  $\theta_i$  como um valor fixo (não variável) e tentam estimar esse valor, por exemplo, maximizando uma função de verossimilhança. Os Bayesianos, por outro lado, também veem os parâmetros como valores fixos e, por serem desconhecidos, são aleatórios e, assim, estão associados a uma distribuição de probabilidades (ou função densidade de probabilidade – FDP), a qual resume o conhecimento que se tem sobre essas quantidades. Os Bayesianos também acreditam que existe um valor verdadeiro para o parâmetro e utilizam sua correspondente distribuição a *posteriori* para obter estimativas pontuais de um determinado parâmetro. Assim, à medida que o conhecimento sobre o parâmetro cresce, espera-se que a incerteza sobre ele diminua.

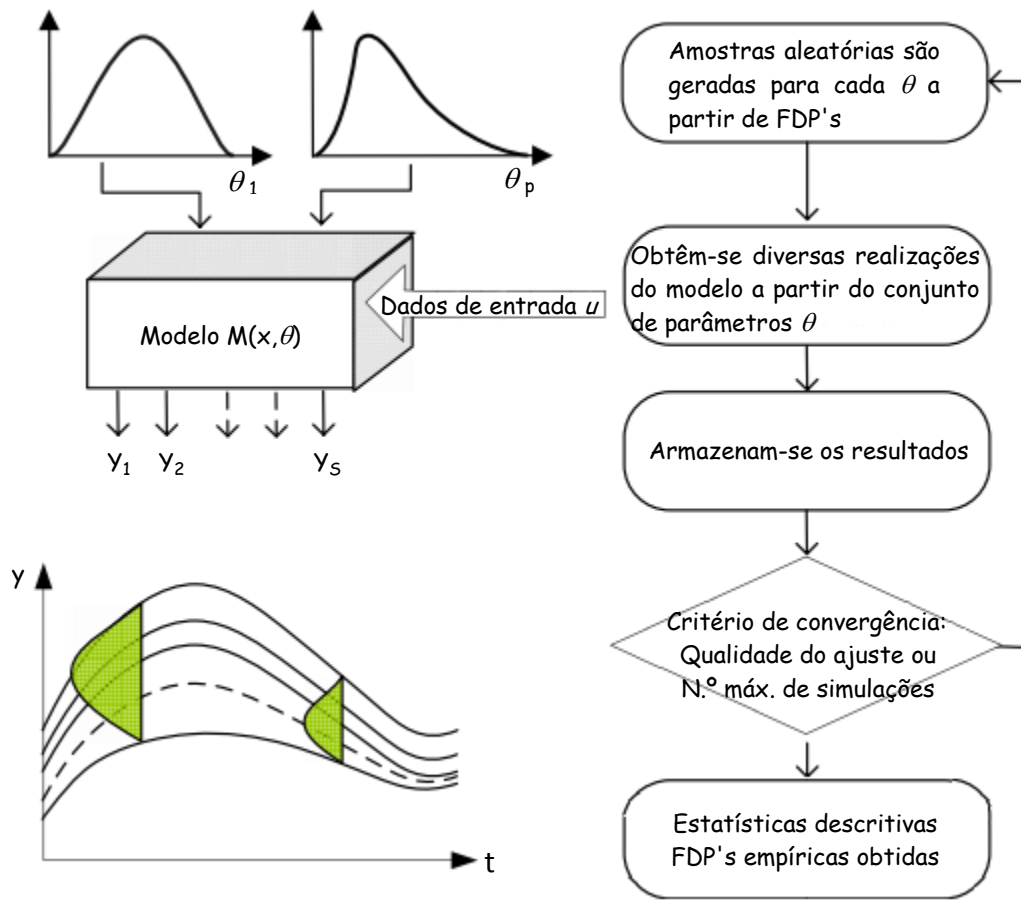
#### Abordagem Bayesiana

Na literatura técnica são descritas diversas técnicas para avaliação da incerteza em modelos chuva-vazão construídos sob o paradigma Bayesiano. Entre essas, as técnicas de simulação e amostragem baseadas em métodos de Monte Carlo têm merecido especial atenção na literatura hidrológica (BEVEN e BINLEY, 1992 *apud* BEVEN, 2009; THIEMANN *et al.*, 2001; WAGENER *et al.*, 2003; MORADKHANI *et al.*, 2005; VRUGT *et al.*, 2005). Simulações de Monte Carlo são métodos de amostragem extremamente flexíveis e robustos, o que os torna aplicáveis a uma diversa gama de problemas. No caso dos modelos chuva-vazão, as principais vantagens estão associadas à inerente dificuldade em se desenvolver métodos analíticos que permitam lidar com a complexidade e não linearidade de múltiplas variáveis presentes na estrutura desses modelos.

Nas simulações de Monte Carlo, valores aleatórios de cada variável desconhecida (ex.: parâmetros  $\theta$ ) são gerados de acordo com suas respectivas distribuições de probabilidade a *priori*. Com isso, é possível obter-se múltiplas realizações dos resultados do modelo (ex.: vazão associada a cada conjunto de  $\theta$ ). Conhecendo-se uma realização do fenômeno modelado (ex.: vazão observada), e admitindo-se um critério para avaliação de sua similaridade em relação aos resultados obtidos via simulação (ex.: função de verossimilhança), é possível sumariar, por meio de propriedades estatísticas (ex.: média, desvio padrão) e de uma distribuição de probabilidades empíricas (obtida a *posteriori*), o



comportamento exibido pelas variáveis aleatórias e pelos resultados do modelo. A Figura 3.8 ilustra uma representação esquemática desses procedimentos.



**Figura 3.8** – Representação esquemática da simulação de Monte Carlo para análise de incertezas (adaptado de SHRESTHA, 2009).

Os métodos Bayesianos aplicados à estimação de incertezas de modelos hidrológicos combinam diversas variações de simulações de Monte Carlo com o teorema de Bayes, definido analiticamente como (BOX e TIAO, 1992):

$$p(\theta|\tilde{y}) = \frac{p(\tilde{y}|\theta).p(\theta)}{p(\tilde{y})} = c.p(\tilde{y}|\theta).p(\theta) \quad (3.1)$$

$$p(\theta|\tilde{y}) \propto L(\theta|\tilde{y}).p(\theta) , \quad (3.2)$$

na qual:

- $p(\theta)$  é a probabilidade da ocorrência de  $\theta$ ,

- $p(\theta / \tilde{y})$  é a probabilidade da ocorrência de  $\theta$  dada a ocorrência de  $\tilde{y}$ ;
- $p(\tilde{y})$  e  $p(\tilde{y} / \theta)$  são definidas analogamente;
- $L(\theta / \tilde{y})$  é denominada função de verossimilhança.

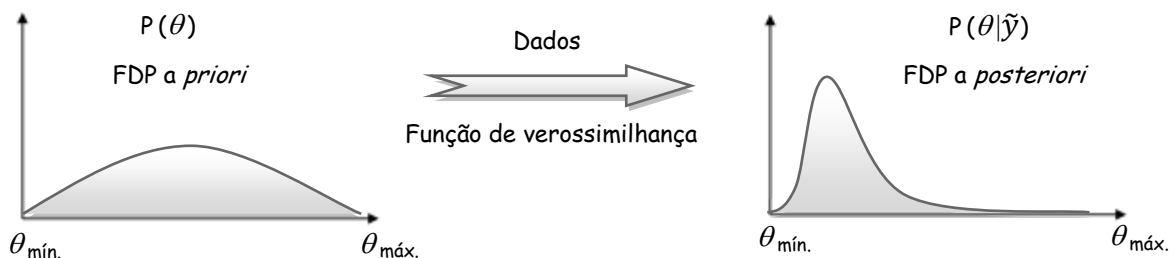
A constante  $c$  é definida como:

$$c^{-1} = p(\tilde{y}) = E[p(\tilde{y}|\theta)] = \begin{cases} \int p(\tilde{y}|\theta).p(\theta)d\theta & ; \theta \text{ contínuo} \\ \sum p(\tilde{y}|\theta).p(\theta) & ; \theta \text{ discreto} \end{cases}, \quad (3.3)$$

na qual a soma ou a integral é tomada em toda a gama de valores admissíveis de  $\theta$ , e  $E[f(\theta)]$  é a esperança matemática de  $f(\theta)$  em relação à distribuição  $p(\theta)$ .

Na expressão (3.1),  $p(\theta)$  é a distribuição *a priori* de  $\theta$  e representa o que é conhecido a respeito de  $\theta$  sem o prévio conhecimento de  $\tilde{y}$ . De forma correspondente,  $p(\theta / \tilde{y})$  reflete o conhecimento acerca dos parâmetros  $\theta$  condicionado à ocorrência de  $\tilde{y}$ , isto é, a distribuição *a posteriori* de  $\theta$ . O fator  $c$  é simplesmente uma constante de normalização, necessária para garantir que a integral completa da distribuição *a posteriori*  $p(\theta / \tilde{y})$  seja limitada à unidade.

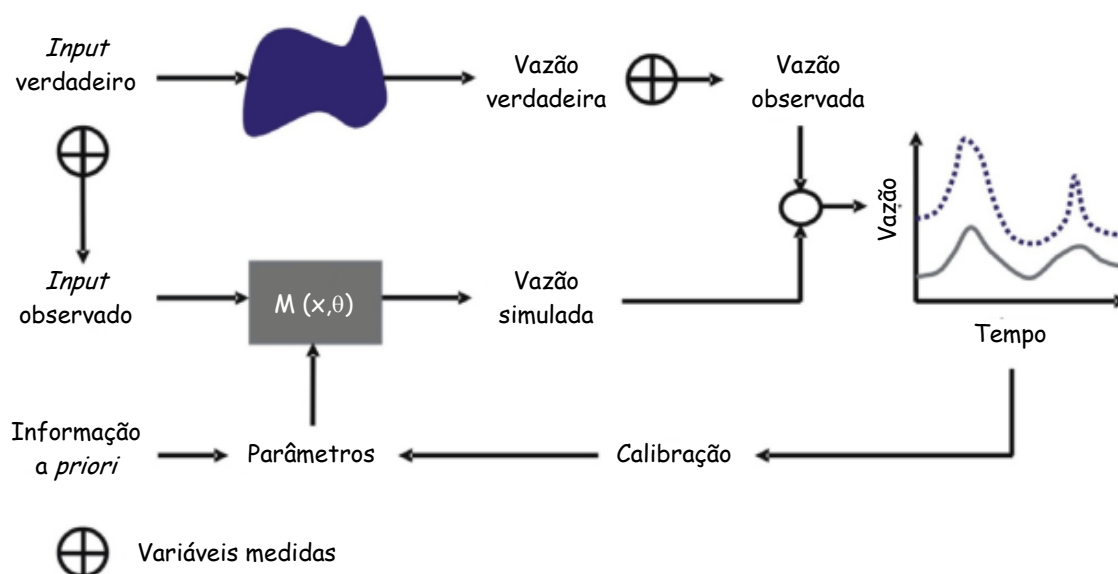
Paulino *et al.* (2003) destacam que a função de verossimilhança  $L(\theta|\tilde{y})$  desempenha um importante papel no teorema de Bayes. É por meio dela que os dados observados  $\tilde{y}$  transformam o conhecimento *a priori* sobre os parâmetros  $\theta$ . Em outras palavras, ela expressa a informação sobre  $\theta$  fornecida pelos dados  $\tilde{y}$ . Segundo Box e Tiao (1992) a função de verossimilhança representa matematicamente uma forma do conhecimento prévio acerca do conjunto de parâmetros ser continuamente atualizado à medida que novas observações do fenômeno são coletadas. A Figura 3.9 ilustra uma representação esquemática desse conceito.



**Figura 3.9** – Representação gráfica do teorema de Bayes (adaptado de KAVETSKI *et al.*, data desconhecida).

De fato, nos métodos de simulação de Monte Carlo fundamentados a partir da análise Bayesiana e utilizados para avaliação da incerteza em modelos chuva-vazão, a função de verossimilhança é empregada como critério para avaliação da convergência da simulação. O objetivo, via de regra, é maximizar (ou minimizar) seu valor a partir da comparação das séries simulada e observada. Para facilidade nos cálculos, é comum maximizar (ou minimizar) o logaritmo natural da verossimilhança, uma vez que o logaritmo é uma função monótona e estritamente crescente. Ademais, tendo em mente o teorema de Bayes, é importante observar que a aplicação do método está condicionada à admissão de um modelo probabilístico para os resíduos  $r$  [ $p(r/\theta)$ ].

Ao se avaliarem esses métodos sobre outra perspectiva, notar-se-á que os parâmetros obtidos pela aplicação dessas técnicas derivam-se de uma análise que busca essencialmente minimizar os resíduos do modelo. Por conseguinte, é comum observar na literatura técnica textos que tratam a estimação de incertezas sob a perspectiva da calibração de modelos. A Figura 3.10 apresenta a relação entre esses tópicos.



**Figura 3.10** – Avaliação das incertezas sobre a perspectiva da calibração de modelos chuva-vazão (adaptado de VRUGT *et al.*, 2008a).

Para que um modelo chuva-vazão seja útil em situações práticas, ele deve ser capaz de reproduzir as vazões observadas em uma seção fluviométrica de interesse, com suficiente grau de acurácia e precisão. Sob esse cenário, admitindo-se que a estrutura conceitual do modelo seja capaz de representar o sistema hidrológico em análise e que os erros nos dados

medidos são de natureza aleatória, o que se busca na calibração é a identificação do conjunto de parâmetros  $\theta$  que conduza à maior equivalência entre vazões observadas e simuladas.

Ao se analisar o desenvolvimento das técnicas de calibração de modelos chuva-vazão nas últimas décadas, observa-se que as pesquisas focaram, sobretudo, nos seguintes aspectos (VRUGT *et al.*, 2013):

- no desenvolvimento de funções objetivo que representem de forma apropriada as diferenças entre vazões observadas e simuladas (resíduos);
- na busca por algoritmos que possam solucionar de forma eficiente o problema da calibração;
- na determinação da quantidade e tipo de informação necessária à aplicação dos métodos e avaliação dos resultados;
- na seleção e desenvolvimento de métodos numéricos capazes de solucionar os sistemas de equações algébricas e diferenciais presentes nos modelos hidrológicos, de maneira eficiente e precisa;
- na compreensão, quantificação e redução das incertezas; e
- no desenvolvimento de técnicas que permitam o refinamento contínuo (ou em tempo real) das soluções.

Não obstante os progressos realizados, é importante destacar que a abordagem tradicional admite duas hipóteses muito fortes (VRUGT *et al.*, 2013). A primeira está associada ao fato de não se considerar que os *inputs*  $u$  dos modelos possuem erros de natureza epistêmica. Por conseguinte, presume-se que os resíduos da simulação exibirão um comportamento estatístico similar (ex.:  $r \sim N(0, \sigma^2)$ ). Há na literatura hidrológica uma série de trabalhos que demonstram que essa premissa é frequentemente violada. A segunda diz respeito ao critério utilizado como identificador de desempenho dos algoritmos de calibração. Independentemente dos fundamentos utilizados em sua definição, admitir que toda informação contida nos dados observados possa ser extraída a partir de um indicador estatístico único (a ser maximizado ou minimizado) é inadequado. Por conseguinte, diferentes conjuntos de parâmetros conduzem a

estatísticas muito similares, o que torna a escolha do “melhor” conjunto de parâmetros subjetiva. Beven e Binley (1992) *apud* Beven (2009) utilizaram o termo “equifinalidade” para explicar essa subjetividade. Ademais, argumentaram que a super parametrização dos modelos hidrológicos, aliada à escassez e/ou baixa representatividade das informações hidrológicas utilizadas durante a calibração, justificariam esse comportamento.

A utilização de métodos de simulação de Monte Carlo e técnicas de inferência Bayesiana permitem uma análise detalhada da primeira hipótese. No entanto, destaca-se que esse não é o objetivo central desta pesquisa. Ademais, ao admitir que os parâmetros sejam variáveis aleatórias que seguem um modelo probabilístico, essa abordagem, senão corrige, relaxa parcialmente a segunda hipótese. Contudo, antecipa-se que a metodologia aqui proposta valer-se-á da premissa de que, preservada a correção no comportamento admitido *a priori* para os resíduos, é possível identificar, no que tange à variação conjunta das variáveis de estado, regiões no espaço paramétrico cuja ocorrência esteja associada a um indicador de desempenho ótimo, respeitando-se, todavia, as incertezas associadas às estimativas dos parâmetros.

Esta pesquisa concentra seus esforços na análise do comportamento *input-state-output*, sob a perspectiva de assimilar (compreender) a influência do padrão de variabilidade das variáveis de estado sobre os resultados de um modelo chuva-vazão. Conceitualmente, é possível empreender essa análise a partir do emprego de técnicas que permitam a incorporação, de maneira dinâmica, de informações que propiciem a melhoria das previsões desses modelos. Esse procedimento de refinamento contínuo da solução é denominado assimilação de dados.

### **3.3 Assimilação de dados**

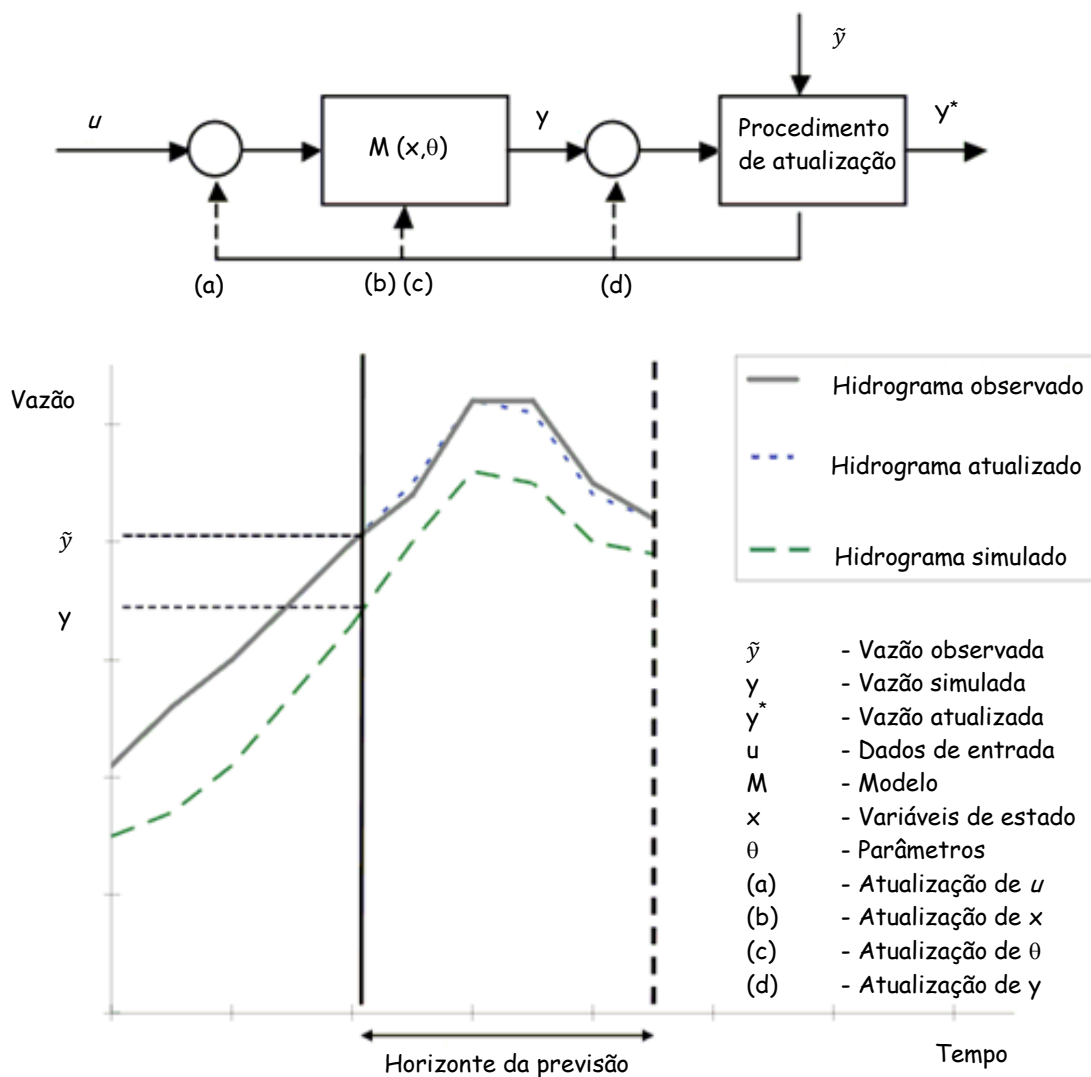
Liu e Gupta (2007) observam que há três principais áreas sobre as quais ações específicas permitiriam a redução das incertezas associadas aos prognósticos de modelos hidrológicos: (i) o desenvolvimento de técnicas de medição e observação mais confiáveis, o que permitiria a aquisição de um maior número de informações de alta qualidade; (ii) a melhoria da representação do processo físico e das técnicas matemáticas utilizadas para sua representação numérica; e (iii) o desenvolvimento de técnicas que possam extrair e assimilar as informações contidas nos dados observados à caracterização do processo. É nesta última categoria que reside a grande motivação do desenvolvimento deste trabalho.

Segundo McLaughlin (2002), os problemas relacionados à assimilação de dados são essencialmente de três tipos: interpolação, suavização e filtragem. Problemas de interpolação frequentemente surgem quando se deseja construir um mapa estático de uma variável (ex.: a precipitação instantânea) espacialmente distribuída a partir de medições pontuais dispersas. Problemas de suavização (ou reanálise) estão associados à reconstrução de condições históricas (ex.: flutuações climáticas ao longo das últimas décadas) a partir de um conjunto de dados existente. Problemas de filtragem decorrem de aplicações de controle em tempo real, ou quando se deseja inicializar modelos para fins de previsão. É este último caso, mais especificamente o problema relacionado à inicialização (ou reinicialização), o foco principal desta pesquisa.

O principal desafio da assimilação de dados é: dado um modelo (ruído) da dinâmica de um sistema, encontrar as melhores estimativas de estados ( $S \Leftrightarrow inputs\ u$ , parâmetros  $\theta$ , variáveis de estado  $x$  e/ ou *output*  $y$ ) desse sistema a partir de observações (ruído) (WALKER e HOUSER, 2005). Há, indiretamente, um segundo propósito (DRÉCOURT, 2004): incorporar informações relativas às incertezas (global ou específicas) das estimativas dos modelos. Sob essa ótica, a modelagem e/ou observação do erro nos estados de um modelo chuva-vazão pode ser utilizada para melhoria das estimativas da vazão prognosticada.

É válido destacar que o termo assimilação de dados não se limita apenas ao problema de estimação de estado, como muitas vezes se observa na literatura, mas sim a um problema mais amplo, o qual presume que a incorporação/assimilação da informação possa ser utilizada para, além da correção de  $u$ ,  $x$ ,  $\theta$ , e/ou  $y$ , a identificação/validação da estrutura conceitual admitida, o que é fundamental para a redução da incerteza na previsões de um modelo (LIU e GUPTA, 2007).

Os métodos de assimilação de dados aplicados à previsão de modelos hidrológicos podem ser organizados, conforme apresentado na Figura 3.11, em quatro categorias (WMO, 1992 *apud* PEREZ, 2009; MELLER, 2012): (a) atualização das variáveis de entrada; (b) atualização das variáveis de estado; (c) atualização dos parâmetros dos modelos; e (d) atualização das variáveis de saída.



**Figura 3.11** – Diagrama esquemático do procedimento de assimilação de dados (adaptado de SERBAN e ASKEW, 1991 *apud* WMO, 2011; e PERRIN, 2003 *apud* WMO, 2011).

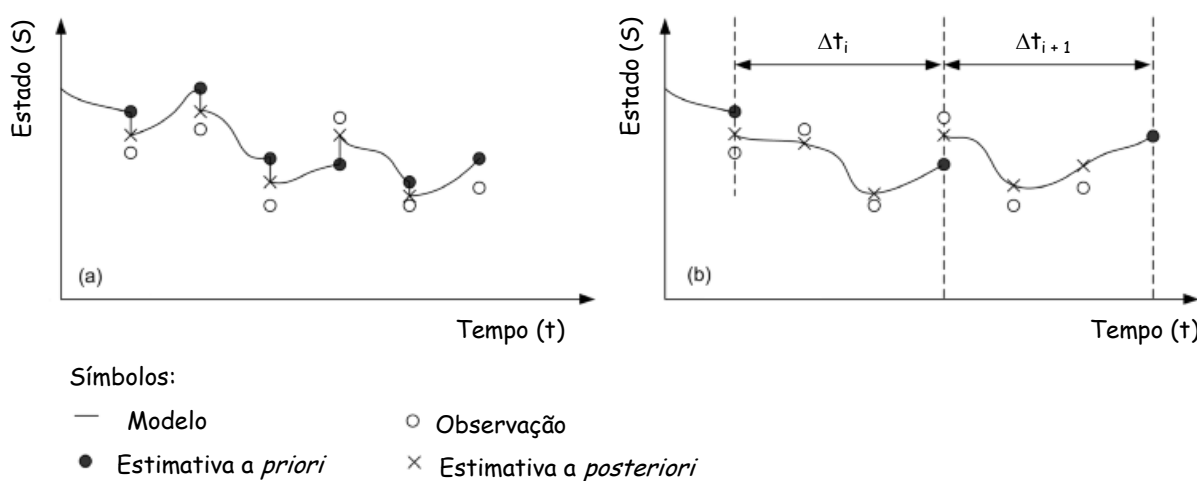
Em geral, a assimilação de dados aplicadas à modelagem chuva-vazão está associada à operação em tempo real, isto é, à previsão de estados futuros de variáveis de interesse (ex.: vazão, volume, nível de água). Tome-se, como exemplo, o problema da previsão de cheias (nível de água – NA) em um rio. No contexto da assimilação de dados, ele pode ser assim definido: conhecida a informação sobre o que está ocorrendo no instante atual,  $NA_t$ , como é possível melhorar a estimativa do que irá ocorrer no futuro ( $NA_{t+\Delta t}$ )? Ou, ainda: identificada a diferença entre a previsão do modelo e as observações disponíveis até o instante atual, como corrigir as estimativas futuras? A hipótese central dessas técnicas é que a magnitude da super ou subestimação da referida quantidade (ex.: NA) pode ser utilizada para o ajuste da previsão, sob a perspectiva de torná-la mais precisa, embora não haja garantia disso, visto que as

incertezas associadas às observações (dos *inputs*, variáveis de estado e/ou *output*) podem ser determinantes sobre o viés identificado na previsão.

Essas técnicas, no tocante à atualização dos estados no tempo, podem ser agrupadas em duas categorias (DRÉCOURT, 2004; NICHOLS, 2010):

- sequencial ou direta: na qual as observações retroalimentam o modelo a cada vez que estão disponíveis, sob o propósito de produzir uma melhor estimativa para o estado atual. Ela é sugerida, em princípio, para situações na qual a acurácia do sistema é controlada pelas condições de contorno; e
- dinâmica: na qual, a partir das observações, busca-se uma trajetória ótima dos estados ao longo de uma janela (intervalo) de tempo a fim de se estimar o estado atual. Nesta abordagem, admite-se que o comportamento do sistema é conduzido pela acurácia das condições iniciais.

A Figura 3.12 apresenta essa classificação.



**Figura 3.12** – Técnicas de assimilação: (a) Sequencial; (b) Dinâmica (adaptado de HOUSER *et al.*, 2012).

Segundo Walker e Houser (2005), as técnicas sequenciais são mais comumente empregadas na modelagem hidrológica. As principais são: (i) inserção direta; (ii) correção estatística; (iii) correção sucessiva (SCM); (iv) interpolação ótima ou estatística; (v) análise corretiva; (vi) *nudging*; (vii) 3D-Var; e (viii) filtros de Kalman e suas variações. Dentre as dinâmicas, destaca-se a denominada 4D-Var, ressaltando-se que é raro verificar na literatura hidrológica



aplicações dessa técnica devido à natureza não linear dos processos hidrológicos (HOUSER *et al.*, 2012).

Uma descrição pormenorizada dessas técnicas extrapola o escopo deste texto. Restringir-se-á aqui àquelas que instigam a abordagem metodológica sugerida. Detalhes adicionais acerca dessas técnicas podem ser vistos em LAHOZ *et al.* (2010).

### 3.3.1 Filtro de Kalman e suas variações

Antes de apresentar o conceito do filtro de Kalman, é válido analisar o algoritmo recursivo de estimação por mínimos quadrados. Seguindo Beven (2009), considere-se como referência o caso da regressão linear múltipla, para o qual a predição de uma variável observável  $\tilde{y}$  é representada pelo produto de um número definido de variáveis independentes  $u$  e um vetor de estados  $x$ , adicionado a um erro aleatório  $e_t$ . Em termos formais, têm-se:

$$\tilde{y}_t = x_1 \cdot u_t + x_2 \cdot u_{t-1} + \dots + x_n \cdot u_{t-n-1} + e_t \quad (3.4)$$

$$\tilde{y}_t = u_t^T \cdot x + e_t \quad (3.5)$$

$$\tilde{y}_t = M[x_{t_{1:n}}, u_{(t-n-1):t}] + e_t, \quad (3.6)$$

na qual o símbolo  $T$  indica o vetor/matriz transposta e:

$$u_t^T = [u_t, u_{t-1}, \dots + u_{t-n-1}]; x = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix}_t. \quad (3.7)$$

O objetivo do problema é atualizar os valores do vetor de estados  $x$  (estacionários). Em um procedimento de previsão em tempo real, busca-se, a partir da identificação da magnitude da diferença entre a estimativa do modelo e a observação, atualizar a estimativa desses estados. Isso pode ser alcançado por meio da correção ponderada dessa diferença, analiticamente representada para o modelo analisado como:

$$x_t^* = x_t + W[\tilde{y}_t - y_t], \quad (3.8)$$

na qual  $W$  é denominada matriz de pesos de inovação e  $\tilde{y}_t$  é a observação no instante  $t$ . Admite-se para tal que os estados sejam independentes e não correlacionados em relação ao

termo do erro  $e_t$ . Assim sendo, a solução é alcançada pela minimização da função custo  $J$ , representada pela soma ao longo do tempo ( $k = 1, 2, \dots, t$ ) dos desvios quadráticos do modelo da predição ( $u_t^T \cdot x$ ) em relação ao valor observado  $\tilde{y}_t$ , isto é:

$$J = \sum_{k=1}^t [\tilde{y}_k - u_k^T \cdot x]^2 = \sum_{k=1}^t e_k^2 . \quad (3.9)$$

$J$  será minimizado quando as derivadas parciais em relação a cada estado  $x$  forem iguais a zero, ou seja:

$$-2 \cdot \sum_{k=1}^t u_k \cdot \tilde{y}_{kt} + \{2 \sum_{k=1}^t u_{kt} \cdot u_k^T\} \cdot x = 0 \quad (3.10)$$

ou

$$x = \sum_{k=1}^t u_k \cdot \tilde{y}_k \cdot [\sum_{k=1}^t u_k \cdot u_k^T]^{-1} \quad (3.11)$$

$$x = P_t \cdot b_t , \quad (3.12)$$

na qual

$$P_t = [\sum_{k=1}^t u_k \cdot u_k^T]^{-1}; b_t = \sum_{k=1}^t u_k \cdot \tilde{y}_k . \quad (3.13)$$

Essa solução pode ser implementada recursivamente, atualizando os estados a cada instante em que uma nova observação  $\tilde{y}_{t+1}$  estiver disponível, da seguinte forma:

$$P_{t+1}^{-1} = P_t^{-1} + u_{t+1} u_{t+1}^T ; b_{t+1} = b_t + u_{t+1} \tilde{y}_{t+1} . \quad (3.14)$$

O vetor de estados pode ser assim atualizado:

$$x_{t+1} = x_t + P_t u_{t+1} [1 + u_{t+1}^T P_t u_{t+1}]^{-1} \{\tilde{y}_{t+1} - u_{t+1}^T x_t\} . \quad (3.15)$$

Essa expressão pode ser reescrita conforme a seguinte equação:

$$x_{t+1} = x_t + W \{\tilde{y}_{t+1} - u_{t+1}^T x_t\} , \quad (3.16)$$

na qual:

$$W = P_t u_{t+1} [1 + u_{t+1}^T P_t u_{t+1}]^{-1} . \quad (3.17)$$

É válido notar que, caso seja presumido que o erro  $e_t$  é distribuído normalmente, devido à linearidade admitida pelo sistema, as estimativas dos estados  $x_t$  também exibirão um comportamento Gaussiano e, então,  $P_t$  será uma matriz de covariância simétrica e positiva, definida como o valor esperado dos erros quadráticos estimados para os estados no instante  $t$ , isto é:

$$P_t = E[(x_t - x_t^*) \cdot (x_t - x_t^*)^T] , \quad (3.18)$$

na qual, o valor esperado é determinado para todas as combinações possíveis de erros conjuntos entre os valores verdadeiros ( $x^*$ ) e estimados dos estados no tempo  $t$ . Os valores ao longo da diagonal da matriz representam a variância do  $i$ -ésimo estado (BEVEN, 2009).

Ainda segundo Beven (2009), no algoritmo recursivo de estimação por mínimos quadrados admite-se que os estados  $x_t$  são invariantes (estacionários). Por conseguinte, caso seja fornecida uma quantidade suficiente de dados, os estados convergirão para valores constantes. O Filtro de Kalman não assume esta hipótese, mas admite que a estrutura que conecta os valores dos estados à variável observada seja fixa. Ele estende o algoritmo recursivo de estimação por mínimos quadrados para permitir que os estados do modelo possam ser escritos como funções lineares (MORADKHANI *et al.*, 2005) da seguinte forma:

$$x_{t+1} = M[x_t, u_t, \theta] + w_t , \quad (3.22)$$

$$y_{t+1} = H[x_{t+1}, \phi] + v_{t+1} , \quad (3.23)$$

nas quais o vetor de estados atualizado  $x_{t+1}$  é determinado pelo operador de modelo  $M(\cdot)$  a partir dos estados  $x$  (ex.: umidade média do solo na bacia) e dos *inputs*  $u$  (ex.: precipitação) no instante de tempo anterior  $t$ , e do conjunto de parâmetros  $\theta$ . O operador de observações  $H(\cdot)$  relaciona as variáveis de estado e um conjunto adicional de parâmetros de medição  $\phi$  à variável observável (ex.: vazão). Os termos  $w_t$  e  $v_{t+1}$  são erros aleatórios.

Caso o conjunto de parâmetros  $\phi$  seja invariante, os operadores  $M(\cdot)$  e  $H(\cdot)$  sejam funções lineares fixas, e os erros aleatórios  $w_t$  e  $v_{t+1}$  sejam mutuamente independentes e modelados probabilisticamente por uma distribuição Normal com média zero e desvio padrão dado pelas covariâncias  $W_t$  e  $V_{t+1}$   $\{w_t \sim N(0, W_t) ; v_{t+1} \sim N(0, V_{t+1})\}$ , ter-se-á que a correção (ou

inovação) será definida, analogamente ao algoritmo recursivo de estimação por mínimos quadrados, como:

$$x_{t+1}^* = x_{t+1} + K[\tilde{y}_{t+1} - Hx_{t+1}], \quad (3.24)$$

onde  $K$  é denominado filtro de Kalman e é definido como:

$$K_{t+1} = P_{t+1}H^T[V_{t+1} + HP_{t+1}H^T]^{-1}, \quad (3.25)$$

na qual  $P_{t+1}$  é obtido a partir da equação 3.18.

O filtro de Kalman pode ser estendido para lidar com modelos não lineares ( $M$ ). Na abordagem denominada EKF (*Extended Kalman Filter*) isso é realizado a partir da linearização da variância prevista (ou covariância no caso da atualização de múltiplos estados) para cada instante  $t$  (BEVEN, 2009). Em razão das aproximações envolvidas, essa abordagem frequentemente introduz instabilidades ou até mesmo divergências na simulação (MORADKHANI *et al.*, 2005). Em razão desse fato, Evensen (1994) *apud* Kalnay (2010) introduziu o método EnKF (*Ensemble Kalman Filter*), uma extensão do algoritmo na qual um conjunto de soluções (*ensemble*) é gerado (via simulações de Monte Carlo) por meio de uma aproximação *a priori* da matriz de covariância do erro do estado, a qual é construída a partir dos estados atualizados em intervalos de tempo anteriores. MORADKHANI *et al.* (2005) ressaltam que essa abordagem requer que a cada geração o erro aleatório introduzido ao *ensemble* seja Gaussiano, com média zero e covariância igual à variância do erro observado.

Uma descrição detalhada dessas extensões pode ser vista em LAHOZ *et al.* (2010).

### 3.3.2 Filtros bayesianos

Sob uma perspectiva probabilística Bayesiana, o problema da estimação recursiva requer a construção de uma distribuição *a posteriori*  $p_\theta(x_{1:t}|\tilde{y}_{1:t})$ . Admitindo-se que o vetor de estados inicial possa ser aproximado como  $p_\theta(x_0|\tilde{y}_0) \sim p_\theta(x_0)$ , a FDP pode ser construída recursivamente em dois estágios: previsão e atualização.

A evolução da distribuição *a posteriori*  $p_\theta(x_{1:t}|\tilde{y}_{1:t})$ , para as expressões 3.22 e 3.23, pode ser escrita recursivamente como (VRUGT *et al.*, 2013):

$$p_{\theta}(x_{1:t}|\tilde{y}_{1:t}) = \overbrace{p_{\theta}(x_{1:t-1}|\tilde{y}_{1:t-1})}^{\text{priori}} \frac{\overbrace{M_{\theta}(x_t|x_{t-1})}^{\text{modelo}} \cdot \overbrace{L_{\theta}(\tilde{y}_t|x_t)}^{\text{função de verossimilhança}}}{\underbrace{p_{\theta}(\tilde{y}_t|\tilde{y}_{1:t-1})}_{\text{constante de normalização}}}, \quad (3.26)$$

na qual  $M_{\theta}(x_t|x_{t-1})$  representa a função densidade de probabilidade de transição do modelo de estados (operador de modelo) e  $L_{\theta}(\tilde{y}_t|x_t)$  é a função densidade de probabilidade das observações dadas as variáveis de estado. Essa função utiliza o operador de observações da expressão 3.23 para relacionar  $x_t$  às variáveis observáveis.

A função marginal de distribuição de probabilidades,  $p_{\theta}(x_t|\tilde{y}_{1:t})$ , é obtida a partir da integração e manipulação de  $x_{1:t-1}$  na equação 3.26 (VRUGT *et al.*, 2013):

$$\overbrace{p_{\theta}(x_t|\tilde{y}_{1:t})}^{\text{atualização}} = \frac{L_{\theta}(\tilde{y}_t|x_t) \cdot p_{\theta}(x_t|\tilde{y}_{1:t-1})}{p_{\theta}(\tilde{y}_t|\tilde{y}_{1:t-1})}, \quad (3.27)$$

na qual:

$$\overbrace{p_{\theta}(x_t|\tilde{y}_{1:t-1})}^{\text{previsão}} = \int M_{\theta}(x_t|x_{t-1}) \cdot p_{\theta}(x_{t-1}|\tilde{y}_{1:t-1}) dx_{t-1}, \quad (3.28)$$

As etapas de previsão e atualização resolvem numericamente a distribuição a *posteriori* dos estados dada a informação das observações até o tempo atual  $t$ . Caso os parâmetros sejam desconhecidos, é possível inferí-los, simultaneamente aos estados, a partir da função densidade a *priori* dos parâmetros  $p(\theta)$  (VRUGT *et al.*, 2013):

$$p(\theta, x_{1:t}|\tilde{y}_{1:t}) \propto p(\theta) \cdot p_{\theta}(x_{1:t}|\tilde{y}_{1:t}), \quad (3.29)$$

onde  $p(\theta, x_{1:t}|\tilde{y}_{1:t})$  é a função densidade de probabilidade conjunta.

VRUGT *et al.* (2013) observam que, caso  $H(\cdot)$  seja linear e  $w_t$  e  $v_t$  (equações 3.22 e 3.23) sejam Gaussianos, a abordagem do filtro de Kalman encontrará a distribuição a *posteriori* exata  $\{p_{\theta}(x_{1:t}|\tilde{y}_{1:t})$  ou  $p(\theta, x_{1:t}|\tilde{y}_{1:t})\}$ . No caso de problemas que envolvam modelos não lineares e erros não Gaussianos, é necessário recorrer a técnicas baseadas em métodos de simulação de Monte Carlo (ex.: *EnKF*). Uma abordagem construída sob a perspectiva Bayesiana e que é capaz de lidar com esse tipo de problema é o método SMC (*Sequential Monte Carlo*). VRUGT *et al.* (2013) introduziram uma variação desse método baseada no

algoritmo admitido para o desenvolvimento da presente tese. Esse filtro Bayesiano sequencial permite a estimação de parâmetros e/ou variáveis de estado para o instante subsequente ao passo de tempo atual.

É notório no problema da previsão hidrológica, e na assimilação de dados aplicada à mesma, a importância da antecedência temporal. Não obstante esse fato, Beven (2009) observa que há diversos trabalhos na literatura hidrológica em que essa análise ocorre somente para um passo de tempo subsequente ao da previsão ( $t+1$ ). Apesar de revelarem, em geral, uma significativa melhoria nos prognósticos, a análise torna-se discutível, uma vez que o problema da tomada de decisão (antecedência da previsão) no âmbito do sistema de interesse é tratado de forma secundária.

### **3.4 Comentário geral**

Os tópicos de predição da incerteza e da assimilação de dados aplicadas à modelagem chuva-vazão, embora tratem-se de ações com objetivos distintos, se sobrepõem parcialmente, principalmente quando analisa-se o problema sob a ótica da simulação e admite-se *a priori* o conhecimento pleno do *input* e do *output*. Essa foi a abordagem metodológica assumida em uma série de publicações recentes. Nota-se, ademais, que há uma prevalência do primeiro tópico no tocante ao desenrolar da discussão. Nesta tese, as análises e argumentações, também, delinearam-se nesse sentido.

Na última década diversos autores se dedicaram a explorar a interface entre esses temas. De forma geral, as publicações que representam o estado da arte na atualidade, sob a perspectiva da predição de incerteza, podem ser assim categorizadas: (i) as que indicam as incertezas aos dados de entrada/ saída; (ii) as que preconizam a estimação conjunta de parâmetros e variáveis de estado; e (iii) as que se propuseram ao desenvolvimento de algoritmos mais eficientes/robustos. Em muito menor número, foram observadas abordagens que procuravam conciliar diferentes correntes de pensamento, sendo essas, possivelmente, as mais “propositivas” no tocante a avanços “reais” nesses temas.

Isso posto, presume-se que seja relevante apresentar ao leitor uma descrição das abordagens metodológicas observadas na literatura atual, as quais integram elementos de ambos os tópicos apresentados ao longo deste capítulo. No entanto, optou-se por realizá-la na forma de uma resenha crítica, no estágio final deste documento, mais precisamente durante a discussão

dos resultados. Há duas razões para isso. A primeira apoia-se na observação de que a predição da incerteza na modelagem hidrológica é uma das áreas de pesquisa mais ativas em hidrologia atualmente (MONTANARI, 2011). Corrobora para essa observação o crescente número de publicações dedicadas a esse tema nos últimos anos (KUCZERA e PARENT, 1998; VRUGT *et al.*, 2003; WAGENER, *et al.*, 2003; KUCZERA *et al.*, 2006; SIROSKA *et al.*, 2012; SUN e BERTRAND-KRAJWESKI, 2013). Por conseguinte, uma apresentação sumária, possivelmente, seria incapaz de cingir todos os aspectos metodológicos envolvidos, ainda que se restringisse àquelas em consonância com a metodologia aqui proposta, descrita adiante neste documento. Ademais, acredita-se que, ao optar-se por realizá-la àquela altura no texto e no referido formato, possibilite-se uma leitura mais fluida e não enviesada. É este o segundo ponto: a crença de que a clareza do documento não será comprometida e que o leitor não será persuadido à restringir-se, *a priori*, ao olhar do autor acerca desses temas.

Na sequência deste documento, a metodologia proposta é apresentada.

## 4 METODOLOGIA

Para assimilar o padrão de variabilidade das variáveis de estado de um modelo chuva-vazão conceitual em esquemas de simulação/previsão hidrológica adotou-se a seguinte sequência metodológica:

- identificação de uma bacia hidrográfica que possua monitoramento hidrológico adequado às necessidades da pesquisa;
- seleção de um modelo chuva-vazão de base conceitual, determinístico e concentrado (semi-distribuído);
- escolha de um método de simulação de Monte Carlo que se valha de técnicas de inferência bayesiana e cuja aplicabilidade à avaliação de incertezas e à calibração de modelos hidrológicos tenha sido anteriormente verificada;
- calibração dos parâmetros e estimação das incertezas associadas;
- identificação de períodos nos quais observe-se uma dissimilaridade “persistente” entre as séries de vazões observadas e modeladas;
- estimação das variáveis de estado ao início dos períodos de dissimilaridades identificados;
- análise do padrão de variabilidade das trajetórias das variáveis de estado, considerando-se as incertezas associadas à modelagem hidrológica.

Pretende-se, inicialmente, determinar a distribuição conjunta de parâmetros  $\theta$  que melhor expliquem as vazões observadas. Em contraste à abordagem convencional, não se busca aqui uma solução única, ótima, mas sim conjuntos que reflitam as incertezas associadas a essas estimativas. Planeja-se, então, identificar períodos (intervalos de tempo) nos quais se observe uma dissimilaridade contínua entre as séries (de vazões) observadas e modeladas. Na sequência, identificar-se-ão as variáveis de estado que, ao início desses períodos, minimizem os resíduos calculados em um horizonte de previsão pré-fixado. Admitir-se-á, para isso, que as precipitações posteriores ao instante  $t$  são as alturas de chuva observadas.



A hipótese da abordagem aqui proposta é que as variáveis de estado presentes nos modelos chuva-vazão podem ser tratadas como aleatórias. Por conseguinte, elas podem ser modeladas valendo-se das mesmas técnicas e hipóteses utilizadas na determinação do conjunto ótimo de parâmetros  $\theta$ .

As variáveis de estado  $x$ , admitidas *a priori* em função da estrutura conceitual (do modelo) considerada, constituem-se no eixo que interliga os fatos (a percepção do modelador acerca dos principais processos físicos que representam o fenômeno a ser modelado) às observações (os dados empregados para descrição do fenômeno à escala da bacia). Elas incorporam tanto a subjetividade inerente à conceitualização do fenômeno, quanto a complexidade da representação espaço-temporal das variáveis hidrológicas. Ademais, sob a ótica matemática, representam, também, a memória do sistema modelado.

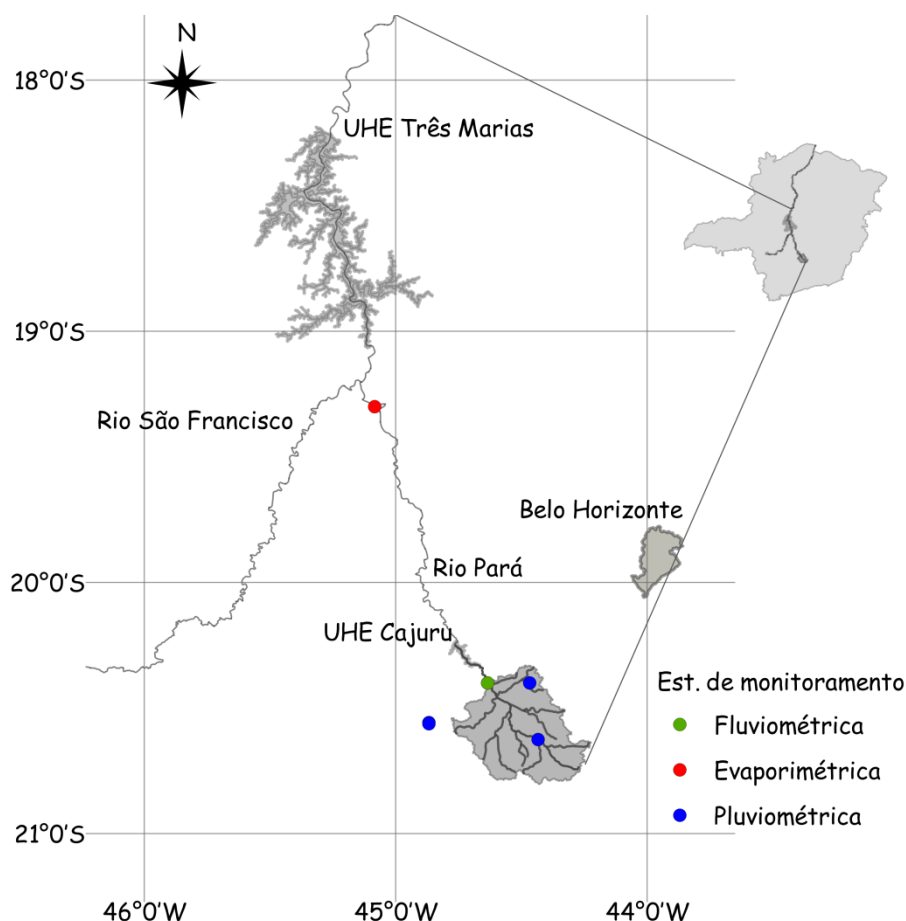
Pressupõe-se na presente abordagem metodológica que a estrutura conceitual considerada é capaz de representar o sistema hidrológico de interesse com suficiente grau de precisão e acurácia. Assim sendo, eventuais vieses em períodos específicos seriam impostos ora por erros e/ou baixa representatividade (espaço-temporal) dos dados observados empregados (*input*  $u$  e/ou *output*  $\tilde{y}$ ), ora por simplificações admitidas na representação do fenômeno modelado, ou, ainda, pela ação de ambos. A hipótese primordial é que o comportamento (trajetória) das variáveis de estado é capaz de caracterizar esses vieses, já que elas assimilam, a cada passo de tempo, todas as fontes de incertezas associadas à modelagem.

De fato, admite-se que essas trajetórias representem uma espécie de assinatura da incerteza global. Ao alterá-las, atualizá-las em um instante de tempo específico sob o prisma de garantir uma maior similaridade entre as séries (de vazão) modeladas e observadas, poder-se-á avaliar o comportamento *input-state-output* do sistema. A perspectiva é validar, contrapor, discutir algumas das impressões/conclusões descritas na literatura hidrológica recente, em especial, a dedicada à avaliação das incertezas na modelagem hidrológica.

Nos subitens que se seguem as informações básicas e métodos utilizados são apresentados de maneira pormenorizada.

#### 4.1 *Bacia hidrográfica e informações hidrológicas*

O principal critério adotado para a seleção da bacia hidrográfica para o caso de estudo foi a disponibilidade e pré-consistência das informações hidrológicas empregadas. Uma sub-bacia do rio Pará (área de drenagem  $\approx 1620 \text{ km}^2$ ; precipitação anual  $\approx 1580 \text{ mm}$ ), distante cerca de 90 km de Belo Horizonte, foi selecionada para o desenvolvimento dos estudos. A Figura 4.1 apresenta a localização da bacia.

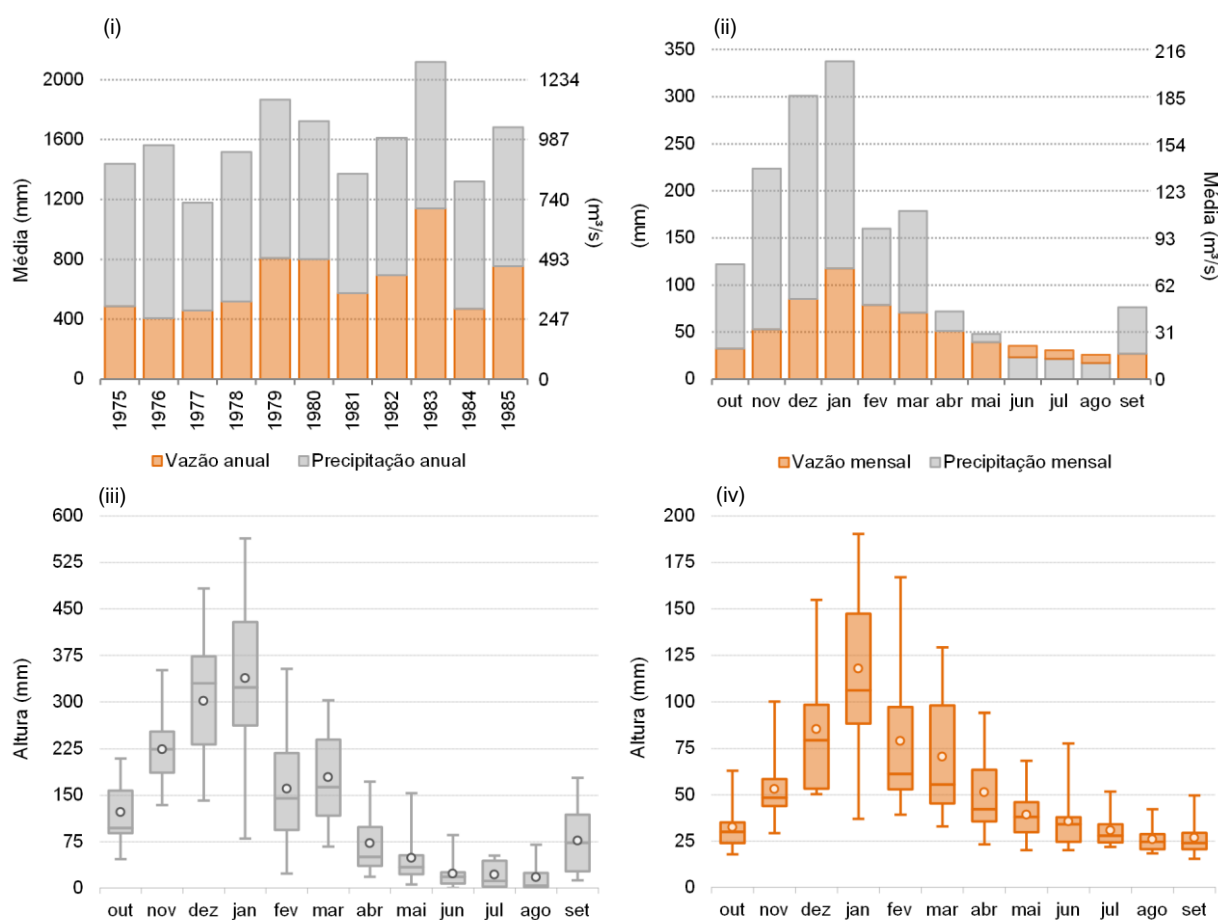


**Figura 4.1** – Localização da bacia e estações de monitoramento analisadas (*Datum: WGS84*).

Na Tabela 4.1 é apresentada uma síntese das informações hidrológicas utilizadas. A escala temporal dos dados é diária. Destaca-se que esses dados foram previamente analisados por Lima (2004) e Fernandes (2006). As vazões e precipitações médias características do período/bacia analisada são apresentadas, na sequência, por meio da Figura 4.2.

**Tabela 4.1** – Informações hidrológicas empregadas no estudo.

Informação	Entidade responsável	Código ANA	Nome	Período	
				Calibração	Validação
Pluviométrica	ANA	02044009	Fazenda Campo Grande	Out/1975 a Set/1980	Out/1981 a Set/1985
	CEMIG	02044036	Itaguara		
	ANA	02044042	Carmo da Mata		
Evaporimétrica	CEMIG	01945017	Porto Pará		
Fluviométrica	CEMIG	40130002	Ponte do Vilela – Jusante		



**Figura 4.2** – Vazões e precipitações médias no período na bacia analisada: (i) média anual; (ii) média mensal; (iii) variabilidade mensal da precipitação; (iv) variabilidade mensal da vazão.

## 4.2 Modelo chuva-vazão

Os estudos aqui descritos fizeram uso do modelo de simulação hidrológica Rio Grande, desenvolvido pelo Departamento de Engenharia Hidráulica e Recursos Hídricos da UFMG (EHR/UFMG), com a finalidade primeira de controle operacional das usinas hidrelétricas da Companhia Energética de Minas Gerais (CEMIG) (EHR/UFMG, 2001). Note-se que qualquer

outro modelo conceitual de simulação chuva-vazão poderia ser igualmente empregado para os objetivos desta pesquisa, desde que sua estrutura e experiência de utilização sugerissem um suficiente realismo com os processos hidrológicos naturais.

O Rio Grande compreende um modelo matemático completo das funções de produção e transferência do ciclo hidrológico, cujas equações procuram sintetizar as relações entre as variáveis do estado atual, permitindo o prognóstico dos estados futuros. Ademais, o modelo contempla um módulo de propagação de vazões, o que permite construir diferentes cenários de um sistema hidrológico. Por conseguinte, o modelo é classificado como conceitual, determinístico e semidistribuído (EHR/UFMG, 2001).

Nesse modelo, a síntese determinística contínua da transformação chuva-vazão é realizada de modo concentrado, por meio de duas funções: a primeira, de produção, a qual executa a separação de um certo pulso de chuva bruta em deflúvio total por unidade de tempo, enquanto a segunda, de concentração, faz a distribuição temporal dos pulsos de deflúvios, calculados em diferentes intervalos de tempo, concentrando-os e transformando-os em vazões na seção exutória.

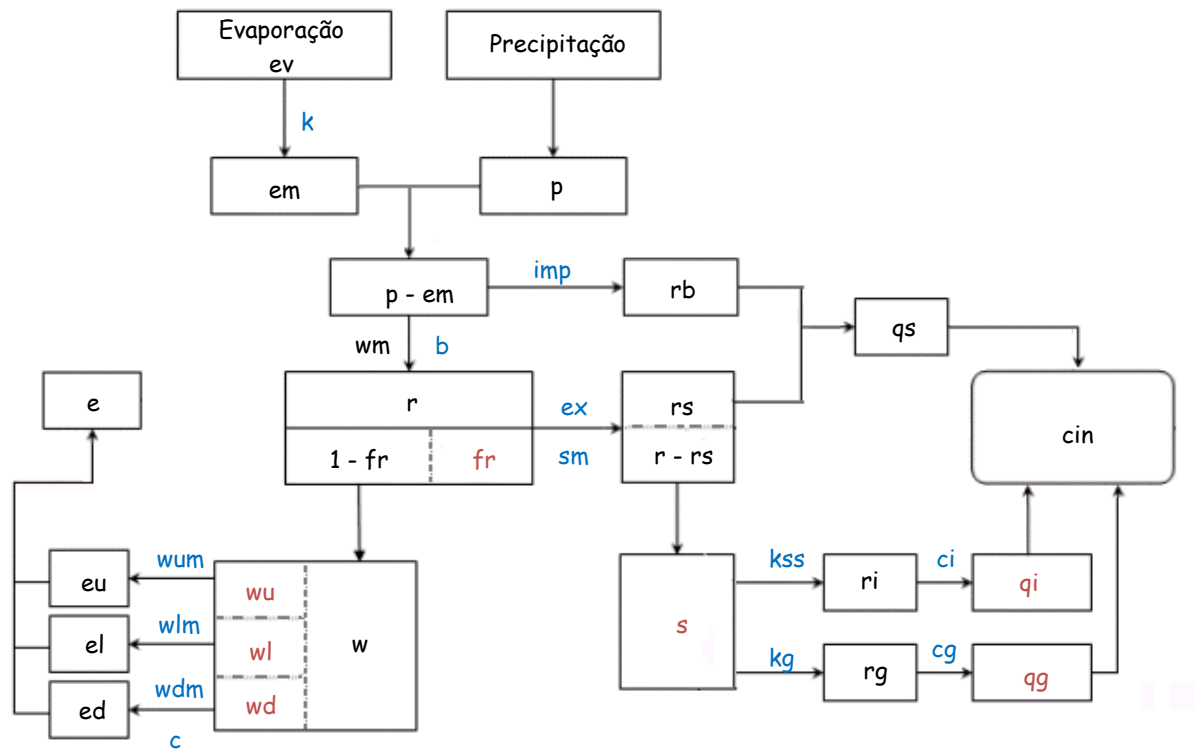
A função de produção tem como base a estrutura do balanço de umidade do modelo conceitual Xinanjiang, descrito por Zhao (1992). A função de concentração utiliza o conceito do hidrograma unitário de Clark, na versão simplificada descrita por HEC (1981), a qual necessita apenas da especificação da forma predominante da bacia.

Nos subitens que se seguem os fundamentos teóricos, a conceituação e a formulação matemática dos módulos que compõem o modelo utilizado são descritos.

### **Módulo de produção**

O conceito de formação do escoamento no modelo baseia-se na hipótese de que em áreas úmidas e semi-úmidas a intensidade da precipitação não representa o fator mais importante para se determinar o volume de escoamento superficial, em consonância com as concepções introduzidas por Dunne *et al.* (1975) *apud* EHR/UFMG (2001). O escoamento superficial e, por extensão, o escoamento subsuperficial irão depender, além do volume de precipitação, da umidade inicial presente no solo. A premissa é que não ocorre escoamento até que a umidade do solo da zona aerada atinja a chamada capacidade de campo, a qual representa o

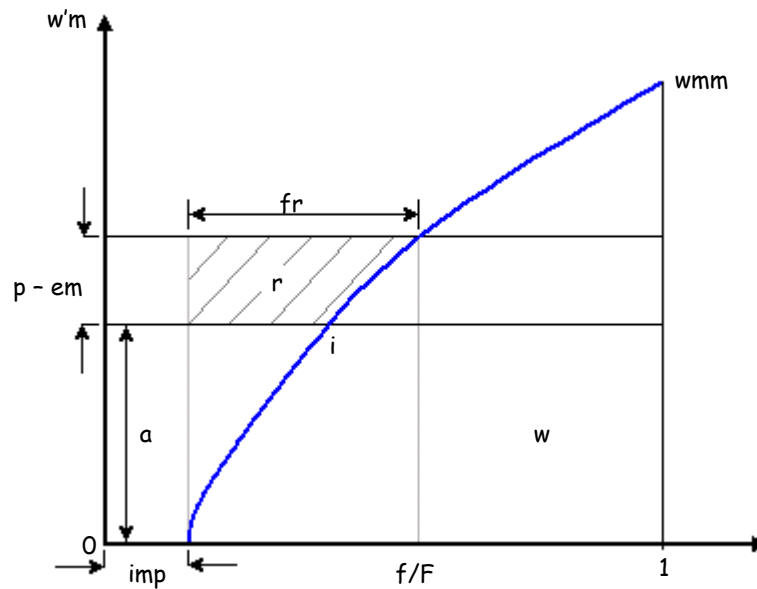
armazenamento proporcionado pela capacidade do solo em reter, contra a ação da gravidade, uma determinada quantidade de água. Por definição, a água contida nesses reservatórios hipotéticos não produz escoamento e pode deles ser extraída apenas pelas ações da evaporação das camadas superficiais do solo e/ou da transpiração pela vegetação.



Em linhas gerais, pode-se sintetizar o módulo de produção do modelo Rio Grande nas seguintes etapas:

### Produção do escoamento

Na modelagem hidrológica de uma bacia deve-se observar o fato de que a distribuição espacial da capacidade de retenção de umidade do solo varia de maneira complexa na extensão de sua área. No modelo Rio Grande, essa não uniformidade espacial é representada por meio da curva paramétrica ilustrada pela Figura 4.4.



**Figura 4.4** – Distribuição espacial da capacidade de retenção de água do solo.

Nessa curva  $w'm$  representa a capacidade de retenção pontual,  $f/F$  a fração de área permeável ( $f$ ) em relação à área total ( $F$ ) da bacia, cuja capacidade de retenção é menor ou igual a  $w'm$ , e  $imp$  corresponde à fração de área impermeável.

A capacidade de retenção de água do solo, em um ponto qualquer da bacia, varia de zero a um valor máximo, de acordo com a relação:

$$\left(1 - \frac{f}{F}\right) = \left(1 - \frac{w'm}{wmm}\right)^b \cdot (1 - imp) \quad (4.1)$$

na qual os termos  $wmm$  e  $b$  representam a máxima capacidade de retenção pontual de umidade do solo na bacia e a não uniformidade da curva, respectivamente.

A capacidade média espacial de retenção de umidade no solo  $w_m$  relaciona-se à máxima pontual por meio do parâmetro  $b$ , tal como apresentado na equação a seguir, obtida pela integração da expressão 4.1:

$$w_{mm} = w_m \cdot \left( \frac{1+b}{1+imp} \right) \quad (4.2)$$

O estado de umidade da bacia, em um dado instante, é representado pela curva da Figura 4.4. O ponto  $i$ , correspondente à ordenada  $a$ , representa o estado inicial. A área à direita da curva e abaixo do ponto  $i$  é proporcional ao teor de umidade média espacial do solo. Essa hipótese implica que cada ponto da bacia ou já atingiu a capacidade de retenção específica (pontos à esquerda de  $i$ ) ou está com um teor de umidade constante (pontos à direita de  $i$ ).

Quando, em um dado intervalo de tempo, a precipitação é maior que a evapotranspiração potencial ( $em$ ), o teor de umidade média espacial no solo sofre um aumento, provocado pelo acréscimo de volume na ordenada da Figura 4.4, e pelo deslocamento do ponto  $i$ , correspondente à umidade inicial, para a direita (sobre a curva). Dessa forma, a área hachurada à esquerda da curva representa o escoamento  $r$  gerado no intervalo.

Caso  $(p - em) + a \leq w_{mm}$ :

$$r = (p - em) - w_m + w + w_m \cdot \left( 1 - \frac{a+(p-em)}{w_{mm}} \right)^{1+b} \quad (4.3)$$

Em caso contrário:

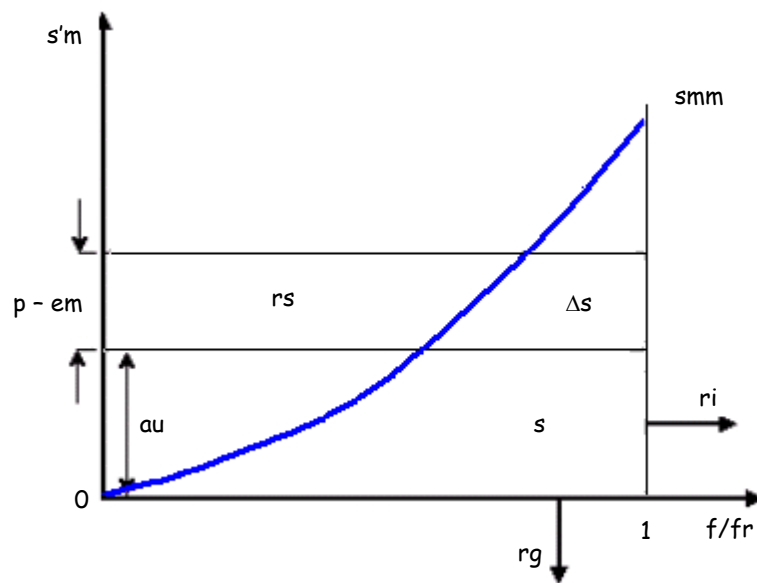
$$r = (p - em) + w - w_m \quad (4.4)$$

Quando a evapotranspiração supera a precipitação, o teor médio espacial de água do solo é reduzido e o ponto  $i$  desloca-se para baixo, consoante à curva da Figura 4.4. Esse deslocamento, representa a diminuição do teor de umidade do solo ( $w$ ), que corresponde à área à direita da curva e abaixo do ponto  $i$ . É importante ressaltar que essa movimentação ao longo da curva implica em uma redistribuição da água dentro da bacia.

### Separação do escoamento

O escoamento gerado em períodos chuvosos, nas áreas permeáveis, conforme mostrado na Figura 4.4, é posteriormente separado em três parcelas: escoamento superficial ( $r_s$ ), escoamento subterrâneo ( $r_g$ ) e escoamento sub-superficial ( $r_i$ ). Essa separação do escoamento se dá a partir do conceito de armazenamento da água que não está retida no meio poroso por ação da capilaridade, às vezes denominada água gravitacional ou, ainda, água livre.

A água livre é aquela disponível para o escoamento após ter sido atingida a capacidade de campo. No módulo de produção, admite-se que a capacidade de água livre é distribuída de maneira não uniforme na extensão da fração de área permeável da bacia ( $f_r$ ). Admite-se que a capacidade de armazenamento de água livre ( $s'm$ ) pode ser representada por uma curva de distribuição com valores variando entre zero e um valor máximo ( $s_{mm}$ ). A forma geral dessa curva está ilustrada esquematicamente na Figura 4.5.



**Figura 4.5** – Distribuição espacial da capacidade de água livre.

A referida curva pode ser expressa analiticamente como:

$$\left(1 - \frac{f}{f_r}\right) = \left(1 - \frac{s'm}{s_{mm}}\right)^{ex} \quad (4.5)$$



na qual  $f/r$  é a fração da área permeável produtora do escoamento, cuja capacidade de armazenamento de água livre é menor ou igual à  $s'm$ , e  $ex$  é um parâmetro que indica o grau de não uniformidade espacial.

Tomando-se o estado inicial do armazenamento da água livre como o ponto da curva correspondente à ordenada  $au$  (Figura 4.5), tem-se que a porção da área à esquerda já atingiu a capacidade de armazenamento de água livre, e é responsável por gerar parte do escoamento superficial (à exceção do escoamento gerado na área impermeável), enquanto que a porção à direita do referido ponto representa a área cuja capacidade de armazenamento de água livre ainda não foi alcançada, e que contribui para formação dos escoamentos sub-superficial e subterrâneo.

A capacidade média espacial de armazenamento de água livre ( $sm$ ) relaciona-se à máxima pontual ( $smm$ ) a partir da seguinte relação:

$$sm = \frac{smm}{(1+ex)} \quad (4.6)$$

Pode-se determinar, num instante qualquer, o armazenamento da água livre ( $sm$ ) sobre a área produtora do escoamento ( $fr$ ) integrando-se a equação 4.5 ( $s'm$ ) e substituindo-se  $sm$  por  $smm$ , (equação 4.6). Em termos analíticos, tem-se:

$$1 - \frac{s}{sm} = 1 - \frac{au}{smm}^{(1+ex)} \quad (4.7)$$

O escoamento  $r$  é gerado pelo volume  $(p - em)$  na sub-bacia, de acordo com a Figura 4.5. Em seguida, é aplicado sobre a curva de distribuição da capacidade de água livre somando-se à ordenada inicial  $au$  e produzindo o escoamento superficial ( $rs$ ) da área permeável.

Caso  $au + p - em < smm$ :

$$rs = \left\{ (p - em) - sm + s + sm \cdot \left[ 1 - \frac{au + (p - em)}{smm} \right]^{1+ex} \right\} \cdot fr \quad (4.8)$$

Em caso contrário,

$$rs = [(p - em) + s - sm] \cdot fr \quad (4.9)$$

Uma vez retirada a parcela que forma o escoamento superficial ( $rs$ ) do escoamento total ( $r$ ), o restante do escoamento ( $\Delta s$ ) é acrescido ao armazenamento da água livre ( $s$ ), de onde provém o escoamento sub-superficial ( $ri$ ) e o escoamento subterrâneo ( $rg$ ), calculados pelas seguintes equações:

$$ri = s.kss.fr \quad (4.10)$$

$$rg = s.kg.fr \quad (4.11)$$

nas quais  $kss$  e  $kg$  são, respectivamente, os coeficientes do fluxo dos escoamentos sub-superficial e subterrâneo, segundo o intervalo de tempo de simulação.

#### Afluxos à rede de drenagem

Esta fase representa o movimento da água proveniente das colunas de solo em direção ao sistema de drenagem. Os escoamentos superficial e das áreas impermeáveis fluem diretamente até os elementos do sistema de drenagem e de forma muito mais rápida do que o fazem os escoamentos sub-superficial e subterrâneo. A inclusão desse fato no modelo é realizada por meio da propagação dos fluxos sub-superficial e subterrâneo através de um reservatório linear.

Os afluxos superficiais à rede de drenagem  $qs$  são calculados pela soma dos escoamentos superficiais gerados pelas áreas impermeáveis ( $rb$ ) e permeáveis ( $rs$ ):

$$qs = rb + rs \quad (4.12)$$

na qual:

$$rb = imp.(p - em). \quad (4.13)$$

Os afluxos sub-superficiais à rede de drenagem, aqui denotados por  $qi$ , são calculados pela expressão:

$$qi_t = qi_{t-1}.ci + ri_t.(1 - ci) \quad (4.14)$$

na qual  $ci$  representa o coeficiente de recessão do fluxo sub-superficial diário.

Similarmente, os afluxos subterrâneos à rede de drenagem ( $qg$ ) são determinados pela equação:

$$qg_t = qg_{t-1} \cdot cg + rg_t \cdot (1 - cg) \quad (4.14)$$

na qual  $cg$  representa o coeficiente de recessão do fluxo subterrâneo diário.

Os termos  $qs$ ,  $qi$  e  $qg$ , todos expressos em mm/km<sup>2</sup>, somam-se para que seja obtido o deflúvio total ( $cin$ ) afluente à rede de drenagem durante o intervalo de tempo de cálculo:

$$cin = qs + qi + qg \quad (4.15)$$

### Evapotranspiração

À medida que a coluna de solo é submetida à ação da evapotranspiração, inicia-se a depleção da umidade armazenada sob forma de água sob tensão capilar. A capacidade média espacial de armazenamento de umidade do solo, de retenção específica ( $w_m$ ) é dividida em três partes, correspondentes às zonas superior ( $w_{um}$ ), inferior ( $w_{lm}$ ) e profunda ( $w_{dm}$ ) do solo. A umidade ou quantidade de água retida nessas três zonas do solo, no instante considerado, é representada pelas variáveis de estado  $wu$ ,  $wl$  e  $wd$ .

A evapotranspiração potencial ( $em$ ) pode ser estimada diretamente a partir de dados de tanque evaporimétrico ( $ev$ ) ou indiretamente por meio de dados meteorológicos. No primeiro caso, emprega-se o coeficiente de tanque ( $k$ ,  $k \leq 1$ ), conforme apresentado na equação 4.16:

$$em = k \cdot ev . \quad (4.16)$$

Se o conteúdo de água capilar na zona superior for maior que a evapotranspiração potencial, ou seja,  $wu > em$ , admite-se que a evapotranspiração real ou efetiva na zona superior ( $eu$ ) é igual à evapotranspiração potencial, ou seja:

$$eu = em . \quad (4.17)$$

Na zona superior do solo, a evapotranspiração ocorre em nível potencial até que todo seu armazenamento ( $wu$ ) seja esgotado. Após esgotado, uma parcela da evapotranspiração potencial é ainda capaz de atuar na zona inferior. Assim, a evapotranspiração real nessa zona

é considerada como sendo igual à evapotranspiração potencial ( $em$ ) multiplicada pela razão entre o conteúdo atual de umidade da zona inferior ( $wl$ ) e sua capacidade de armazenamento ( $wlm$ ), ou seja:

$$el = em \cdot \frac{wl}{wlm} \quad (4.18)$$

Na zona profunda, a evapotranspiração é uma fração da evapotranspiração potencial ( $em$ ), que é proporcional à área da bacia coberta com vegetação de raízes profundas. Logo, se a umidade das zonas superior e inferior se esgotarem ( $wu = wl = 0$ ), então:

$$ed = c \cdot em \quad (4.19)$$

na qual  $c$  representa o coeficiente de evapotranspiração da zona aerada profunda.

Se a evapotranspiração real da zona inferior é menor do que evapotranspiração real da zona profunda, admite-se que  $el$  seja igual  $ed$  e, nessas condições:

$$el = c \cdot em \quad (4.20)$$

A evapotranspiração efetiva total, denotada por  $e$ , é calculada somando os termos correspondentes às zonas superior, inferior e profunda, isto é:

$$e = eu + el + ed \quad (4.21)$$

### **Módulo de concentração**

No modelo Rio Grande, o módulo de produção resulta em uma série de pulsos discretos de lâminas de escoamento total  $cin$ , incluindo os fluxos superficial, sub-superficial e subterrâneo, sobre uma área unitária da sub-bacia em estudo, em  $N$  intervalos de tempo definidos a partir das datas inicial e final de simulação e do intervalo de cálculo. Em cada um desses intervalos de tempo, a unidade usual de  $cin$  é  $mm/(\Delta t \cdot km^2)$ .

Embora se presuma a uniformidade dessas lâminas de escoamento sobre a sub-bacia em estudo, há que se considerar que diferentes áreas irão concentrar suas respectivas contribuições em diferentes tempos, na seção exutória pertinente. No modelo, essa consideração é posta em prática a partir de uma função de transferência baseada no

hidrograma unitário sintético de Clark, tal como formulação original do *Hydrologic Engineering Center* do *U. S. Army Corps of Engineers* (HEC, 1981).

O hidrograma unitário sintético de Clark requer os seguintes parâmetros:

- $t_c$ : tempo de concentração da sub-bacia, aqui fixado como um número inteiro de horas e usado para se ter em conta os efeitos de translação;
- $R$ : coeficiente de armazenamento da sub-bacia, também expresso em um número não necessariamente inteiro de horas e empregado para se incorporar os efeitos de armazenamento aos de translação.
- Histograma Tempo-Área ( $HTA$ ), o qual representa a distribuição temporal das áreas da sub-bacia que contribuem para a formação das vazões na seção exutória.

O tempo de concentração  $t_c$  pode ser obtido por meio de fórmulas empíricas ou por análise de hidrogramas observados. Uma vez conhecido, o coeficiente de armazenamento  $R$  pode ser estimado a partir da formulação empírica  $t_c + R = 0,625 \cdot \sqrt{\text{Área da bacia}}$ , sugerida por HEC (1981), na qual a área da bacia é expressa em  $\text{km}^2$ .

O histograma tempo-área pode ser estimado diretamente a partir de mapas topográficos e curvas hipsométricas da sub-bacia em questão. Pode-se também adotar um histograma tempo-área estimado a partir da análise dos escoamentos na bacia ou o histograma tempo-área sintético, com base na forma determinante da bacia, conforme sugerido por HEC (1981).

HEC (1981) propõe uma formulação sintética para o histograma tempo-área que é suficientemente genérica para ser incluída em um modelo de simulação. Essa formulação, descrita na sequência, pressupõe o conhecimento de  $t_c$  e da forma dominante da bacia (retangular, losangular ou intermediária).

Denotando-se por  $T$  a fração do tempo de concentração, ou seja:

$$T = \frac{t}{t_c} \quad (0 \leq T \leq 1) \quad (4.22)$$

e  $AI$  a área acumulada, fração da área total  $AT$ ,

$$AI = \frac{A}{AT} \quad (0 \leq AI \leq 1) \quad (4.23)$$

o histograma tempo-área sintético é dado pela seguinte função potencial:

$$AI = a \cdot T^n \quad \text{se } (0 \leq T \leq 0,5) \quad (4.24)$$

$$AI = 1 - a \cdot (1 - T)^n \quad \text{se } (0,5 < T \leq 1) \quad (4.25)$$

onde o expoente  $n$ , função da forma dominante da sub-bacia, assume os seguintes valores:

- $n = 1 \Rightarrow$  bacia de forma predominantemente retangular;
- $n = 2 \Rightarrow$  bacia de forma predominantemente losangular;
- $n = 1,5 \Rightarrow$  bacia de forma elipsoidal ou intermediária entre as duas primeiras.

O parâmetro  $a$  pode ser estimado tomando-se os limites estabelecidos na expressão a seguir:

$$AI = 0,5 \text{ e } T = 0,5 \Rightarrow 0,5 = a \cdot 0,5^n \Rightarrow a = 0,5^{(1-n)}, \quad (4.26)$$

assim:

- $a = 1 \Rightarrow$  bacias retangulares
- $a = 2 \Rightarrow$  bacias losangulares
- $a = 1,414 \Rightarrow$  bacias elipsoidais.

Uma vez obtido o histograma tempo-área, o mesmo deve ser transformado em ordenadas incrementais, ou seja  $[AI_i - AI_{i-1}]$  para  $\frac{t_i}{tc}$ , e posteriormente em unidades de  $\frac{m^3}{s.mm.km^2}$ . Isso corresponderá ao HU de translação, cujas ordenadas são denotadas por  $y_i$  no tempo  $t_i$ .

O método de Clark pressupõe ainda uma propagação do HU de translação por um reservatório linear de constante  $R$  (expressa em horas). A propagação é feita da seguinte forma:

$$q_i = ca \cdot y_i + cb \cdot q_{i-1} \quad (4.27)$$

na qual:

$$ca = \frac{\Delta t}{R+0,5\Delta t} \quad (4.28)$$

e

$$cb = 1 - ca \quad (4.29)$$

onde:

$\Delta t$  representa o intervalo de tempo (horas) dos pulsos de chuva efetiva ou lâminas de escoamento total em mm/km<sup>2</sup>.

Finalmente, as ordenadas do HU de Clark, denotadas por  $x_i$  no tempo  $i$ , são calculadas por:

$$x_i = \frac{q_{i-1} + q_i}{2} \quad (4.30)$$

lembrando-se que  $x_i = 0$  (para  $t = 0$ ) e que, por construção, o número total de ordenadas será dado quando o volume total do HU superar 0,995.

### **Síntese dos parâmetros e variáveis de estado**

A função de produção do modelo Rio Grande é composta por 13 parâmetros e 7 variáveis de estado. O módulo de concentração incorpora dois outros parâmetros à estrutura conceitual do modelo. As Tabelas 4.2 e 4.3 sumarizam as informações acerca desses parâmetros e variáveis de estado do modelo.

**Tabela 4.2 – Parâmetros do modelo Rio Grande.**

<b>Módulo</b>	<b>Símbolo</b>	<b>Descrição</b>
Produção	<i>k</i>	Coeficiente de tanque
	<i>imp</i>	Fração de área impermeável
	<i>wum</i>	Capacidade de retenção na zona superior do solo
	<i>wlm</i>	Capacidade de retenção na zona inferior do solo
	<i>wdm</i>	Capacidade de retenção na zona profunda do solo
	<i>sm</i>	Capacidade de água livre ou gravitacional
	<i>b</i>	Expoente da curva de distribuição em área sob tensão capilar
	<i>ex</i>	Expoente da curva de distribuição espacial de capacidade de água livre
	<i>c</i>	Coeficiente de evapotranspiração profunda
	<i>kss</i>	Coeficiente do fluxo de escoamento subsuperficial
	<i>kg</i>	Coeficiente do fluxo de escoamento subterrâneo
	<i>ci</i>	Coeficiente de recessão do fluxo subsuperficial
Concentração	<i>cg</i>	Coeficiente de recessão do fluxo subterrâneo
	<i>tc</i>	Tempo de concentração
	<i>n</i>	Forma predominante da bacia

**Tabela 4.3 – Variáveis de estado do modelo Rio Grande.**

<b>Módulo</b>	<b>Símbolo</b>	<b>Descrição</b>
Produção	<i>wu</i>	Quantidade de água armazenada na zona superior do solo
	<i>wl</i>	Quantidade de água armazenada na zona inferior do solo
	<i>wd</i>	Quantidade de água armazenada na zona profunda do solo
	<i>s</i>	Quantidade de água contida no solo disponível para o escoamento
	<i>qi</i>	Afluxo subsuperficial à rede de drenagem
	<i>qg</i>	Afluxo subterrâneo à rede de drenagem
	<i>fr</i>	Fração de área permeável geradora do escoamento

A calibração e incertezas associadas às estimativas foram determinadas para 7 dos 13 parâmetros relativos ao módulo de produção (*k*, *wum*, *wlm*, *sm*, *kss*, *kg*, *ci*). Os parâmetros do módulo de concentração foram estimados a partir de uma formulação empírica (*tc*) e da inspeção visual da forma da área de drenagem, delimitada a partir de um modelo digital de elevação do terreno.



### 4.3 Inferência Bayesiana por meio de simulações de Monte Carlo

O algoritmo selecionado para estimar a função densidade de probabilidade a *posteriori* dos parâmetros é denominado DREAM (*Differential Evolution Adaptive Metropolis*). Esse método de simulação de Monte Carlo via cadeias de Markov (MCMC) admite distribuições a *priori* para os parâmetros  $\theta$  e simula diferentes cadeias simultaneamente, ajustando a escala e a forma dos modelos distributivos propostos durante a evolução das distribuições a *posteriori* (VRUGT *et al.*, 2008a).

Os fundamentos teóricos e a conceituação do método são detalhados a seguir. Na sequência, apresentam-se os modelos probabilísticos admitidos para os resíduos.

#### 4.3.1 O método DREAM

A base de qualquer método de simulação de Monte Carlo via cadeias de Markov (MCMC) é produzir uma busca aleatória no espaço paramétrico multidimensional e sucessivamente buscar soluções que denotem uma frequência estável e estacionária da distribuição de probabilidade ( $\pi$ ), associada a cada uma das variáveis aleatórias em análise. Para tal, o algoritmo produz movimentos aleatórios, a partir do estado atual (anterior) da cadeia,  $x_{t-1}$ , a um novo estado  $z$  (VRUGT *et al.*, 2008a).

O método de MCMC mais antigo e, provavelmente o mais conhecido, denomina-se *Random Walk Metropolis* (RWA). Ele pode ser sumariado em três etapas (VRUGT *et al.*, 2009): admitindo-se que a busca aleatória tenha produzido uma amostra de pontos  $\{x_0, \dots, x_{t-1}\}$ , (i) um ponto candidato  $z$ , o qual depende da posição atual  $x_{t-1}$ , é amostrado de uma distribuição proposta  $q(\cdot)$  simétrica,  $q(x_{t-1}, z) = q(z, x_{t-1})$ ; (ii) na sequência, o ponto candidato é aceito ou rejeitado a partir da probabilidade de aceitação de *Metropolis*, formalizada por:

$$\alpha(x_{t-1}, z) = \begin{cases} \min\left(\frac{\pi(z)}{\pi(x_{t-1})}, 1\right) & \text{se } \pi(x_{t-1}) > 0 \\ 1 & \text{se } \pi(x_{t-1}) = 0 \end{cases} \quad (4.31)$$

na qual  $\pi(\cdot)$  denota a densidade da distribuição alvo; e finalmente, (iii) caso o ponto candidato tenha sido aceito, a cadeia move-se para  $z$ , caso contrário, ela permanece em  $x_{t-1}$ .

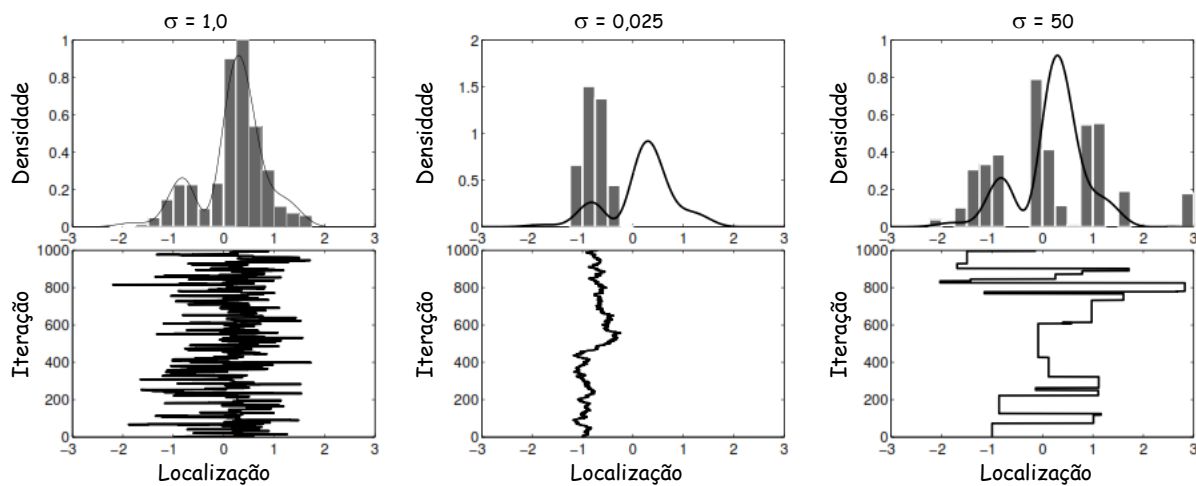
O algoritmo *RWA* original foi construído sob a perspectiva de manter o balanço detalhado em relação a  $\pi(\cdot)$  a cada etapa na cadeia, ou seja:

$$\pi(x_{i-1}).p(x_{i-1} \rightarrow z) = \pi(z).p(z \rightarrow x_{i-1}) \quad (4.32)$$

na qual,  $\pi(x_{i-1})$  representa a probabilidade do sistema encontrar-se no estado  $x_{i-1}$  e  $p(x_{i-1} \rightarrow z)$  a probabilidade do sistema mover-se deste estado para  $z$ . Os termos  $\pi(z)$  e  $p(z \rightarrow x_{i-1})$  são definidos analogamente. A condição de balanço detalhado está associada à garantia da existência de uma distribuição de equilíbrio. Em síntese, caso o equilíbrio tenha sido alcançado, a probabilidade do sistema ir do estado  $x_{i-1}$  para  $z$  deve ser igual à do estado  $z$  mover-se para  $x_{i-1}$  (TIERNEY, 1994; CHIB e GREENBERG, 1995).

Hastings (1970) estendeu a equação 4.31 para inclusão de distribuições não simétricas [  $q(x_{t-1}, z) \neq q(z, x_{t-1})$  ], para as quais um movimento ao estado  $z$  ( $x_{i-1} \rightarrow z$ ) não possui a mesma probabilidade no caso inverso ( $z \rightarrow x_{i-1}$ ). Essa extensão originou o algoritmo de *Metropolis-Hastings* (MH), o qual tornou-se o cerne de diversos esquemas de amostragem via cadeia de Markov hoje existentes.

Diversos experimentos demonstraram a convergência a uma distribuição de equilíbrio de esquemas MCMC construídos a partir do algoritmo MH, embora a taxas extremamente lentas. Segundo Haario *et al.* (2001, 2006), a eficiência do método é afetada pela escala e orientação da distribuição proposta,  $q(x_{t-1}, .)$ , utilizada na geração de movimentos aleatórios (transições) na cadeia de Markov. Quando a função proposta é muito ampla, diversos pontos candidatos são rejeitados e a convergência para a distribuição alvo é muito lenta. Em contrapartida, quando a distribuição proposta é muito restrita, todos os pontos candidatos são aceitos, mas a distância do movimento é tão reduzida que é necessário um número muito elevado de atualizações para que o amostrador convirja para a distribuição alvo. A Figura 4.6 ilustra essa observação.



**Figura 4.6** – Influência da distribuição proposta sobre as amostras obtidas no algoritmo de *Metropolis*: painel a esquerda:  $N(\bar{x}, \sigma = 1)$ ; (ii) painel central: distribuição muito restrita –  $N(\bar{x}, \sigma = 0,025)$ ; e (iii) painel a direita: distribuição muito ampla –  $N(\bar{x}, \sigma = 50)$  (adaptado de: NAVARRO e PERFORS, data desconhecida).

O ajuste automático da função proposta pode contornar muito dos problemas associados à escolha de uma distribuição candidata adequada (HAARIO *et al.*, 2001; VRUGT *et al.*, 2009). Nessa abordagem, a informação relativa ao histórico da amostragem é utilizada continuamente para se adaptar a forma e o comprimento da distribuição proposta, a fim de aumentar a eficiência na convergência a uma distribuição de equilíbrio (HAARIO *et al.*, 2006). As técnicas pertinentes podem ser agrupadas em duas categorias, definidas segundo o número de cadeia analisadas simultaneamente (LALOY e VRUGT, 2012):

- Única: na qual a covariância  $C_t$  da função Gaussiana proposta  $[q(x_{i-1}, \cdot) = N_d(x_{i-1}, C_t)]$  é adaptada a partir da trajetória admitida pelas amostras na cadeia. Algebricamente,  $C_t = s_d \cdot \text{Cov}(x_0, \dots, x_{i-1}) + s_d \cdot \varepsilon \cdot I_d$ , onde  $s_d$  representa o fator de escala e depende somente da dimensão  $d$  do problema (número de variáveis aleatórias em análise),  $I_d$  é a matriz identidade de dimensão  $d$  e  $\varepsilon$  é uma constante multiplicativa, cujo propósito é aumentar a covariância real, de modo a garantir que todo o espaço paramétrico seja amostrado. Incluem-se nesta categoria os métodos *Adaptive Metropolis* (AM) (HAARIO *et al.*, 2001) e *Delayed Rejection Adaptive Metropolis* (DRAM) (HAARIO *et al.*, 2006); e
- Múltiplas: as quais utilizam diferentes trajetórias simuladas paralelamente para exploração da distribuição alvo *a posteriori*. Vrugt *et al.* (2009) argumentam que, no

caso de distribuições a *posteriori* complexas, parâmetros correlacionados, multimodalidade e ocorrência de diversos ótimos locais, esta abordagem revela-se muito mais robusta do que a anterior. Ademais, argumentam que a utilização de múltiplas cadeias previne a convergência prematura e permite a aplicação de uma variedade de métricas estatísticas para avaliação da convergência à distribuição de equilíbrio. Dentre as técnicas que integram esta categoria destacam-se: *Shuffled Complex Evolution Metropolis Algorithm* (SCEM-UA) (DUAN *et al.*, 1992; SOROOSHIAN *et al.*, 1993), cujo número de aplicações à ciência hidrológica é significativo, e o *Differential Evolution Markov Chain* (DE-MC) (ter BRAAK, 2006). Este último é a base do método empregado nesta pesquisa.

No método DE-MC (ter BRAAK, 2006),  $N$  diferentes cadeias  $\{x^i, i = 1, \dots, N\}$  são simuladas em paralelo. No estado atual, a população  $X$  é armazenada em uma matriz  $N \times d$ , na qual  $d$  refere-se à dimensão de  $\pi$ . Movimentos em cada cadeia  $\{i=1, \dots, N\}$  são gerados tomando-se um número fixo de diferenças entre dois membros (cadeias) aleatoriamente escolhidas de  $X$  (sem reposição) com índices  $r_1$  e  $r_2$ . Formalmente, tem-se a seguinte expressão:

$$z^i = (x_{t-1}^i) + \gamma \cdot (X_{t-1}^{r_1} - X_{t-1}^{r_2}) + \epsilon \quad (r_1 \neq r_2 \neq i), \quad (4.33)$$

na qual  $\gamma$  é um escalar cujo valor é definido a partir das orientações do método RWM para estimativa do parâmetro  $s_d \left( \frac{2,38}{\sqrt{2d}} \right)$ , e  $\epsilon$  é obtida a partir de uma distribuição  $d$ -dimensional simétrica com variância pequena (GELMAN *et al.*, 1996; LALOY e VRUGT, 2012). A probabilidade de aceitação de de *Metropolis* é utilizada para determinar se os pontos candidatos são aceitos ou rejeitados.

Segundo Vrugt *et al.* (2009), esse método soluciona dois importantes problemas relacionados à amostragem. O primeiro está associado à seleção automática da escala e da orientação da distribuição proposta. O segundo associa-se às distribuições multimodais e de caudas pesadas, as quais são eficientemente exploradas, uma vez que o algoritmo utiliza a posição atual na cadeia (armazenada em  $X$ ), em contraposição à covariância adaptativa,  $Cov(X)$ , para gerar pontos candidatos, permitindo movimentos diretos entre uma e outra moda de  $X$ .

O algoritmo MCMC utilizado na presente pesquisa denomina-se *Differential Evolution Adaptive Metropolis* (DREAM) e baseia-se no método DE-MC. Seguindo Vrugt *et al.* (2009),

ele pode ser assim sumariado, observando-se que: (i) os subscritos relativos à geração no tempo  $t$  foram suprimidos; e, assim, (ii) o estado atual da  $i$ -ésima cadeia,  $i^{th}$  é dada pelo vetor  $x_i$   $\{i = 1, \dots, N\}$  de dimensão  $d$  em seu  $j$ -ésimo elemento,  $x_j^i$ :

1. Amostra-se uma distribuição inicial da população  $\{x_i, i = 1, \dots, N\}$  a partir de uma distribuição *a priori*;
2. Calcula-se a densidade  $\pi(x^i)$  para cada elemento  $i = 1, \dots, N$ ;
3. Para  $i = 1, \dots, N$  (evolução da cadeia):
  - Gera-se o vetor candidato,  $z^i$  na cadeia  $i$ :

$$z^i = x^i + (1_d + e) \cdot \gamma(\delta, d') \cdot \left[ \sum_{j=1}^{\delta} x^{r_1(j)} - \sum_{n=1}^{\delta} x^{r_2(n)} \right] + \epsilon \quad (4.34)$$

onde  $r_1(j), r_2(n) \in \{1, \dots, N\}$ ;  $r_1(j) \neq r_2(n) \neq i$  para  $j, n = 1, \dots, \delta$ ;  $z^i$  representa o vetor candidato de dimensão  $d$ ;  $x^i$  é o estado atual da  $i$ -ésima cadeia que compõe a população;  $1_d$  é um vetor unitário com dimensão  $d$ ; e  $\gamma(\delta, d')$  é um fator de ponderação que é função do número de vetores da população atual utilizados para ajustar a função proposta. Os valores de  $e$  e  $\epsilon$  são obtidos amostrando-se uma distribuição Uniforme  $U_d(-b, b)$ , onde  $|b| < 1$ , e Normal  $N_d(0, b^*)$ , respectivamente. Nessas expressões,  $b$  e  $b^*$  são pequenos quando comparados à escala da distribuição alvo;

- Substitui-se cada elemento  $\{j = 1, \dots, d\}$  do ponto candidato  $z_j^i$ , utilizando-se um esquema de cruzamento binomial com probabilidade  $1-CR$ , onde  $CR$  representa a probabilidade de cruzamento e é determinada a partir da expressão:

$$z_j^i = \begin{cases} x_j^i & , \text{ se: } U \leq 1 - CR, d' = d' - 1 \\ z_j^i & \text{ caso contrário} \end{cases} \quad (4.35)$$

Em problemas com múltiplas dimensões, é frequente que a amostra não seja composta por valores ótimos simultâneos em todos os  $d$  elementos  $\{z_j^i, j = 1, \dots, d\}$  de  $z^i$ . Essa estratégia de amostragem aleatória no

subespaço paramétrico modifica cada dimensão por meio da probabilidade  $CR$ , a cada instante de tempo que um ponto candidato é gerado;

- Calcula-se  $\pi(z^i)$  e  $\alpha(x^i, z^i)$ ;
  - A partir da probabilidade de aceitação de *Metropolis*, determina-se a rejeição do vetor candidato ( $x^i$  mantém-se inalterado) ou aceitação ( $x^i = z^i$ );
4. Caso o algoritmo esteja no período de aquecimento (*burn-in*), removem-se as cadeias potencialmente atípicas (*outliers*), utilizando-se a estatística denominada IQR (*inter-quartile range*);
  5. Calcula-se a estatística de Gelman e Rubin ( $\hat{R}_j$ ) para diagnóstico da convergência. Caso esse valor seja inferior a um limite estabelecido [Vrugt *et al.*, (2009) sugerem 1,2] para todas as dimensões, finaliza-se o algoritmo, caso contrário, retorna-se à etapa 3.

Vrugt *et al.* (2008a, 2009) destacam os três pontos que diferenciam os algoritmos DREAM e DE-MC:

- A atualização dos pontos candidatos por meio de cruzamento binomial, a qual aumenta a eficiência do algoritmo quando o problema envolve várias dimensões. Ademais, durante o *burn-in*, na tentativa de acelerar a convergência das cadeias à distribuição de equilíbrio, o algoritmo tende a privilegiar movimentos com maior comprimento em cada cadeia;
- No algoritmo DE-MC, os vetores candidatos são funções de apenas dois membros de  $N$  cadeias. Já no algoritmo DREAM esses vetores candidatos incluem um número elevado de combinações na tentativa de aumentar a diversidade. Não obstante esse fato, o fator de ponderação  $\gamma(\delta, d')$ , tal como implementado, facilita o movimento entre diferentes modas da distribuição *a posteriori*;
- No algoritmo DREAM, as cadeias *outliers* são removidas durante o *burn-in* para facilitar a convergência à distribuição alvo.

O algoritmo utilizado para calibração e estimativa das incertezas dos parâmetros e variáveis de estado do modelo Rio Grande foi implementado como uma biblioteca para o *software* R (R CORE TEAM, 2013) por Guillaume e Andrews (2012). Em relação à versão original, a única limitação está associada à especificação da distribuição *a priori*, a qual é admitida como sendo uniforme.

A Tabela 4.4 faz um sumário das principais informações necessárias à utilização do algoritmo.

**Tabela 4.4 – Informações necessárias à utilização do método DREAM.**

Símbolo	Descrição
<i>func.type</i>	Especificação da função de verossimilhança a ser considerada
<i>pars</i>	Lista com o nomes e respectivos intervalos de incerteza das variáveis aleatórias
<i>nseq</i>	Número de cadeias avaliadas em paralelo
<i>ndraw</i>	Número máximo de iterações
<i>burn-in</i>	Número de iterações desconsideradas após o início da simulação
<i>thin.t</i>	Intervalos entre avaliações da função de verossimilhança utilizados para composição da amostra
<i>Rthres</i>	Estatística de Gelman e Rubin ( $\hat{R}_j$ ) para diagnóstico da convergência

É válido ressaltar que há na literatura uma série de referências à utilização do algoritmo para fins similares ao do presente pesquisa (HE *et al.*, 2010; VRUGT e ter BRAAK, 2011; MINASNY, *et al.*, 2011; LALOY e VRUGT, 2012).

#### 4.3.2 Modelos probabilísticos para os resíduos

A abordagem Bayesiana para inferência de parâmetros em modelos chuva-vazão pressupõe o conhecimento *a priori* da estrutura de erros do modelo. A avaliação do espaço paramétrico via função de verossimilhança permite, além da identificação de um intervalo plausível para um conjunto de parâmetros, a predição das incertezas associadas às estimativas desses parâmetros.

Ao longo dos últimos anos, ocorreu um intenso debate na comunidade técnica acerca da especificação da função de verossimilhança para fins semelhantes ao propósito da presente pesquisa (MANTOVAN e TODINI, 2006 *apud* SCHOUPS e VRUGT, 2010; BEVEN *et al.*, 2008, VRUGT *et al.*, 2008b; STEDINGER *et al.*, 2008; McMILLAN e CLARK, 2009). As

abordagens sugeridas, identificadas como formal e informal, distinguíam-se quanto ao rigor estatístico admitido em sua concepção.

Na abordagem formal é necessário admitir um modelo probabilístico para os erros. Isso é feito de forma explícita, *a priori* [ex.:  $r \sim N(0, \sigma)$ ]. Por conseguinte, é possível validar *a posteriori*, a hipótese assumida. Na abordagem informal, não há referência direta a um modelo probabilístico para os resíduos (ex.: *Nash-Sutcliffe*). Assim sendo, essa validação não pode ser procedida *a posteriori*.

Tendo clara a percepção que ambas as abordagens são empregadas para fins de inferência Bayesiana, pode-se sumariar essa discussão a partir da seguinte pergunta: qual o limite admissível para o relaxamento das hipóteses em um método probabilístico, visto que o objetivo da análise é beneficiar-se dos resultados estatísticos derivados desse método de inferência? Diversos argumentos foram utilizados, dentre os quais destacam-se a complexidade do fenômeno de transformação da chuva e vazão (e, por conseguinte, do comportamento exibido pelos resíduos: heterocedasticidade, autocorrelação etc.), a não aleatoriedade dos erros produzidos por esses modelos (sobretudo devido às diversas fontes de incerteza associadas) e a possibilidade de se combinar medidas de verossimilhança (sob a perspectiva de melhor caracterizar as diferenças impostas pela sazonalidade e não linearidade da relação entre chuvas e vazões).

Na impossibilidade de uma resposta inequívoca, diversos hidrólogos dedicaram-se ao desenvolvimento de funções de verossimilhança que se revelassem capazes de lidar com a complexa estrutura exibida pelos erros dos modelos hidrológicos. Dentre estas, a presente pesquisa optou pela avaliação de dois modelos formais distintos:

- Normal (NL): o qual admite que os resíduos da modelagem são independentes e identicamente distribuídos segundo a distribuição Normal, com média zero e desvio padrão constante, e cuja função log de verossimilhança é representada pela expressão:

$$L(\theta|y) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\sum_{t=1}^N(\tilde{y}_t - y_t)^2, \quad (4.36)$$

onde  $\tilde{y}$  é a vazão observada,  $y$  é a vazão modelada e  $\sigma$  é o desvio padrão dos erros na vazão observada. O valor de  $\sigma^2$  pode ser especificado *a priori*, com base no



conhecimento acerca dos erros nas observações da vazão, ou inferido simultaneamente aos parâmetros  $\theta$  do modelo (VRUGT *et al.*, 2013).

Esse modelo é comumente admitido para explicar o comportamento dos resíduos na modelagem hidrológica e conduz o problema da estimação de parâmetros à abordagem sugerida pelo método dos mínimos quadrados ordinário ou padrão (*SLS – standard least squares*).

- Generalizado (GL) (SCHOUPS e VRUGT, 2010): na qual se pressupõe que os resíduos são heterocedásticos, autocorrelacionados e modelados probabilisticamente segundo a função densidade de probabilidade exponencial potência assimétrica.

Os fundamentos utilizados para construção do modelo generalizado são apresentados na sequência.

### **Função de verossimilhança generalizada**

Segundo Schoups e Vrugt (2010), o desenvolvimento baseia-se em um modelo de regressão não linear aditivo de forma:

$$\tilde{y} = E + e, \quad (4.37)$$

na qual  $\tilde{y}$  é um vetor com  $N \{t = 1, \dots, N\}$  observações da vazão;  $E$  é um vetor com as vazões médias correspondentes, e  $e$  é um vetor de resíduos aleatórios com média zero.

As vazões médias correspondentes às observadas ao longo do tempo,  $E_t$ , são obtidas a partir das vazões médias modeladas,  $y$ , por meio do modelo hidrológico  $M$  por meio da seguinte expressão:

$$E_t = y_t(u|\theta) \cdot \mu_t, \quad (4.38)$$

onde  $y_t$  é função dos *inputs*  $u$  e dos parâmetros  $\theta$ , e  $\mu_t$  é um fator multiplicativo que representa o viés introduzido à vazão modelada devido aos erros nos dados observados ( $u$ ,  $\tilde{y}$ ) e na estrutura conceitual admitida. Ele varia com o tempo e é função da vazão modelada. A parametrização admitida foi:

$$\mu_t = \exp(\mu_M \cdot y_t), \quad (4.39)$$

na qual  $\mu_M$  é um parâmetro que representa um viés estimado a partir dos dados. Esse parâmetro permite amplificar a não-linearidade da resposta hidrológica modelada ao longo do tempo, o que pode ser útil no caso da ocorrência de resíduos com magnitude elevada (SCHOUPS e VRUGT, 2010). É válido notar que, no caso de  $\mu_t$  ser igual a unidade ( $\mu_M = 0$ ), nenhum viés é introduzido à vazão modelada.

Os resíduos na equação 4.37 (e) são caracterizados por uma função densidade de probabilidade e um conjunto de parâmetros  $\theta_e$ . Tradicionalmente, esses erros são admitidos normais, entretanto, diversas aplicações à modelagem chuva-vazão revelaram que essa premissa é frequentemente violada. A fim de lidar com esse fato, admite-se que os resíduos comportem-se probabilisticamente segundo o modelo:

$$\Phi_p(B)e_t = \sigma_t \cdot a_t, \text{ na qual } a_t \sim SEP(0,1,\xi,\beta), \quad (4.40)$$

onde  $\Phi_p(B) = 1 - \sum_{i=1}^p \phi_i \beta^i$  é um polinômio autorregressivo com  $p$  parâmetros  $\phi_i$ ,  $B$  é o operador de defasagem ( $\beta^i e_t = e_{t-i}$ ),  $\sigma_t$  é o desvio padrão no tempo  $t$ , e  $a_t$  denota erros aleatório *iid* (independentes e igualmente distribuídos), com média nula e desvio padrão unitário, descritos por uma função densidade de probabilidade exponencial potência assimétrica (*SEP – Skew Exponential Power*), com parâmetros  $\xi$  (assimetria) e  $\beta$  (curtose) para acomodar a não-normalidade:

$$p(a_t|\xi,\beta) = \frac{2\sigma_\xi}{\xi+\xi^{-1}} w_\beta \cdot \exp\left\{-c_\beta |a_{\xi,t}|^{2/(1+\beta)}\right\}, \quad (4.41)$$

onde  $a_{\xi,t} = \xi^{-\text{sign}(\mu_\xi + \sigma_\xi \cdot a_t)} \cdot (\mu_\xi + \sigma_\xi \cdot a_t)$ , e  $\mu_\xi$ ,  $\sigma_\xi$ ,  $c_\beta$ ,  $w_\beta$  são valores calculados a partir da assimetria,  $\xi$ , e da curtose,  $\beta$ . A densidade será simétrica caso  $\xi = 1$ , e assimétrica positivamente (negativamente) caso  $\xi > 1$  ( $\xi < 1$ ). Ademais, no caso em que a densidade é simétrica, o modelo probabilístico assume a forma da distribuição uniforme quando  $\beta = -1$ , Gaussiana para  $\beta = 0$  e de Laplace se  $\beta = 1$ .

O modelo autoregressivo,  $AR(p)$ , é admitido na equação 4.40 sob a perspectiva de descrever autocorrelação, em geral, verificada na série de resíduos de modelos chuva-vazão.

Similarmente, a heterocedasticidade é explicitamente considerada, admitindo-se que o desvio padrão varie linearmente com a vazão média  $E_t$ , isto é:

$$\sigma_t = \sigma_0 + \sigma_1 \cdot E_t \quad (4.42)$$

na qual  $\sigma_0$  e  $\sigma_1$  são estimados a partir dos dados. Essa relação entre o aumento do desvio padrão dos resíduos com as vazões modeladas expressa as incertezas associadas aos ramos superiores das curvas-chave, os quais, em geral, possuem um número inferior de medições de descarga líquida, relativamente aos ramos inferiores.

Schoups e Vrugt (2010) apresentam a seguinte função log de verossimilhança derivada desse modelo de resíduos:

$$L(\theta|y) = n \cdot \log \frac{2\sigma_\xi w_\beta}{\xi + \xi^{-1}} - \sum_{t=1}^n \log \sigma_t - c_\beta \cdot \sum_{t=1}^n |a_{\xi,t}|^{2/(1+\beta)}, \quad (4.43)$$

Um algoritmo em R foi implementado a partir da versão original em MATLAB disponibilizada pelos autores (SCHOUPS e VRUGT, 2010). Essa função foi utilizada como argumento na simulação do método DREAM. Nela são admitidos modelos autorregressivos de, no máximo, ordem 4. Ademais, é permitido selecionar quais características das séries de resíduos são explicitamente consideradas (autocorrelação, heterocedasticidade e/ou viés na vazão modelada).

É válido destacar que ambas as funções de verossimilhança baseiam-se em parâmetros que devem ser estimados simultaneamente aos parâmetros do modelo Rio Grande. Para tal, a exemplo dos parâmetros do modelo hidrológico, o algoritmo admite uma distribuição uniforme *a priori* para sua estimação. A Tabela 4.5 lista e nomeia esses parâmetros.

**Tabela 4.5 – Parâmetros das funções de verossimilhança que devem ser estimados.**

Modelo	Símbolo	Descrição
NL	$\sigma$	Desvio padrão
	$\sigma_0$	Heterocedasticidade: intercepto
	$\sigma_1$	Heterocedasticidade: inclinação
GL	$\beta$	Curtose
	$\xi$	Assimetria
	$\phi_i$	Coefficientes de autocorrelação
	$\mu_M$	Viés na vazão modelada

Schoups e Vrugt (2010) destacam um aspecto relevante associado à função verossimilhança generalizada: ela é capaz de sumariar diversas formas de funções de verossimilhança desenvolvidas previamente para fins de inferência de parâmetros de modelos chuva-vazão e estimação de incertezas associadas, a partir da variação dos parâmetros admitidos em sua conceituação. Por exemplo, caso assumam-se  $\sigma_1 = 0$ ,  $\phi_i = 0$ ,  $\beta = 0$ ,  $\xi = 1$  e  $\mu_M = 0$ , o modelo GL se equivalerá ao modelo NL.

#### **4.4 Identificação de períodos de dissimilaridade entre vazões modeladas e observadas**

Ao analisar-se a similaridade entre as vazões observadas e as modeladas a partir da aplicação de um modelo chuva-vazão, é relativamente frequente observarem-se períodos nos quais ocorrem diferenças expressivas entre elas. Acerca desse fato, é válido destacar algumas justificativas, usualmente observadas na literatura e no exercício da engenharia. São elas:

- Erros durante a observação da variável hidrológica decorrentes de imperfeições no processo de leitura/ aquisição da informação, seja por imperícia do observador, seja por falha no equipamento e/ou manutenção da estação (ex.: acúmulo de sedimentos sobre o 1º lance de régua, elevada/ baixa taxa de sedimentos em suspensão, descargas atmosféricas próximas aos equipamentos etc.);
- Baixa representatividade espaço-temporal dos *inputs*  $u$ , mais comumente da precipitação, devido ao exíguo número de equipamentos de monitoramento, a frequência de ocorrência de eventos convectivos, entre outros fatores;

- Escassez de medições nos ramos superiores das curvas-chave empregadas para obtenção das vazões (*output*  $\tilde{y}$ ) na seção de interesse; e ainda, menos frequentemente
- Incorreção ou insuficiência da representação do sistema hidrológico imposta pela estrutura conceitual do modelo e/ou dificuldades na identificação de seus parâmetros.

Esse quadro de incertezas é inerente à maioria das aplicações de modelos chuva-vazão, em especial, no território nacional.

Tendo clara a percepção da necessidade do emprego desses modelos e a impossibilidade do controle absoluto de todas as fontes de incerteza, há muito os hidrólogos têm se dedicado ao desenvolvimento de técnicas que permitam “corrigir” essa dissimilaridade persistente em intraperíodos das séries. Elas são denominadas técnicas de assimilação de dados.

Em síntese, essas técnicas baseiam-se na ideia de acrescentar informações à análise, sob a perspectiva de elevar o nível de conhecimento acerca de quantidades características do sistema hidrológico. A magnitude da diferença entre as quantidades estimada e observada deve guardar, em geral, uma relação de dependência com a diferença entre as vazões observadas e simuladas (erros da simulação/previsão hidrológica).

Em outra vertente e, sobretudo na última década, a comunidade hidrológica tem se dedicado ao desenvolvimento de técnicas que permitam “estimar” a magnitude dessa dissimilaridade. Neste caso, procura-se mensurar a influência dessas fontes de incerteza sobre o erro da modelagem.

Ambas as abordagens permitem uma análise do comportamento *input-state-output* do sistema, entretanto, essa avaliação é, em geral, preterida. No caso da assimilação de dados, sobretudo, porque o objetivo primordial é corrigir a estimativa e, no caso da análise das incertezas, pois a avaliação *input-output* permite caracterizar aquela que se acredita representar a principal fonte de incerteza da modelagem: os dados observados (vejam-se as justificativas enumeradas).

Diante desse quadro, é possível elaborar o seguinte questionamento: qual é a real influência da dinâmica do modelo, isto é, do comportamento *input-state-output*, definido quando concebida sua estrutura conceitual, sobre os erros? É verossímil admitir que, ao caracterizá-la,

seria possível tanto estimar a magnitude dos erros, quanto corrigir eventuais vieses impostos pelos dados observados e/ou pela estrutura conceitual admitida.

De fato, não se observou durante a revisão de literatura uma resposta direta a esse questionamento. Em geral, a análise da estrutura conceitual costuma se dar, sobretudo, por meio da comparação: (i) de modelos e métricas que denotam o grau de similaridade entre vazões observadas e modeladas, a partir de abordagens que definem o problema sob a ótica da otimização ou da avaliação de incertezas (TIAN *et al.*, 2013; WILLEMS *et al.*, 2014); ou, (ii) de estruturas flexíveis, as quais são construídas pelo modelador através da representação de um mesmo processo físico a partir de diferentes formulações pré-idealizadas (CLARK *et al.*, 2008; FENICIA *et al.*, 2011; KAVETSKI e FENICIA, 2011; VAN ESSE, 2012; LING, 2013; EUSER *et al.*, 2013). Observou-se, ainda, com o intuito de melhor caracterizar a incerteza na estimativa da vazão, o emprego complementar de diferentes estruturas conceituais, seguido da análise de conjunto (*ensemble*) dos resultados (XIONG *et al.*, 2001; BROCHERO, 2013; LIANG *et al.*, 2013.; MADADGAR e MORADKHANI, 2014).

O objetivo aqui é avaliar o comportamento *input-state-output* à luz das incertezas inerentes à modelagem. Pretende-se verificar a possibilidade de caracterizar a incerteza na estimativa da vazão a partir do comportamento das variáveis de estado. Ao propor a correção de vieses em períodos específicos, busca-se compreender melhor a dinâmica de todo o sistema e, por conseguinte, possibilitar a melhoria da estrutura conceitual admitida. Ademais, essa correção permitirá contrapor algumas abordagens metodológicas/ conclusões observadas na literatura recente, mais especificamente as que se propõem a caracterizar as incertezas associadas aos dados observados a partir da análise *input-output* do sistema.

Para identificação dos períodos de dissimilaridade persistente, empregou-se a seguinte métrica:

$$\sqrt{\frac{\frac{1}{n} \sum_{i=ti}^{tf} (\tilde{y}_i - y_i)^2}{\frac{1}{n} \sum_{i=ti}^{tf} \tilde{y}_i}} = \left( \frac{RMSE}{vazão\ observada} \right)_{ti:tf} \quad (4.44)$$

na qual  $ti$  ( $tf$ ) é o instante inicial (final) e  $\tilde{y}$  ( $y$ ) a vazão observada (simulada).

Em função da estrutura do modelo Rio Grande, a qual emprega um hidrograma unitário (HU) para distribuição e transformação temporal dos pulsos de deflúvios em vazões na seção de

interesse, admitiu-se que  $tf$  possui aproximadamente o mesmo número de ordenadas do HU (no caso de estudo descrito no próximo capítulo; 77 horas/ordenadas  $\approx$  4 dias), sob o propósito de presumir uma janela de tempo cuja extensão configurasse uma característica intrínseca do sistema hidrológico analisado.

A partir do incremento sequencial de  $ti$  (correspondente à escala temporal admitida, ou seja, o dia), determinou-se o índice descrito pela equação 4.44. Os valores obtidos foram dispostos em ordem decrescente. Avaliou-se, então, a sequência temporal dos maiores índices (persistência), selecionando-se janelas cuja extensão mínima equivalesse a 10 vezes a escala temporal considerada (10 dias).

Essa análise considerou somente o período de calibração, uma vez que é ele que condiciona a estimação do conjunto de parâmetros  $\theta$  considerados aceitáveis para a representação do sistema hidrológico avaliado.

O desenvolvimento de uma análise baseada em uma janela móvel apoiou-se na abordagem metodológica sugerida por Choi e Beven (2007) para avaliação multicritério de um modelo chuva-vazão.

#### **4.5 Estimação das variáveis de estado e análise das alterações em suas trajetórias**

A premissa da presente abordagem metodológica é que as alterações na trajetória das variáveis de estado possam ser empregadas para corrigir eventuais vieses impostos por incertezas nos dados observados e/ou pela estrutura do modelo. Embora a referida abordagem se assemelhe à admitida pelas técnicas de assimilação de dados aplicadas à modelagem chuva-vazão, ela é conceitualmente díspar.

Nas técnicas de assimilação de dados a correção é realizada sequencialmente por meio do acréscimo da informação monitorada à análise. Em esquemas de previsão hidrológica em tempo real, essa acepção é direta. Quando o sistema presta-se à simulação, a análise é similar, todavia é comum presumir-se o conhecimento dos *inputs*  $u$  (ex.: precipitação) e/ou da vazão quando o objetivo é avaliar a “eficácia” de determinada técnica. A presente análise vale-se desse segundo cenário, contudo não se baseia na inserção de novas informações. Pressupõe-se aqui, que as variáveis de estado possam ser tratadas como aleatórias e, assim sendo, possam ser determinadas a partir das mesmas técnicas empregadas para estimativa dos parâmetros.

Ademais, contrapondo-se à análise sequencial, avaliar-se-á cada evento de dissimilaridade de forma independente, em uma janela temporal pré-estabelecida.

Não obstante essas diferenças, é necessário observar que, no caso de modelos conceituais, as variáveis de estado nem sempre possuem uma interpretação física direta e, ainda que possuíssem, caso pudessem ser medidas, não há garantia de que seriam representativas (das condições da bacia), em especial, quando obtidas a partir de amostras pontuais. Assim sendo, é substancial notar que o número de trabalhos que optam pela análise *input-output*, seja através da assimilação de dados (via atualização de  $u$  ou  $\tilde{y}$ ), seja por meio análise das incertezas, é expressivamente superior à abordagem aqui sugerida.

De forma sumária, deseja-se avaliar:

- Se há algum padrão no comportamento exibido pelas variáveis de estado que permita identificar esses períodos de dissimilaridade *a priori*.
- Se as alterações nas trajetórias (das variáveis de estado  $x$ ) permitem o alcance de um maior grau de similaridade entre as séries (de vazões) observadas e modeladas em um intervalo de tempo pré-fixado;
- A magnitude dessas alterações;
- Se a dispersão em torno do valor mais frequente é expressiva e segue algum modelo distributivo;
- Por quanto tempo essa alteração pontual é percebida na vazão modelada;
- Caso a alteração ocorresse defasada no tempo (em relação ao instante definido a partir da métrica descrita anteriormente), a magnitude e a incerteza das alterações sofreriam algum impacto; e
- Se essa atualização introduz algum viés nos intervalos de tempo subsequentes ao considerado.

A hipótese admitida na presente abordagem metodológica é que a magnitude da alteração das vazões modeladas (provocada pela atualização de  $x$ ) não é capaz de alterar a estrutura



probabilística do erro ao longo de todo o período de simulação e, por conseguinte, os conjuntos de parâmetros estimados durante a etapa de calibração/ estimativa de incertezas continua válido. Ela é necessária, pois ao admitir uma janela de tempo relativamente curta, não é possível estimar com precisão uma função densidade de probabilidade que permita o confronto do comportamento do erro (nessa janela) e o modelo admitido *a priori*.

Para construção da análise foram avaliados dois procedimentos para atualização das variáveis de estado  $x$ . O primeiro admitiu que elas poderiam assumir quaisquer valores, desde que respeitados os limites impostos pelo conjunto de parâmetros. No segundo, admitiu-se que a variável de estado poderia sofrer uma alteração (redução ou aumento) de até 50% do valor assumido previamente (quando não considerada qualquer atualização). Em ambos os procedimentos elas são consideradas independentes.

No primeiro caso, o valor atualizado da variável de estado independe do valor assumido por ela no instante de tempo anterior à mudança. Assim, o problema aproxima-se daquele do início da simulação/previsão hidrológica, ou seja, de inicialização dos estados. Já no segundo, a atualização é função do valor calculado em  $t-1$  e, por conseguinte, procura-se validar a hipótese de que a introdução de um ruído (caracterizado pela alteração da trajetória do estado) permitiria corrigir eventuais erros impostos pelos dados e/ou estrutura conceitual.

O conjunto de parâmetros considerado a cada iteração é determinado aleatoriamente a partir da amostragem das distribuições *a posteriori* obtidas quando da calibração/ estimativa das incertezas dos parâmetros.

A janela que caracteriza o início e o final da dissimilaridade ( $t_i:t_f$ ; *número de intervalos de tempo considerados* =  $t_a$  (dias) =  $t_f - t_i + 1$ ) foi definida a partir da análise sequencial da métrica descrita anteriormente. Ela é variável para cada evento e possui extensão mínima igual a 10 dias. Sabe-se, *a priori*, que sua extensão condiciona tanto a eficiência do procedimento de atualização quanto a possibilidade de introdução de um novo viés. Por conseguinte, a fim de permitir uma análise mais detalhada de sua influência, admitiram-se 3 variações da mesma, as quais foram definidas a partir de múltiplos da janela identificada previamente: (i)  $1 \times t_a$  ( $t_i:t_f$ ); (ii)  $2,5 \times t_a$ ; e (iii)  $5 \times t_a$ . Sua variação permitirá, também, a avaliação da hipótese assumida na presente abordagem metodológica (a atualização pontual de  $x$  não altera a estrutura probabilística do erro, quando considerada toda a série).

Similarmente, empregar-se-ão três variações para definição do instante inicial de atualização. Todas baseiam-se na diferença entre  $t_f$  e  $t_i$  empregada para cálculo da métrica (4 dias  $\approx$  n.º de ordenadas do HU): (i)  $t_i$  ; (i)  $t_i - 4$  dias ; (ii)  $t_i - 10$  dias. A introdução desse *lag* permitirá uma avaliação comparativa da magnitude da alteração dos estados ou, mais especificamente se sua admissão implica em uma redução da magnitude da alteração, uma vez que há mais “tempo” para a correção da trajetória exibida por  $x$ .

É necessário observar que, a princípio, os valores assumidos pelos estados no primeiro procedimento de atualização (no qual eles independem do comportamento exibido em instantes de tempo anteriores) equivalem àqueles considerados ótimos para a janela de tempo considerada. A questão central é quão incerta é essa estimativa. Outra questão imediata é se o aumento na similaridade entre vazões observadas e modeladas é superior ao obtido através do segundo procedimento, uma vez que, nesse caso, a alteração é condicionada a 50% do valor assumido por  $x$  no instante anterior à atualização.

Na abordagem metodológica admitida, cada evento de dissimilaridade acarretará em nove simulações para cada procedimento de atualização (2), as quais são definidas pela combinação dos lags (3) e janelas (3) consideradas.

## 5 RESULTADOS E DISCUSSÃO

Os resultados obtidos são apresentados ao longo deste capítulo, segundo a sequência metodológica proposta.

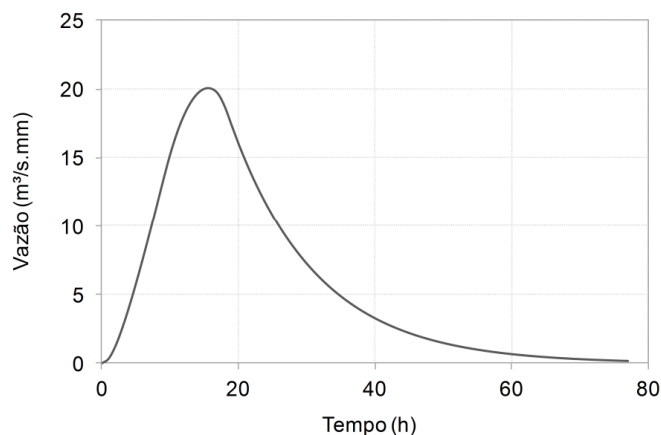
### 5.1 Dados de entrada e particularidades da simulação

Para calibração de um modelo chuva-vazão é necessário identificar, *a priori*, os intervalos de incertezas respectivamente associados aos parâmetros a serem estimados. Note-se que, conforme menção anterior, na presente utilização do modelo Rio Grande as estimativas restringiram-se aos parâmetros relativos ao módulo de Produção. Diversas simulações foram realizadas a fim de se estabelecerem os limites dos referidos intervalos. Para tal, tomaram-se como referência os períodos da série fluviométrica nos quais verificavam-se os extremos sazonais (vazões de pico e fim da recessão). Ademais, um modelo digital de elevação foi utilizado para extração automática das características físicas da bacia do rio Pará, em Ponte do Vilela, necessárias à estimação dos parâmetros do módulo de Concentração. Os limites estabelecidos e os parâmetros fixos, como empregados na análise, estão apresentados na Tabela 5.1.

**Tabela 5.1** – Parâmetros do modelo Rio Grande admitidos *a priori*.

Módulo	Símbolo	Limite inferior	Limite superior	Valor fixo
Produção	<i>k</i>	0,5	1	
	<i>imp</i>	0	0,03	
	<i>wum</i>	5	150	
	<i>wlm</i>	50	250	
	<i>wdm</i>	5	150	
	<i>sm</i>	20	120	
	<i>b</i>	0,1	1	
	<i>ex</i>	0,1	2	
	<i>c</i>	0,01	0,4	
	<i>kss</i>	0,05	0,35	
	<i>kg</i>	0,05	0,65	
	<i>ci</i>	0,1	0,99	
Concentração	<i>cg</i>	0,8	0,999	
	<i>tc</i>			17,4 h
	<i>n</i>			1,5

A concentração dos pulsos discretos de escoamento total à seção de interesse foi realizada por meio da convolução desses pulsos e do hidrograma unitário de Clark, estabelecido a partir do histograma tempo-área sintético desenvolvido pelo HEC (1981). O HU obtido está ilustrado na Figura 5.1.



**Figura 5.1** – Hidrograma unitário de duração igual a 1h calculado para o caso de estudo.

Uma dificuldade inerente ao emprego de modelos que utilizam algum tipo de HU em sua estrutura está relacionada à adequação do intervalo de tempo empregado na simulação. É necessário que a duração do HU seja condizente com tempo de concentração da bacia. Ademais, as séries hidrológicas empregadas deverão ser discretizadas segundo esse mesmo intervalo de tempo.

Não obstante esse fato, grande parte das séries hidrológicas disponíveis para o território nacional refere-se a dados em intervalo de tempo diário. No presente caso, os dados de chuva e evaporação são totais acumulados em 24 h (às 7 e 9 h, respectivamente) enquanto, os dados de vazão correspondem a médias diárias calculadas a partir de duas observações ao longo do dia (às 7 e 17 h).

A fim de contornar essa incompatibilidade de intervalos de tempo, relativamente comum, o modelo Rio Grande permite que os intervalos de simulação dos módulos de Produção e Concentração sejam diferentes. Para tal, os pulsos discretos de deflúvio total são uniformemente desagregados, segundo a duração admitida para o HU. As vazões obtidas, após a convolução, são regularizadas segundo o intervalo de tempo dos dados amostrados, a partir da média de  $n$  passos de cálculo, onde  $n$  é a razão entre os intervalos de tempo dos dados (módulo de Produção) e do HU (módulo de concentração).

É fato que as séries empregadas na modelagem se propõem a representar os processos hidrológicos na escala espaço-temporal da bacia. A premissa é de que elas sejam capazes de reproduzi-los com suficiente grau de acurácia e precisão. A regularização dos intervalos de tempo agrega, de modo intrínseco, incertezas às vazões modeladas. Contudo, se por um lado esta situação é díspar relativamente às verificadas em estudos correlatos a este, ela é representativa da realidade brasileira.

No tocante aos modelos de resíduos analisados (NL e GL), é necessário estabelecer os limites de incertezas associados aos parâmetros a serem estimados concomitantemente aos do módulo de produção. Todos eles, à exceção de  $\sigma$ ,  $\sigma_0$  e  $\sigma_1$ , foram definidos a partir dos limites numéricos de validade das formulações nas quais são utilizadas.

No caso do parâmetro  $\sigma_1$  (coeficiente angular da reta definida pela equação 4.42), o valor admitido refere-se ao empregado no artigo que descreve o método (SCHOUPS e VRUGT, 2010). Para o parâmetro  $\sigma_0$  foram realizados diversos testes variando o valor descrito em Schoups e Vrugt (2010). A limitação é que o desvio padrão deve necessariamente ser superior a zero. O limite inferior estabelecido somente é plausível caso uma correção explícita do desvio padrão seja realizada no algoritmo. No presente caso, o valor mínimo permitido é 1E-06. Essa correção também é realizada no caso do limite inferior de  $\sigma$ . Já em relação ao limite superior, admitiu-se que o valor seria aproximadamente 10% da vazão máxima observada.

Os limites aqui empregados estão relacionados na Tabela 5.2.

**Tabela 5.2** – Intervalos de incertezas admitidos *a priori* para os parâmetros das funções de verossimilhança avaliadas.

Modelo	Símbolo	Limite inferior	Limite superior
NL	$\sigma$	0	50
	$\sigma_0$	-2	1
	$\sigma_1$	0	1
GL	$\beta$	-0,99	1
	$\xi$	0,1	10
	$\phi_i$	0	1
	$\mu_M$	-	-

Por fim, ainda é necessário estabelecer os parâmetros do método de simulação de Monte Carlo via cadeia de Markov. Alguns testes foram realizados para a definição do número de cadeias (*nseq*), no entanto, não se verificaram ganhos ou anomalias introduzidas por esse parâmetro. O número máximo de simulações (*ndraw*) foi definido tendo-se como certo que o algoritmo cessaria somente quando da convergência das cadeias. O *burn-in* utilizado foi o sugerido pelos autores do algoritmo. Em contrapartida, o valor da estatística de Rubin e Gelman, empregada para verificação da convergência (*Rthres*), é significativamente inferior ao sugerido por Vrugt *et al.* (2009) (1,2), mas similar aos descritos em Marshall *et al.* (2004). De fato, quando utilizado o valor sugerido pelos referidos autores, o algoritmo denotava a convergência prematura das cadeias, revelada pela diferença significativa das distribuições *a posteriori* obtidas. O valor aqui utilizado foi adotado após alguns testes, os quais também serviram para definir o número de intervalos entre avaliações da função de verossimilhança, para composição da amostra (*thin.t*). Note-se que o valor adotado garante que a autocorrelação entre parâmetros da amostra não seja significativa.

A Tabela 5.3 apresenta os valores dos parâmetros aqui empregados, na aplicação do método DREAM.

**Tabela 5.3 – Valores dos parâmetros empregados na utilização do método DREAM.**

Símbolo	Valor
<i>nseq</i>	5
<i>ndraw</i>	500.000
<i>burn-in</i>	50% iniciais
<i>thin.t</i>	100
<i>Rthres</i>	1,01

É válido destacar que a taxa de aceitação do algoritmo variou entre 20 e 40 %. Nos resultados apresentados na sequência, foram necessárias aproximadamente 120.000 (260.000) simulações para a convergência das cadeias de Markov, segundo o modelo de resíduos NL (GL). Após verificá-la, as amostras eram produzidas a partir 1E+06 novas simulações, o que produziu 10.000 conjuntos de parâmetros.

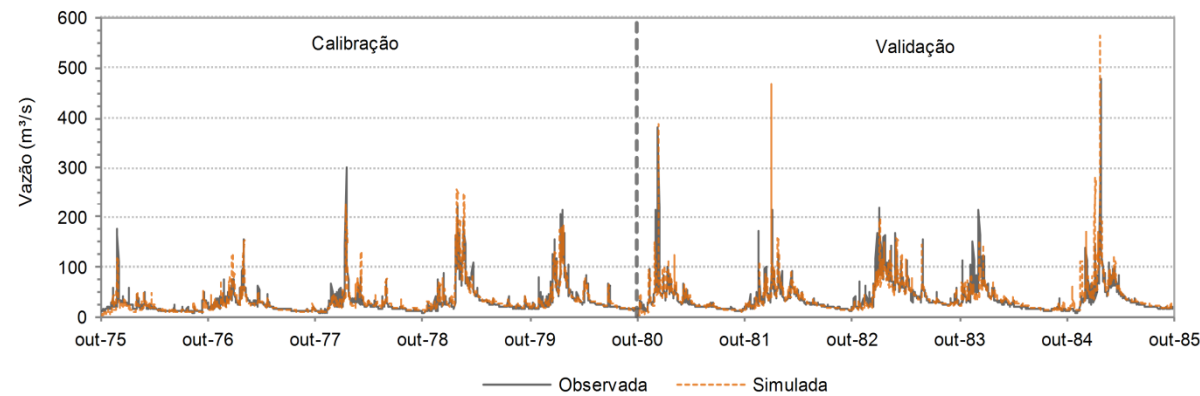
A fim de compreender a influência de cada parâmetro sobre a qualidade do ajuste do modelo realizou-se também uma análise de sensibilidade dos parâmetros do modelo Rio Grande. Essa avaliação revelou que os parâmetros *k*, *wum*, *wlm*, *sm*, *kss*, *kg* e *ci* impõem ao hidrograma

modelado significativas variações quando da alteração de seus valores. Por conseguinte, as análises que se sucederam ficaram restritas a esses parâmetros. É válido, no entanto, destacar que se observou uma forte interação não-linear entre os parâmetros do modelo, o que produzia efeitos similares ou compensatórios sobre diferentes segmentos dos hidrogramas modelados, conforme a variação conjunta desses parâmetros.

### 5.2 Avaliação dos modelos probabilísticos admitidos para os resíduos

O primeiro modelo probabilístico analisado pressupôs que os erros do modelo eram independentes e identicamente distribuídos, de acordo com a distribuição Normal  $[N(0, \sigma^2)]$ . Essa premissa, comumente admitida para explicar o comportamento dos resíduos na modelagem hidrológica, conduz o problema da estimação de parâmetros à abordagem sugerida pelo método dos mínimos quadrados padrão (SLS – *standard least squares*).

A Tabela 5.4 faz um sumário dos resultados obtidos e a Figura 5.2 os ilustra.



**Figura 5.2** – Série temporal de vazões observadas e simuladas admitindo-se a normalidade dos resíduos.

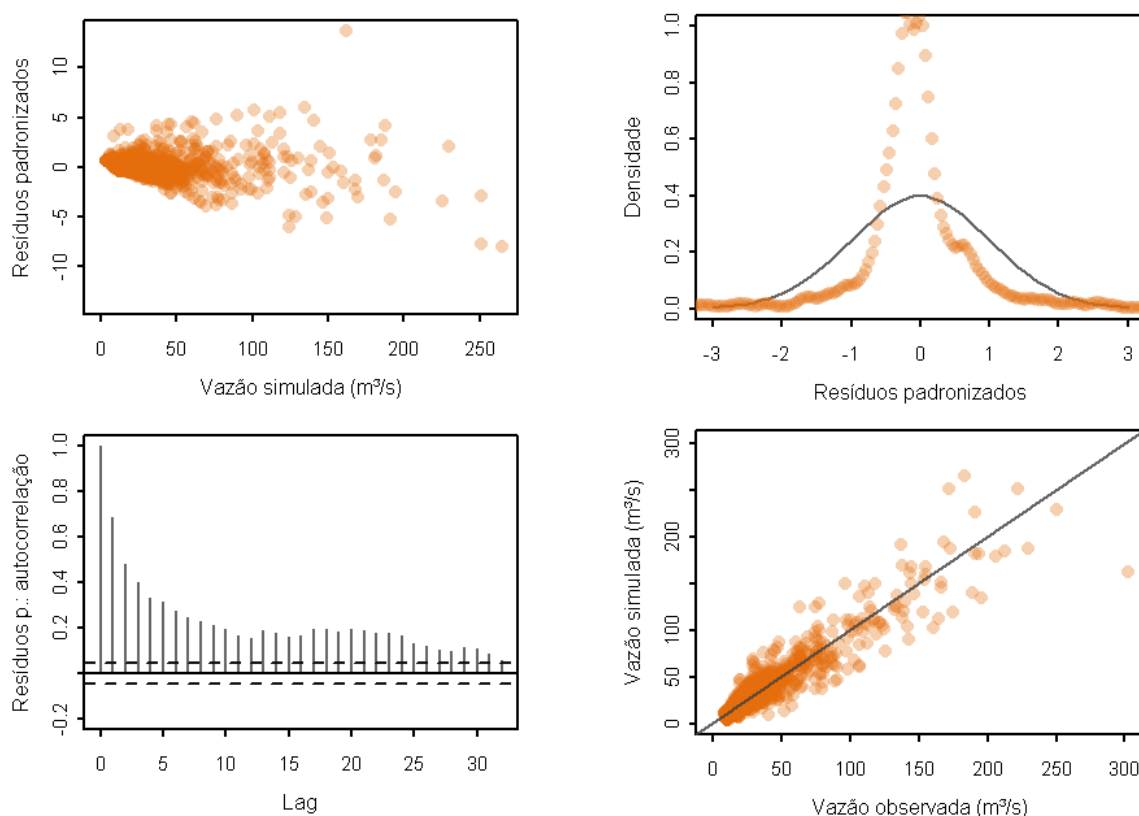
**Tabela 5.4** – Estatísticas descritivas dos resultados admitindo-se a normalidade dos resíduos.

Estatística	Símbolo	Calibração	Validação
Coefficiente de <i>Nash-Sutcliffe</i>	NS	0,871	0,797
Relação de volumes	RV	0,988	1,023
Raiz do erro quadrático médio	RMSE	10,2	14,8
Coefficiente de correlação de Pearson	$\rho$	0,934	0,912

\* As estatísticas referem-se aos parâmetros determinados a partir da moda das distribuições *a posteriori*

Observa-se que o modelo foi capaz de bem reproduzir as vazões observadas, tanto no período de calibração quanto no de validação, ainda que neste as diferenças tenham sido ligeiramente acentuadas.

A análise dos resíduos, no entanto, revelou que os mesmos não seguem a hipótese de normalidade admitida *a priori*, conforme demonstrado pela Figura 5.3.



**Figura 5.3** – Comportamento dos resíduos admitidos normais durante a calibração.

Nos painéis da Figura 5.3, evidencia-se que: (i) a variância aumenta com a magnitude da vazão simulada; e que (ii) os resíduos são significativamente autocorrelacionados. Isso contradiz as hipóteses admitidas de variância constante e de independência. Ademais, o confronto entre a distribuição dos resíduos e o modelo Gaussiano corrobora para a afirmação de que, apesar do elevado grau de similaridade entre vazões observadas e simuladas, o modelo de erros admitido é inadequado.

Outros estudos reportaram um comportamento dos resíduos análogo ao obtido (KUCZERA, 1983 *apud* SCHOUPS e VRUGT, 2010; THYER *et al.*, 2009). Por ser uma hipótese frequentemente admitida (e violada), observou-se na última década um intenso debate na



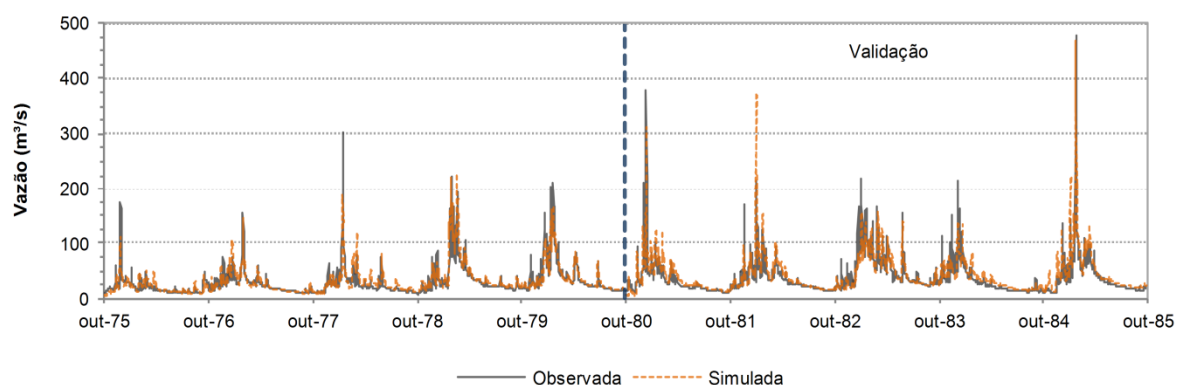
comunidade técnica acerca da especificação, desenvolvimento e avaliação (em uma abordagem formal ou informal) de funções de verossimilhança mais adequadas às técnicas de estimativas de incertezas na modelagem hidrológica construídas sob o paradigma Bayesiano (BEVEN *et al.*, 2008; VRUGT *et al.*, 2008b; SMITH *et al.*, 2010; EVIL *et al.*, 2013; EVIL *et al.*, 2014).

O modelo generalizado (GL) de erros (SCHOUPS e VRUGT, 2010) insere-se neste contexto. Os resultados obtidos a partir de sua utilização estão sintetizados na Tabela 5.5 e nas Figura 5.4 e 5.5.

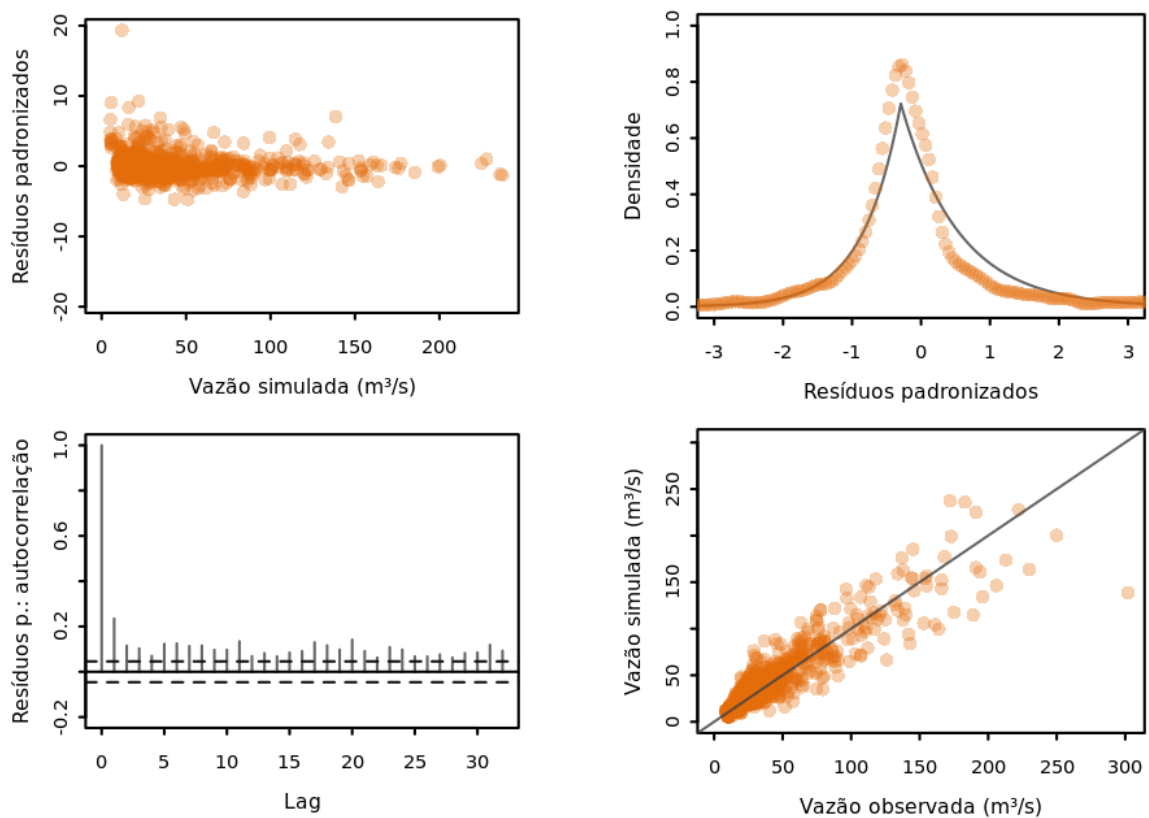
**Tabela 5.5** – Estatísticas descritivas dos resultados admitindo-se o modelo GL.

Estatística	Símbolo	Calibração	Validação
Coeficiente de <i>Nash-Sutcliffe</i>	NS	0,848	0,813
Relação de volumes	RV	1,033	1,091
Raiz do erro quadrático médio	RMSE	11,1	14,2
Coeficiente de correlação de Pearson	$\rho$	0,922	0,915

\* As estatísticas referem-se aos parâmetros determinados a partir da moda das distribuições *a posteriori*



**Figura 5.4** – Série temporal de vazões observadas e simuladas admitindo-se o modelo GL.



**Figura 5.5** – Comportamento dos resíduos admitidos heterocedásticos, autocorrelacionados e modelados probabilisticamente segundo a FDP de Laplace assimétrica durante a calibração.

A exemplo do anterior modelo de erros, as estatísticas descritivas da qualidade do ajuste reportaram um elevado grau de similaridade entre as vazões observadas e simuladas. Constata-se, entretanto, que o ajuste é levemente inferior ao observado anteriormente, ainda que a diferença entre os períodos de calibração e validação tenha sido atenuada. Em contrapartida, tanto a heterocedasticidade quanto a autocorrelação dos resíduos foram praticamente removidas (uma leve autocorrelação com *lag* 1 ainda é observada). A distribuição de Laplace assimétrica ajustou-se muito bem à distribuição empírica dos resíduos, em especial ao pronunciado pico observado em sua forma, o que a torna significativamente díspar da distribuição Normal.

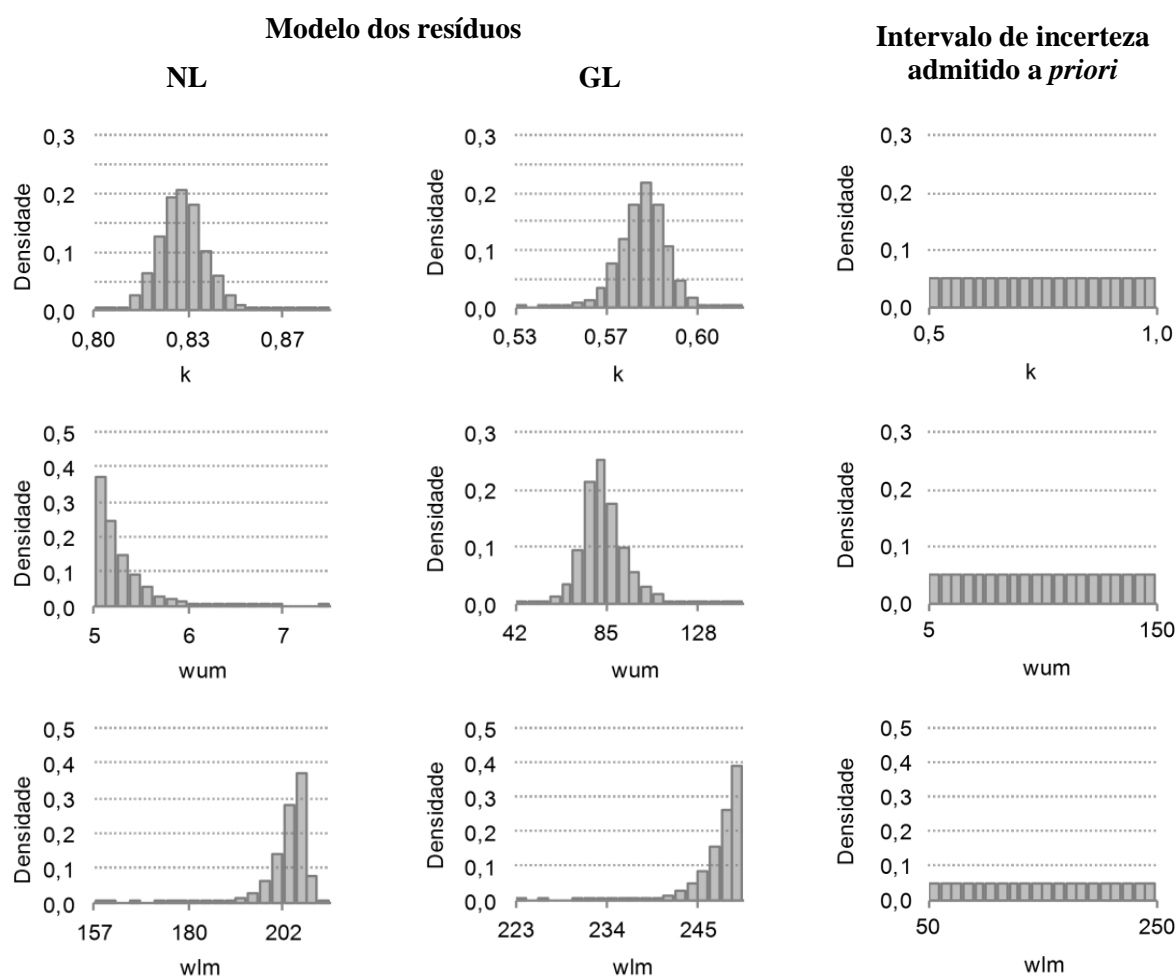
A adoção do modelo generalizado de erros implica na necessidade de estimar novos parâmetros simultaneamente aos do modelo hidrológico (resultados não apresentados). Apesar da maior complexidade, o método DREAM foi capaz de estimar esses parâmetros sem que tenham sido observados problemas de convergência. Exceção se faz ao parâmetro de autocorrelação  $\phi_I$ , cuja presença representou uma elevação significativa no número de

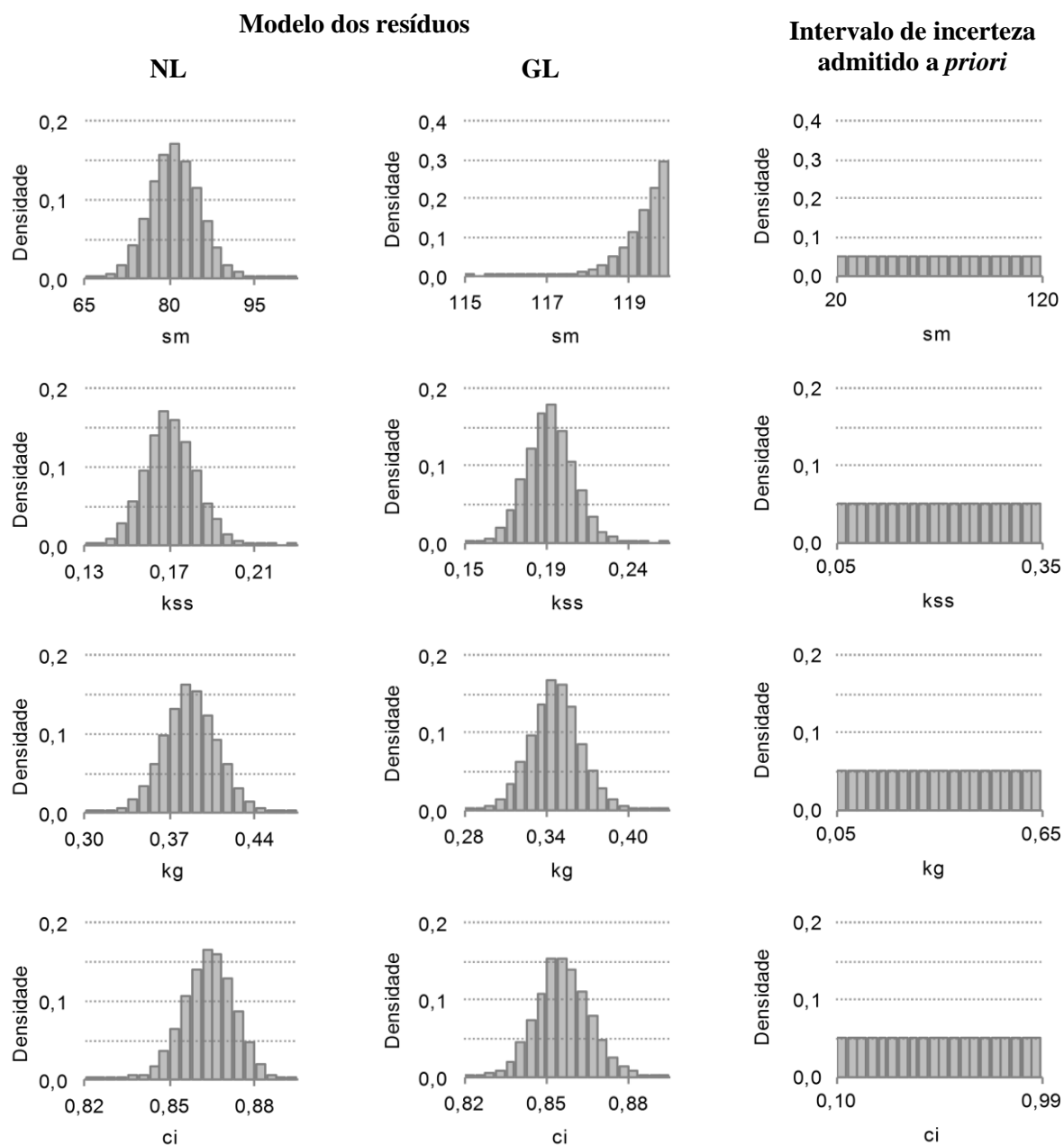
simulações necessárias à convergência do algoritmo. Por essa razão, optou-se por fixá-lo em 0,70, a exemplo do parâmetro de curtose  $\beta$  (1,0), que, em função da forma exibida pela distribuição empírica dos resíduos, tendia invariavelmente para a unidade.

O confronto entre as hipóteses admitidas pelos modelos de erros e os resíduos obtidos é essencial para a obtenção de estimativas não enviesadas das incertezas. Há que se observar, entretanto, exemplos na literatura nos quais essa avaliação foi menosprezada (FEYEN *et al.*, 2007). O decorrente impacto nas estimativas é destacado nos resultados descritos na sequência.

### 5.3 Incertezas nas estimativas dos parâmetros

As respectivas distribuições *a posteriori* dos parâmetros estimados foram obtidas para ambos os modelos de resíduos avaliados, ainda que o modelo NL tenha-se revelado inadequado. O objetivo é demonstrar o viés introduzido nos resultados. O comportamento exibido pelos parâmetros do modelo Rio Grande está ilustrado na Figura 5.6.





**Figura 5.6** – Distribuições *a posteriori* dos parâmetros do modelo Rio Grande.

É importante observar a expressiva alteração dos intervalos das estimativas dos parâmetros  $k$ ,  $wum$  e  $sm$ . A esse respeito, é oportuno avaliar o viés introduzido na estimativa da incerteza. Considerando-se a análise de sensibilidade dos referidos parâmetros e as informações hidrológicas utilizadas, pode-se inferir que:

- $k$ : há uma incerteza nos dados de evaporação imposta, sobretudo, pela distância da estação evaporimétrica à bacia e pela dificuldade operacional no monitoramento da variável (tanque evaporimétrico). Por conseguinte, essa variação reflete um efeito

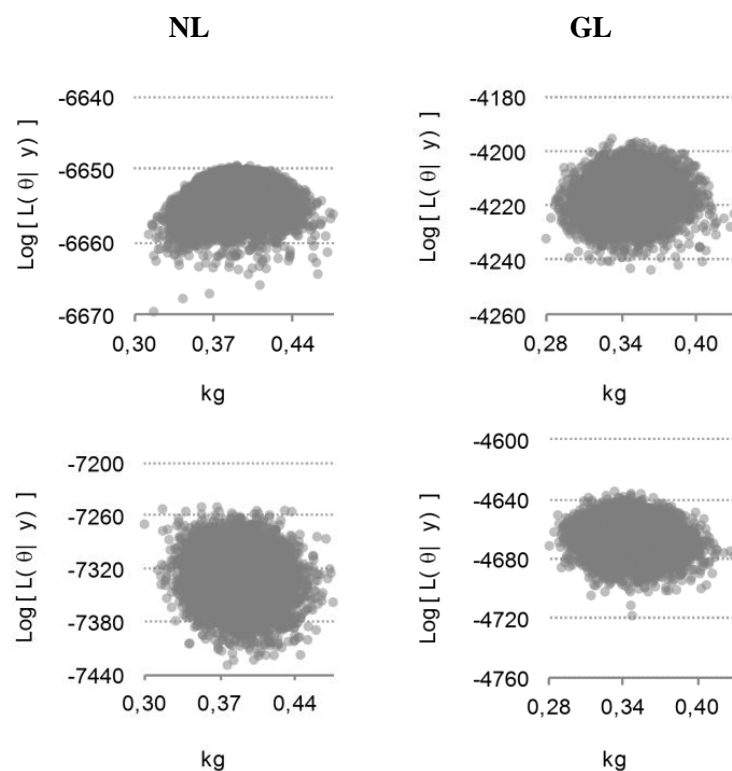
compensatório da incerteza nos dados de entrada (não considerada nesta análise) sobre a estimativa do parâmetro;

- *sm*: sua alteração implica em uma variação pronunciada da magnitude das vazões de pico. A incerteza da estimativa, refletida na diferença entre os modelos, decorre da observação de que o modelo GL enfatiza menos o ajuste das vazões de pico, devido à heterocedasticidade dos erros;
- *wum*: ao efeito combinado da alteração das variáveis anteriores, as quais implicam necessariamente em variações no balanço de umidade do solo ou, em outras palavras, na maior ou menor disponibilidade de água para evapotranspiração e/ou geração de escoamento superficial.

O maior padrão de regularidade exibido pelos demais parâmetros pode induzir o modelador a acreditar que o viés introduzido, neste caso, é insignificante. No entanto, é fundamental, na discussão da adequação do modelo de erros e das incertezas nas estimativas nos parâmetros, se avaliar a diferença entre o comportamento exibido pela função de verossimilhança durante os períodos de calibração e validação. As Figuras 5.7 e 5.8 ilustram a variação observada da função log de verossimilhança correspondente aos parâmetros *kg* e *ci*, sob os dois modelos de resíduos, para as fases de calibração e validação.

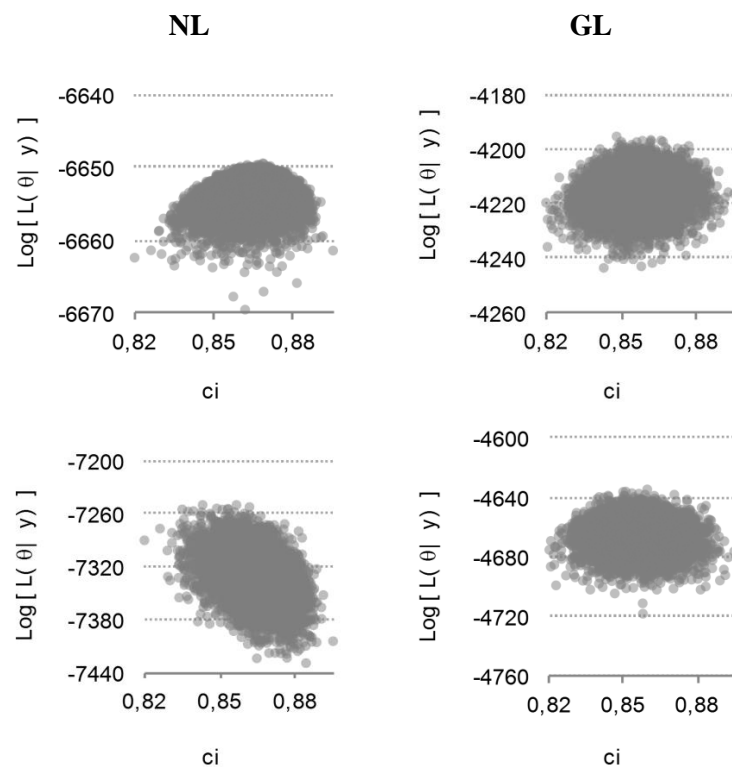
Durante a calibração, nota-se que a função de verossimilhança possui um valor máximo em ambos os modelos, ressaltando-se que esse comportamento é mais explícito no modelo NL. Já na etapa de validação, na qual ambos os modelos apresentaram desempenho inferior à etapa de calibração (observe-se a diferença no valor máximo do logaritmo da função de verossimilhança), esse comportamento somente é observado no modelo GL. Esse fato, evidenciado também pela maior regularidade exibida nas estatísticas descritivas do modelo GL durante os períodos de calibração e validação, reafirma a adequação do modelo que pressupõe que os erros são heterocedásticos, autocorrelacionados e modelados probabilisticamente por uma FDP exponencial potência assimétrica.

### Modelo de resíduos



**Figura 5.7** – Comportamento da função de verossimilhança para o parâmetro  $kg$ : calibração (acima) e validação (abaixo).

### Modelo de resíduos

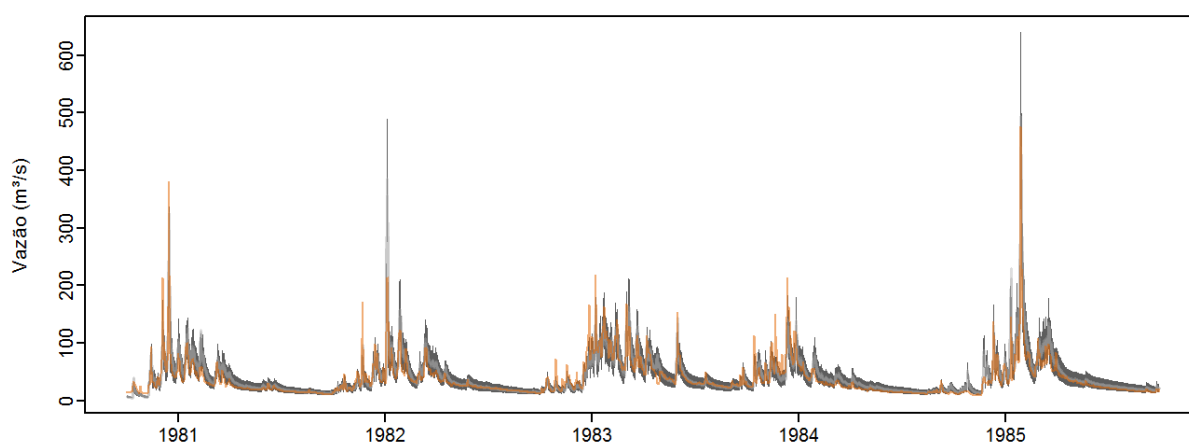


**Figura 5.8** – Comportamento da função de verossimilhança para o parâmetro  $ci$ : calibração (acima) e validação (abaixo).

Independentemente desse fato, é notório que diferentes estimativas do parâmetro conduzem a valores muito similares da função de verossimilhança. Em decorrência, a hipótese de um conjunto ótimo, largamente difundida na literatura que trata da calibração de parâmetros sob a perspectiva de um problema de otimização, não pode ser verificada. De fato, há múltiplas estimativas aceitáveis dos parâmetros, as quais não podem ser rejeitadas facilmente. A tese da equifinalidade (BEVEN e BINLEY, 1992 *apud* BEVEN, 2009), confirmada pela Figura 5.6, além de reiterar a importância de se avaliar as incertezas associadas aos prognósticos dos modelos hidrológicos, reforça a forte influência da medida de verossimilhança sobre os resultados da modelagem.

#### 5.4 Predição da incerteza na vazão modelada

É possível, a partir da metodologia empregada, prever a incerteza na vazão modelada. Os intervalos com 95% de incerteza estão apresentados na Figura 5.9.



**Figura 5.9** – Incerteza na predição da vazão modelada durante o período de validação utilizando-se o modelo de resíduos GL – série temporal observada (laranja) e intervalo com 95% de incerteza: (i) decorrente das estimativas nos parâmetros (cinza claro); e (ii) total (cinza escuro).

A inspeção do gráfico da Figura 5.9 revela que as vazões observadas durante o período de validação estão contidas entre os limites associados à incerteza total em cerca de 89% do tempo. Apesar do elevado percentual, é notório que algumas das vazões de pico observadas extrapolam o limite superior calculado. Esse fato indica, em uma primeira análise, que é necessária uma série mais extensa do que a utilizada para a calibração do modelo (5 anos). Não obstante essa inferência, é necessário ponderar que a presente análise restringiu-se à avaliação das incertezas associadas aos parâmetros, desconsiderando as demais fontes (dados

e estrutura do modelo), as quais devem representar outra importante parcela da incerteza na predição da vazão modelada, uma vez que, em apenas 9% do período de validação, os limites associados à incerteza nos parâmetros continham as vazões observadas.

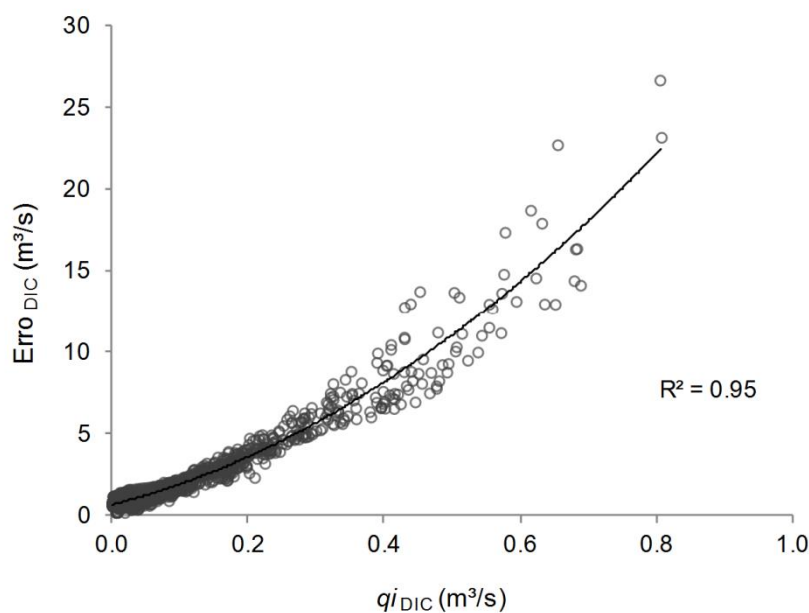
Nota-se que, apesar do reduzido percentual no caso dos limites associados à incerteza dos parâmetros, as estatísticas empregadas para mensurar a qualidade do ajuste são relativamente elevadas. Outro ponto que cabe destaque é a observação de que os referidos percentuais, em especial o associado à incerteza dos parâmetros, raramente são apresentados em estudos correlatos. Sadegh & Vrugt (2014) é uma das exceções. Eles reportaram valores iguais a 19 e 13% para incertezas associadas aos parâmetros, nos períodos de calibração e validação, respectivamente, considerando-se o emprego do algoritmo DREAM e de uma função de verossimilhança que admitia que os resíduos eram Gaussianos. Esses valores são similares aos obtidos nesta tese.

Tendo em mente esses resultados, parece substancial a busca por técnicas que permitam a redução dos erros impostos por incertezas associadas aos dados e à estrutura conceitual do modelo chuva-vazão. Este estudo orientou-se por essa percepção. No entanto, contrapondo-se ao reconhecimento explícito, *a priori*, dessas outras fontes de incerteza, as quais são requeridas por algumas das técnicas descritas na literatura recente da especialidade (AJAMI *et al.*, 2007; RENARD *et al.*, 2010), optou-se por caracterizá-las a partir da análise do comportamento *input-state-output* do modelo empregado. Em parte, essa opção decorre da crença de que é impossível dissociá-las. Ademais, acredita-se que o comportamento das trajetórias das variáveis de estado de um modelo chuva-vazão, as quais são calculadas a partir do conjunto de parâmetros considerados aceitáveis, represente uma espécie de assinatura da incerteza total. Assim sendo, ao assimilá-la, seria possível não apenas estimá-la, mas, também, corrigir eventuais desvios caracterizados por elas.

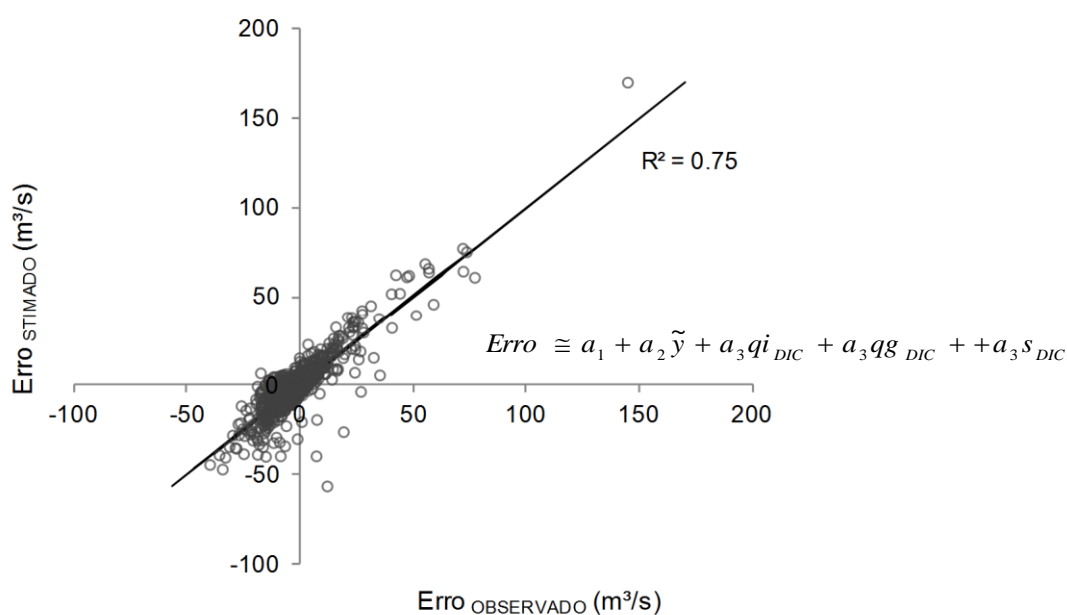
## **5.5 Atualização das variáveis de estado**

Antes de identificar os períodos de dissimilaridade persistente, investigou-se a hipótese de que as variações na evolução da trajetória das variáveis de estado possam ser empregadas para caracterizar o aumento/redução do erro ao longo do tempo de simulação/previsão. A fim de verificá-la, foram desenvolvidas diversas análises gráficas, sendo as mais expressivas as apresentadas nas Figuras 5.10 e 5.11.





**Figura 5.10** – Diferença entre os limites superior (97,5%) e inferior (2,5%) dos intervalos de credibilidade (*DIC*): da variável de estado *qi* versus a do erro na vazão simulada.



**Figura 5.11** – Dispersão dos resíduos de um modelo de regressão linear múltipla construído para a estimação da mediana da série de erros.

Na Figura 5.10, a diferença entre os limites superior (97,5%) e inferior (2,5%) dos intervalos de credibilidade (*DIC*) da variável de estado *qi*, a qual representa o afluxo subsuperficial à rede de drenagem, e do erro, calculado a partir da diferença entre as vazões obtidas por meio

do conjunto de parâmetros  $\theta$  considerados aceitáveis e as vazões observadas, são grafadas nos eixos das abcissas e das ordenadas, respectivamente. Nota-se que um aumento/redução da incerteza no comportamento exibido por  $qi$  está diretamente relacionado a uma maior variabilidade do erro. A qualidade do ajuste ainda é expressiva, caso admita-se um *lag* de 1 dia entre essas variáveis:  $R^2_{1\text{ DIA}} \approx 0,88$ . Assim sendo, é verossímil admitir que seja possível prever a variação da incerteza na série de erros, com relativa antecedência, a partir do comportamento de uma variável de estado.

É importante ressaltar que o fato da incerteza nos erros elevar-se em um instante de tempo  $t$  qualquer, não garante que os erros calculados sejam maiores para todo conjunto de parâmetros  $\theta$ . Para mensurar a mediana da série de erros, pode-se, por exemplo, empregar a expressão sugerida na Figura 5.11. Nela são apresentados os resíduos de um modelo de regressão múltipla construído a partir da combinação linear das vazões observadas e das *DIC* das variáveis de estado  $qi$ ,  $qg$  e  $s$ . De fato, é possível obter-se expressivos índices de ajuste a partir da combinação de diferentes *DIC* das variáveis de estado  $x$ . Essa observação continua válida, ainda que, empregue-se a mediana das séries de  $x$ 's.

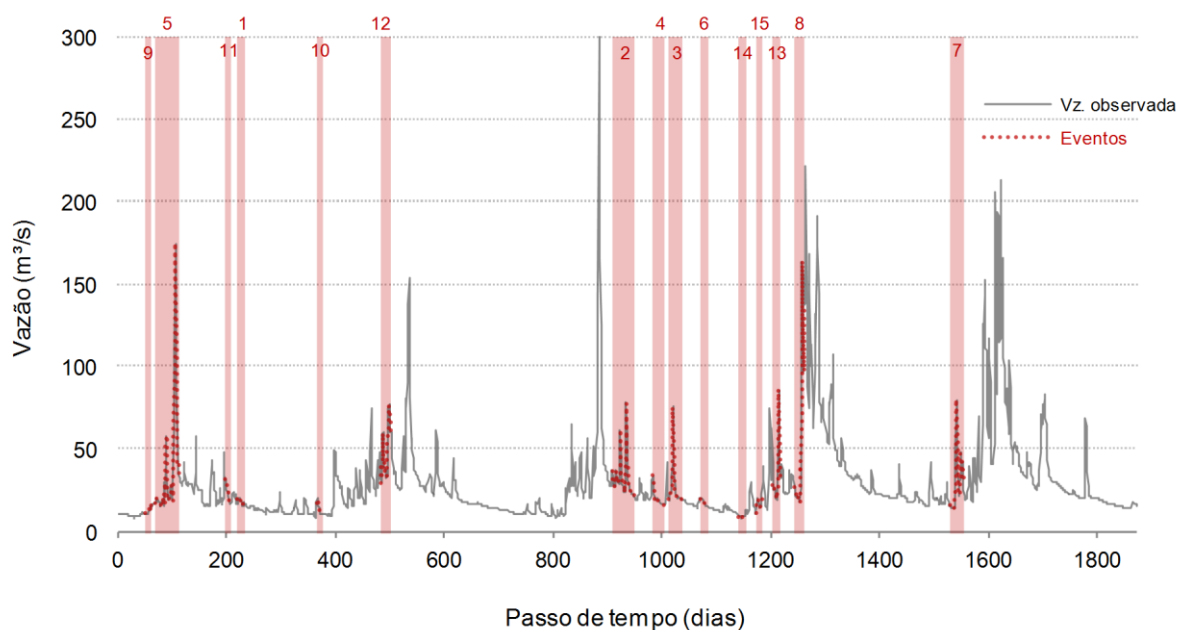
O fato relevante nessas relações é a constatação de que a análise do comportamento do conjunto das trajetórias assumidas pelos estados  $x$ , ao longo do tempo de simulação, permite não apenas a estimação do erro, o que já é difundido pelas técnicas de assimilação de dados, mas, também, a identificação prévia do aumento da dissimilaridade entre as séries de vazão observada e de, ao menos, parte das simuladas. Assim sendo, é plausível admitir que a incerteza sobre a estimação dos parâmetros represente uma parcela menor da incerteza total em intraperíodos da série caracterizados por esse aumento de variabilidade dos estados  $x$ . Nesse contexto, a introdução de um ruído nas trajetórias de  $x$  poderia “compensar” eventuais incertezas decorrentes dos dados observados e/ou da estrutura conceitual do modelo chuva-vazão. Nos subitens que se seguem, essa assertiva é verificada.

### **5.5.1 Identificação dos períodos de dissimilaridade persistente**

Para identificação dos períodos de dissimilaridade persistente empregou-se a métrica descrita pela equação 4.44. A Tabela 5.6 e a Figura 5.12 apresentam um resumo dos períodos selecionados.

**Tabela 5.6** – Períodos de dissimilaridade selecionados para atualização das variáveis de estado.

Evento	Data	Alvo	Janela considerada			
	Início Fim	Passo de tempo inicial	N.º de dias	Vazão média obs. ( $\bar{y}$ m³/s)	RMSE (m³/s)	$\frac{RMSE}{\bar{y}}$
01	22 mar 05 abr 76	221	14	17,8	13,6	0,76
02	08 fev 22 mar 78	909	42	32,1	22,5	0,70
03	24 mai 19 jun 78	1014	26	26,2	14,8	0,56
04	24 abr 15 mai 78	984	21	19,1	10,3	0,54
05	21 out 05 dez 75	68	45	38,4	18,7	0,49
06	22 jul 05 ago 78	1073	14	16,6	8,0	0,48
07	21 out 15 nov 79	1529	25	28,4	13,4	0,47
08	09 jan 27 jan 79	1244	18	55,5	22,8	0,41
09	05 out 15 out 75	52	10	12,2	5,0	0,41
10	14 ago 25 ago 76	366	11	13,7	5,1	0,37
11	28 fev 11 mar 76	198	12	22,2	8,2	0,37
12	08 dez 28 dez 76	482	20	48,6	17,7	0,36
13	01 dez 15 dez 78	1205	14	34,7	12,6	0,36
14	29 set 11 out 78	1142	12	9,1	3,2	0,35
15	30 out 11 nov 78	1173	12	16,3	5,3	0,32



**Figura 5.12** – Eventos de dissimilaridade selecionados para atualização das variáveis de estado.

Nota-se que as vazões do conjunto de eventos selecionados apresentam uma variabilidade apreciável. São observados tanto períodos de cheia quanto de vazante. A razão entre a raiz do erro médio quadrático e a vazão média observada variou entre 32 e 76% nesses intraperíodos. Há uma concentração expressiva de eventos no período compreendido entre Fev/78 e Fev/79, destacando-se que o mês de Jan/79 caracteriza-se por registrar as máximas vazões em diversas estações fluviométricas localizadas no alto rio São Francisco. Ademais, é necessário mencionar que a diferença entre o *warmup* utilizado (1,5 meses = 47 passos de tempo) e o início dos eventos 9 e 5 (52 e 68, respectivamente) é relativamente pequena. Acredita-se, por consequência, que o viés identificado, ao menos em parte, decorra dos valores assumidos pelo conjunto de  $x$  no início da simulação.

### 5.5.2 Estimação das variáveis de estado e análise das alterações em suas trajetórias

A estrutura conceitual do modelo Rio Grande, apresentada no capítulo 4, possui sete variáveis de estado, as quais estão sumariadas na Tabela 4.3. Destas, três se referem ao conteúdo dos reservatórios empregados para representar o teor de umidade média espacial no solo; são eles os reservatórios superior ( $wu$ ), intermediário ( $wl$ ) e inferior ( $wd$ ). A evolução dos seus conteúdos é complexa, pois, apesar de haver uma relação entre os mesmos, o deplecionamento de um reservatório em um nível inferior não implica na ausência de

conteúdo nos níveis superiores. Isso possibilita uma representação mais robusta do processo de formação do escoamento, em especial no tocante à evolução da frente de molhamento no solo entre eventos pluviométricos distintos subsequentes.

Para que os procedimentos de atualização das variáveis de estados respeitem a síntese conceitual proposta para representação do fenômeno de transformação da chuva em vazão, deve-se garantir que os limites impostos pelos parâmetros sejam respeitados e que as relações entre estados, caso existam, também o sejam. Entretanto, em função da complexidade da estrutura conceitual admitida, não é possível assegurar essa correção *a priori*, no caso dos estados dos reservatórios do modelo Rio Grande.

A fim de contornar esse fato, supôs-se que a atualização ocorreria sobre um único reservatório, cujo limite e o conteúdo eram representados pela soma dos três ( $wu+wl+wd$ ). A distribuição do volume de água armazenado nesses reservatórios supôs-se linear, isto é, a existência de água em um reservatório em nível superior ( $wl$  ou  $wu$ ) implica que o inferior esteja completamente preenchido ( $wd$  ou  $wl$ ). Essa suposição do comportamento do volume de água no solo, apesar de simplista, não viola a estrutura conceitual admitida. Esse reservatório hipotético foi representado genericamente pelo símbolo  $w$ .

Os resultados obtidos são descritos na sequência. Ao apresentá-los, por simplicidade, adotou-se uma notação resumida para expressar os múltiplos das janelas empregadas para avaliação dos procedimentos de atualização (apresentadas na Tabela 5.6) removendo-se o número de dias considerados para avaliação dos procedimentos de atualização (1x, 2,5x e 5x). Ademais, sob a mesma perspectiva, referiu-se aos procedimentos propostos para atualização das variáveis de estado como aqueles baseados em multiplicadores (aplicados sobre os valores assumidos pelos estados  $x$  no instante anterior à atualização) ou percentuais (dos limites impostos pelos parâmetros).

#### Comportamento probabilístico exibido pelos resíduos após os procedimentos de atualização

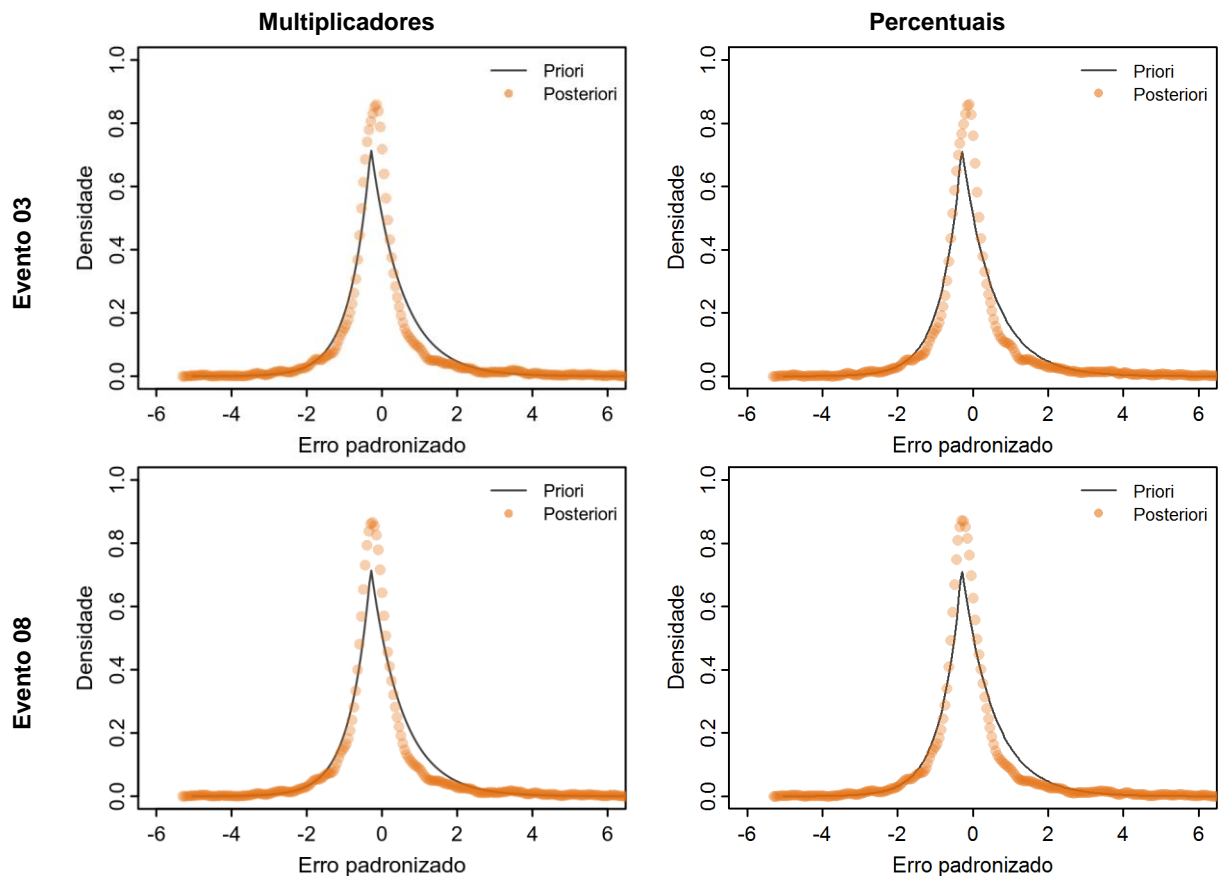
A técnica de inferência Bayesiana empregada para estimação das incertezas nos parâmetros do modelo Rio Grande presume, a exemplo de outras contruídas sob esse paradigma, a admissão de um modelo probabilístico, *a priori*, para os resíduos obtidos. Essa hipótese deve ser verificada *a posteriori*, conforme apresentado no subitem 5.2.

Os procedimentos formulados para atualização das variáveis de estado foram construídos a partir da identificação de uma janela de dissimilaridade. Essa janela e, múltiplos dela, foram assumidas para estimação das incertezas nas variáveis de estado no início desses períodos, sob a suposição de que essa “correção” em seus comportamentos permitiria incorporar à simulação incertezas impostas pelos dados observados e/ou pela estrutura conceitual admitida.

Se, por um lado, essa construção elimina a necessidade de suposições acerca do comportamento do erro sobre os dados/modelo empregados, ao valer-se da mesma técnica empregada para estimação das incertezas nos parâmetros, exige-se, analogamente, a admissão de um modelo probabilístico para os resíduos. Entretanto, a análise de uma janela relativamente curta (quando comparada à extensão temporal da simulação) não permite uma estimação precisa da função densidade de probabilidade. Por conseguinte, valendo-se dos resultados indicados no item 5.2, nos quais demonstrou-se a influência da função de verossimilhança sobre os parâmetros/ resíduos obtidos, supôs-se que esse comportamento deveria permanecer inalterado. Dessa forma, os conjuntos de parâmetros obtidos previamente permaneceriam aceitáveis.

O comportamento exibido pelos resíduos da simulação, quando incorporados os procedimentos de atualização, são apresentados na Figura 5.13 para 2 dos 15 eventos analisados. Os parâmetros da função de verossimilhança (GL) empregados foram obtidos quando da estimação dos parâmetros.

Nota-se que a aderência entre os modelos probabilísticos admitido *a priori* e obtido *a posteriori* permaneceu elevada. Essa observação continua válida quando considerados outros eventos, *lags* e janelas. Esses resultados são apresentados para cada um dos eventos analisados no Anexo II.



**Figura 5.13** – Comportamento probabilístico exibido pelos resíduos da simulação após a atualização das variáveis de estado:  $lag = 0$  dias e janela admitida = 1x.

### Qualidade do ajuste

A fim de sumariar os inúmeros resultados numéricos obtidos e permitir uma comparação direta entre as combinações de lags e janelas admitidas, optou-se por apresentar as principais estatísticas descritivas através de *boxplots*, dispostos de forma matricial, segundo os lags e janelas considerados. A nomenclatura empregada nesses gráficos foi:

Em relação aos lags:

- 0 dias: quando a atualização dos estados ocorreu no instante em que se iniciou a dissimilaridade identificada ( $t_{ALVO}$ );
- -4 dias: quando a atualização ocorreu quatro dias antes do instante  $t_{ALVO}$ ; e
- -10 dias: quando a atualização ocorreu dez dias antes do instante  $t_{ALVO}$ ;

e em relação as janelas temporais:

- 1x: quando, para o cálculo da função de verossimilhança, considerou-se os resíduos produzidos somente durante o período de dissimilaridade identificado;
- 2,5x: quando presumiu-se uma janela temporal igual a 2,5 vezes o período de dissimilaridade identificado, para cálculo da função de verossimilhança; e
- 5x: quando os resíduos produzidos em uma janela temporal igual a 5 vezes a da dissimilaridade identificada foram considerados para o cálculo da função de verossimilhança.

A variação do instante de atualização foi procedida sob a perspectiva de avaliar a magnitude da alteração. A hipótese é que ao admitir uma defasagem temporal em relação ao início da dissimilaridade observe-se uma menor alteração na trajetória dos estados  $x$ , uma vez que há um tempo maior para “correção” dos mesmos. Já a variação da extensão temporal empregada no cálculo da função de verossimilhança foi admitida: (i) pois, se desconhece, a *priori*, a influência, em termos de extensão temporal, de uma alteração pontual nos estados  $x$  sobre as vazões modeladas; (ii) para avaliação da possibilidade de introdução de um novo viés; e (iii) para relaxar a influência da métrica empregada na identificação do período de dissimilaridade.

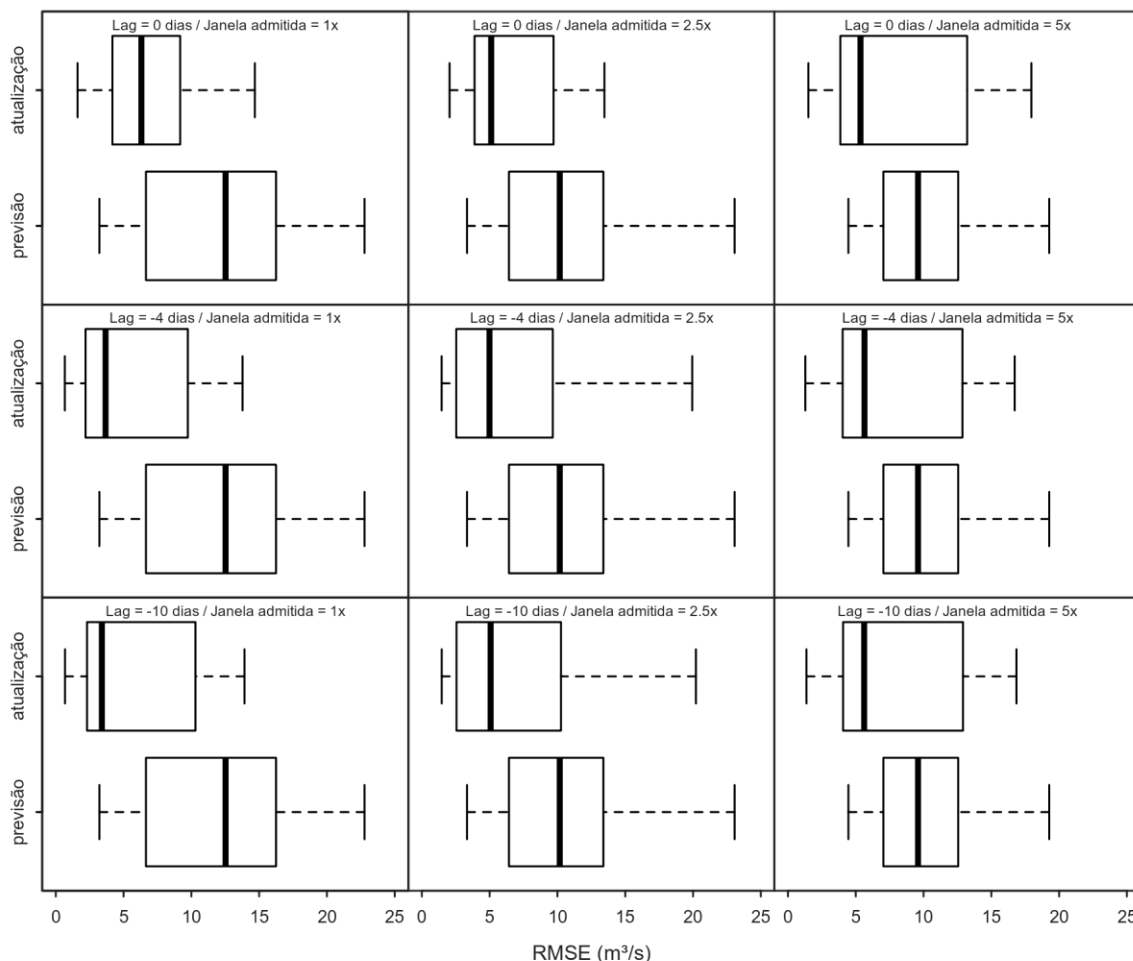
A verificação da “eficácia” dos procedimentos propostos para atualização das variáveis de estado (elevação/redução da similaridade entre vazões observadas e modeladas) foi realizada por meio da comparação da raiz do erro médio quadrático (*RMSE*) das séries prevista, isto é, da mediana das vazões simuladas obtidas a partir do emprego do conjunto de parâmetros  $\theta$ , e atualizada, calculada de forma análoga, considerando-se a alteração pontual, no instante alvo, das trajetórias do conjunto de  $x$ . Essa comparação foi realizada para os instantes subsequentes ao da alteração, restringindo-se, inicialmente, à janela admitida.

A síntese dos resultados para o procedimento que previam a alteração percentual dos estados, isto é, quando a atualização independe do valor assumido por  $x_i$  no instante de tempo anterior à alteração, é apresentada na Figura 5.14. Nota-se que a atualização de  $x$  reduziu a mediana do *RMSE* para todos os *lags* e janelas admitidas. As janelas 1x e 2,5x apresentaram percentis muito similares, entretanto, observa-se uma suave elevação do 3º quartil, quando considerada a janela igual a 5x. Credita-se esse fato à menor ênfase ao período de dissimilaridade,



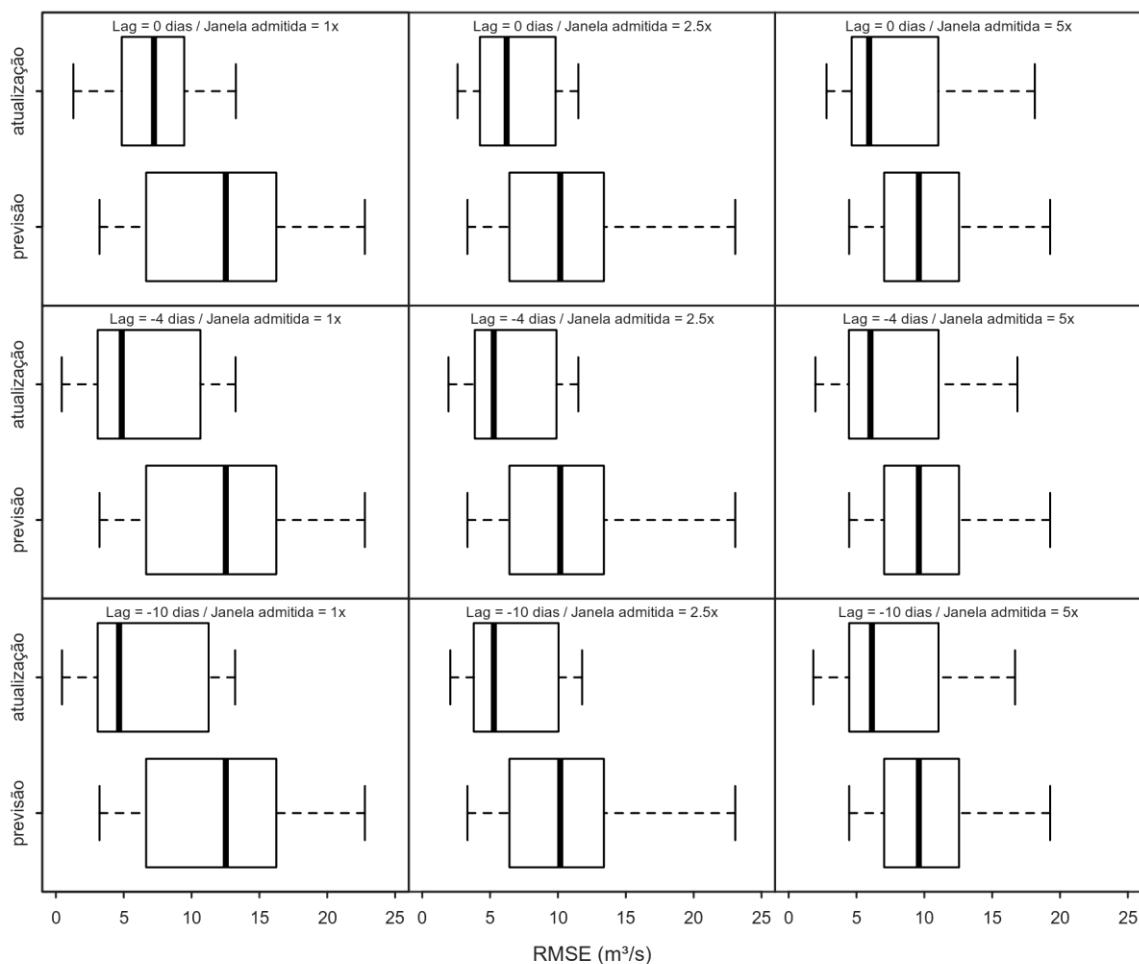
representado pela maior extensão temporal admitida nesse caso. No tocante aos *lags*, não se observou uma superioridade de quaisquer dos valores avaliados, à exceção da janela 1x, na qual se observa uma ligeira elevação da mediana dos erros.

As extensões médias aproximadas das janelas de dissimilaridades avaliadas são: (i) 20 dias (1x); (ii) 49 dias (2,5x); e (iii) 99 dias (5x).



**Figura 5.14** – Comparação do comportamento exibido pela raiz do erro médio quadrático, nos intraperíodos selecionados, quando admitido o procedimento para atualização das variáveis de estado baseado em percentuais.

Esses resultados estão em consonância com aqueles difundidos pelas técnicas de assimilação de dados. Essa observação continua válida quando considerado que a atualização se dá pela alteração da trajetória dos estados no instante alvo, obtida por meio da utilização de multiplicadores (para cada  $x$ ), estimados através do algoritmo DREAM. Na Figura 5.15 o comportamento do *RMSE* é descrito para o referido caso.



**Figura 5.15** – Comparação do comportamento exibido pela raiz do erro médio quadrático, nos intraperíodos seleccionados, quando admitido o procedimento para atualização das variáveis de estado baseado em multiplicadores.

Os resultados obtidos seguem a mesma tendência do procedimento anterior, destacando-se que houve uma pequena redução do 3º quartil no caso da janela 5x.

A respeito desses resultados, é necessário destacar que o procedimento que se vale de multiplicadores para estimativa das variáveis de estado no instante alvo, apresentou *RMSE's* muito similares aos do procedimento que atualiza  $x$  por meio da estimativa de percentuais dos limites impostos pelos parâmetros. A princípio, era de se esperar uma superioridade do procedimento baseado nos percentuais, uma vez que sua estimação não é condicionada pelos valores de  $x$  em  $t_{ALVO}-1$  e pelo limite imposto aos multiplicadores (0,5 a 1,5).

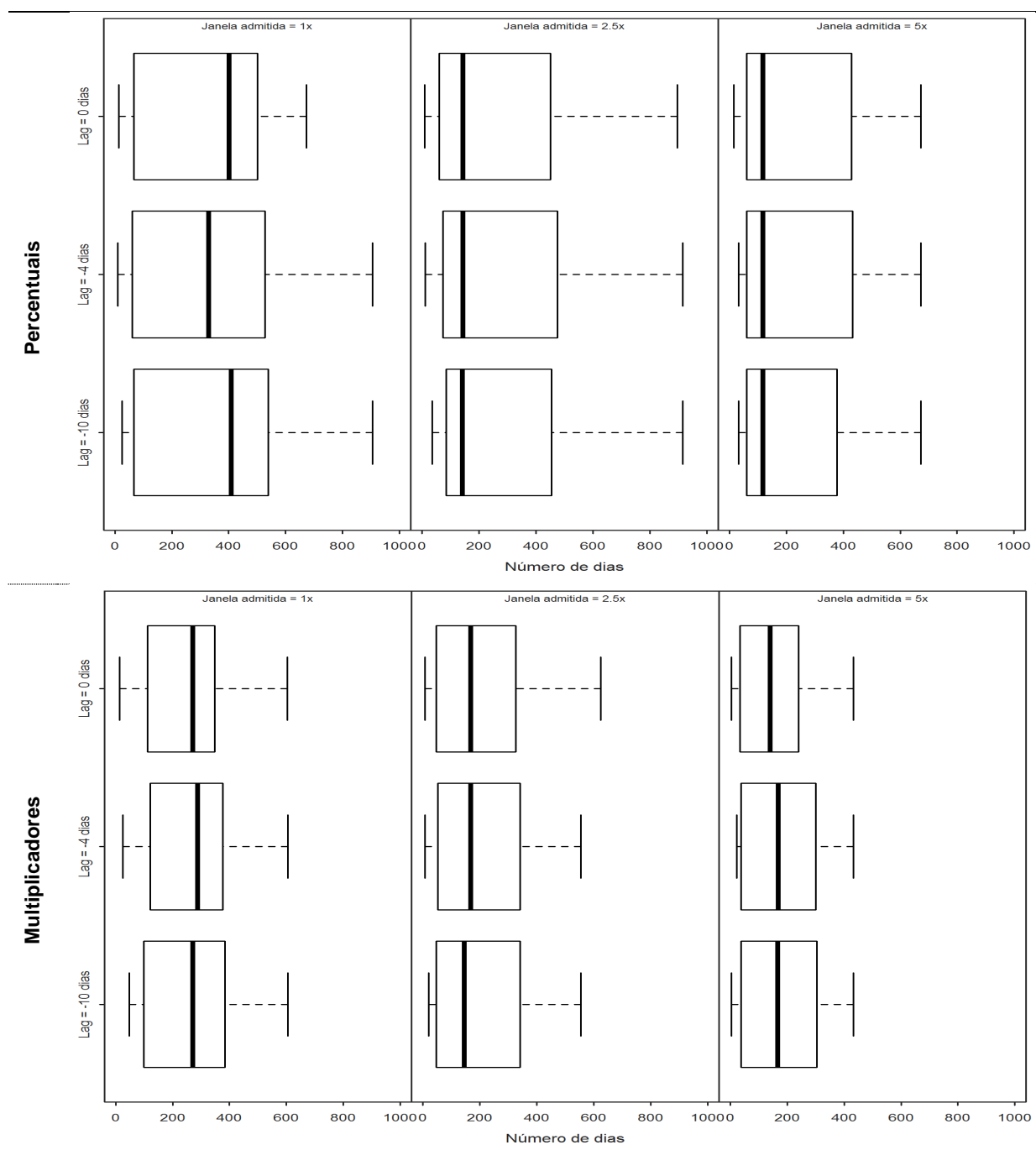
Esse fato ressalta o quão complexo é o processo de identificação de parâmetros e variáveis de estado. Sabe-se, considerando-se os resultados descritos na literatura e ratificados pelos

obtidos nesta tese, que diferentes conjuntos de parâmetros  $\theta$  podem produzir resultados (vazões) muito similares. Presume-se agora que a incerteza na evolução dos estados  $x$  é capaz de produzir os mesmos efeitos.

Nesse contexto, torna-se relevante avaliar a qualidade do ajuste sob o prisma da persistência temporal das alterações. Ela foi definida como o instante de tempo subsequente à atualização no qual a razão entre vazões simuladas pelo modelo, previstas e atualizadas, está contida no intervalo 0,95 – 1,05, durante 10 dias consecutivos. Essa diferença inferior a 5% durante um período especificado é um indicador de que as alterações nos estados  $x$  já não influenciam as vazões produzidas pelo modelo.

Há duas questões imediatas: (i) qual o tempo de persistência da alteração; e (ii) se a atualização produziu uma melhora ou piora da similaridade nesse período. A Figura 5.16 esclarece a primeira indagação.

A mediana da persistência temporal da alteração é superior a todas as janelas admitidas. Nota-se que essa persistência é maior quando considera-se na estimação dos percentuais/multiplicadores a janela igual à do período de dissimilaridade (1x). Não há diferenças expressivas entre as janelas 2,5x e 5x. A diferença dos *lags* parece, também, não influenciar nos percentis obtidos. Há um ligeiro aumento da persistência temporal quando emprega-se o procedimento baseado nos multiplicadores, o que, a princípio, parece contraditório, uma vez que se supõe que as alterações nos estados sejam suavizadas, ou menos abruptas, nesse caso.



**Figura 5.16 – Persistência da alteração pontual da variável de estado no tempo.**

Esses resultados revelam que uma alteração pontual na dinâmica dos estados possui um longo efeito sobre as vazões obtidas. Observa-se que a menor mediana calculada foi de aproximadamente 120 dias (4 meses). Assim sendo, quaisquer ações que impliquem em perturbações nas trajetórias do estados, incluindo, por consequência, alterações nos *inputs*, devem ser analisadas à luz desses resultados. Ressalta-se, ainda, que alterações sequenciais em períodos relativamente curtos serão, necessariamente, de difícil interpretação.

A análise do desempenho dos procedimentos durante todo o período no qual a influência da alteração é percebida permite uma avaliação mais precisa do potencial de “correção” dos vieses identificados. A fim de permitir uma leitura da evolução temporal desse ajuste, empregou-se o seguinte procedimento:

- calculou-se a raiz do erro médio quadrático para as vazões previstas e atualizadas em janelas subsequentes de 10 dias;
- identificou-se o maior valor dentre os calculados e determinou-se a razão entre eles, admitindo-se que o numerador era o maior valor;
- a essa razão foi subtraída a unidade, ou seja,  $[(RMSE_{10 \text{ dias}})_{\text{Maior}} / (RMSE_{10 \text{ dias}})_{\text{Menor}}] - 1$

Agrupou-se esse índice em dois grupos distintos. Se o *RMSE* das vazões atualizadas tenha sido inferior ao das previstas, uma redução do erro foi verificada e, por conseguinte, houve uma melhora (1º grupo) na performance do modelo. Caso contrário, o procedimento de atualização além de ineficaz, reduz a similaridade entre as vazões observadas e modeladas (piora – 2º grupo). O índice calculado permite uma leitura direta de quão superior (inferior) foi a performance do modelo atualizado, em uma janela de 10 dias. A Figura 5.17 sintetiza essa análise para o evento 04, para um *lag* e janelas iguais a 0 dias e 1x, respectivamente. Os demais eventos, incluindo todas as variações de *lags* e janelas, estão apresentados no Anexo II.



**Figura 5.17** – Avaliação da qualidade do ajuste, em janelas de 10 dias, durante o período no qual a atualização das variáveis de estado impõe as vazões previstas e atualizadas diferenças superiores a 5%: evento 04, lag = 0 dias, janela admitida = 1x, procedimento de atualização baseado em multiplicadores.

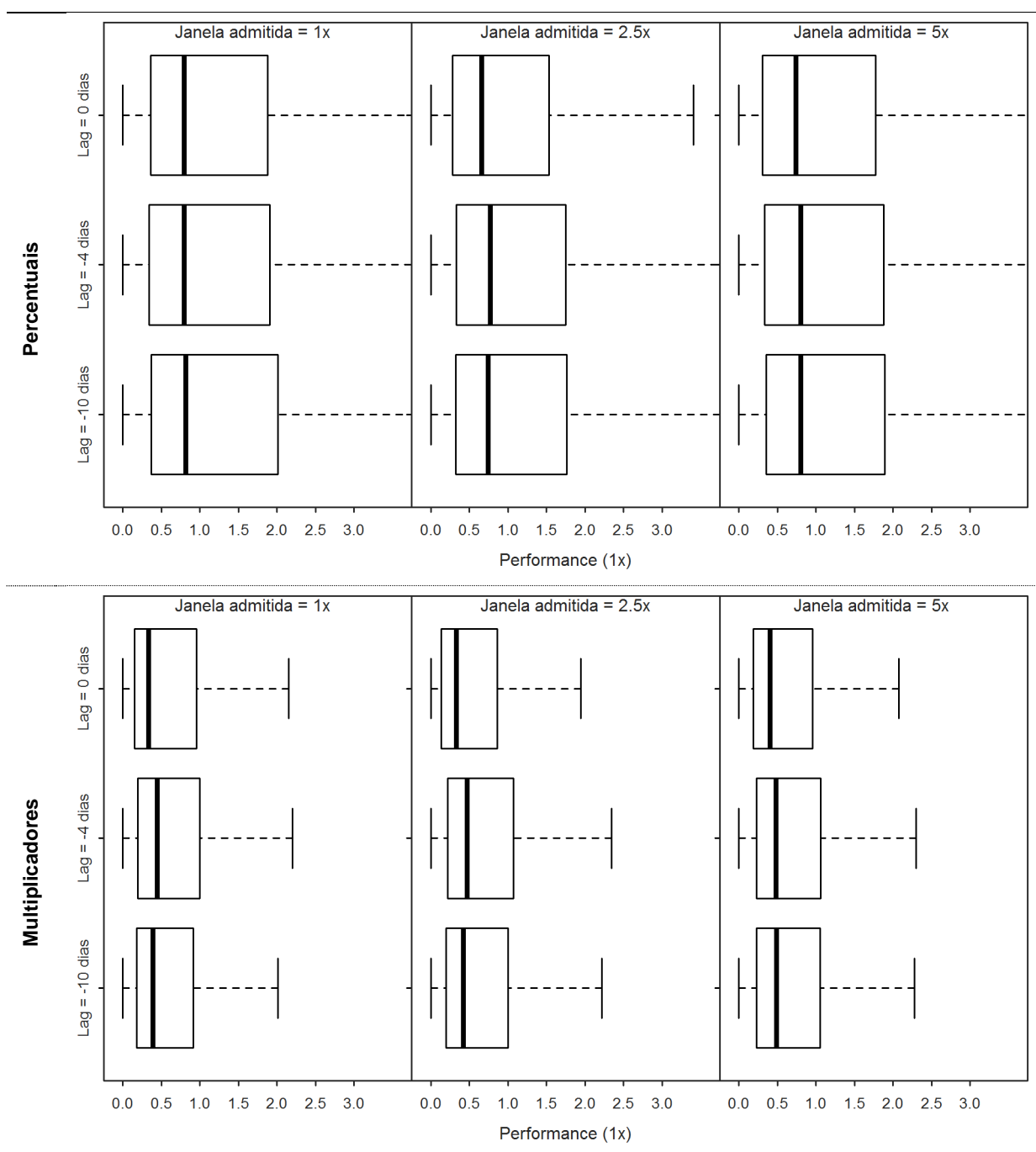
No gráfico da Figura 5.17, o início do eixo das abscissas refere-se ao instante alvo. Os percentuais referem-se a aumento (melhora)/redução (piora) do erro no período definido entre o alvo e aquele no qual a influência da alteração pontual nos estados produz diferenças inferiores a 5% entre as vazões previstas e atualizadas. O índice das ordenadas representa a alteração na performance do modelo. Tomando-se como exemplo um instante de tempo no qual o *RMSE* em um intervalo de 10 dias foi igual à 10 m<sup>3</sup>/s para as vazões previstas e 5 m<sup>3</sup>/s para as atualizadas. O índice calculado seria igual a unidade ( $10/5 - 1 = 1$ ). Isto indica que o procedimento utilizado para atualização das variáveis de estado proporcionou uma melhoria na performance igual à 1x a do modelo alternativo. A interpretação inversa (piora) é análoga.

Uma análise global desse índice, somente para os intervalos de tempo nos quais houve uma melhora da performance do modelo devido à atualização dos estados  $x$ , é apresentada na Figura 5.18. Nela, uma comparação da performance dos procedimentos de atualização é apresentada. Nota-se, conforme já mencionado, que não há uma superioridade de um procedimento, independente do *lag* ou janela considerada, uma vez que os índices de melhoria são muito similares. Esse fato merece destaque, já que a alteração no caso do procedimento que se baseia em multiplicadores é condicionada a um limite pré-estabelecido.

Uma comparação dos índices de performance no tempo é apresentada na Tabela 5.7.

**Tabela 5.7** – Percentuais de melhora da performance do modelo no tempo para períodos subsequentes à atualização dos estados  $x$ , no qual a diferença entre vazões previstas e atualizadas são superiores à 5%.

Lag (dias)	Janela admitida	Melhora no tempo (%)	
		Percentual	Multiplicador
0 dias	1x	59	52
0 dias	2,5x	60	54
0 dias	5x	65	68
-4 dias	1x	59	53
-4 dias	2,5x	59	57
-4 dias	5x	65	68
-10 dias	1x	60	54
-10 dias	2,5x	61	58
-10 dias	5x	67	67
Média		62	59



**Figura 5.18** – Comparação do comportamento exibido pela raiz do erro médio quadrático, calculado em janelas de 10 dias, para os períodos em que ocorreu uma melhora na performance do modelo devido ao emprego de um procedimento para atualização dos estados  $x$ .

É notório que os percentuais de melhora no tempo são muito similares em ambos os procedimentos de atualização. Nota-se uma maior variabilidade nos percentis para o caso do

procedimento baseado em um percentual. Não há diferenças significativas entre os *lags* e janelas considerados. Ademais, a taxa de piora no tempo (as quais são iguais ao complemento à 100 dos valores descritos na Tabela 5.7) variou entre 32 e 48% do tempo. Esses percentuais revelam que a possibilidade de se introduzir um novo viés aos resultados da simulação é significativamente elevada.

### Magnitude da alteração

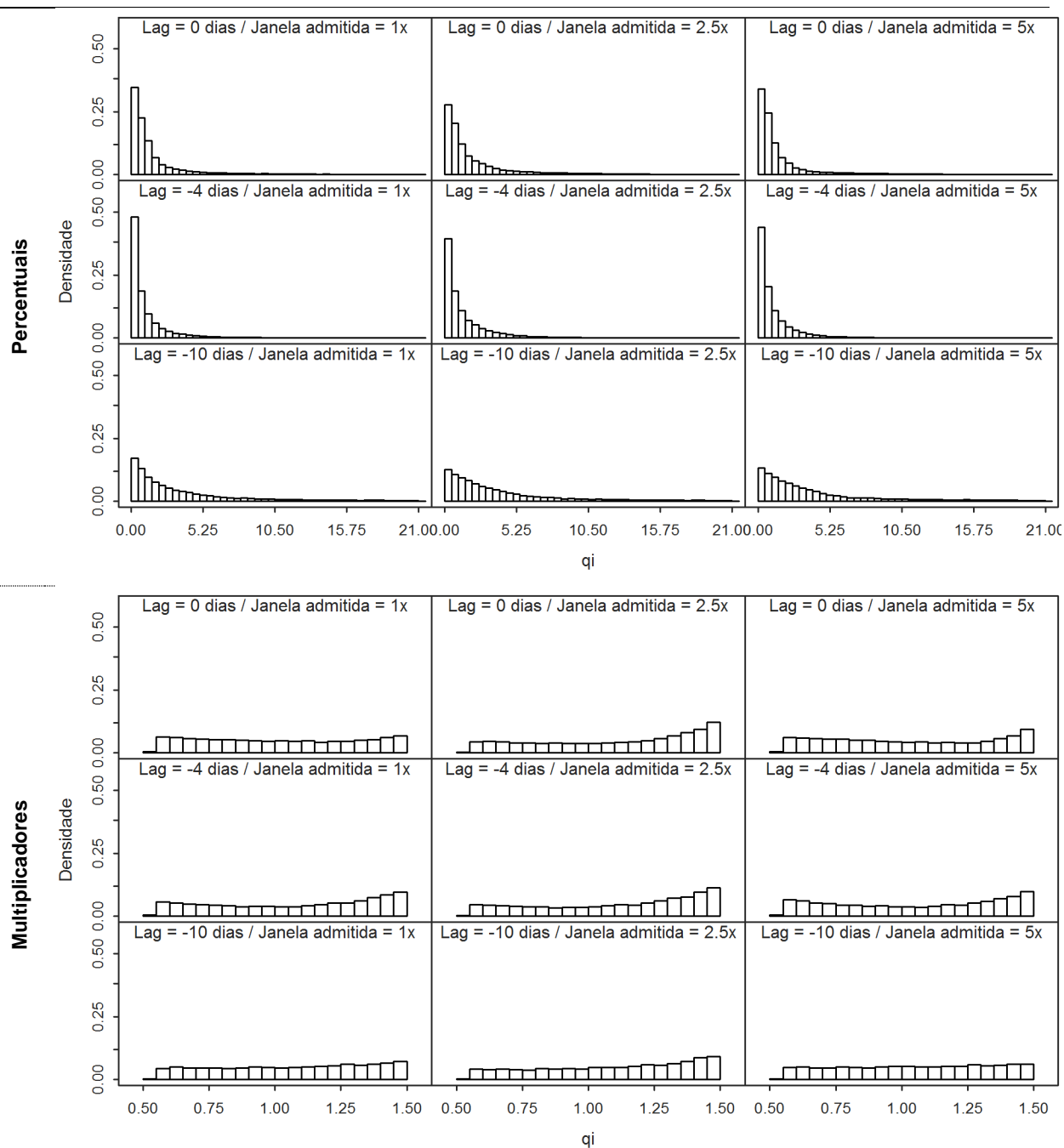
A fim de se possibilitar a comparação da magnitude da alteração imposta às variáveis de estados por ambos os procedimentos de atualização propostos, procedeu-se um reescalonamento dos resultados obtidos a partir do procedimento baseado nos percentuais dos limites dos parâmetros. Para esses, determinou-se a razão entre os valores da variável de estado estimada e aquele assumido por ela no instante anterior à atualização ( $t_{ALVO} - I$ ).

No caso do procedimento baseado em um multiplicador, a magnitude da alteração variou entre 0,5 e 1,5, uma vez que se restringiu a um limite de 50% do valor de  $x$  em  $t_{ALVO} - I$ . No caso da atualização baseada em percentuais, não há um limite superior para alteração. Deve-se observar que quanto mais díspar da unidade, maior a magnitude da alteração.

As Figuras 5.19 e 5.20 apresentam os resultados obtidos para as variáveis de estado que representam os fluxos à rede de drenagem. No Anexo I, essa análise global é apresentada para todas as variáveis de estado  $x$ . No Anexo II, os resultados obtidos individualmente, isto é, para cada evento e  $x$ , são apresentados.

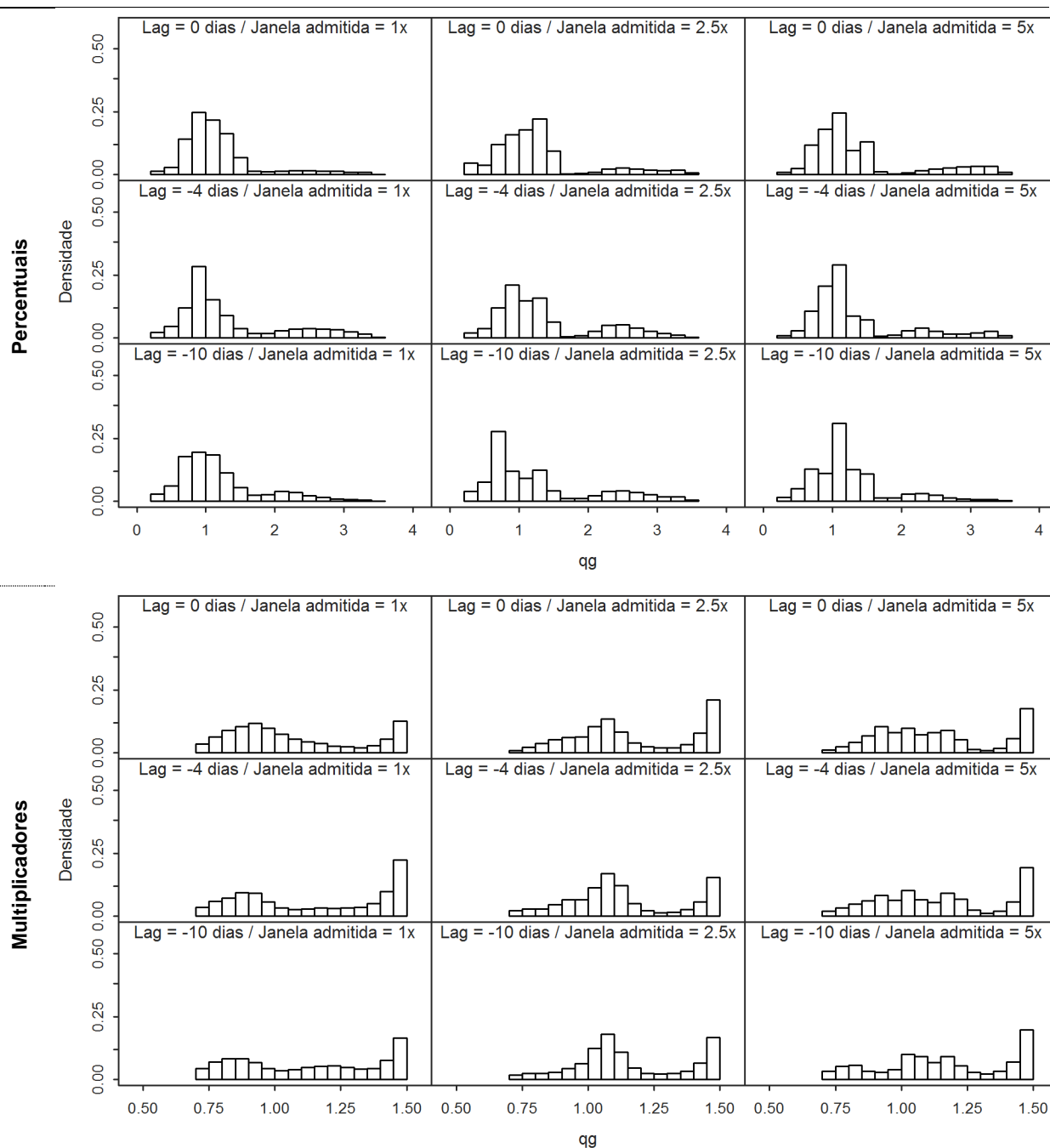
É importante avaliar essas figuras à luz dos resultados descritos previamente. Nota-se que a variação da magnitude da alteração é muita mais expressiva para o procedimento baseado em percentuais. Entretanto, como há um limite imposto à variação no caso do procedimento baseado em multiplicadores, pode-se questionar, em um primeiro momento, se os limites aplicados foram muito restritivos. Esse parece ser o caso do fluxo subsuperficial  $qi$ . Ele revela que, de fato, há uma limitação imposta pela construção do procedimento. Se por um lado é possível condicionar a alteração das trajetórias a limites, em princípio, plausíveis, por outro lado, caso os valores de  $x$  se aproximem muito dos extremos inferiores dos intervalos dos parâmetros e a diferença entre os limites (superiores e inferiores) seja significativa, a alteração será, necessariamente, pouco expressiva.





**Figura 5.18** – Magnitude da alteração dos afluxos subsuperficiais à rede de drenagem obtidos a partir dos procedimentos de atualização das variáveis de estado analisados.

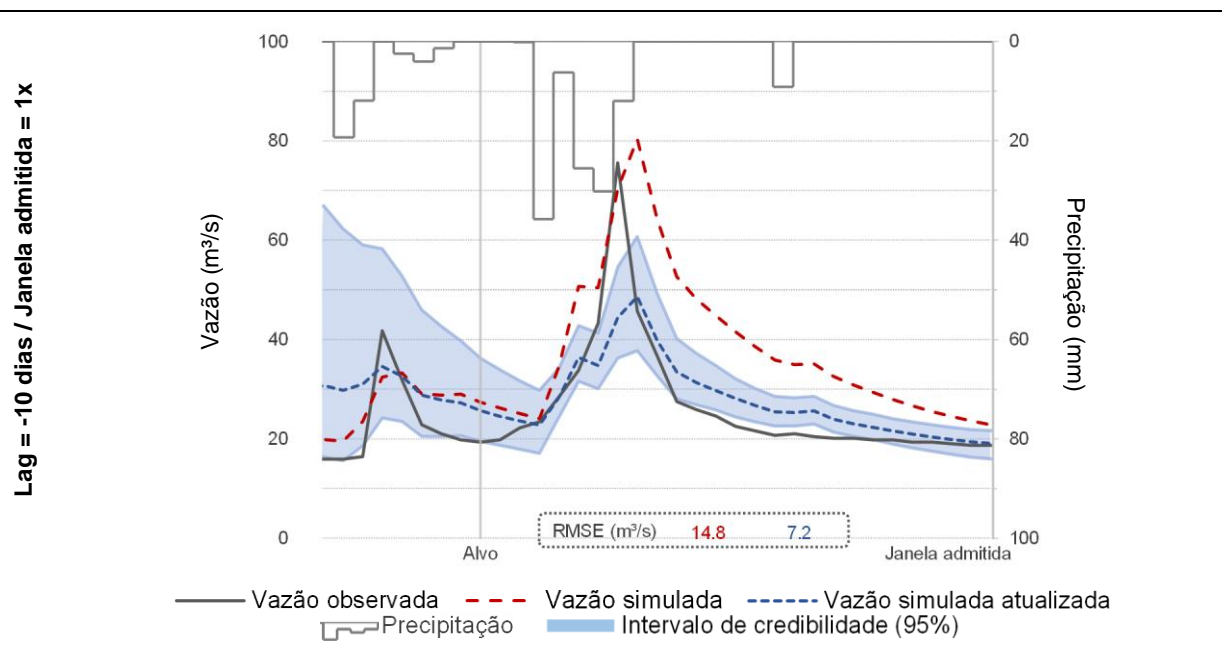
Há que se ponderar, no entanto, que a incerteza associada às estimativas é muito elevada, para ambas as variáveis, independentemente do procedimento utilizado para atualização dos estados  $x$ . Em contraposição a esse fato, a estatística empregada para aferição da qualidade do ajuste ( $RMSE$ ) apresentou magnitudes muito similares.



**Figura 5.19** – Magnitude da alteração dos afluxos subterrâneos à rede de drenagem obtidos a partir dos procedimentos de atualização das variáveis de estado analisadas.

Essa incerteza na estimação das variáveis de estado reforça o problema da equifinalidade (BEVEN e BINLEY, 1992 *apud* BEVEN, 2009): diferentes conjuntos de parâmetros e variáveis de estado podem produzir resultados (vazões) muito similares. Assim, quaisquer interpretações de ações que impliquem em perturbações nas trajetórias dos estados, incluindo alterações nos *inputs*, devem ser analisadas com cautela. Corroboram para essa afirmação a

forte influência da medida de verossimilhança descrita anteriormente, a qual é enfatizada na Figura 5.20



**Figura 5.20** – Evento 03: variação das vazões previstas e atualizadas por meio de um procedimento baseado em percentuais.

Nota-se na Figura 5.20 que o comportamento das variáveis de estado, estimadas a partir da admissão de uma medida de verossimilhança às vazões observadas, é capaz de impor uma expressiva variação ao intervalo de credibilidade das vazões simuladas (atualizadas), sobretudo nos instantes imediatamente posteriores à atualização. É notório, também, que a redução da raiz do erro médio quadrático ampliou a diferença em relação a vazão de pico observada na janela temporal admitida.

Esses resultados indicam que a medida de verossimilhança, a qual integra todas as incertezas no *output*, independentemente de sua natureza (dados observados, parâmetros e/ou estrutura do modelo), condiciona a trajetória dos estados  $x$  ao longo do tempo de simulação. Ademais, é necessário reconhecer que a variação contínua desses estados, em uma escala de tempo relativamente curta (dia), impõe novos desafios à avaliação da similaridade entre séries observadas e modeladas.

Conforme já mencionado, a predição da incerteza na modelagem hidrológica é uma das áreas de pesquisa mais ativas em hidrologia atualmente (MONTANARI, 2011). Esse tema se

sobrepõe parcialmente ao da assimilação de dados aplicada à modelagem chuva-vazão, quando analisa-se o problema sob a ótica da simulação e admite-se, *a priori*, o conhecimento pleno do *input* e do *output*. As análises desenvolvidas valeram-se desse cenário e demonstraram a complexidade na interação *input-state-output* de um modelo chuva-vazão conceitual. Diversos pesquisadores analisaram esses temas a partir de diferentes abordagens metodológicas. As principais metodologias empregadas, assim como as análises/ conclusões derivadas de suas aplicações, são discutidas no próximo item.

## **5.6 *Predição de incerteza aplicada à modelagem chuva-vazão: sinopse das abordagens atuais***

Uma resenha crítica das principais abordagens encontradas na literatura atual sobre o tema de predição de incertezas, no contexto da presente tese, é apresentada na sequência. Ela também incorpora aspectos relacionados ao tema assimilação de dados. O objetivo é ressaltar os principais aspectos metodológicos e conclusões descritos pelos autores.

### **5.6.1 *Abordagens que indicam as incertezas associadas aos dados de entrada/saída***

Essas são abordagens que reconhecem que a calibração e a predição, no âmbito da modelagem hidrológica, são afetadas pelas incertezas presentes nos dados de entrada, nas respostas do modelo, assim como na estrutura conceitual admitida para a transformação chuva-vazão. Elas refutam a cômoda suposição de que o fenômeno de transformação da chuva em vazão ocorra de modo determinístico e admitem na modelagem do fenômeno a estocasticidade dos fluxos e armazenamentos, devido à noção de que essas quantidades, embora variem no tempo e no espaço, são tomadas, em situações práticas, como médias temporais e espaciais.

Uma referência representativa dessas abordagens é o trabalho de Kuczera *et al.* (2006), no qual os autores formalizam a hipótese de que os erros estruturais de um modelo chuva-vazão conceitual podem ser caracterizados pela variação aleatória, mas dependente do evento pluvial, de um ou mais de seus parâmetros. Depois de identificarem os parâmetros mais sensíveis aos diferentes eventos pluviais, os autores empregaram um modelo hierárquico Bayesiano, o qual denominaram BATEA (*Bayesian Total Error Analysis*), para diferenciar explicitamente as três fontes de incertezas, nomeadamente, os erros de entrada, de resposta e do modelo. Na abordagem descrita por Kuczera *et al.* (2006) os próprios autores reconhecem as deficiências da incompleta, e talvez simplista, caracterização da estocasticidade das

incertezas da modelagem chuva-vazão e das dificuldades de se identificarem inequivocamente as escalas temporais nas quais se dá a referida estocasticidade.

Conforme Beven (2012), além de apresentarem parcial ou totalmente as deficiências já mencionadas, as abordagens que particularizam a variação de um ou mais parâmetros de um modelo a um certo episódio de chuva, por meio de condicionantes ou pela aplicação dos chamados “multiplicadores de chuva”, parecem determinar fortemente os resultados, sem a devida justificativa conceitual. Com efeito, no caso dos multiplicadores de chuva, se (i) a vazão simulada supera a vazão observada, então há um excesso de chuva que se corrige mediante a aplicação de um multiplicador inferior a 1; (ii) ou superior a 1, em caso contrário. Embora tal ação possa produzir índices de desempenho melhores, ela não parece ter sustentação conceitual, pois, afinal, pressupõe-se (i) que um modelo de simulação hidrológica represente uma sùmula matemática que apresente analogia e realismo físico verossímeis com a transformação natural chuva-vazão e (ii) que os erros decorrentes da inadequada representação da chuva sobre a bacia não devem ser compensados por alterações fortuitas nos valores dos parâmetros que produzem a referida sùmula.

Em trabalho posterior, Kuczera *et al.* (2010) empregaram o algoritmo BATEA para calibrar um modelo chuva-vazão conceitual, a partir da constatação de que os erros presentes na representação da chuva sobre a bacia, a partir dos dados pluviométricos, coexistem e, muitas vezes, superam as incertezas provenientes da estrutura conceitual do modelo. Depois de identificarem o parâmetro mais sensível do modelo utilizado, os autores realizaram o experimento de descrever sua variação por meio de uma distribuição lognormal, em separado e, depois, em conjunto, com a aplicação de multiplicadores para os eventos de chuva. Os autores analisaram uma anomalia observada no balanço hídrico do rio Vesonne (França), não lhes sendo possível apontar a verdadeira causa da referida anomalia, com base apenas nos dados que dispunham, sem a descrição prévia das incertezas neles contidas. Os autores advertem que as conclusões de qualquer quadro lógico de inferência estatística estão forçosamente condicionadas às informações disponíveis. Consoante o paradigma Bayesiano, as informações são fornecidas pelos dados, que determinam a função de verossimilhança, pelo conhecimento *a priori* sobre as incertezas intrínsecas aos dados e aos parâmetros, e, finalmente, pelas incertezas pertinentes à estrutura do modelo. Kuczera *et al.* (2010) finalizam alertando que sem o escrutínio completo das referidas fontes de incertezas, a discussão perde a sustentação científica.

Embora Kuczera *et al.* (2010) tenham estabelecido sólidos argumentos para uma discussão profícua sobre os temas de predição das incertezas e assimilação de dados, no âmbito da modelagem hidrológica, do ponto de vista prático, as alternativas por eles empregadas parecem conter muitas das deficiências encontradas no trabalho anterior de Kuczera *et al.* (2006). Primeiramente, o emprego de multiplicadores de chuva tem o efeito de condicionar fortemente os resultados. Parece fazer sentido permitir variações sazonais em um ou mais parâmetros, mas permitir variações a cada novo evento de chuva não parece ter a devida justificativa conceitual. Em segundo lugar, percebe-se, em Kuczera *et al.* (2010), que a justificativa para a variação do parâmetro mais sensível do modelo por eles empregado é o comportamento da distribuição *a posteriori*, depois de aplicados os multiplicadores de chuva. Diante disso, é legítimo questionar se o referido comportamento da distribuição *a posteriori* não teria sido condicionado pela aplicação dos multiplicadores de chuva.

### **5.6.2 Abordagens que preconizam a estimação conjunta de parâmetros e variáveis de estado**

Em outra vertente das pesquisas recentes sobre as incertezas decorrentes da modelagem chuva-vazão, encontram-se os trabalhos que preconizam a estimação conjunta (ou simultânea) de parâmetros e variáveis de estado do modelo. Um trabalho representativo dessa vertente é o descrito em Vrugt *et al.* (2005), no qual os autores apresentam o método SODA (*Simultaneous Optimization and Data Assimilation*), que combina as virtudes de eficiência de um robusto algoritmo de otimização de parâmetros, nomeadamente, o SCEM-UA (*Shuffled Complex Evolution Metropolis*), descrito em Vrugt *et al.* (2003), com o filtro de Kalman por conjuntos, ou EKF (*Ensemble Kalman filter*), tal como em Evensen (1994), para prover, segundo os autores, um tratamento mais adequado das incertezas presentes no *input*, no *output*, nos parâmetros e na estrutura do modelo de simulação empregado.

Vrugt *et al.* (2005) observam, no entanto, que os parâmetros identificados pelo método SODA só fazem sentido se combinados com uma técnica simultânea de assimilação sequencial de dados, ou SDA, do termo em língua inglesa *Sequential Data Assimilation*, uma vez que o conjunto de parâmetros não mais é visto como único e capaz de gerar as melhores predições, sem o concomitante ajustamento das variáveis de estado. Os autores apregoam que essa concepção permite a obtenção (i) de intervalos de incertezas mais significativos para as predições do modelo; e (ii) de séries temporais de ajustamentos recursivos dos estados e do *output*, as quais têm potencial valor para o aperfeiçoamento dos modelos chuva-vazão e para

o melhor entendimento do *modus operandi* dos sistemas hidrológicos. Embora empregue outros algoritmos, destinados principalmente a relaxar as premissas de erros Gaussianos do filtro de Kalman, essa parece ser também a filosofia subjacente a outros trabalhos recentes de Vrugt e colegas, como o descrito em Vrugt *et al.* (2013).

Para além das questões associadas à maior ou menor eficiência relativa entre algoritmos e implementações computacionais, cabem certas objeções a serem postas no âmbito da aplicação das abordagens que preconizam a estimação simultânea de parâmetros e variáveis de estado. A primeira refere-se à simultaneidade das estimações, realizadas por meio de *loops* independentes, e que potencialmente pode resultar em uma desconexão entre as posições instantâneas dos reservatórios fictícios que compõem o modelo e suas capacidades nominais, dadas por certos parâmetros, em favor, de uma descrição mais robusta das incertezas globais. Observa-se que, independentemente da estimação ser conjunta ou não, o *desideratum* é que os resíduos sejam suficientemente informativos, o que, *a priori*, sabe-se que não ocorrerá, pois haverá sempre erros intrínsecos aos dados empíricos utilizados.

Além da já mencionada desconexão entre estimações simultâneas, há que se reconhecer também (i) a questão da equifinalidade imposta por diferentes modelos e conjuntos de parâmetros, e (ii) a dificuldade, ou mesmo a impossibilidade, de uma noção quantitativa objetiva *a priori* das diferentes incertezas intrínsecas à modelagem hidrológica. Em resposta, deve-se procurar contornar (ou abordar) essas questões em um quadro lógico de inferência que permita integrar e descrever todas as fontes de incerteza. Este é certamente o paradigma Bayesiano, no qual é possível incluir as noções subjetivas apriorísticas do analista, advindas de seu conhecimento e experiência de utilização de um ou mais modelos, e que podem ser confirmadas e/ou atualizadas durante o processo de inferência. Tome-se o exemplo da definição do limite superior da variância para estimação da incerteza via função de verossimilhança Normal. Conforme apresentado, na implementação do método aqui proposto, com base no paradigma de Bayes, o referido limite superior foi arbitrado aprioristicamente como uma certa fração da vazão máxima observada, o que se confirmou adequado nas etapas posteriores da inferência.

### **5.6.3 Abordagens que apregoam o desenvolvimento de algoritmos cada vez mais eficientes/robustos**

Desde sua introdução por Vrugt *et al.* (2009), o método DREAM, descrito em detalhes no subitem 4.3.1, tem sido uma referência para os estudos de análise das incertezas da modelagem hidrológica, em decorrência de se apoiar no paradigma Bayesiano de inferência e em comprovados e robustos métodos de otimização, e de simulação de cadeias de Markov. O método tem sido desdobrado em novas versões, tais como em Vrugt e ter Braak (2011) e Laloy e Vrugt (2012), que buscam maior capacidade de processamento, maior eficiência computacional e aplicações multi-critérios e multi-modelos, embora os resultados nem sempre indiquem essa superioridade de modo patente.

Apesar de ser inegável a importância da busca por procedimentos mais eficientes e robustos, em diversos domínios da ciência e, particularmente, nos campos da estatística e das técnicas de otimização, ela não parece tão óbvia no ponto atual de desenvolvimento dos métodos de análise de incertezas da modelagem hidrológica. Dada a incrível e crescente capacidade de processamento dos microcomputadores atuais, esta não parece constituir um fator determinante ou limitador dos estudos dessa natureza. Parece mais urgente ou prioritário reconhecer e procurar contornar (ou resolver) as limitações devido (i) à insuficiência espaço-temporal e à qualidade duvidosa dos dados hidrológicos; (ii) à incompleta, ou mesmo simplista, representação dos fenômenos hidrológicos por modelos chuva-vazão; e (iii) ao correto tratamento estatístico a ser dado às incertezas epistêmicas e aleatórias, implícitas na ação de síntese da variação de fenômenos naturais.

No contexto exposto, a busca pelo algoritmo “perfeito” parece retirar o foco sobre essas importantes questões e postergar eventuais avanços que sua discussão poderia suscitar, com o agravante de poder trazer alguma confusão aos usuários de modelos hidrológicos menos experientes ou menos afeitos a elaborações de cunho mais acadêmico. Neste particular, é notável o embate filosófico entre as correntes de pensamento lideradas pelos cientistas Jasper Vrugt e Keith Beven, materializadas na última década por diversos artigos científicos de suas respectivas autorias e co-autorias, e resumidas, embora com indisfarçada parcialidade, no capítulo 7 do livro *Rainfall-Runoff Modelling – The Primer*, de Beven (2012).



#### 5.6.4 Abordagem “propositiva”

Em resumo, ao longo de seu desenvolvimento histórico, o campo de estudos da modelagem hidrológica enfocou os seguintes temas: (i) o desenvolvimento de funções-objetivo particularizadas para sumariar os erros entre observações e predições; (ii) a busca por algoritmos de otimização eficientes para resolver o problema de calibração dos parâmetros de um modelo; (iii) a determinação do volume apropriado de dados e seus tipos mais informativos; (iv) a seleção do algoritmo numérico mais apropriado para a solução das equações diferenciais e algébricas próprias dos modelos hidrológicos; (v) a representação das incertezas; e (vi) o desenvolvimento de métodos para inferir e refinar a estrutura matemática e processar as equações dos modelos hidrológicos (Vrugt e Sadegh, 2013).

Conforme observam Vrugt e Sadegh (2013), essa sequência histórica de desenvolvimento das pesquisas tem inspiração e apoio em princípios oriundos da abordagem clássica para a agregação dos erros residuais, entre modelos e dados empíricos, introduzida no século XIX pelos matemáticos Adrian Marie Legendre e Carl Friedrich Gauss, que, mais tarde, passou a ser referida como técnica de minimização da soma dos erros quadráticos ou, simplesmente, por técnica dos mínimos quadrados. Em 1822, Gauss demonstrou que essa é válida apenas se os resíduos não são correlacionados entre si, têm média nula e variância constante.

Muitas das referências citadas nesta resenha crítica formalizaram críticas à aludida abordagem clássica por sua incapacidade de lidar com erros inerentes à estrutura do modelo e aos dados de entrada (precipitação e evapotranspiração), os quais são admitidos como “desprezíveis” ou que são “absorvidos” pelos erros residuais. Além disso, diagnósticos *a posteriori* demonstram que os erros residuais exibem substancial variação no viés, na variância e na estrutura de correlação, em diferentes partes (ou componentes) da resposta calculada pelo modelo hidrológico, o que se deve parcialmente aos erros associados à sua estrutura e aos dados de entrada, em geral, maiores do que os associados aos dados de saída ou de calibração.

Como se vê nesta resenha crítica, muitas abordagens adotam o paradigma Bayesiano de inferência e relaxam a premissa de existência de um único conjunto ótimo de parâmetros, em favor de uma distribuição *a posteriori* que reconhece o papel das incertezas devido à estrutura do modelo, aos dados de entrada, aos dados de saída e à calibração de seus parâmetros. A Equação 3.26, que resume o modo de inferência Bayesiano, pode ser posta na forma de proporção, caso se abstraia a constante de normalização de seu denominador, dando origem a

uma função densidade a *posteriori* não padronizada, a qual é o ponto de partida do método proposto por Vrugt e Sadegh (2013). Esses autores observam que na estimação da densidade não padronizada, a função de verossimilhança, vista como o resumo das diferenças entre as observações e as simulações pelo modelo, é, de fato, o ponto chave e objeto de intensos debates na literatura da especialidade. Apontam ainda que a própria construção da função de verossimilhança, em si, como um resumo das diferenças entre observações e simulações, dilui e mistura a informação disponível em um índice único que torna difícil a discriminação entre comportamentos específicos do sistema.

Como alternativa, Vrugt e Sadegh (2013) propõem o emprego do método ABC (*Approximate Bayesian Computation*), que permite a inferência Bayesiana sem a necessidade de prescrição formal de uma função de verossimilhança. Para sua utilização com modelos hidrológicos complexos, os autores sugerem o uso conjunto de diferentes métricas, às quais referem como *hydrologic signatures* dada sua inspiração na teoria de sistemas hidrológicos (Beven, 2012), que podem ser livremente escolhidas pelo usuário do ABC, tomando-se o cuidado de garantir a convergência da distribuição a *posteriori*. Vrugt e Sadegh (2013) apontam que o método ABC é particularmente útil para o diagnóstico da calibração e avaliação de um modelo hidrológico específico. Essa proposta parece fazer convergir as distintas correntes de pensamento lideradas pelos cientistas Jasper Vrugt e Keith Beven, já mencionadas anteriormente.

## **5.7 Discussão geral**

Os resultados descritos nesta tese indicaram que a alteração das variáveis de estado permitem “ajustes” expressivos na similaridade entre as séries observadas e modeladas. Essas alterações poderiam ter sido obtidas, a exemplo de outras abordagens, por meio da perturbação do *input* (ex.: emprego de multiplicadores a eventos de chuva). Espera-se, por consequência, que alterações conjuntas dos parâmetros permitam, também, o alcance de índices de ajustes significativos.

Nota-se, por meio dos resultados apresentados, que há um equilíbrio tênue entre incertezas nos dados observados e na estrutura conceitual admitida, e uma medida de verossimilhança adequada. As inerentes diferenças de escala temporal presentes na estimativa de variáveis de estado e parâmetros impõem desafios à identificação dessas quantidades. Sua avaliação, entretanto, parece encoberta pela busca de estatísticas que denotem a validade e robustez das

abordagens metodológicas empregadas. Todas elas apoiam-se na premissa de que os resíduos do modelo são suficientemente informativos, apesar de argumentarem que é necessário reconhecer as incertezas nos dados observados utilizados.

É notório, ao se analisar a literatura sobre o tema produzido na última década, que a discussão sobre o algoritmo, e das premissas por ele admitidas, sobrepôs-se a uma análise crítica dos dados, da abordagem metodológica e, por vezes, dos resultados obtidos. A incorporação sequencial de dados à análise, tal qual preconizado por algumas das abordagens metodológicas descritas, parece contrastar com a ideia sugerida pela assimilação de dados. Conforme implementada, ela parece ser capaz de condicionar os resultados, sem, no entanto, agregar novos elementos que permitam lidar com a complexidade da representação espaço-temporal das variáveis hidrológicas.

As análises desenvolvidas nesta tese partiram da hipótese, verificada posteriormente, que a assimilação do comportamento dos estados  $x$  de um modelo chuva-vazão conceitual pode ser empregada para identificação de períodos de dissimilaridade entre as séries de vazões observadas e modeladas, quando admitidas as incertezas na estimação dos parâmetros. Demonstrou-se, a exemplo de outros trabalhos, que diferentes conjuntos de parâmetros podem ser considerados aceitáveis. A equifinalidade foi incorporada aprioristicamente à análise na proposta metodológica aqui desenvolvida.

A inferência Bayesiana permite incluir à análise, *a priori*, a percepção do modelador acerca do sistema hidrológico e da estrutura conceitual admitida. Sob uma perspectiva prática, seu emprego permite acomodar, metodologicamente, uma expressão largamente empregada na prática da hidrologia: o “eu acho”. Sua adequação à modelagem chuva-vazão é notória.

A metodologia empregada nesta tese poderia ser expressa, de forma sumária, por meio do seguinte questionamento: a introdução de um viés permitiria melhorar a estimativa da vazão? Em sua natureza, encontra-se o reconhecimento explícito das limitações dos dados/modelos empregado. Ao fazê-lo, de antemão, acredita-se que se cria um cenário mais propício ao desenvolvimento de estruturas conceituais mais robustas (o que não quer dizer mais complexas), sob o ponto de vista da representação do fenômeno.

Por fim, destaca-se que cabe ao modelador reconhecer as limitações dos modelos chuva-vazão e, a partir de técnicas que permitam a estimativa da incerteza (tal qual a aqui aplicada), indicar

o resultado determinístico quando necessário. Esse resultado é, necessariamente, função da incerteza quantificada e dos riscos associados. A técnica (o algoritmo) é um instrumento para a tomada de decisão. Ele, independentemente de quão robusto seja, auxilia o modelador a tomá-la. Parece óbvio, mas essa não é a interpretação imediata ao consultar as publicações que representam o estado da arte na atualidade.

## 6 CONCLUSÕES E RECOMENDAÇÕES

Nesta tese foram avaliadas as incertezas associadas às estimativas do conjunto de parâmetros  $\theta$  de um modelo chuva-vazão conceitual, por meio do emprego de técnicas de inferência Bayesiana via simulações de Monte Carlo. Nos estudos desenvolvidos foram avaliados dois modelos probabilísticos para os resíduos. As hipóteses admitidas pelos mesmos foram analisadas e foi discutido o viés imposto às estimativas dos parâmetros.

A abordagem Bayesiana revelou-se adequada aos propósitos do estudo. No entanto, é necessário ressaltar o impacto da medida de verossimilhança sobre os resultados obtidos. O modelo de resíduos generalizado GL foi capaz de explicar o comportamento admitido *a priori* para a diferença entre as vazões observadas e simuladas.

A incerteza na predição da vazão imposta pela estimação dos parâmetros foi apresentada. O intervalo descrito foi capaz de representar as vazões observadas, apesar da subestimação de algumas vazões de pico. Esse fato enfatizou a necessidade de se avaliar a influência de outras fontes de incerteza sobre a vazão modelada. Nesse sentido, procedeu-se uma extensa avaliação do comportamento *input-state-output* do modelo chuva vazão considerado. Formularam-se dois procedimentos para assimilação do padrão de variabilidade das variáveis de estado do modelo adotado, a partir do emprego da técnica de inferência Bayesiana previamente utilizada. Essas análises permitiram a verificação da hipótese de que o comportamento (trajetória temporal) das variáveis de estado é capaz de caracterizar os erros na série modelada. Ademais, a partir da avaliação detalhada do impacto da atualização das variáveis de estado sobre períodos específicos da série, foi possível contrapor algumas das abordagens metodológicas e conclusões descritas nas principais referências sobre o tema na atualidade.

Os resultados descritos permitiriam, ainda, analisar se a magnitude da correção dos estados do modelo guarda alguma relação com os erros/ dados observados. Isso possibilita o possível desenvolvimento de um método para correção das trajetórias das variáveis de estado, baseado na estimação das incertezas dos parâmetros e total. Essa análise, entretanto, foi preterida em função dos objetivos delineados.

Independentemente do estágio de desenvolvimento, é fato que a identificação objetiva da incerteza específica no processo de modelagem é uma tarefa complexa. Em geral, essa análise

baseia-se na premissa de que os resíduos sejam suficientemente informativos. Há uma relação causal entre a ocorrência de uma cheia e um evento pluviométrico. Não é incomum, entretanto, verificar-se durante a modelagem períodos em que essa causalidade não seja respeitada.

É possível que todas as séries possuam alguns períodos não informativos. Caso haja um período inconsistente de dados sendo utilizado no processo de modelagem (calibração/validação), então a inferência não enviesada dos parâmetros só poderia ser alcançada por um equilíbrio fortuito entre a desinformação e uma medida de verossimilhança adequada (BEVEN, 2012). É certo, entretanto, que a proposta de se corrigir as variáveis de estado, incorporando as incertezas nas estimativas dos parâmetros e respeitando a inerente diferença de escala temporal dos processos representados por  $x$  e  $\theta$ , permite compensar, ao menos em parte, o viés imposto à estimativa das vazões.

O acréscimo de novas informações à análise, seja na perspectiva da assimilação de dados, seja por meio da observância de novas métricas, conforme sugerido pelo método ABC, permite vislumbrar novas abordagens ao problema da predição/redução da incerteza. Dentre esses, merece destaque a possibilidade de se avaliar o problema da escala temporal na estimação dos parâmetros e variáveis de estado. A abordagem Bayesiana permite incluir noções subjetivas apriorísticas do analista, advindas de seu conhecimento e experiência. Essas noções poderiam ser confirmadas e/ou atualizadas durante o processo de inferência.

Não obstante a complexidade do problema analisado, é crescente o consenso na comunidade técnica da necessidade de se avaliar as incertezas envolvidas nos prognósticos dos modelos hidrológicos. Esta tese orientou-se por essa percepção, sob a perspectiva de contribuir para o avanço dessa discussão, analisando um caso representativo da realidade brasileira.

## REFERÊNCIAS

- AJAMI, N., DUAN, Q., SOROOSHIAN, S. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, v. 43, 2007.
- BECKER, A. Criteria for a hydrologically sound structuring of large scale land surface process models. In: O'KANE, J. (ed.). *Advances in theoretical hydrology*. Elsevier, pp. 97-111, Amsterdam, NL, 1992.
- BERTHET, L., RAMOS, M., PERRIN, C., ANDRÉASSIAN, V., LOUMAGNE, C. Can discharge assimilation methods be used to improve flood forecasting when few data are available? *New Approaches to Hydrological Prediction in Data-sparse Regions (Symposium HS.2 at the Joint IAHS & IAH Convention)*, p. 94-100, 2009.
- BEVEN, K., CALVER, A., MORRIS, E. Testing a physically-based flood forecasting model (TOPMODEL) for three UK catchments. *Journal of Hydrology*, v. 69, p. 119-143, 1984.
- BEVEN, K., BINLEY, A. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, v. 6, n. 3, p. 279-298, 1992.
- BEVEN, K., SMITH, P., FREER, J. So just why would a modeller choose to be incoherent? *Journal of Hydrology*, v. 354, n. 1-4, p. 15-32, 2008.
- BEVEN, K. *Environmental modelling: An uncertain future?* Routledge, England, 2009.
- BEVEN, K. *Rainfall-runoff modeling: The primer*. John Wiley & Sons Ltd, 2 ed., 353 p., England, 2012.
- BOX G.E.P., TIAO, G.C. *Bayesian inference in statistical analysis*. John Wiley & Sons Ltd, 588 p., 1992.
- BOYLE, D., GUPTA, H. SOROOSHIAN, S. Towards improved calibration of hydrologic models: model calibration and uncertainty in prediction. *Hydrological Processes*, n. 6, p. 279-298, 2000.
- BROCHERO, D. *Hydroinformatics and diversity in hydrological ensemble prediction systems*. Tese (Doctorat en génie civil et génie des eaux). Université Laval, p. 170, 2013.
- BROOKS, S.P. Bayesian computation: a statistical revolution. *Philosophical Transactions of the Royal Society of London A*, vol. 361, n. 1813, p. 2681-2697, 2003.
- BULYGINA, N. *Model structure estimation and correction through data assimilation*. Tese (Hydrology and Water Resources) – University of Arizona, Tucson, Arizona, USA, 198 p., 2007.
- BULYGINA, N., GUPTA, H. Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. *Water Resources Research*, v. 45, n. 12, 20 p., 2009.

BURNASH, R.J.C., FERRAL, R.L., MCGUIRE, R.A. A generalized streamflow simulation system: conceptual modeling for digital computers. U.S. Department of Commerce, National Weather Service and State of California, Department of Water Resources, 1973.

CASSIE, D., POLLOCK, T., CUNJAK, R. Variation in stream chemistry an hydrograph separation in a small drainage basin. *Journal of Hydrology*, v. 178, n. 1-4, p. 137-157, 1996.

CHIB, S., GREENBERG, E. Understanding the Metropolis-Hasting Algorithm. *The American Statistician*, v. 49, pp. 327-335, 1995.

CHOI, H., BEVEN, K. Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. *Journal of Hydrology*, v. 332, ed. 3-4, p. 316-336, 2007.

CHOW, V., MAIDMENT, D., MAYS, L. *Applied hydrology*. McGraw-Hill, 572 p., New York, 1988.

CLARK, C. O. Storage and the unit hydrograph. *Transactions of the American Society of Civil Engineers*, v. 110, New York, 1943.

CLARK, M., SLATER, A., RUPP, D., WOODS, R., VRUGT, J., GUPTA, H., WAGENER, T., HAY, L. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, v. 44, ed. 12, 14 p., 2008.

DRÉCOURT, J. Data assimilation in hydrological modelling. Tese (Doutorado em Engenharia) – Technical University of Denmark, Denmark, 162 p., 2004.

DUAN, Q., SOROOSHIAN, S., GUPTA, V. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, v. 28, p. 1015-1031, 1992.

DUNNE, T., MOORE, T., TAYLOR, C. Recognition and Prediction of Runoff-Producing Zones in Humid Regions. *Hydrological Sciences Bulletin*, XX, n. 3/9, p. 305-327, 1975.

EHR/UFGM. Modelos de simulação hidrológica para previsão de vazões de curto prazo: Relatório técnico parcial 1. Projeto de pesquisa aplicada, 117 p., Belo Horizonte, 2001.

EVENSEN, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, v. 99, b. C5, p. 10,143-10,162, 1994.

EUSER, T., WINSEMIUS, H., HRACHOWITZ, M., FENICIA, F., UHLENBROOK, S., SAVENIJE, H. A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*, v. 17, ed. 5, p. 1893-1912, 2013.

EVIN, G., KAVETSKI, D., THYER, M., KUCZERA, G. Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. *Water Resources Research*, v. 49, p. 4518-4524, 2013.



EVIN, G., THYER, M., KAVETSKI, D., McINERNEY, D., KUCZERA, G. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research*, v. 50, p. 2350-2375, 2014.

FENICIA, F., KAVETSKI, D., SAVENIJE, H. Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, v. 47, ed. 11, 13 p., 2011.

FERNANDES, W. Método para estimação de quantis de enchentes extremas com o emprego conjunto de análise bayesiana, de informações não sistemáticas e de distribuições limitadas superiormente. Tese (Doutorado em Saneamento, Meio Ambiente e Recursos Hídricos) – Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte, 185 p., 2009.

FEYEN, L., VRUGT, J., NUALLÁIN, B., VAN DER KNIJFF, J., DE ROO, A. Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. *Journal of Hydrology*, v. 332, n. 3-4, p. 276-289, 2007.

GELMAN, A., ROBERTS, G., GILKS, W. Efficient Metropolis jumping rules. In: BERNARDO, J. BERGER, J., DAWID, A., SMITH, A. (eds). *Bayesian statistics 5*. Oxford University Press, USA, 1996.

GONG, W., GUPTA, H., YANG, D., SRICHARAN, K., HERO, A.O. Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. *Water Resources Research*, vol. 49, p. 2253-2273, 2013.

GUILLAUME, J., ANDREWS, F. Dream: DiffeRential Evolution Adaptive Metropolis. R package version 0.4-2, 2012. Disponível em: <http://CRAN.R-project.org/package=dream>. Acesso em: 22 de janeiro de 2014.

GUPTA, H., SOROOSHIAN, S., YAPO, Y. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, v. 34, p. 751-763, 1998.

GUPTA, H., SOROOSHIAN, S., HOGUE, T., BOYLE, D. Advances in Automatic Calibration of Watershed Models. In: DUAN, Q., GUPTA, H., SOROOSHIAN, S., ROUSSEAU, A., TURCOTTE, R. (eds). *Calibration of Watershed Models*. Water Science and Application 6, American Geophysical Union, Washington, USA, 2003.

GUPTA, H., BEVEN, K., WAGENER, T. Model calibration and uncertainty estimation. In: ANDERSON, M., MCDONNELL, J. (eds). *Encyclopedia of Hydrologic Sciences*. Wiley, Chichester, UK, 2005.

HAARIO, H., SAKSMAN, E., TAMMINEN, J. An adaptive Metropolis algorithm. *Bernoulli*, v. 7, ed. 2, p. 223-242, 2001.

HAARIO, H., LAINE, M., MIRA, A., SAKSMAN, E. DRAM: Efficient adaptive MCMC. *Statistics and Computing*, v. 16, ed. 4, p. 339-354, 2006.

HASTINGS, W. Monte-Carlo sampling methods using Markov Chains and their applications. *Biometrika*, v. 57, p. 97–109, 1970.

HE, M., HOGUE, T., FRANZ, K., MARGULIS, S., VRUGT, J. Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes. *Advances in Water Resources*, vol. 34, p. 114-127, 2010.

HYDROLOGIC ENGINEERING CENTER, Hydrologic Engineering Methods for Water Resources Development: hydrograph analysis. U. S. Army Corps of Engineers, v. 4, Davis, 1973.

HYDROLOGIC ENGINEERING CENTER. HEC-1 Flood Hydrograph Package – Reference Manual. U. S. Army Corps of Engineers, Davis, 1981.

HOETING, J., MADIGAN, D., RAFTERY, A., VOLINSKY, C. Bayesian model averaging: A tutorial. *Statistical Science*, v. 14, p. 382-401, 1999.

HOUSER, P., LANNOY, G., WALKER, J. Hydrologic Data Assimilation. In: TIEFENBACHER, J. Approaches to Managing Disaster: Assessing Hazards, Emergencies and Disaster Impacts. InTech, p. 41-64, 2012.

JONES, J.P., SUDICKY, E.A., MCLAREN, R.G. Application of a fully-integrated surface-subsurface flow model at the watershed-scale: a case study. *Water Resources Research*, vol. 44, ed. 3, 2008.

KALNAY, E. Ensemble Kalman Filter: Current status and potential. In: LAHOZ, W., KHATTATOV, B., MENARD, R. (eds). *Data Assimilation: Making Sense of Observations*. Springer, p. 69-92, 2010.

KAVETSKI, D., FRANKS, S., KUCZERA, G. Confronting input uncertainty in environmental modelling. In: DUAN, Q., GUPTA, H., SOROOSHIAN, S., ROUSSEAU, A., TURCOTTE, R. (eds). *Calibration of Watershed Models*. Water Science and Application 6, American Geophysical Union, Washington, USA, 2003.

KAVETSKI, D., KUCZERA, G., FRANKS, S. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, v. 42, 2006a.

KAVETSKI, D., KUCZERA, G., FRANKS, S. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resources Research*, v. 42, 2006b.

KAVETSKI, D., FENICIA, F. Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resources Research*, v. 47, ed. 11, 20 p., 2011.

KRZYSZTOFOWICZ, R. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, v. 35, p. 2739-2750, 1999.

KRZYSZTOFOWICZ, R. The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, v. 249, p. 2-9, 2001.

- KRZYSZTOFOWICZ, R. Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, v. 268, p. 16-40, 2002.
- KUCZERA, G. Improved parameter inference in catchment models. Evaluating parameter uncertainty. *Water Resources Research*, v. 19, n. 5, p. 1151-1162, 1983.
- KUROWICKA, D., COOKE, R. Uncertainty analysis with High Dimensional Dependence Modelling. John Wiley & Sons, 284 p., 2006.
- LAHOZ, W., KHATTATOV, B., MENARD, R. (eds). *Data Assimilation: Making Sense of Observations*. Springer, 732 p., 2010.
- LALOY, E., VRUGT, J. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing. *Water Resources Research*, v. 48, 18 p., 2012.
- LAUDON, H., SLAYMAKER, O. Hydrograph separation using stable isotopes, silica and electric conductivity: an alpine example. *Journal of Hydrology*, v. 201, n. 1-4, p. 82-101, 1997.
- LING, W. Data requirement for model structure identification. Dissertação (MSc thesis in Department of Watermanagement). Delft University of Technology, 63 p., 2013.
- LIANG, Z., WANG, D., GUO, Y., ZHANG, Y., DAI, R. Application of Bayesian Model Averaging Approach to Multimodel Ensemble Hydrologic Forecasting. *Journal of Hydrologic Engineering*, v. 18, ed. 11, p. 1426-1436, 2013.
- LIU, Y., GUPTA, H. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, v. 43, 2007.
- LIMA, A. Metodologia Integrada para Determinação da Enchente de Projeto de Estruturas Hidráulicas por Meio de Séries Sintéticas de Precipitação e Modelos Chuva-Vazão. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) – Escola de Engenharia, UFMG, Belo Horizonte, 2004.
- LINDSTROM, G., JOHANNSON, B., PERSSON, M., GARDELIN, M., BERGSTROM, S. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, vol. 201, ed. 1-4, p. 272-288, 1997.
- MADADGAR, S., MORADKHANI, H. Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. *Water Resources Research*, v. 50, p. 9586-9603, 2014.
- MANTOVAN, P., TODINI, E. Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. *Journal of Hydrology*, v. 330, p. 368-381, 2006.
- MARSHALL, L., NOTT, D., SHARMA, A. A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. *Water Resources Research*, v. 40, 11 p., 2004.

MARSHALL, L. Bayesian analysis of rainfall-runoff models: Insights to parameter estimation, model comparison and hierarchical model development. Tese (Doutorado em Engenharia) – University of New South Wales, Australia, 222 p., 2005.

MATOTT, L., BABENDREIER, J., PURUCKER, S. Evaluating uncertainty in integrated environmental models: A review of concepts and tools. *Water Resources Research*, v. 45, 2009.

McLAUGHLIN, D. An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. *Advances in Water Resources*, v. 25, n. 8-12, p. 1275-1286, 2002.

McMILLAN, H., CLARK, M. Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme. *Water Resources Research*, v. 45, 2009.

MELLER, A. Previsão de cheias por conjunto em curto prazo. Tese (Doutorado em Recursos Hídricos) – Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, 204 p., 2012.

MINASNY, B., VRUGT, J., MCBRATNEY, A. Confronting uncertainty in model-based geostatistics using Markov Chain Monte Carlo simulation. *Geoderma*, v. 163, p. 150-162, 2011.

MONTANARI, A., BRATH, A. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, v. 40, 2004.

MONTANARI, A. Uncertainty of hydrological predictions. In: WILDERER, P. (ed.). *Treatise on Water Science*. Oxford Academic Press, v. 2, p. 459-478, 2011.

MONTANARI, A., KOUTSOYIANNIS, D. A blueprint for process-based modeling of uncertain hydrological systems. *Water Resources Research*, v. 48, n. 9, 2012.

MOORE, R.J. The PDM rainfall-runoff model. *Hydrology & Earth System Sciences*, vol. 11, ed. 1, p. 483-499, 2007.

MORADKHANI, H., SOROOSHIAN, S., GUPTA, H., HOUSER, PAUL R. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, v. 28, n. 2, p. 135-147, 2005.

MORADKHANI, H., SOROOSHIAN, S. General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In: SOROOSHIAN, S., HSU, K-L, COPPOLA, E., TOMASSETTI, B., VERDECCHIA, M., VISCONTI, G (eds). *Hydrological modelling and the water cycle: coupling the atmospheric and hydrological models*. Springer, 2008.

NETO, R., GUETTER, A., MINE, M. Modelo hidrológico com atualização de estado parte I : modelagem hidrológica. *Revista Brasileira de Recursos Hídricos*, v. 12, n. 3, p. 93-103, 2007.

NAVARRO, D., PERFORS, AMY. The Metropolis-Hastings algorithm. COMPSCI 3016: Computational cognitive science. University of Adelaide, data desconhecida.

NETO, R., GUETTER, A., MINE, M. Modelo hidrológico com atualização de estado – Parte I: modelagem hidrológica. *Revista Brasileira de Recursos Hídricos*, v. 12, n.3, p. 93-103, 2007.

NEUMAN, S. Accounting for conceptual model uncertainty via maximum likelihood Bayesian model averaging. *Acta Universitatis Carolinae – Geologica*, 46(2/3), p. 529–534, 2002.

NICHOLS, N. Mathematical Concepts of Data Assimilation. In: LAHOZ, W., KHATTATOV, B., MENARD, R. (eds). *Data Assimilation: Making Sense of Observations*. Springer, p. 16-40, 2010.

PAULINO, C., TURKMAN, A., MURTEIRA, B. *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa, 2003.

PEREZ, G. Hybrid models for hydrological forecasting: integration of data-driven and conceptual modelling techniques. Tese (Doutorado em Hidroinformática) – UNESCO-IHE Institute for Water Education, Delft University of Technology, Netherlands, 2009.

PERRIN, C. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, vol. 279, ed. 1-4, p. 275–289, 2003.

QUEIROGA, Y. Estudo e modelagem dos erros de simulação hidrológica e sua assimilação na previsão de vazões de curto prazo – o caso da bacia do rio Grande na UHE Camargos. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) – Escola de Engenharia, UFMG, 348 p., Belo Horizonte, 2003.

R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

REFSGAARD, J., STORM, B. (eds). *Distributed hydrological modelling*. Kluwer Academic Publishers, Dordrecht, Netherlands, 1996.

RENARD, B., KAVETSKI, D., KUCZERA, G., THYER, M., FRANKS, S. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, v. 46, ed. 5, 2010.

SADEGH, M., VRUGT, J. Approximate Bayesian computation using Markov Chain Monte Carlo simulation: DREAM\_ABC. *Water Resources Research*, v. 50, ed. 8, p. 6767-6787, 2014.

SCHOUPS, G., VRUGT, J. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, v. 10, n. 10, 17 p., 2010.

SERBAN, P., ASKEW, A.J. Hydrological forecasting and updating procedures. *IAHS Publ.*, v. 201, p. 357–369, 1991.

SHRESTHA, D. Uncertainty analysis in rainfall-runoff modelling: Application of machine learning techniques. Tese (Doutorado em Hidroinformática) – UNESCO-IHE Institute for Water Education, Delft University of Technology, 205 p., Netherlands, 2009.

SINGH, V.P., WOOLHISER, D.A. Mathematical Modeling of Watershed Hydrology. *Journal of Hydrologic Engineering*, p. 270-292, 2002.

SIVAPALAN, M., TAKEUCHI, K., FRANKS, S., GUPTA, V., KARAMBIRI, H., LAKSHMI, V., LIANG, X., MCDONNELL, J., MENDIONDO, E., O'CONNELL, P., OKI, T., POMEROY, J., SCHERTZER, D., UHLENBROOK, S., ZEHE, E. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, v. 48, 857-880, 2003.

SKLASH, M., FARVOLDEN, R. The role of groundwater in storm runoff. *Journal of Hydrology*, v. 43, p. 45-65, 1979.

SMITH, T., SHARMA, A., MARSHALL, L., MEHROTRA, R., SISSON, S. Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments. *Water Resources Research*, v. 46, 11 p., 2010.

SOLOMATINE, D., SHRESTHA, D. A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research*, v. 45, 2009.

SOROOSHIAN, S., DUAN, Q., GUPTA, V. Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture accounting model. *Water Resources Research*, v. 29, p. 1185–1194, 1993.

SOROOSHIAN, S., GUPTA, V. Model calibration, In: SINGH, V. *Computer Model of Watershed Hydrology*. WRP, Highlands Ranch, p. 23-68, 1995.

STEDINGER, J., VOGEL, R., LEE, S., BATCHELDER, R. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resources Research*, v. 44, 2008.

ter BRAAK, C. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, v. 16, n. 3, p. 239-249, 2006.

THYER, M., RENARD, B., KAVETSKI, D., KUCZERA, G., FRANKS, S., SRIKANTHAN, S. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resources Research*, v. 45, n. 12, 2009.

TIAN, Y., BOOIJ, M., XU, Y. Uncertainty in high and low flows due to model structure and parameter errors. *Stochastic Environmental Research and Risk Assessment*, v. 28, ed. 2, p. 319-332, 2013.

TIERNEY, L. Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, v. 22, ed. 4, p. 1701-1728, 1994.

- TODINI, E. The ARNO rainfall-runoff model. *Journal of Hydrology*, v. 175, n. 1-4, p. 339-382, 1996.
- TUCCI, C. Modelos Hidrológicos. ABRH, 2º ed., 678 p., Porto Alegre, 2005.
- van ESSE, W. Demystifying hydrological monsters – Can flexibility in model structure help explain monster catchments? Dissertação (Msc thesis in Department of Water Engineering and Management / Hydrosystems and Bioprocesses Research Unit). University of Twente/ Irstea, 110 p., 2012.
- VRUGT, J., ter BRAAK, J., CLARK, M., HYMAN, J., ROBINSON, B. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, v. 44, n. 12, 15 p., 2008a.
- VRUGT, J., ter BRAAK, C., GUPTA, H., ROBINSON, B. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, v. 23, n. 7, p. 1011-1026, 2008b.
- VRUGT, J., ter BRAAK, C., DIKS, C., ROBINSON, B., HYMAN, J., HIGDON, D. Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, v. 10, n. 3, p. 273–290, 2009.
- VRUGT, J., ter BRAAK, C. DREAM<sub>(D)</sub>: an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrology and Earth System Sciences*, v. 15, p. 3701-3713, 2011.
- VRUGT, J., ter BRAAK, C., DIKS, C., SCHOUPS, G. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water Resources*, v. 51, p. 457-478, 2013.
- WAGENER, T., GUPTA, H. Model identification for hydrological forecasting under uncertainty. *Stochastic Environmental Research and Risk Assessment*, vol. 19, ed. 6, p. 378-387, 2005.
- WAGENER, T., MCINTYRE, N., LEES, M.J., WHEATER, H.S., GUPTA, H.V. Towards reduced uncertainty in conceptual rainfall-runoff modeling: Dynamic identifiability analysis. *Hydrological Processes*, vol. 17, ed. 2, p. 455– 476, 2003.
- WAGENER, T., WHEATER, H., GUPTA, H. Rainfall-runoff modelling in gauged and ungauged catchments. Imperial College Press, 306 p., 2004.
- WALKER, J., HOUSER, P. Hydrologic Data Assimilation. In: ASWATHANARAYANA, U. (ed.). *Advances in Water Science Methodologies*. Balkema/ Taylor & Francis Group, p. 25–48, The Netherlands, 2005.
- WILLEMS, P., MORA, D., VANSTEENKISTE, T., TAYE, M., VAN STEENBERGEN, N. Parsimonious rainfall-runoff model construction supported by time series processing and validation of hydrological extremes – Part 2: Intercomparison of models and calibration approaches. *Journal of Hydrology*, v. 510, p. 591-609, 2014.

WMO. Simulated real-time intercomparison of hydrological models. Operational hydrology report (OHR), 38, n. 779, 241 p. 1992.

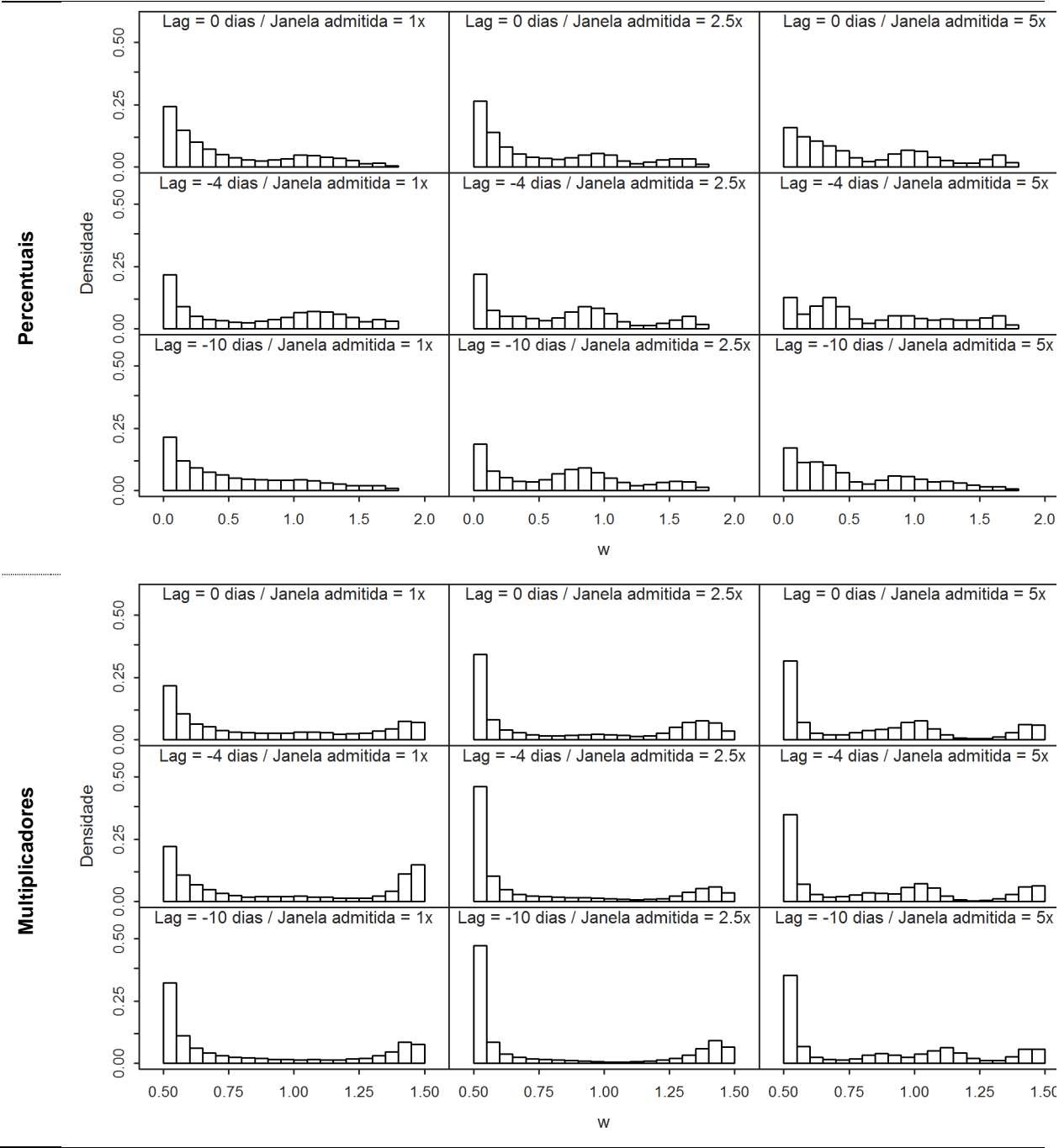
WMO. Manual on flood forecasting and warning. WMO-No. 1072, 2011.

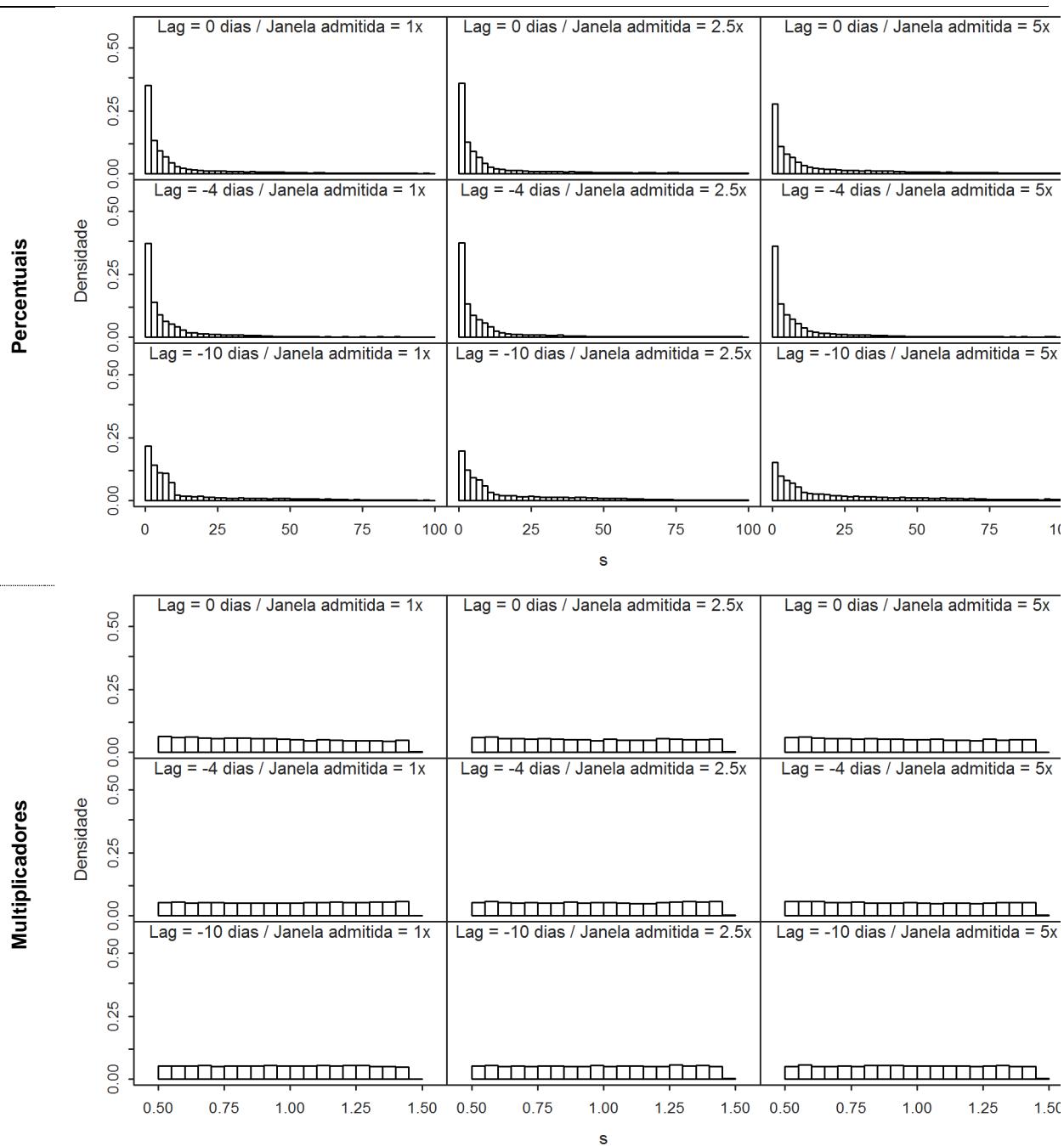
XIONG, L., SHAMSELDIN, A., O'CONNOR, K. A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi–Sugeno fuzzy system. Journal of Hydrology, v. 245. ed. 1-4, p. 196-217, 2001

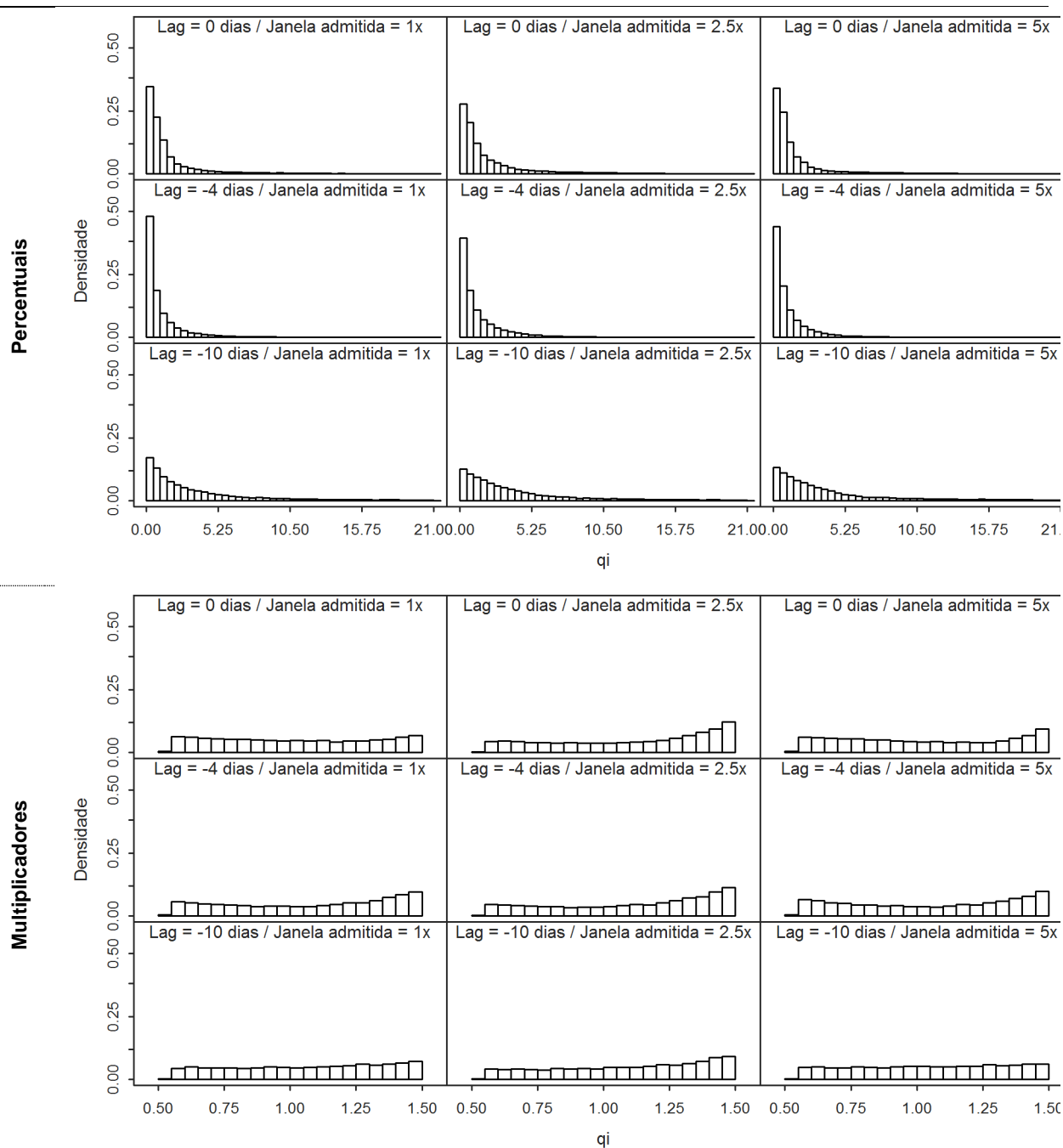
ZHAO, R.J. The Xinanjiang model applied in China. Journal of Hydrology, vol. 135, ed. 1-4, p. 371-381, 1992.

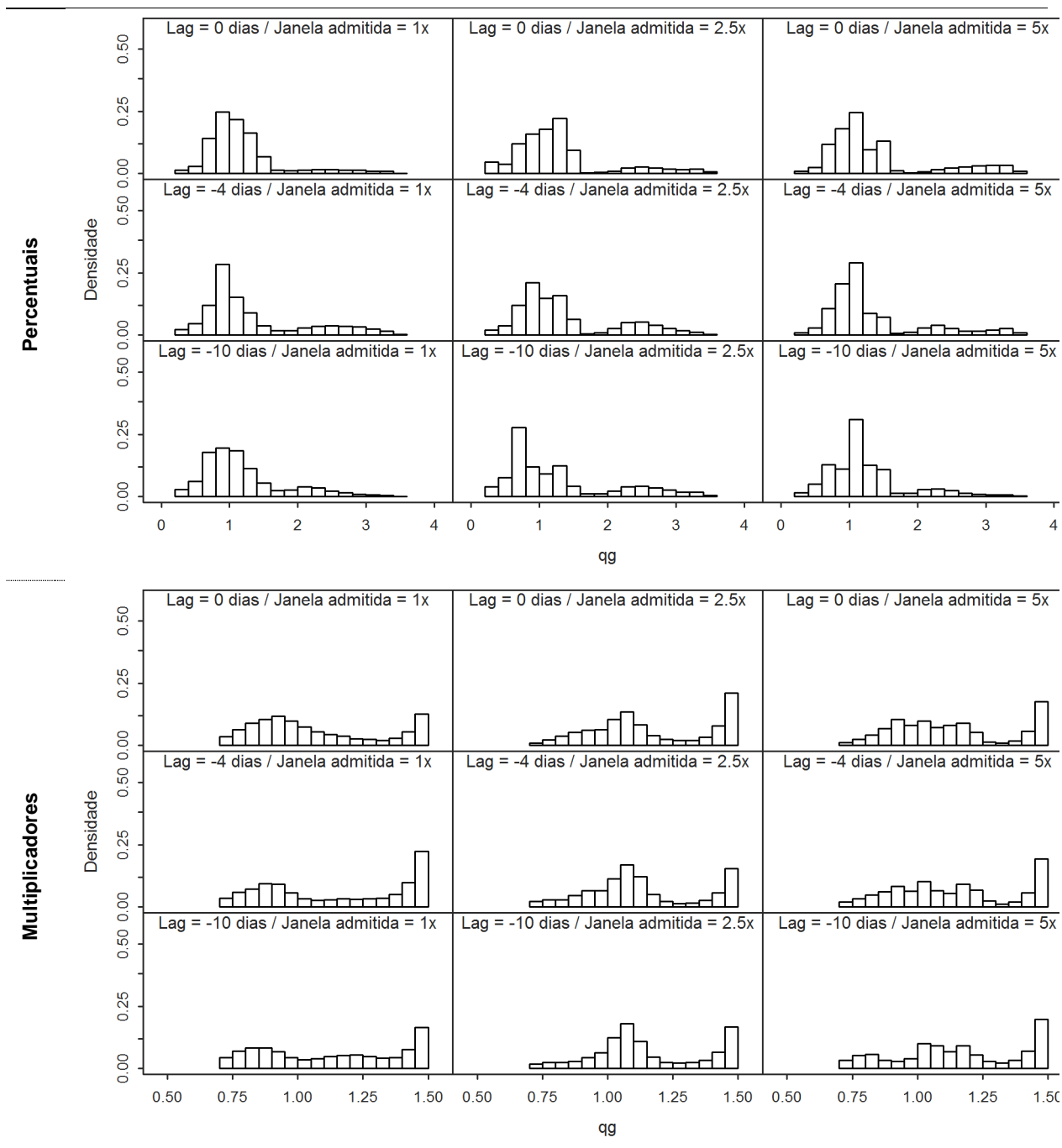


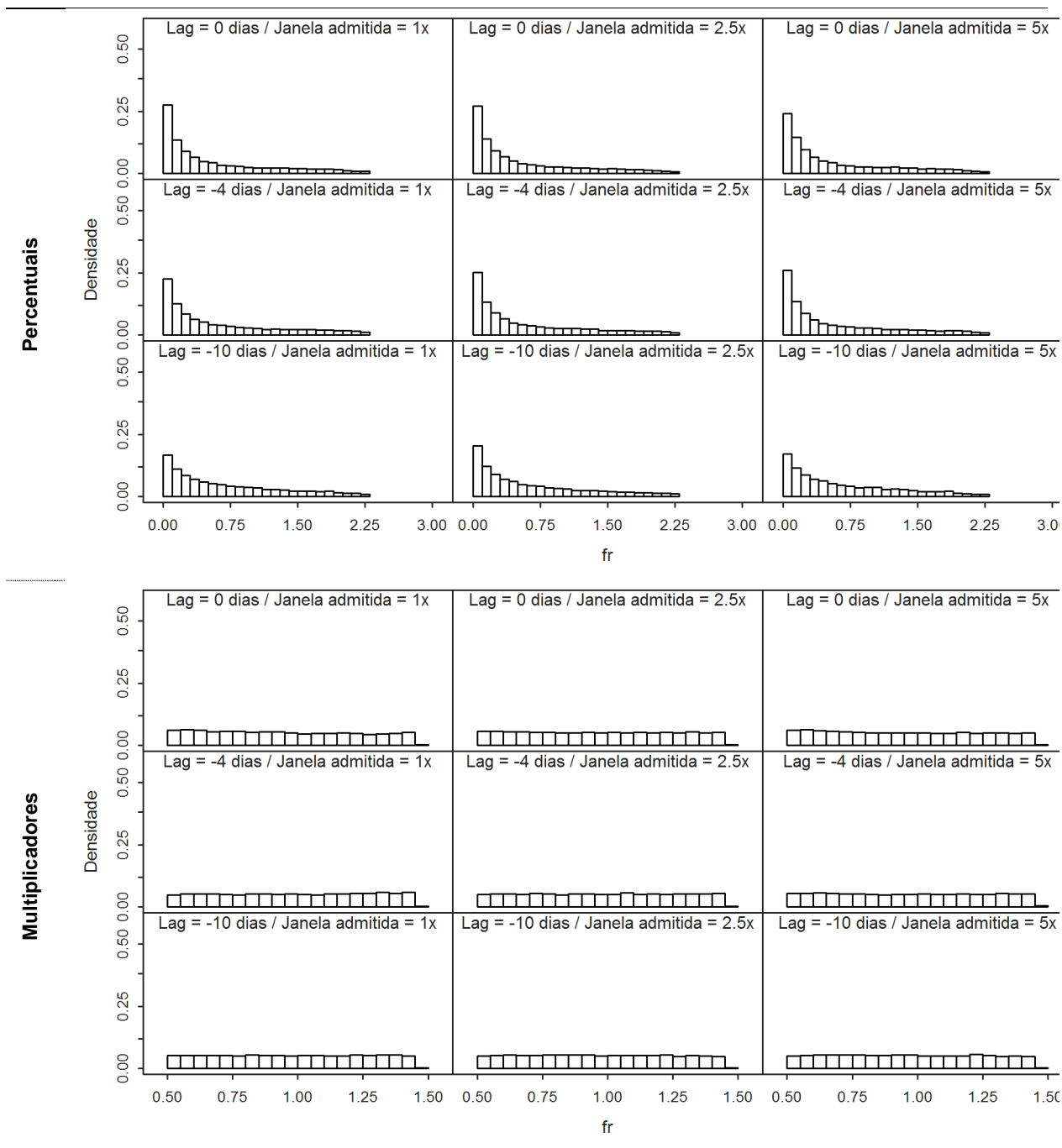
**ANEXO I – MAGNITUDE DA VARIAÇÃO DAS VARIÁVEIS DE ESTADO ATUALIZADAS**





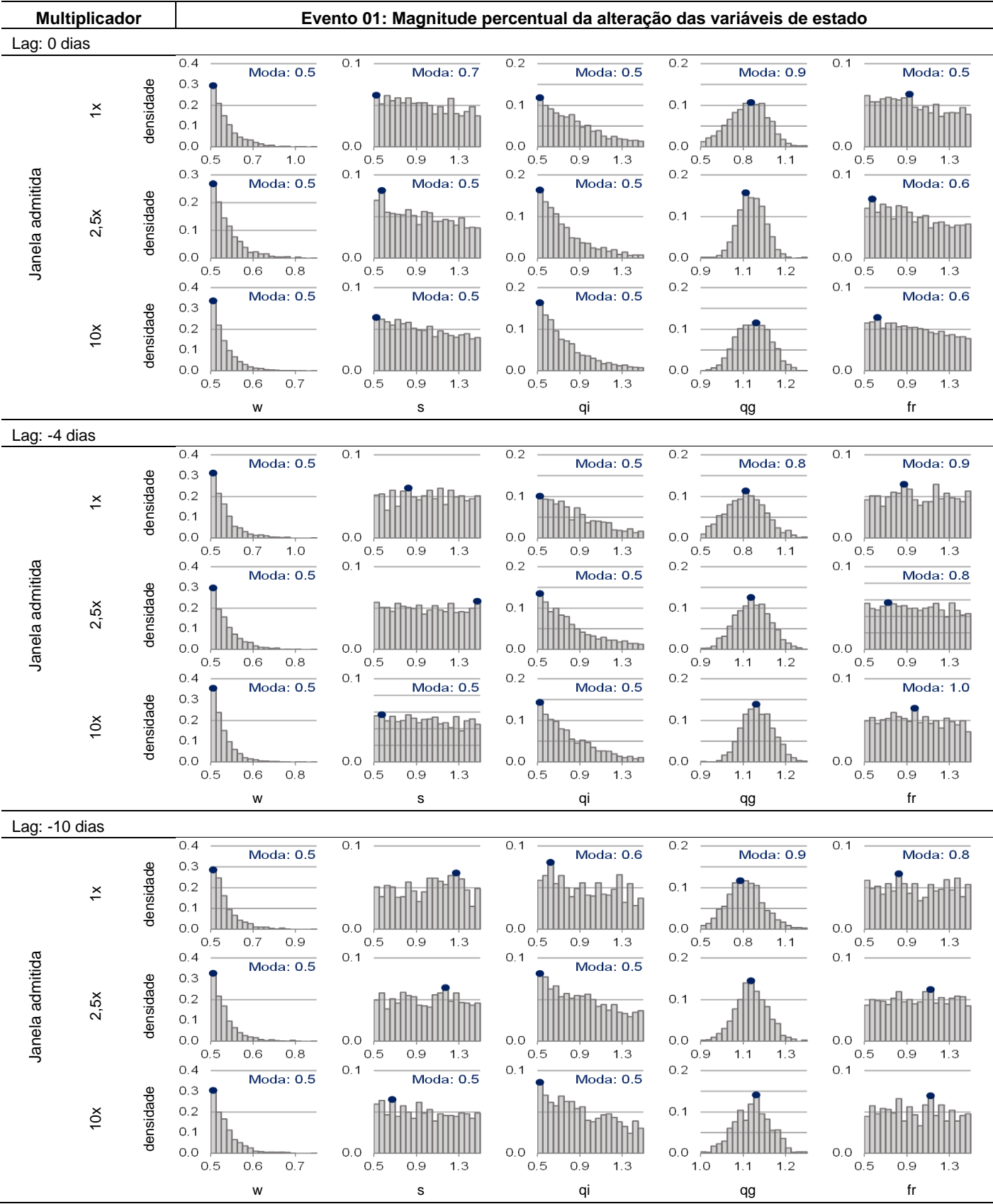
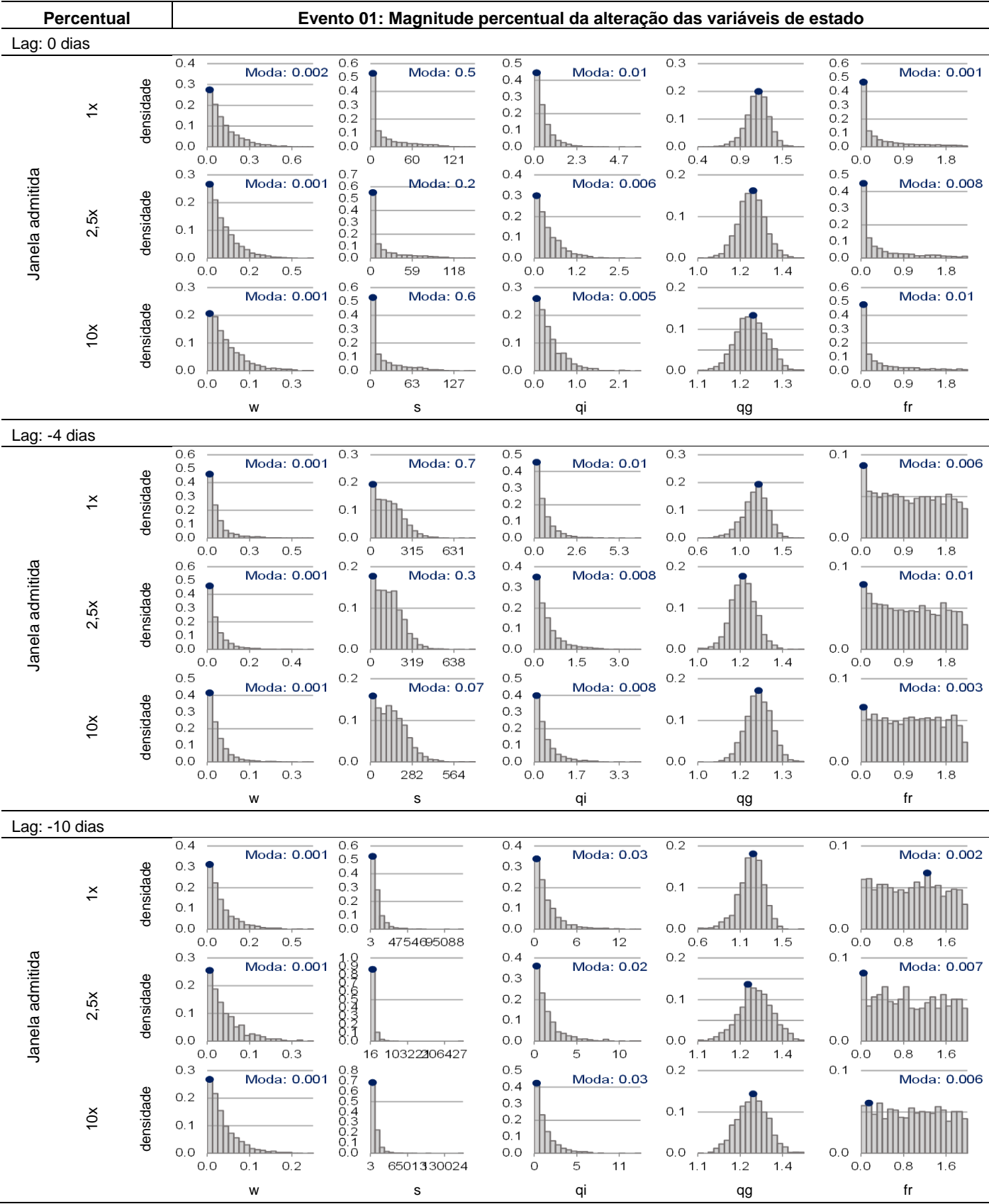


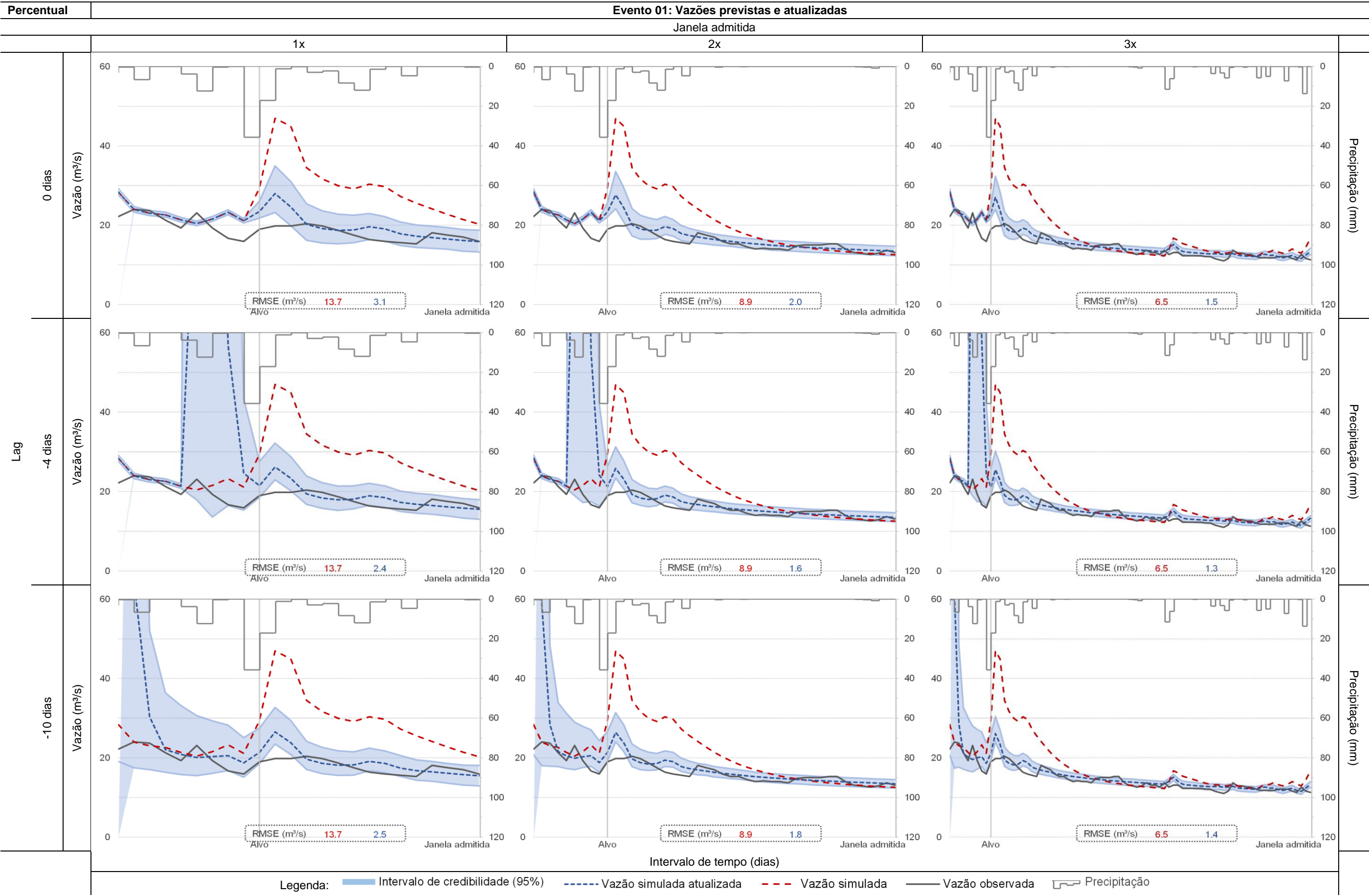




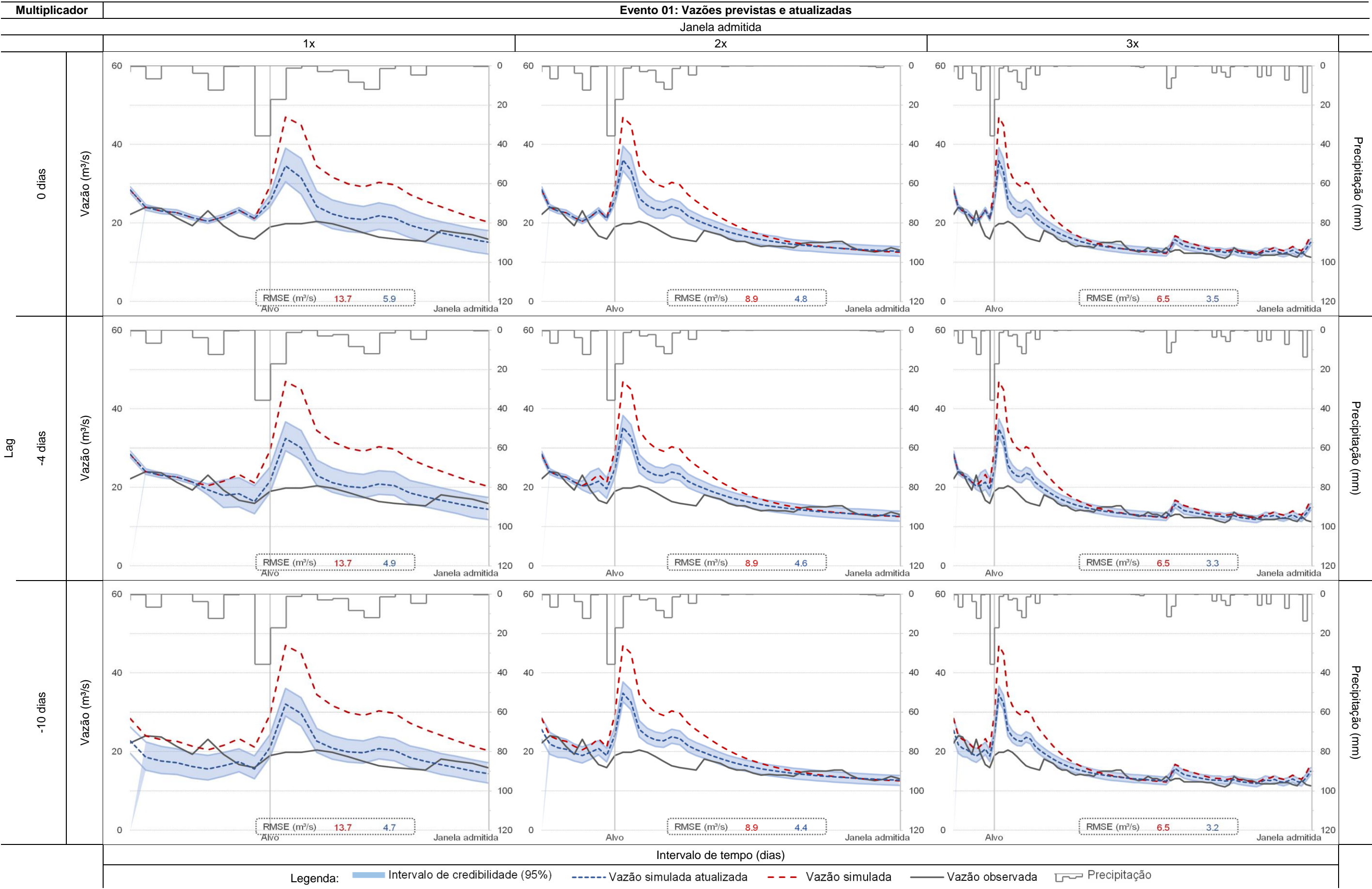
## **ANEXO II – ATUALIZAÇÃO DAS VARIÁVEIS DE ESTADO: RESULTADOS GRÁFICOS**

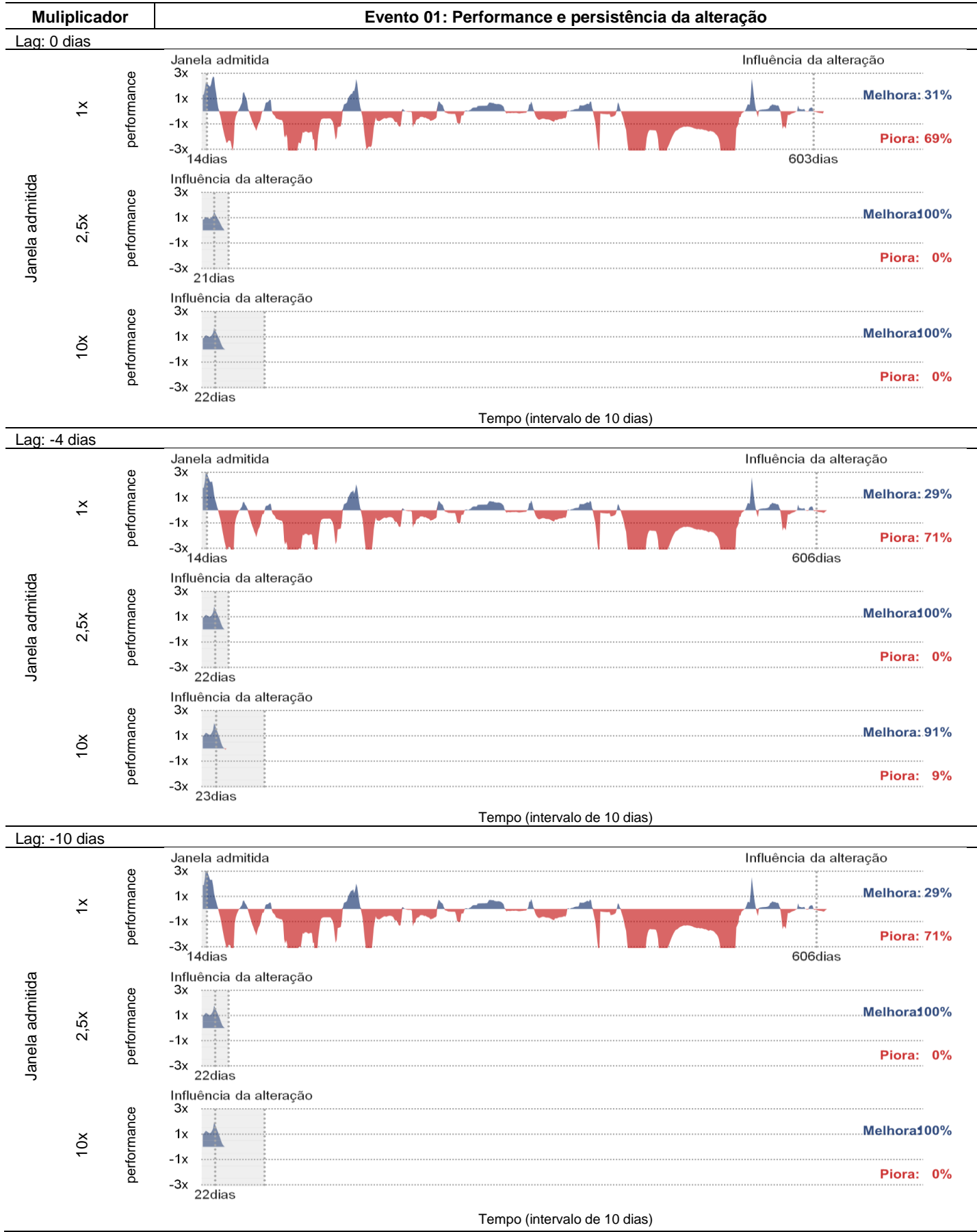
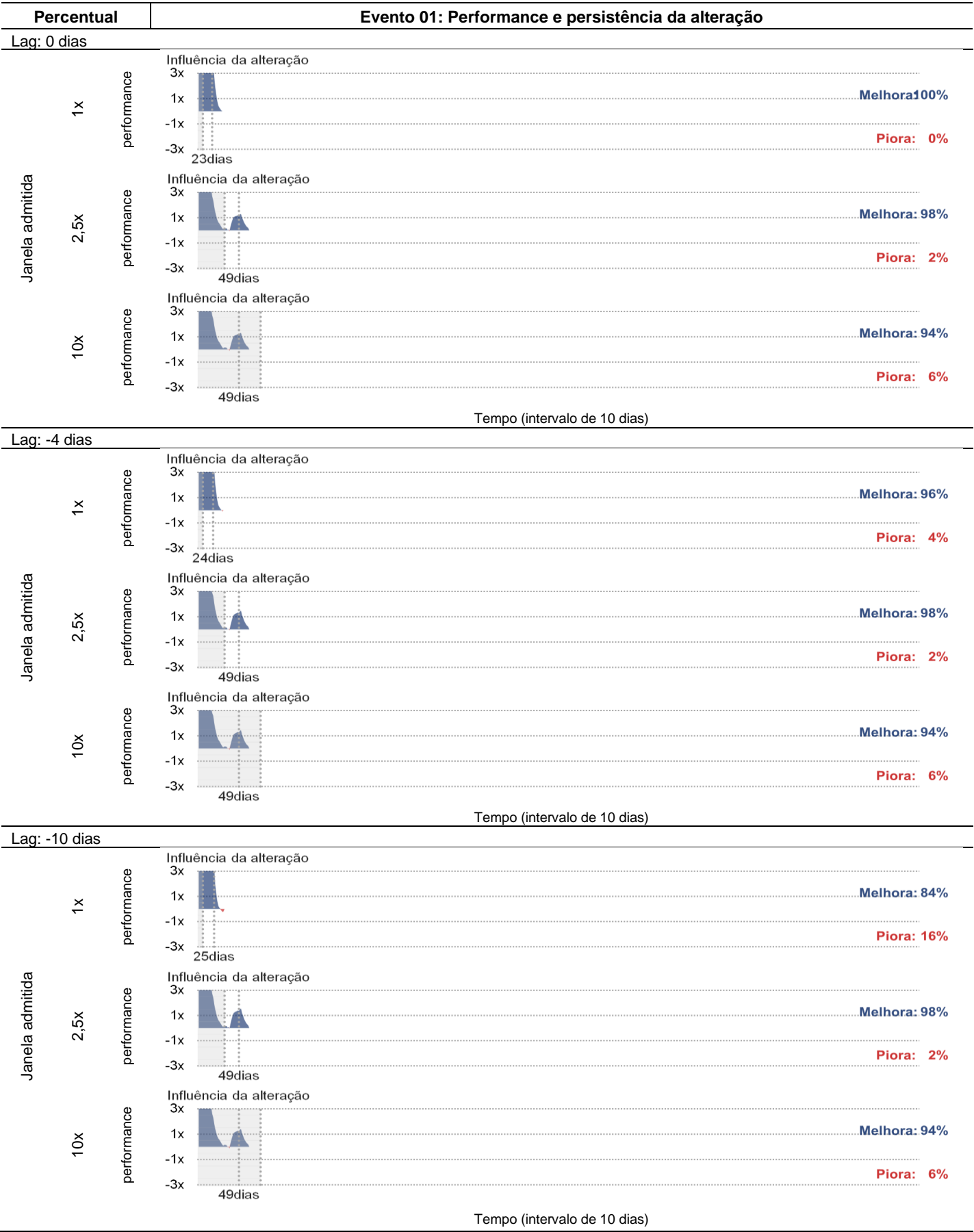
Evento 01 .....	22 mar – 05 abr 1976
Evento 02 .....	08 fev – 22 mar 1978
Evento 03 .....	24 mai – 19 jun 1978
Evento 04 .....	24 abr – 15 mai 1978
Evento 05 .....	21 out – 05 dez 1975
Evento 06 .....	22 jul – 05 ago 1978
Evento 07 .....	21 out – 15 nov 1979
Evento 08 .....	09 jan – 27 jan 1979
Evento 09 .....	05 out – 15 out 1975
Evento 10 .....	14 ago – 25 ago 1976
Evento 11 .....	28 fev – 11 mar 1976
Evento 12 .....	08 dez – 28 dez 1976
Evento 13 .....	01 dez – 15 dez 1978
Evento 14 .....	29 set – 11 out 1978
Evento 15 .....	30 out – 11 nov 1978





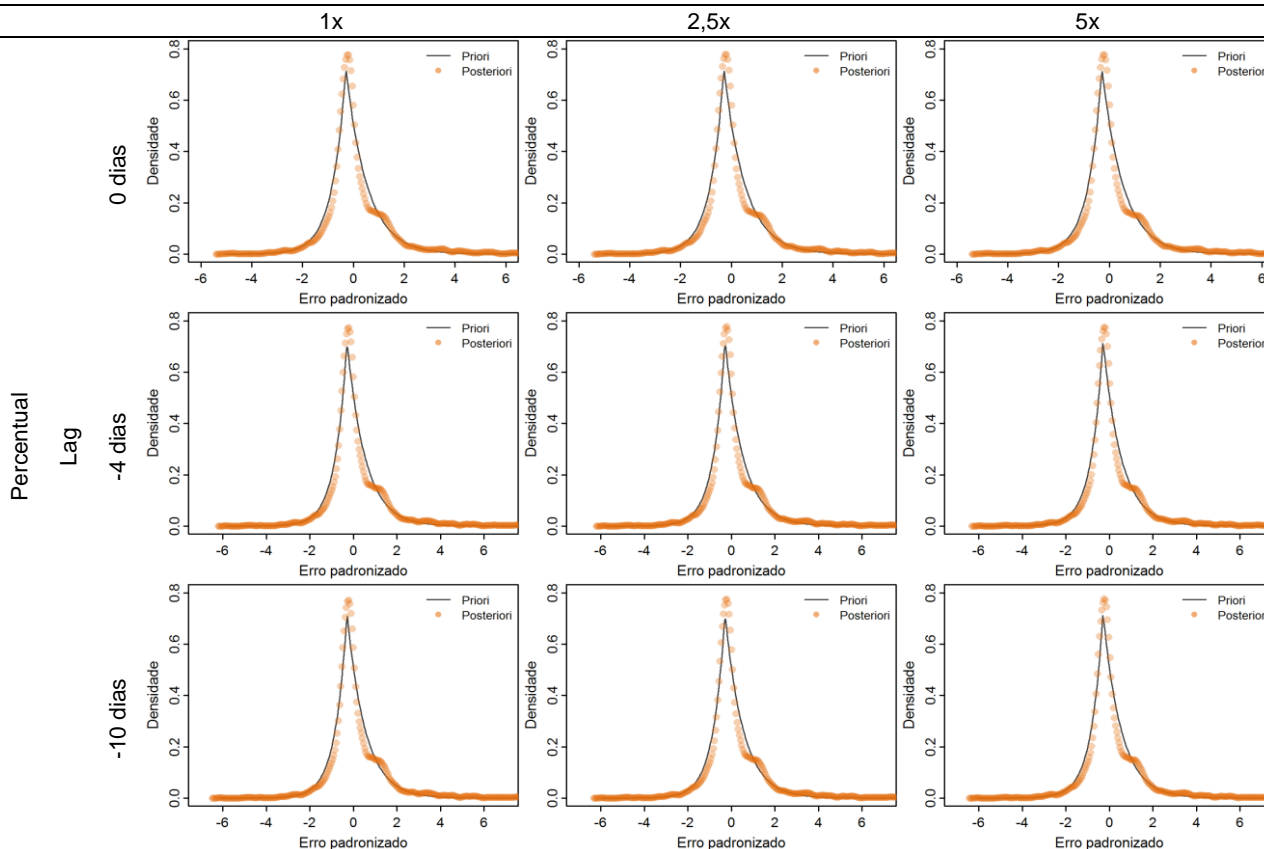




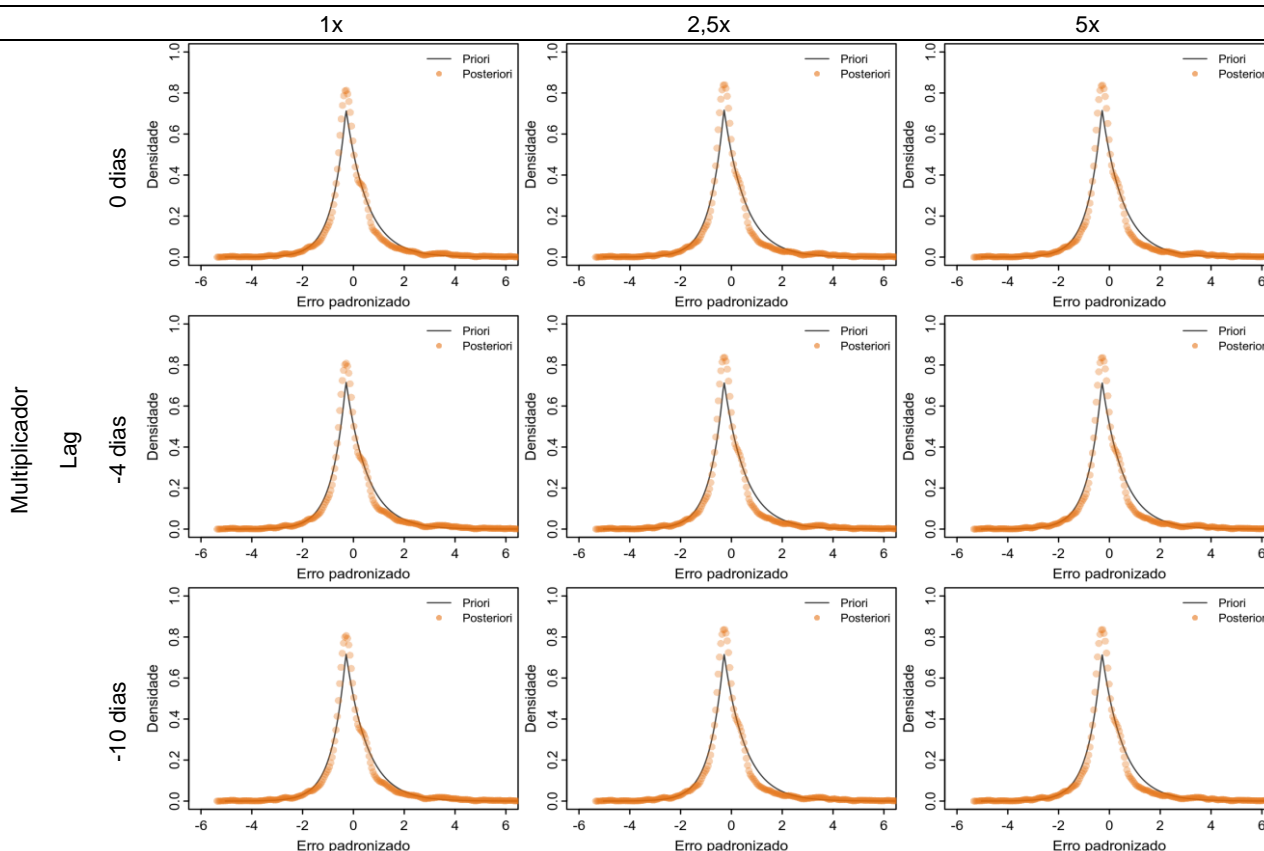


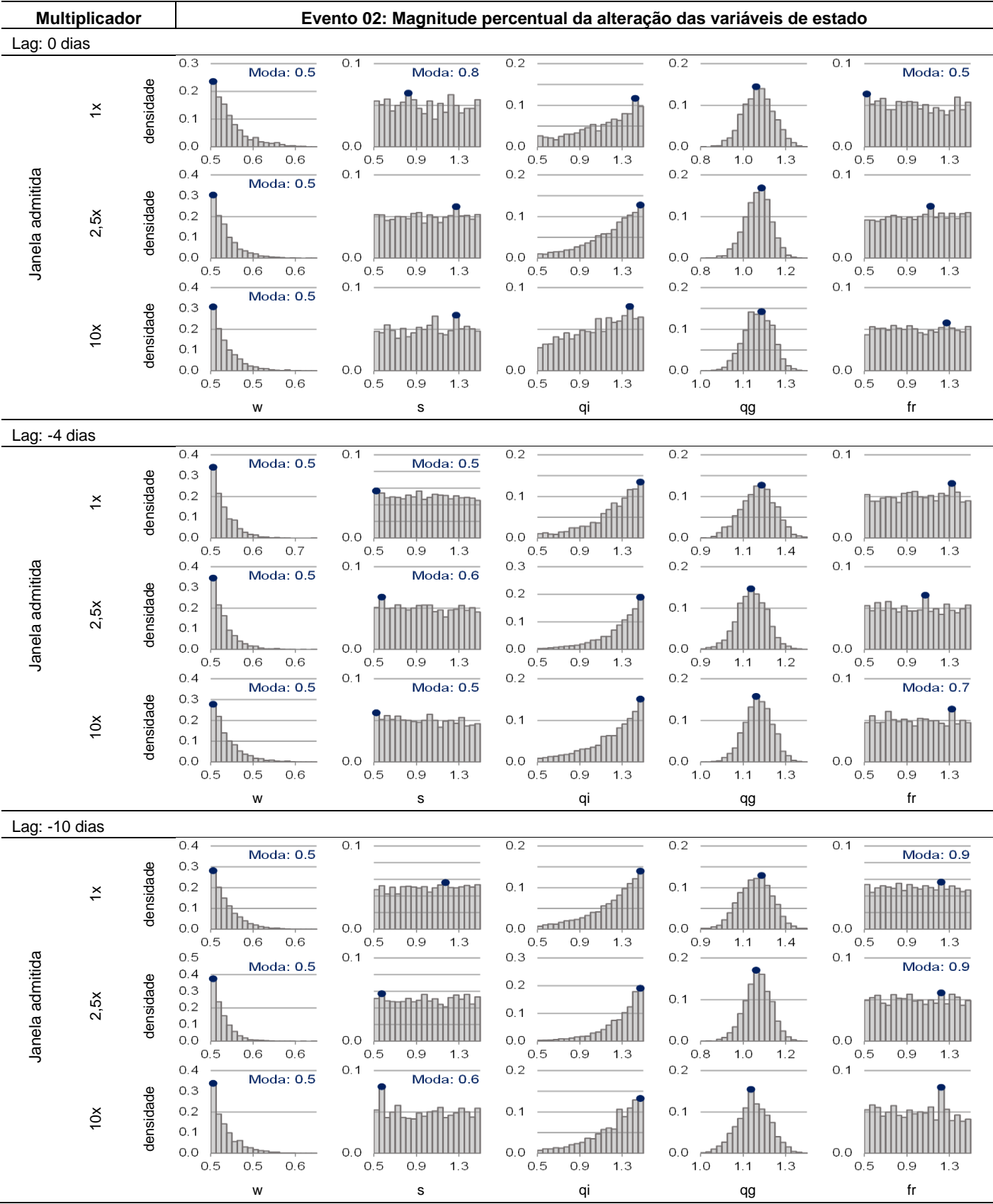
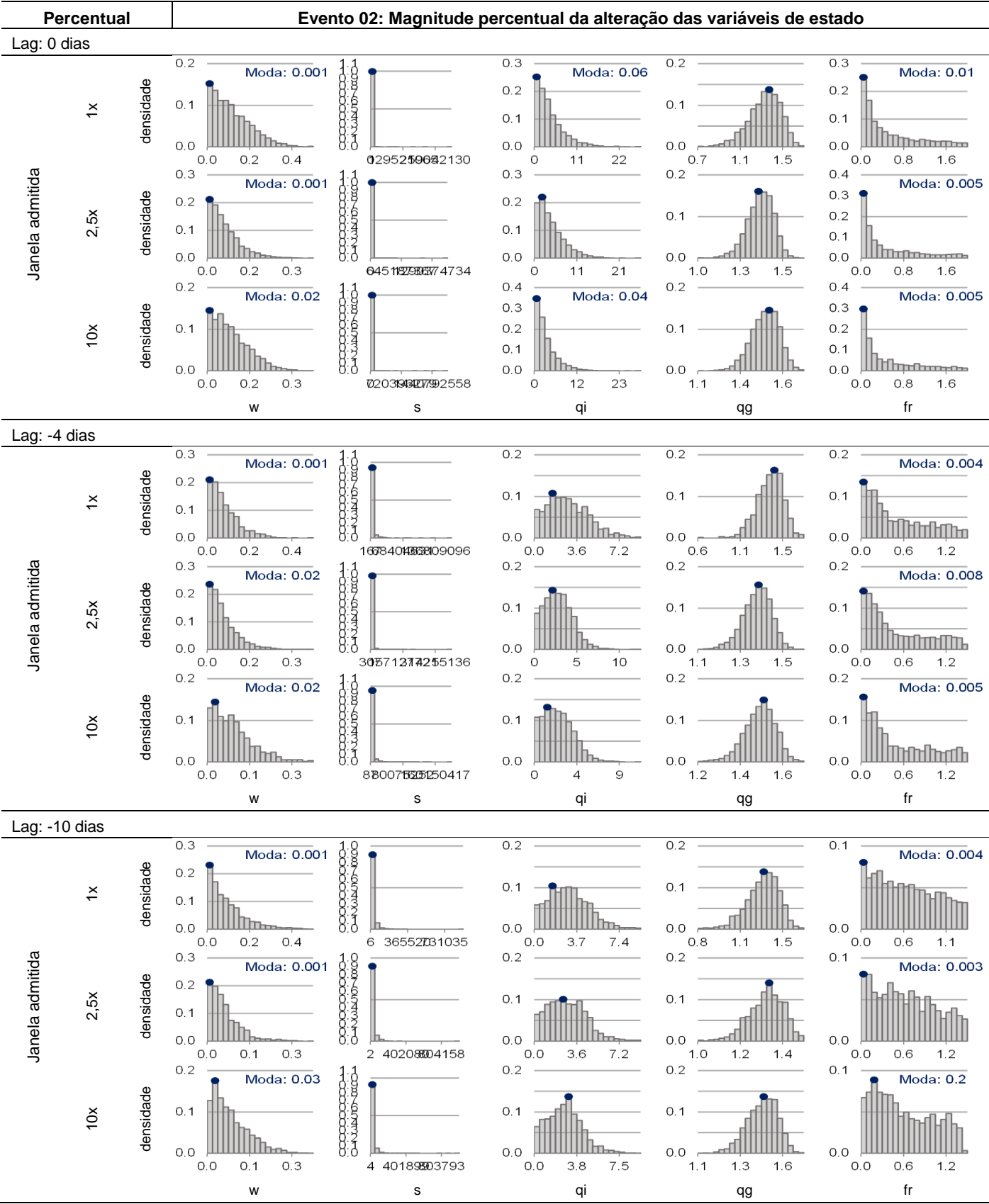
## Evento 01: Comportamento probabilístico dos resíduos da simulação

Janela admitida

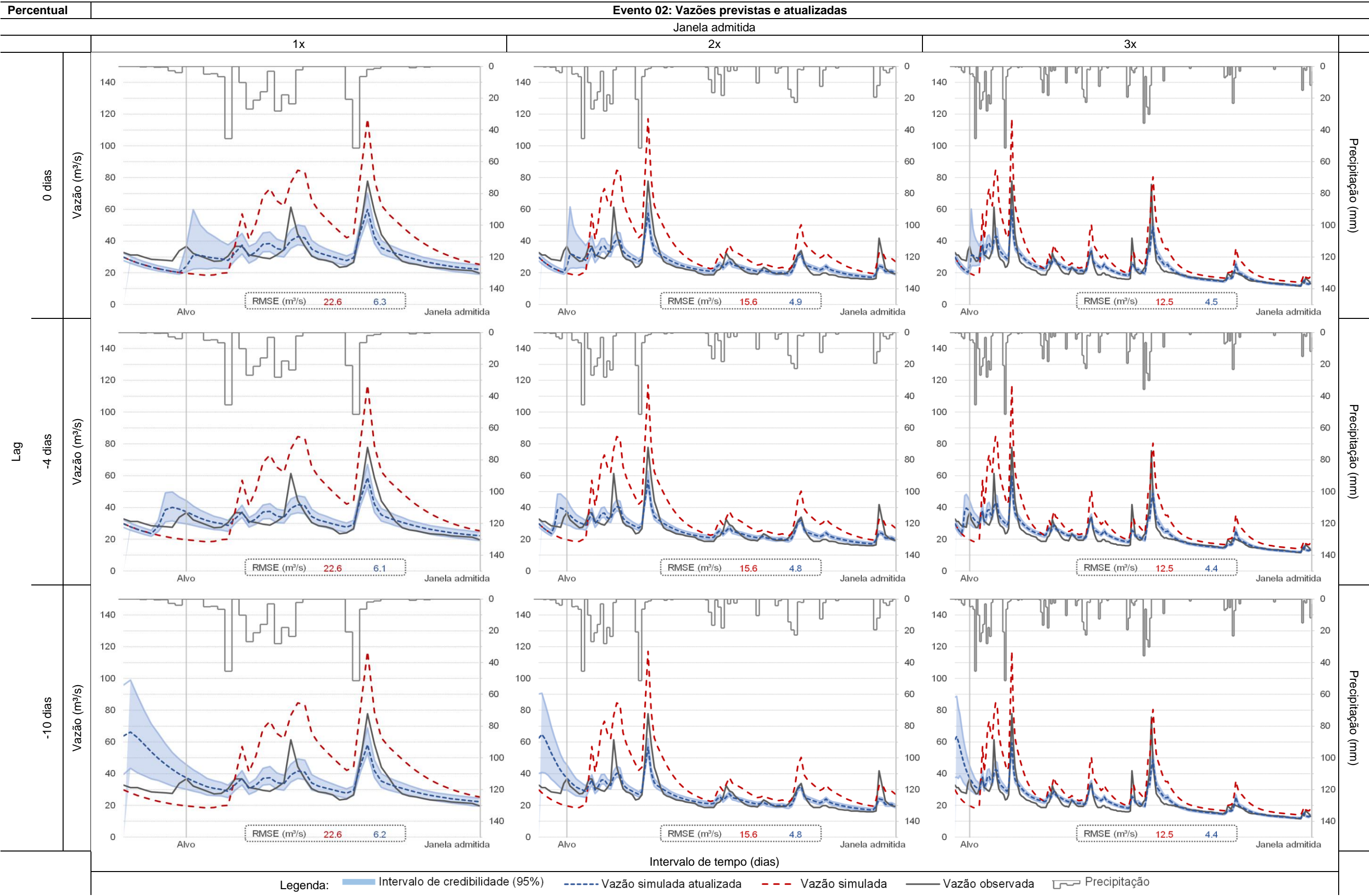


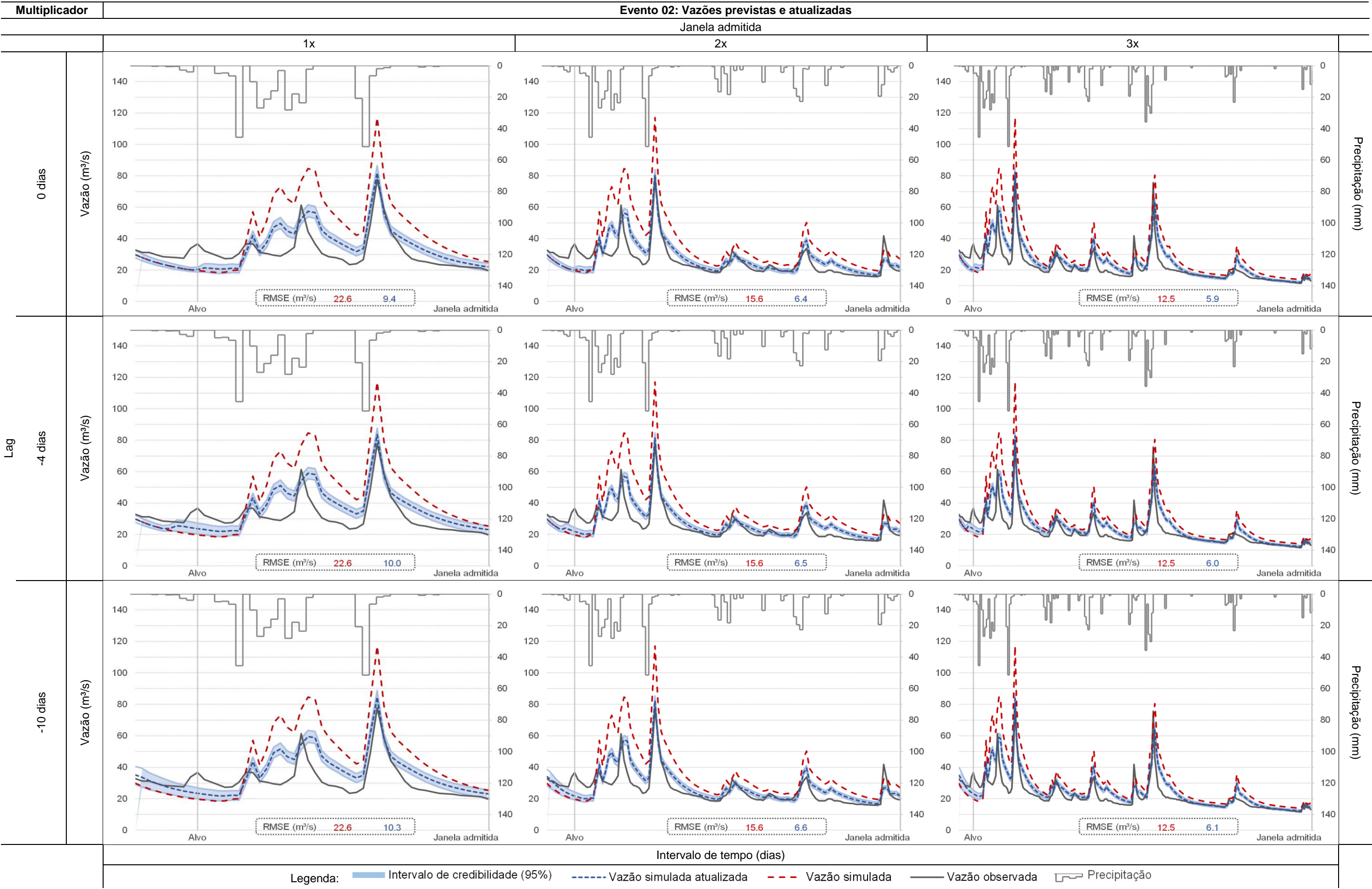
Janela admitida



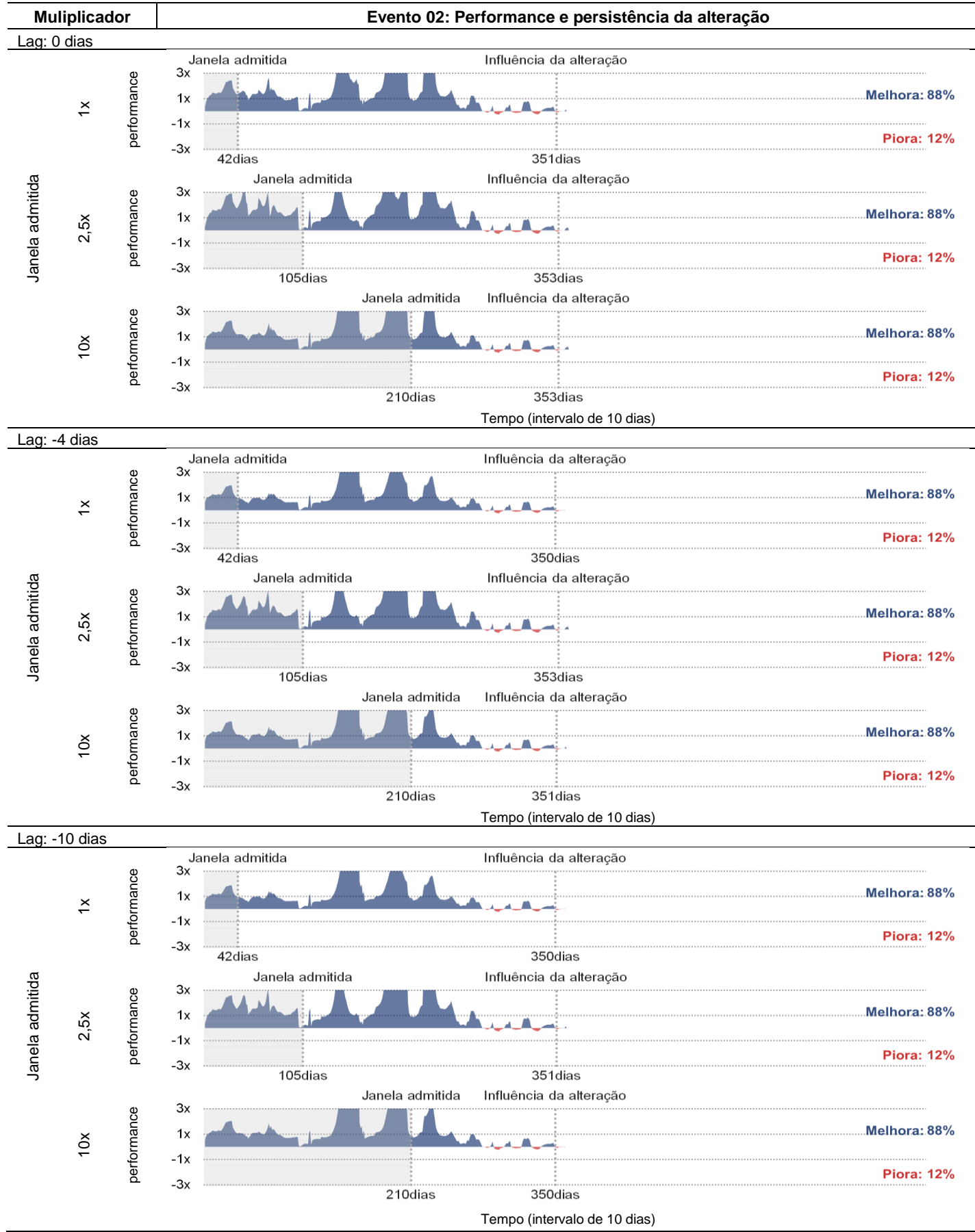
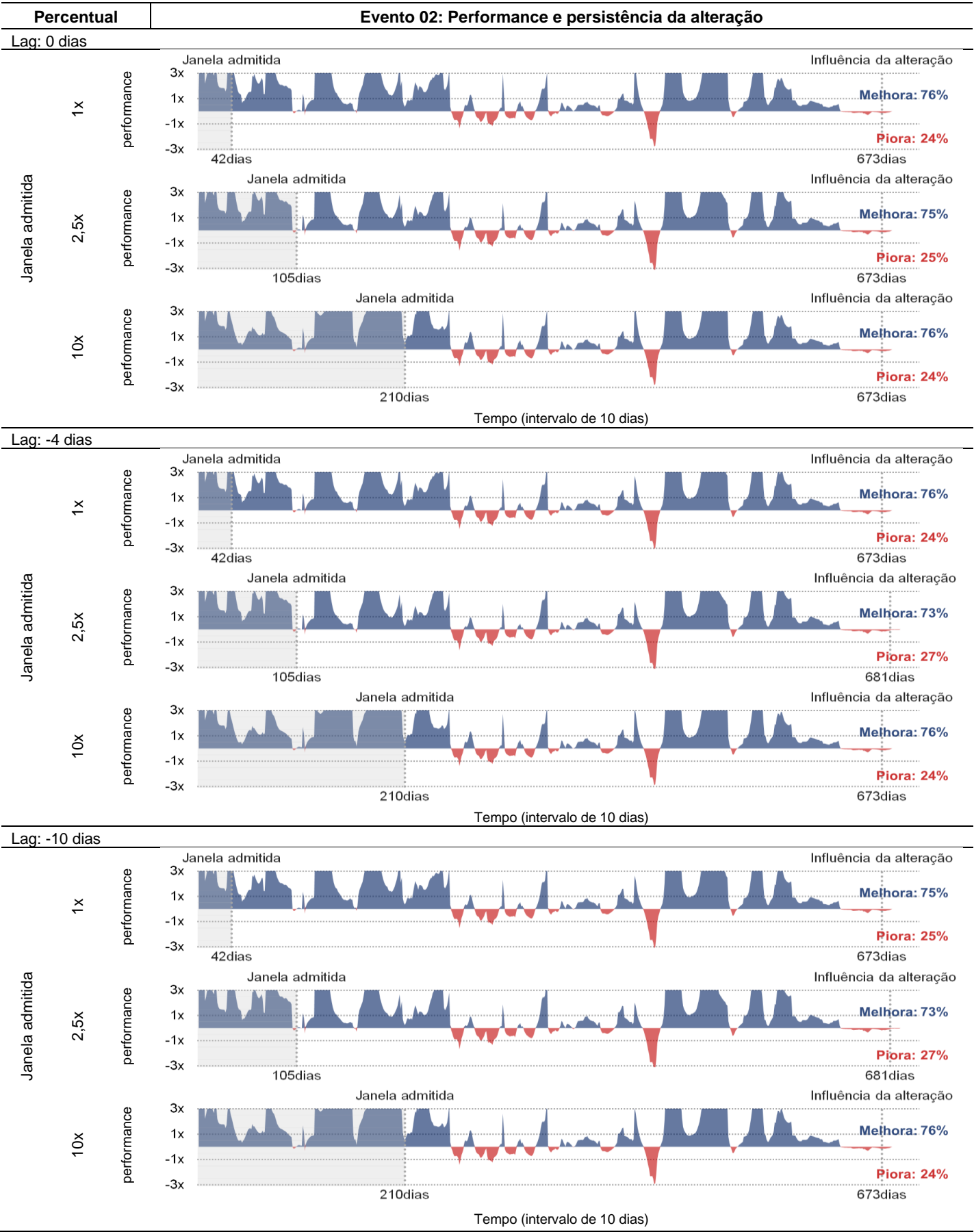






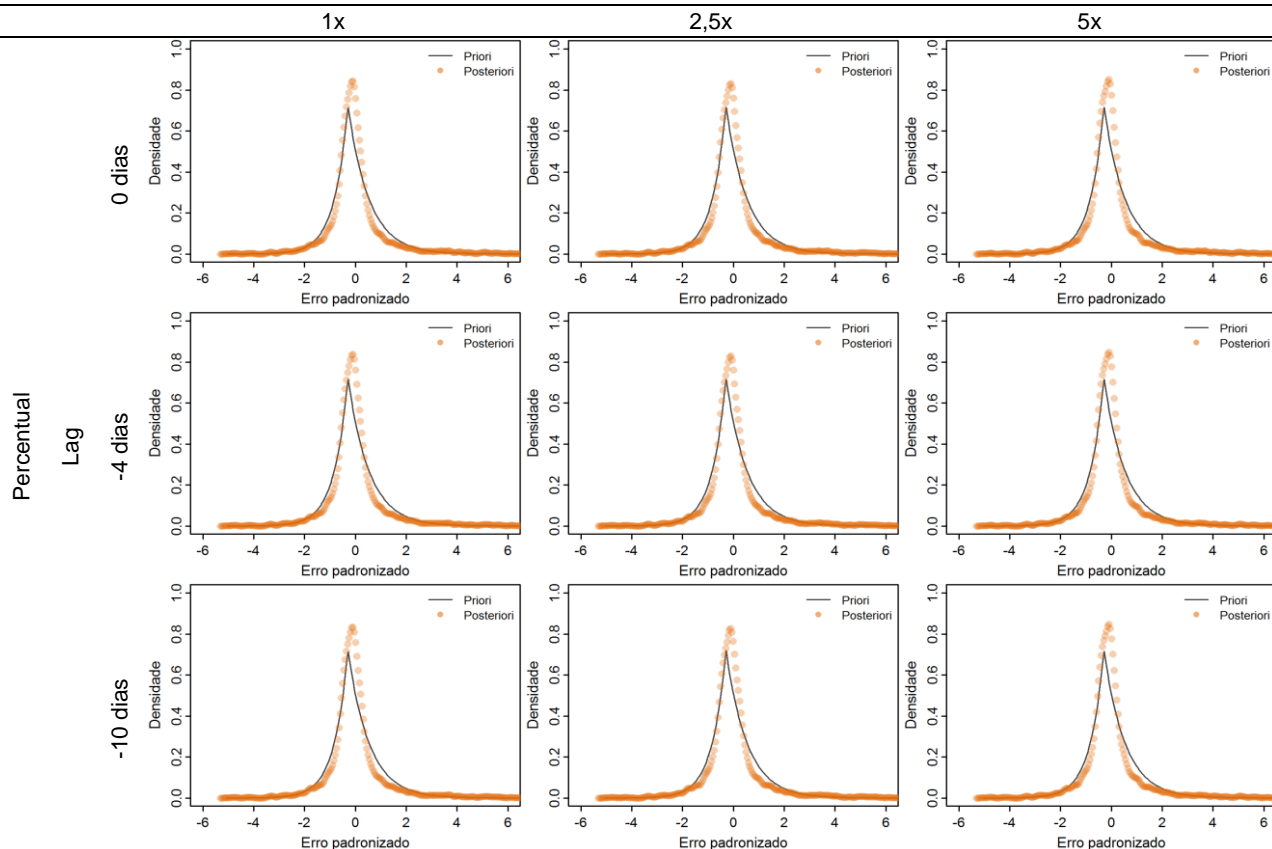




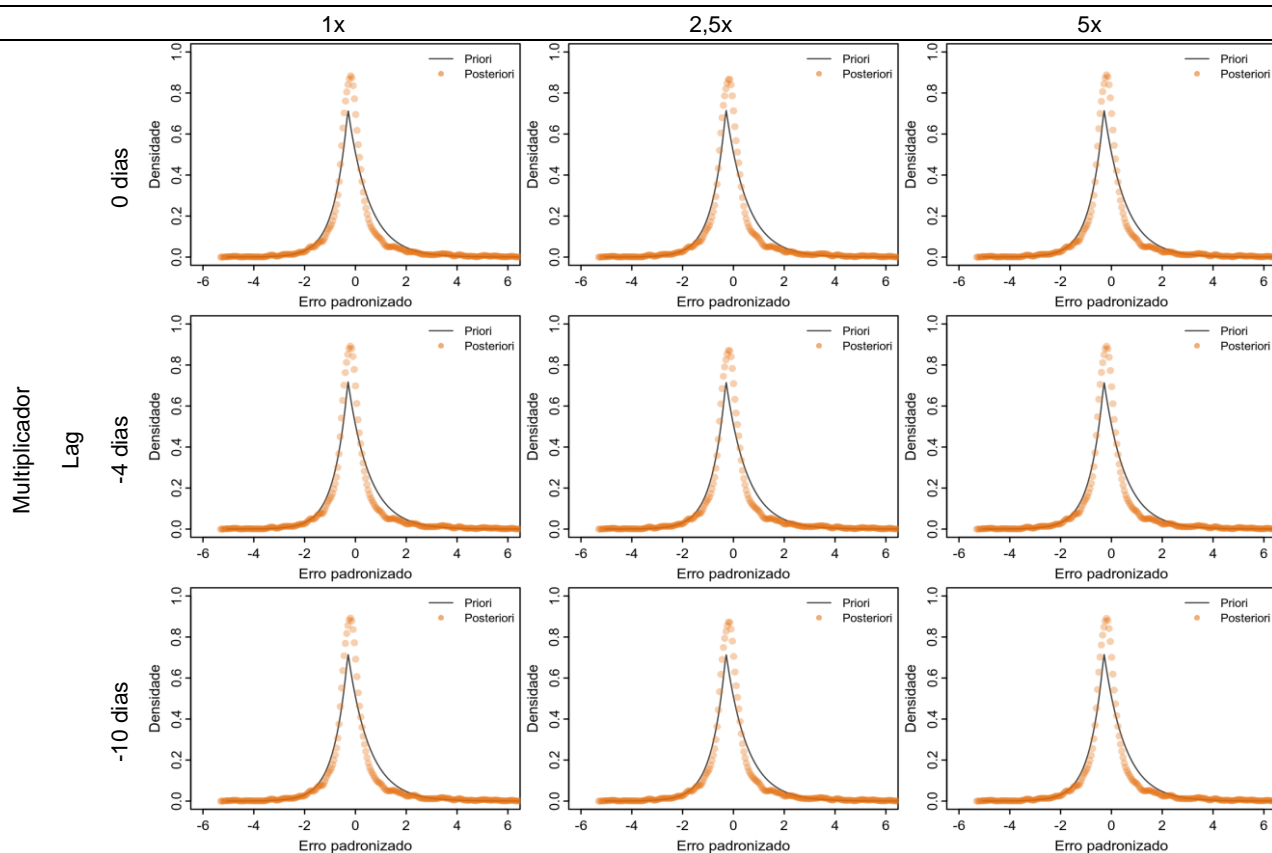


## Evento 02: Comportamento probabilístico dos resíduos da simulação

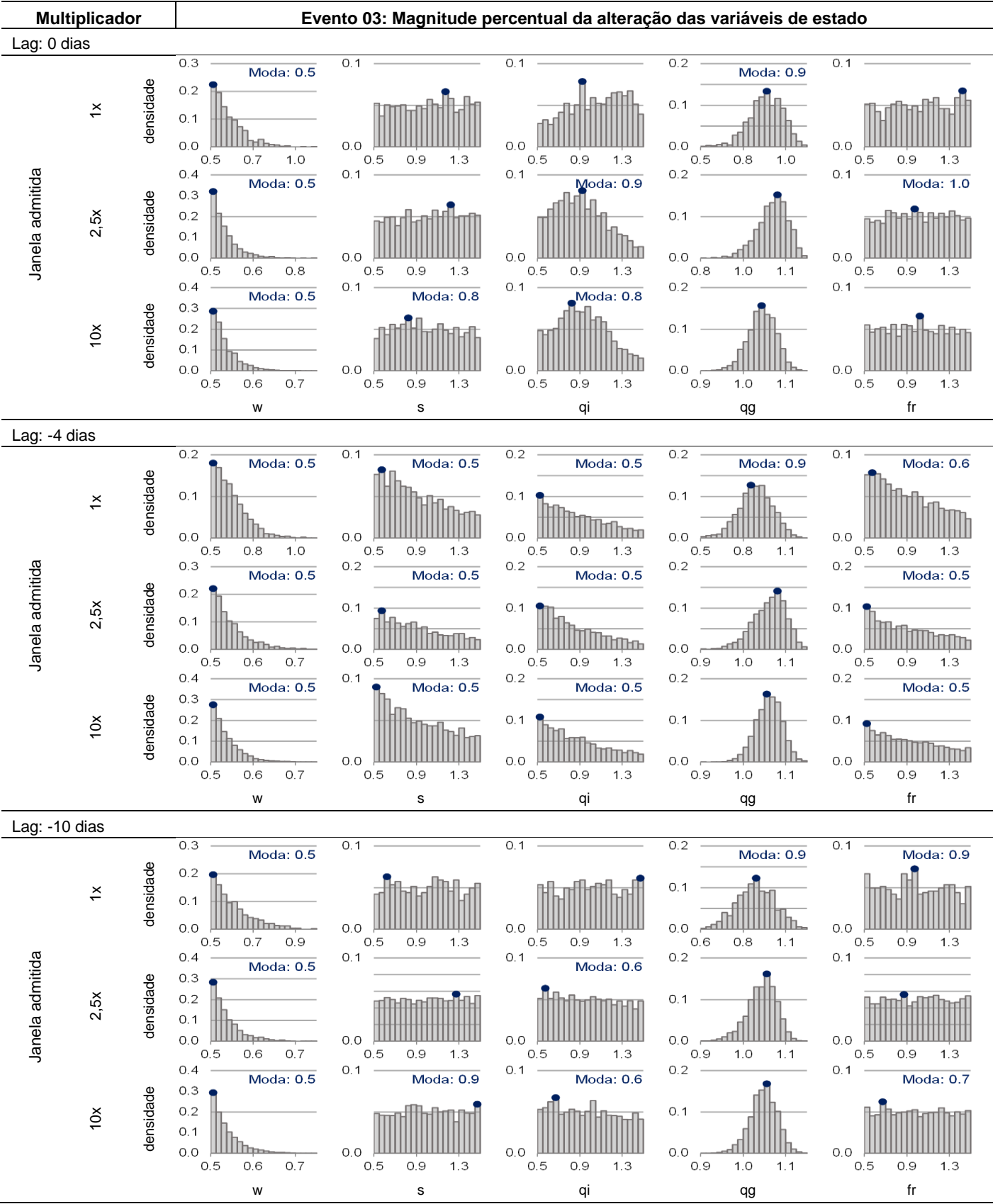
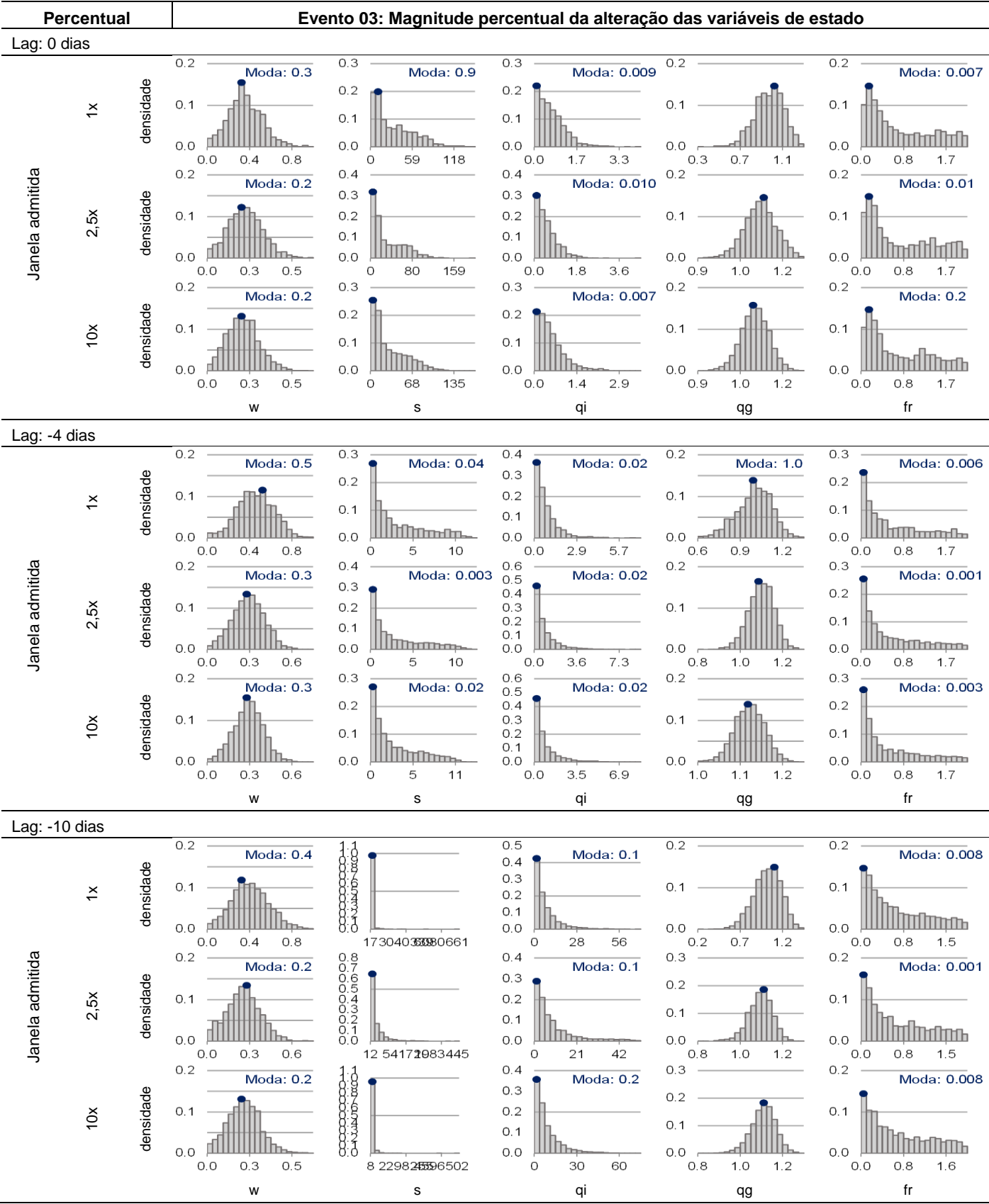
Janela admitida

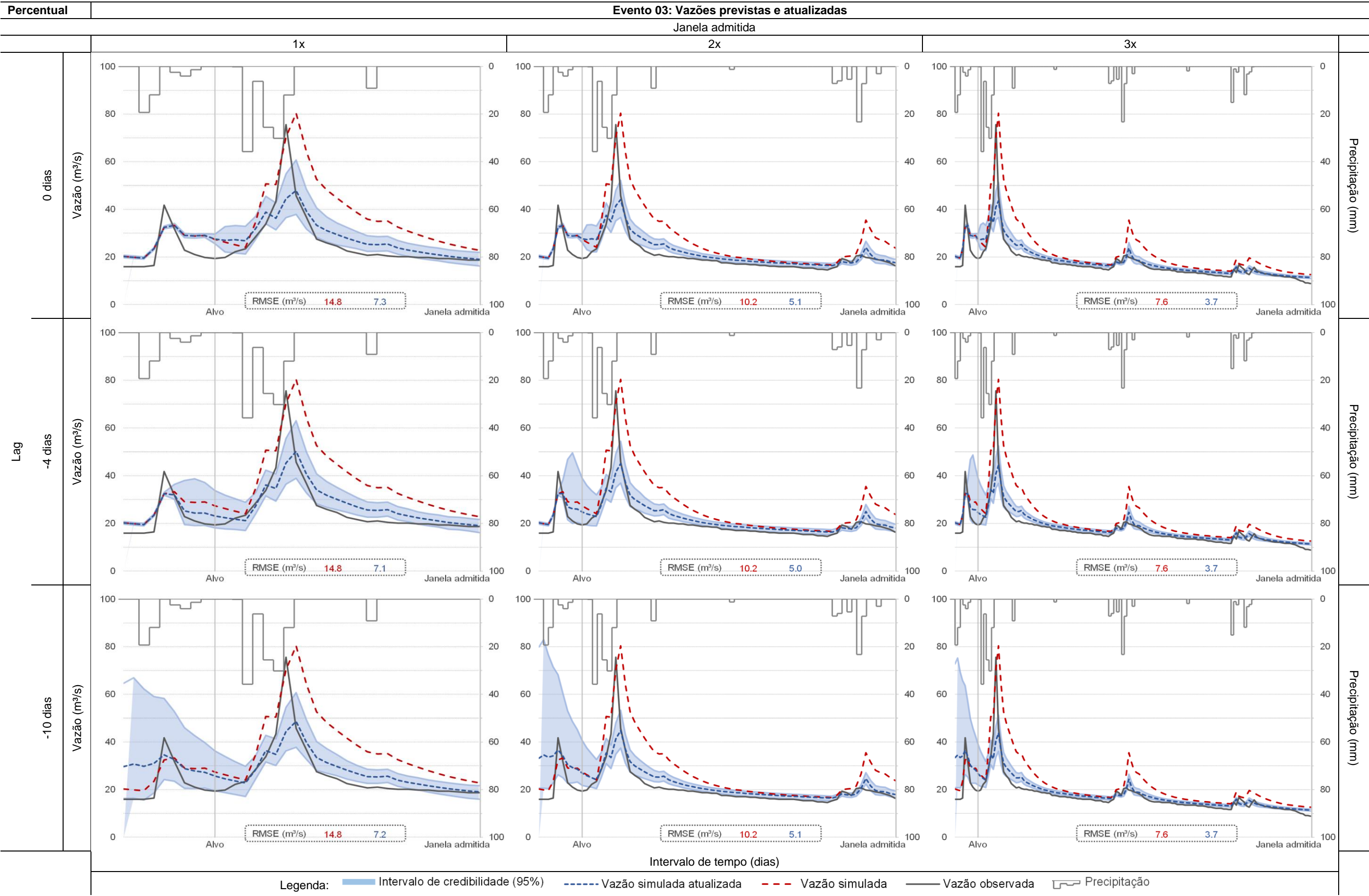


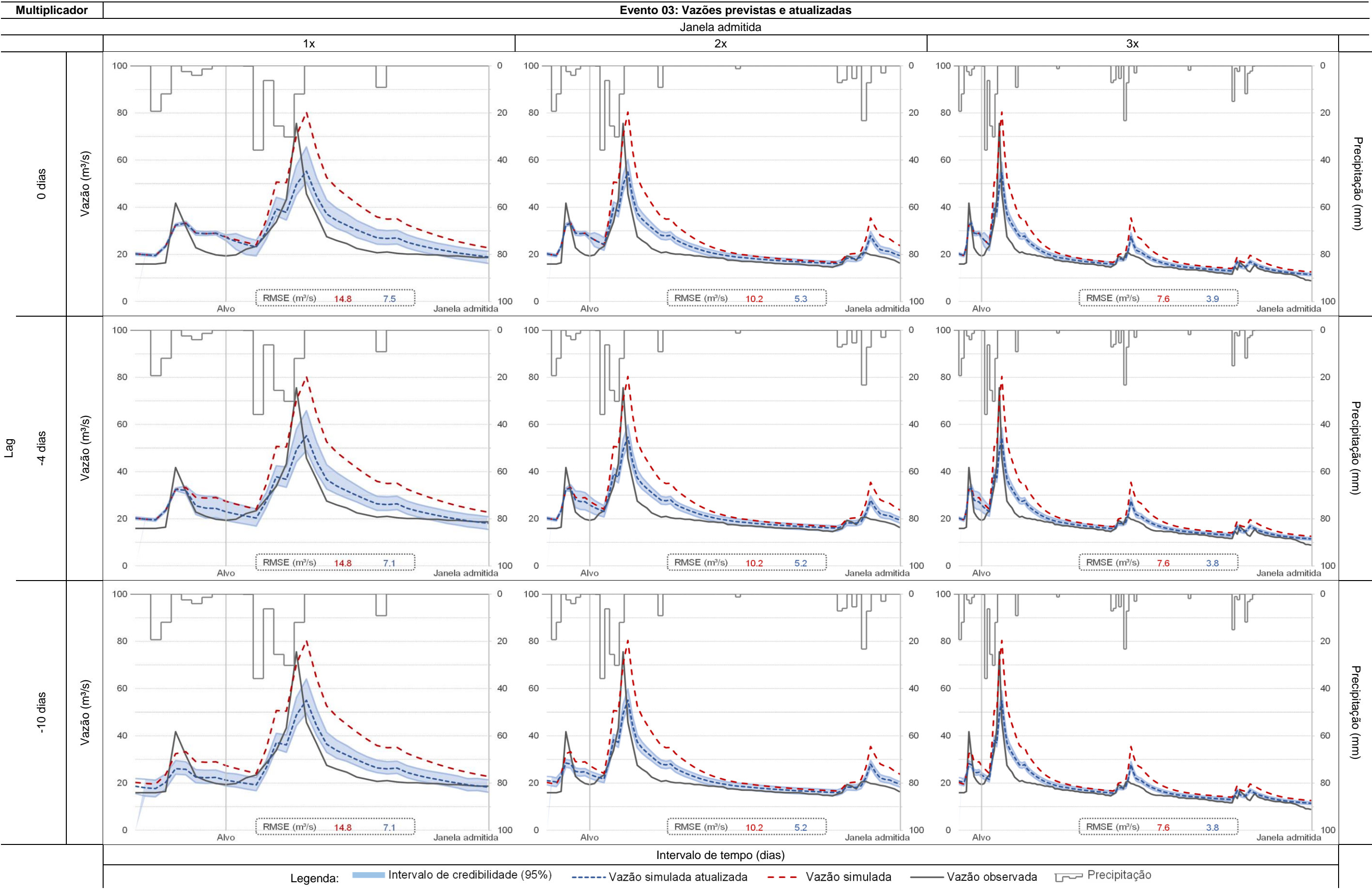
Janela admitida



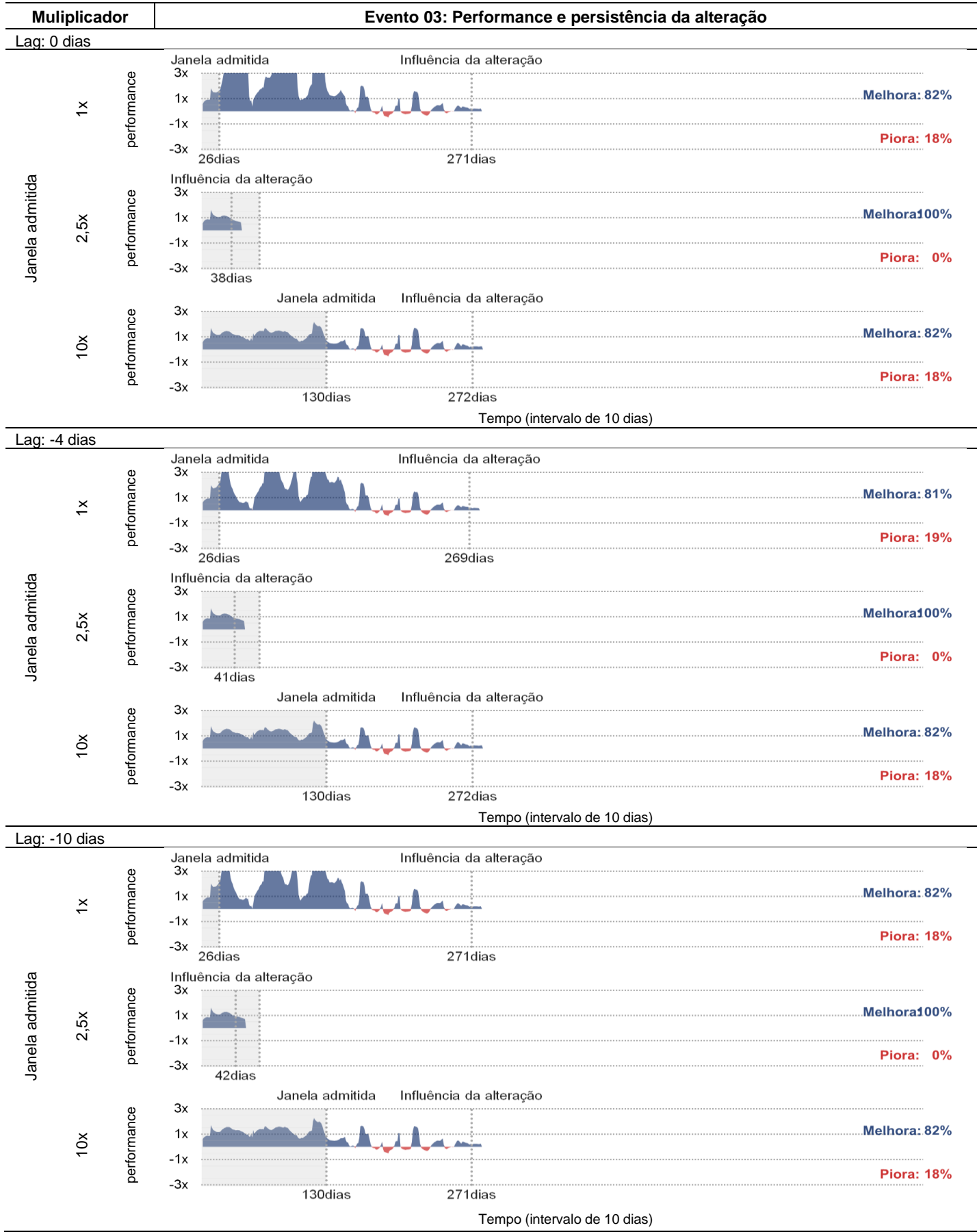
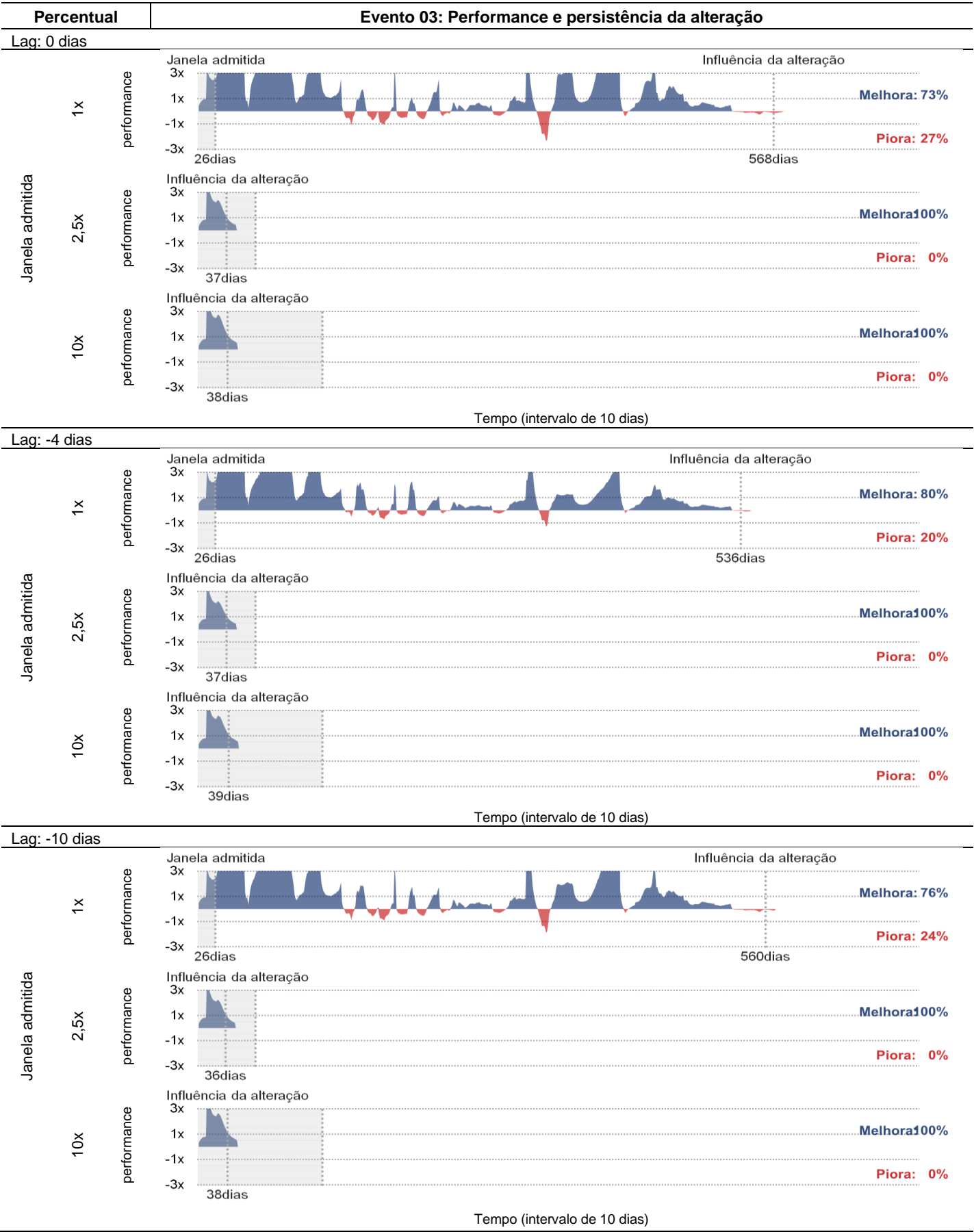






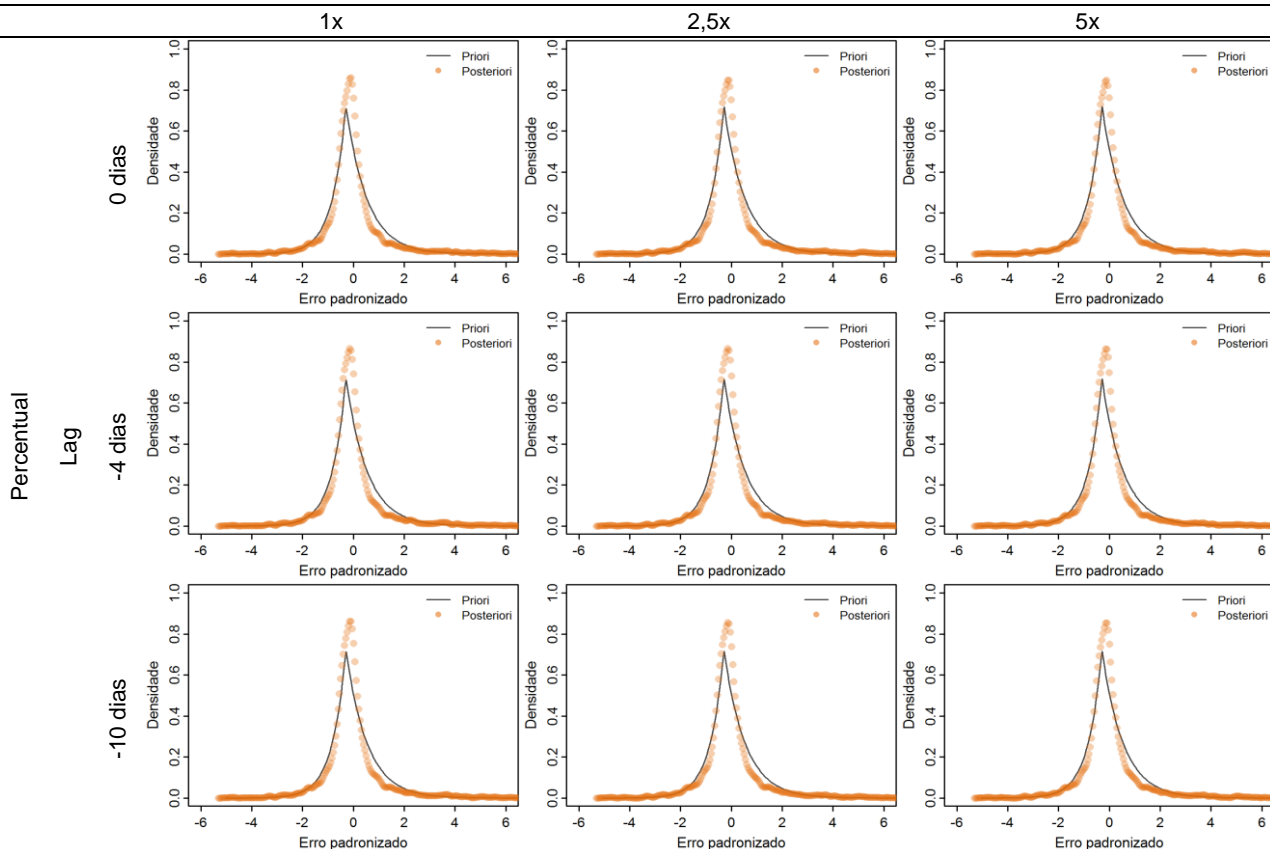




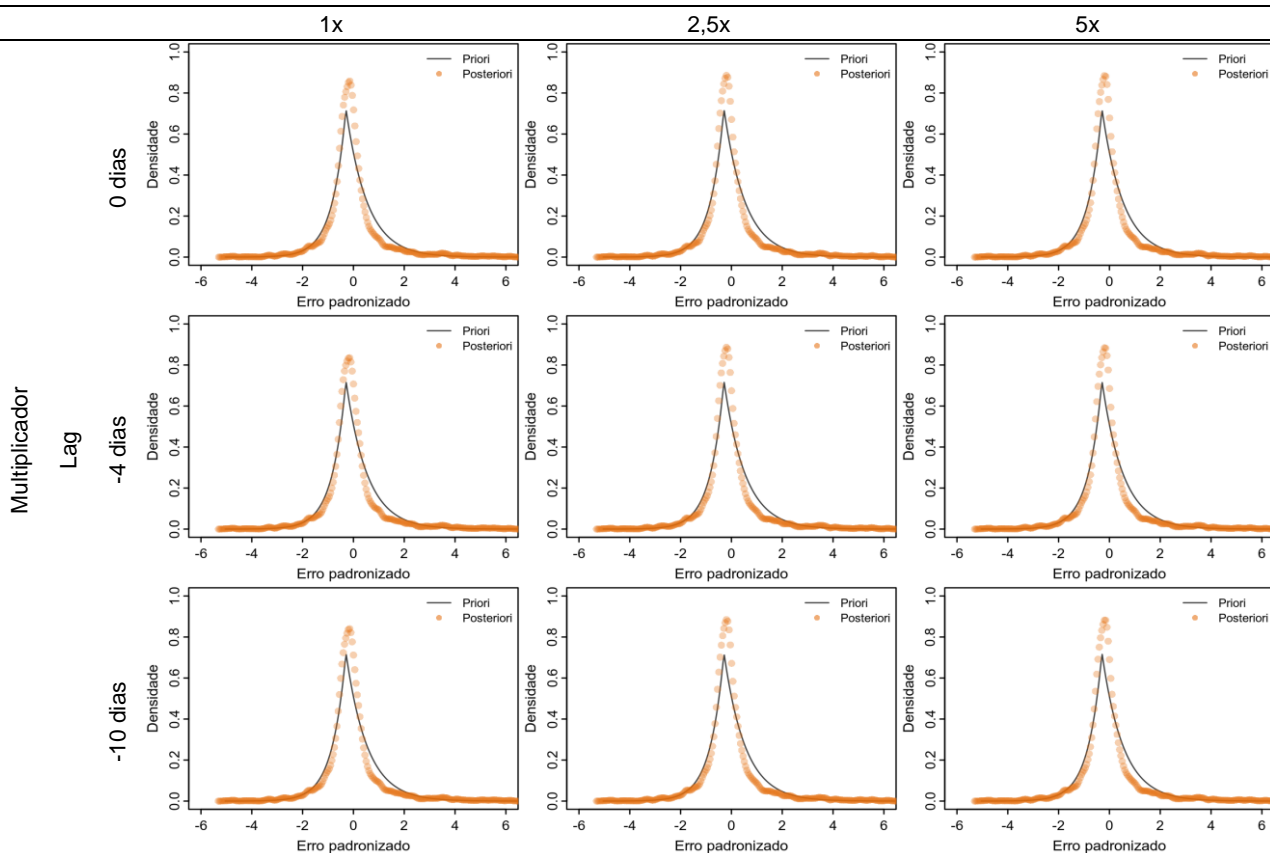


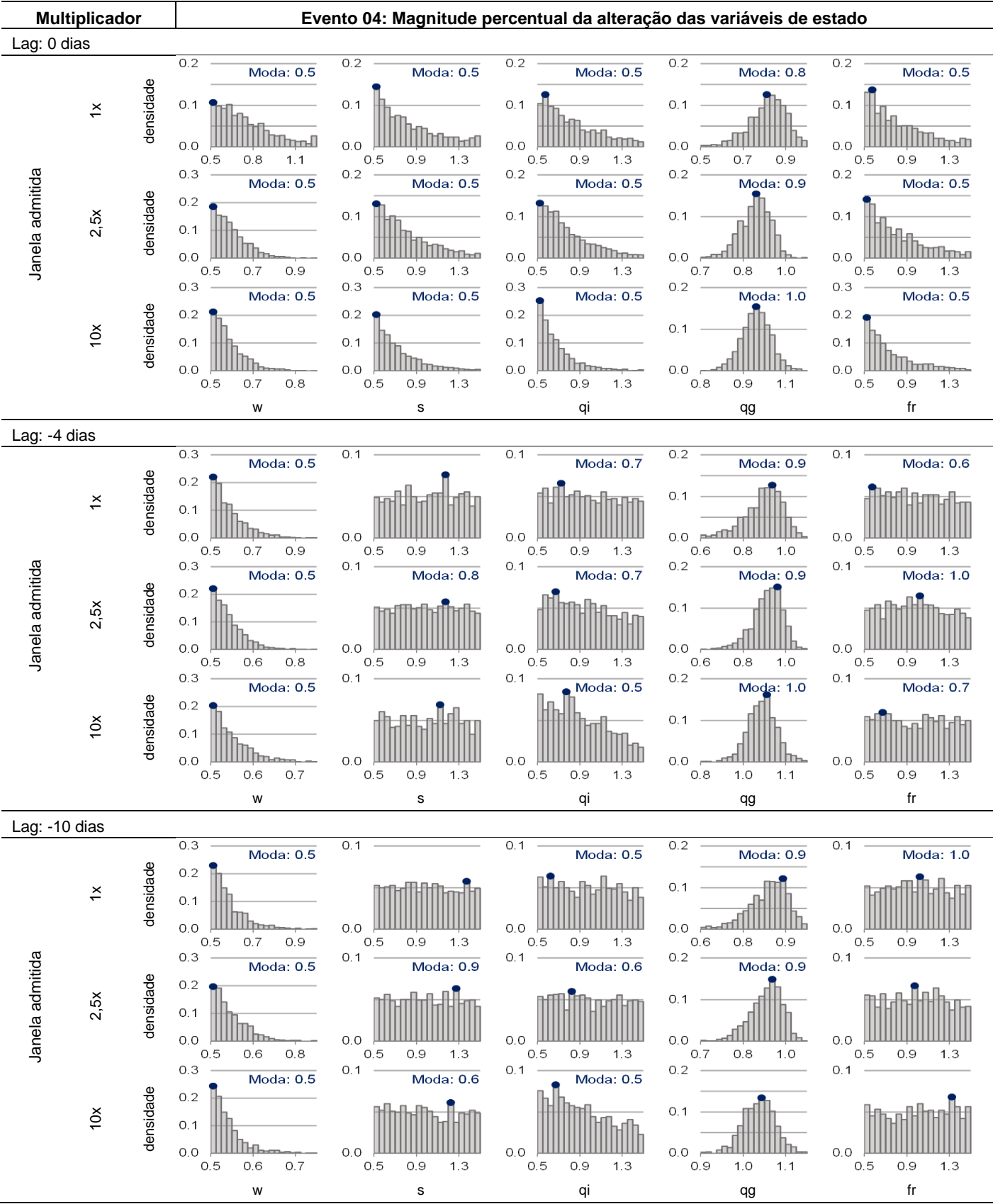
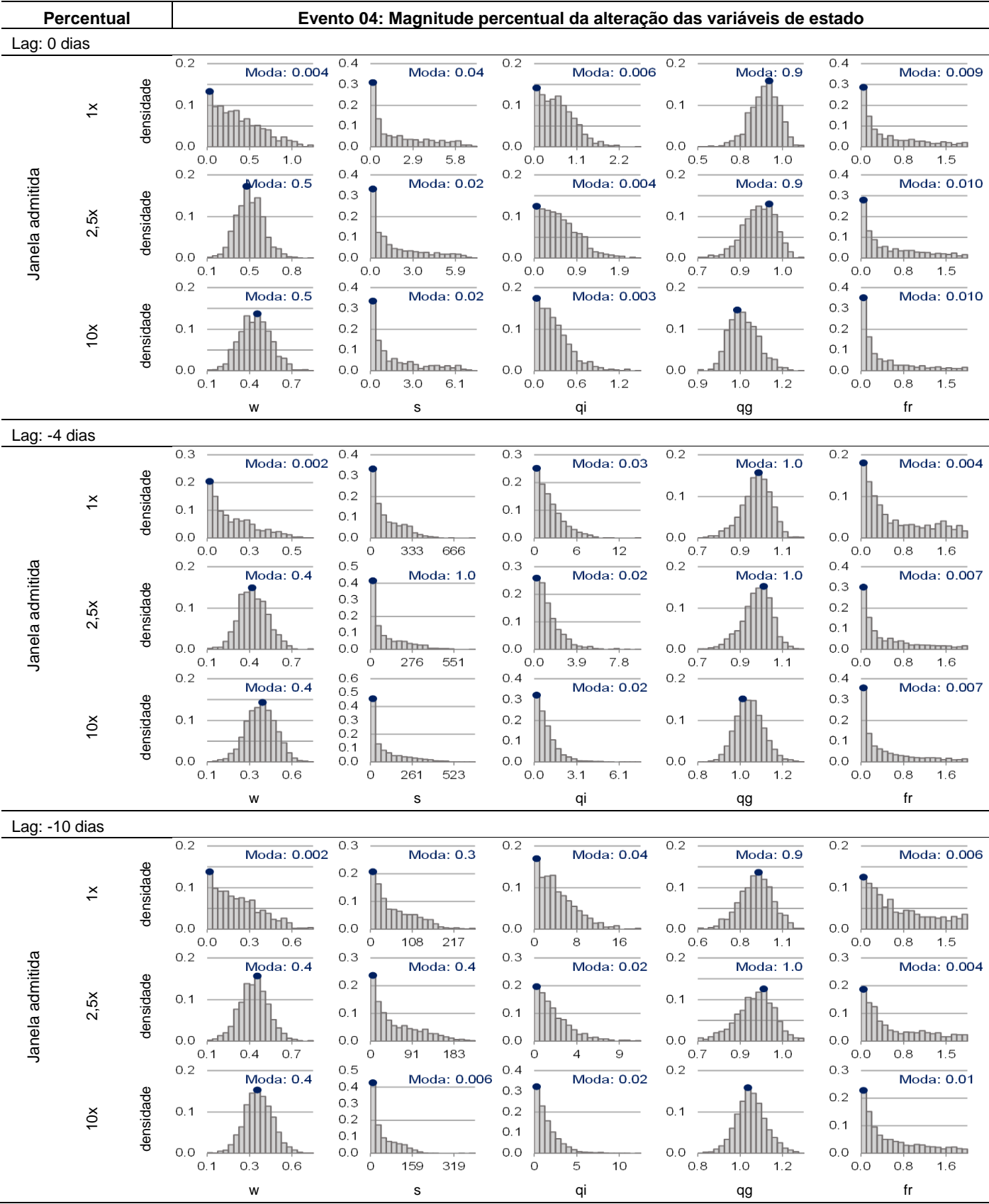
## Evento 03: Comportamento probabilístico dos resíduos da simulação

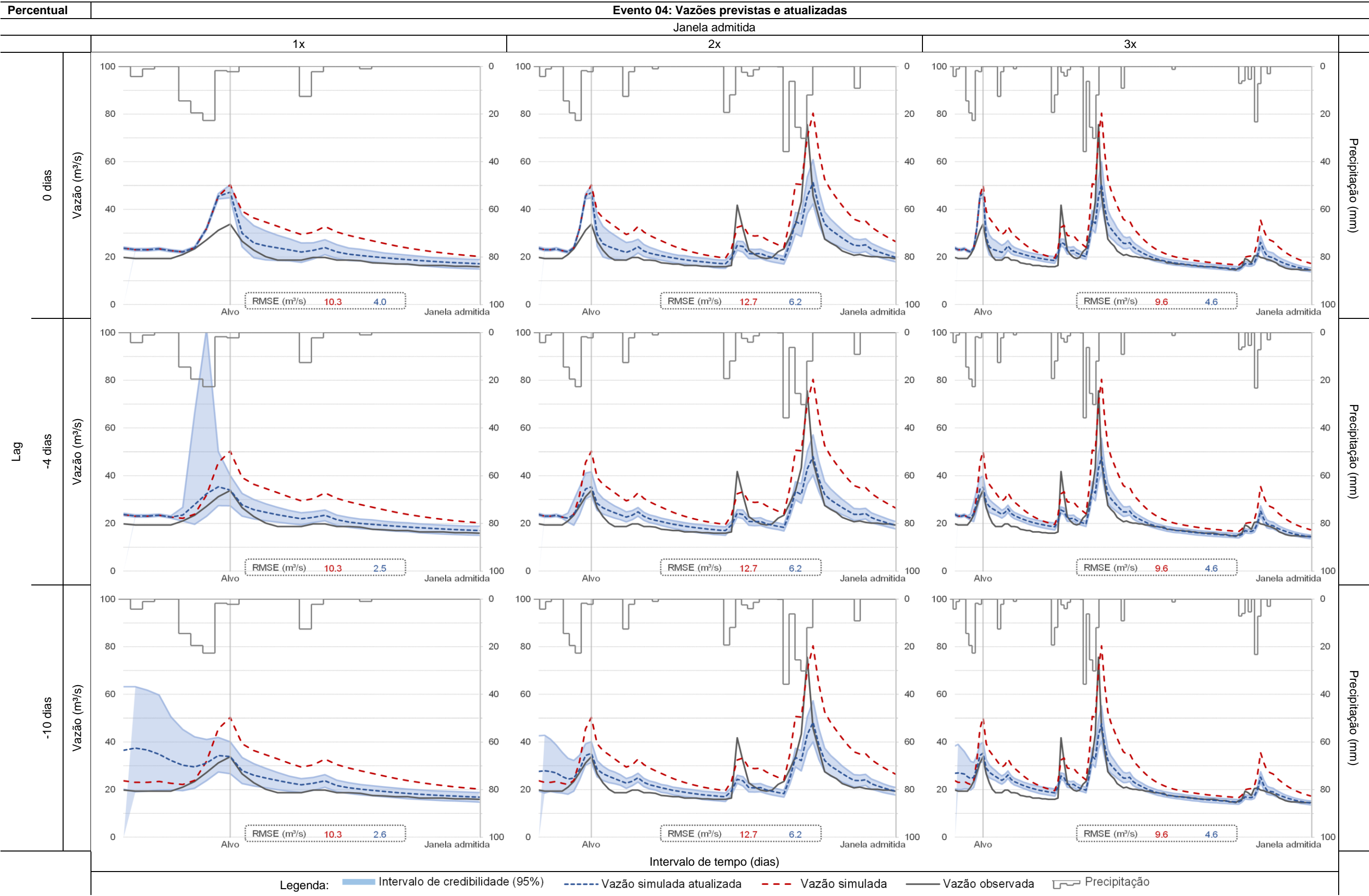
Janela admitida

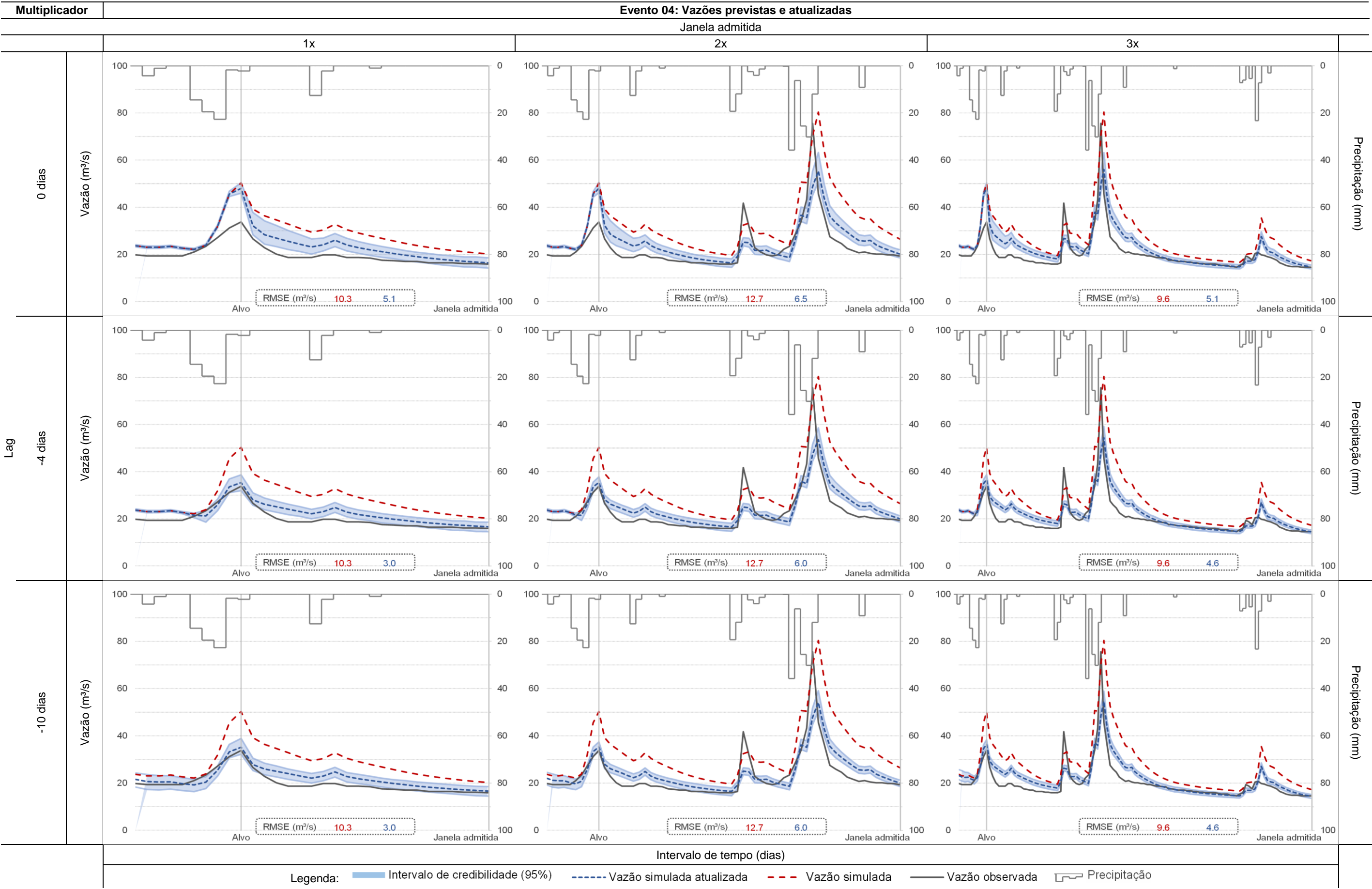


Janela admitida

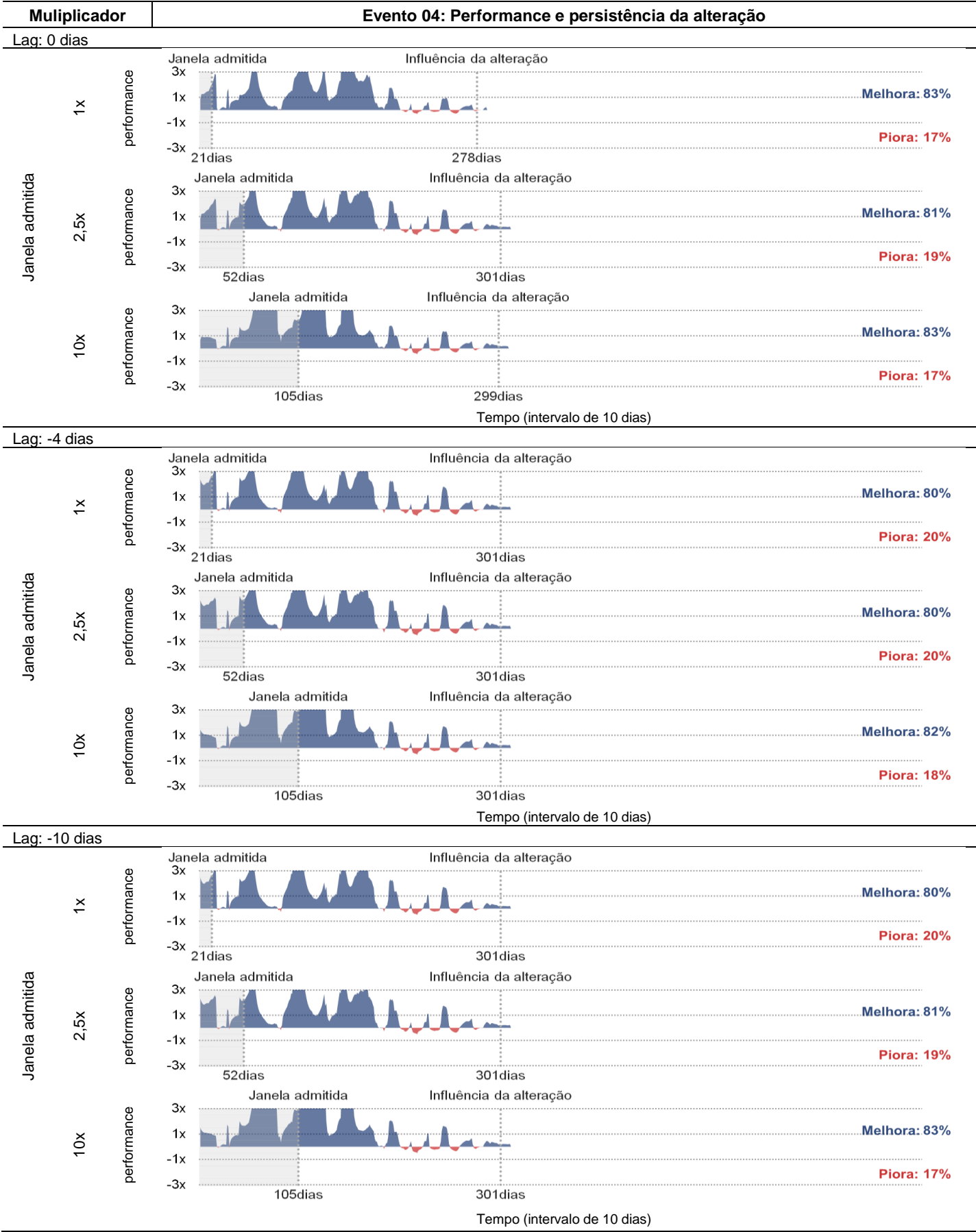
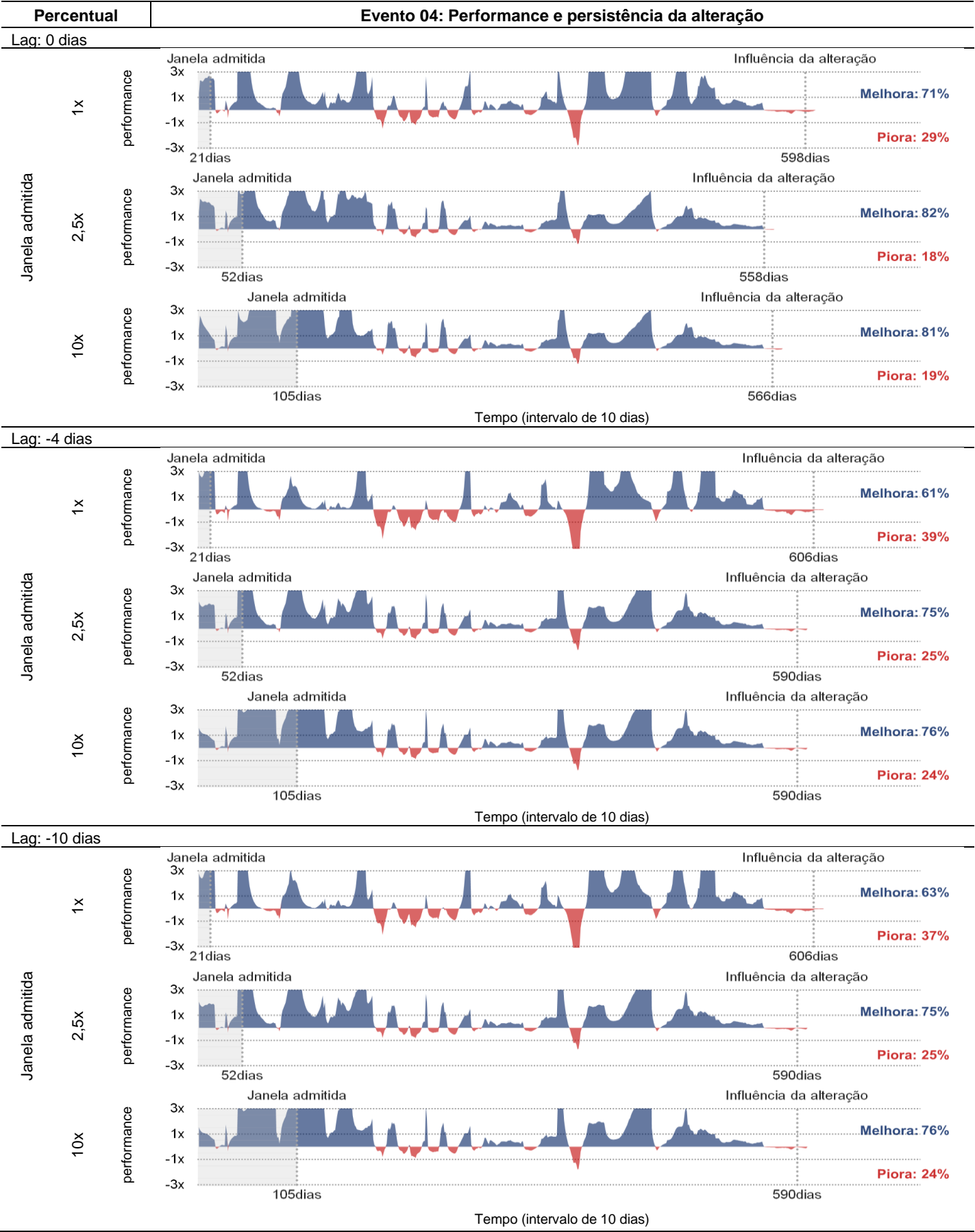






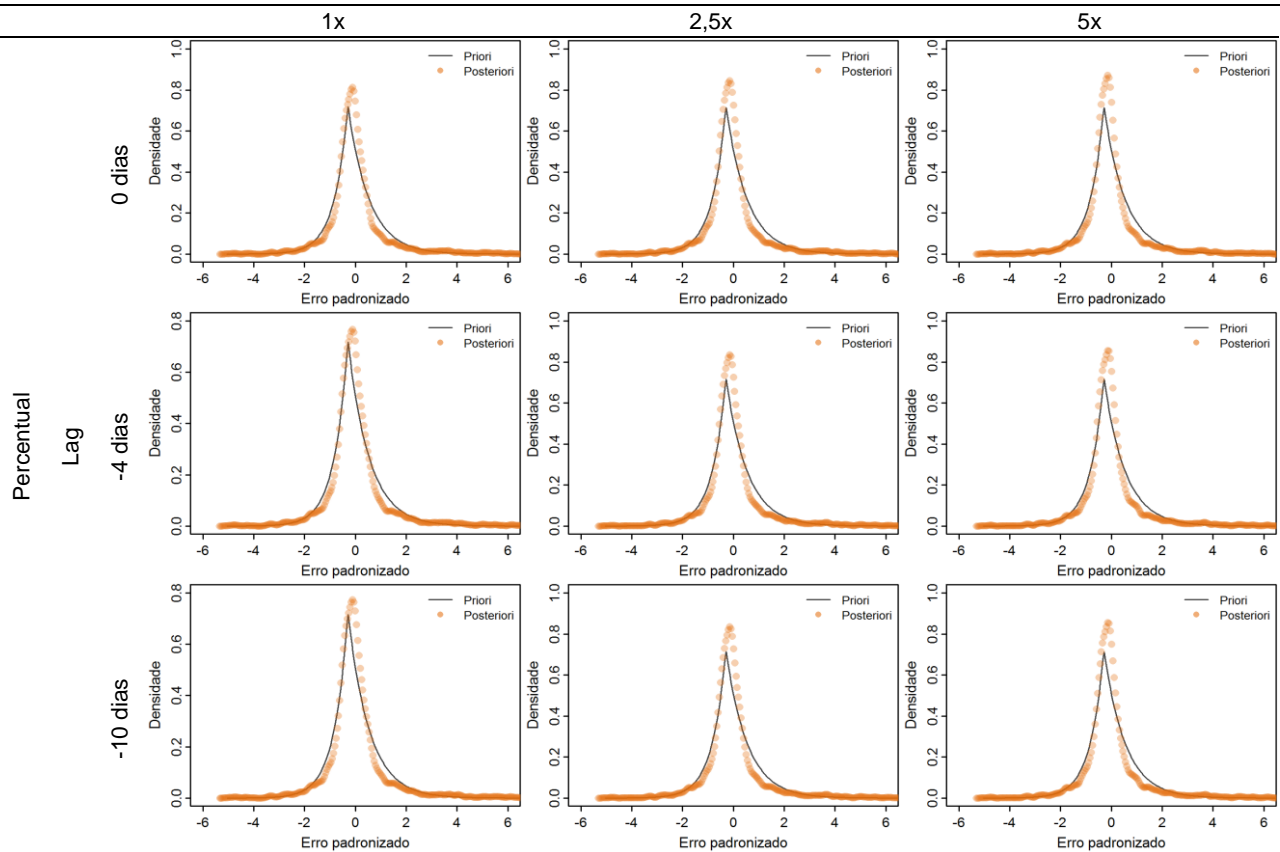




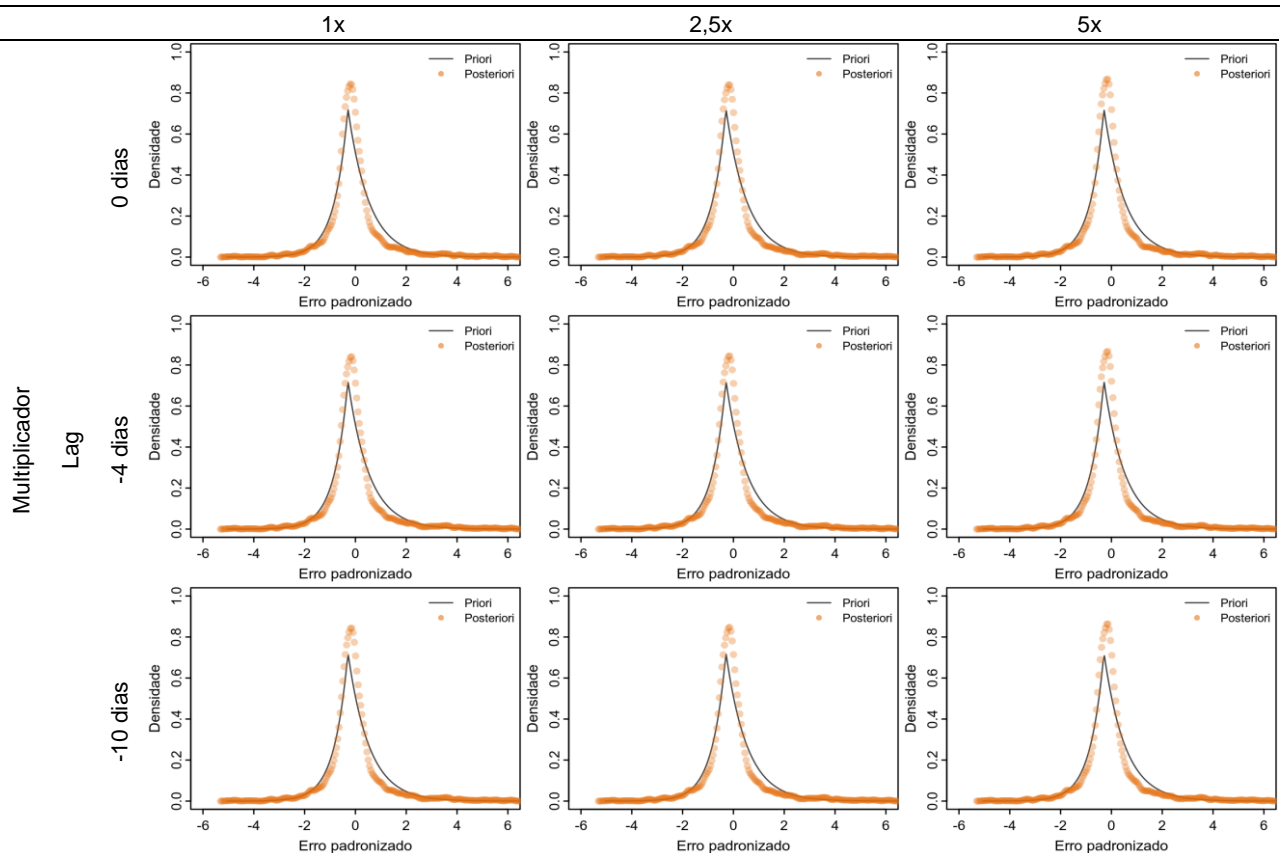


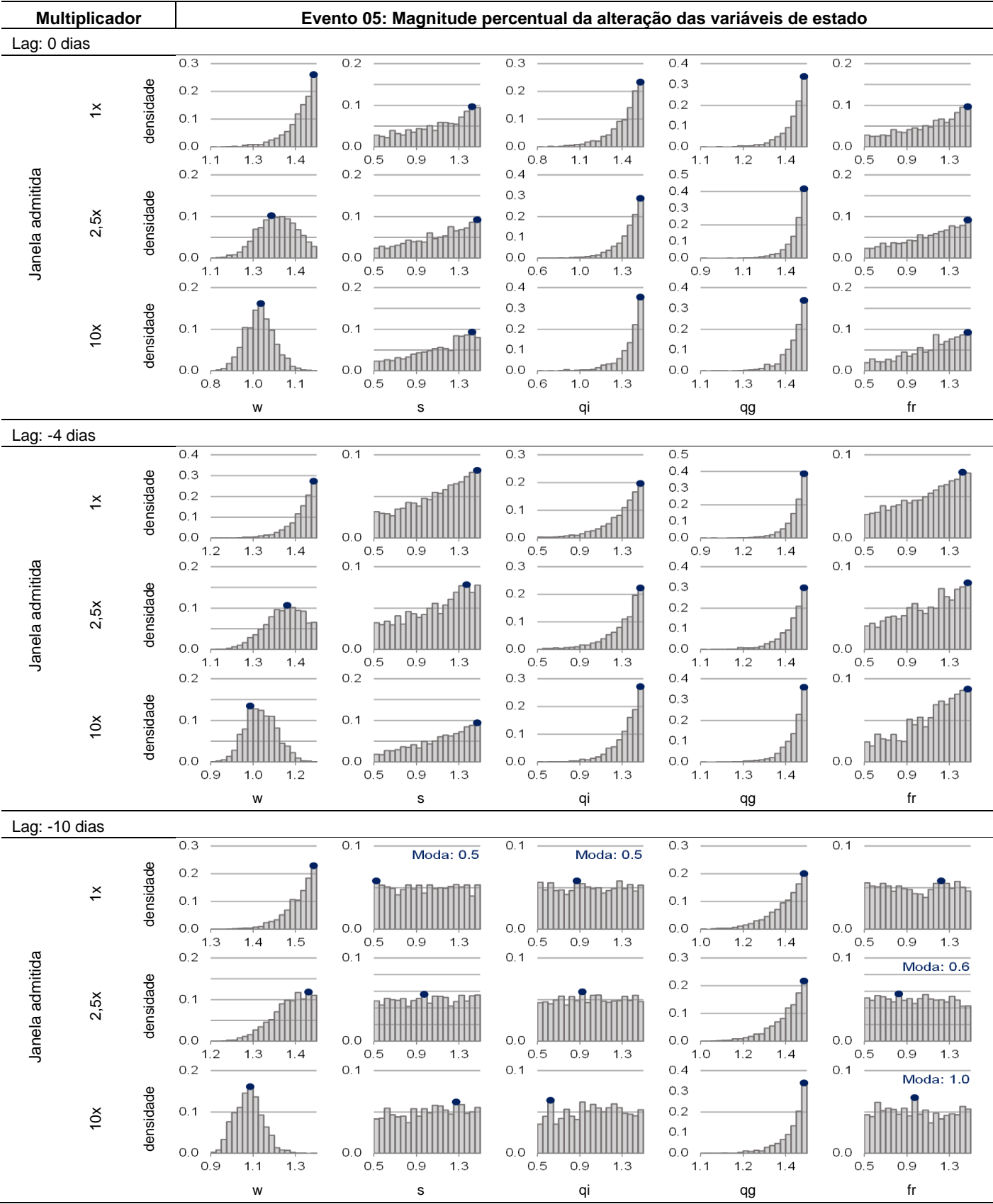
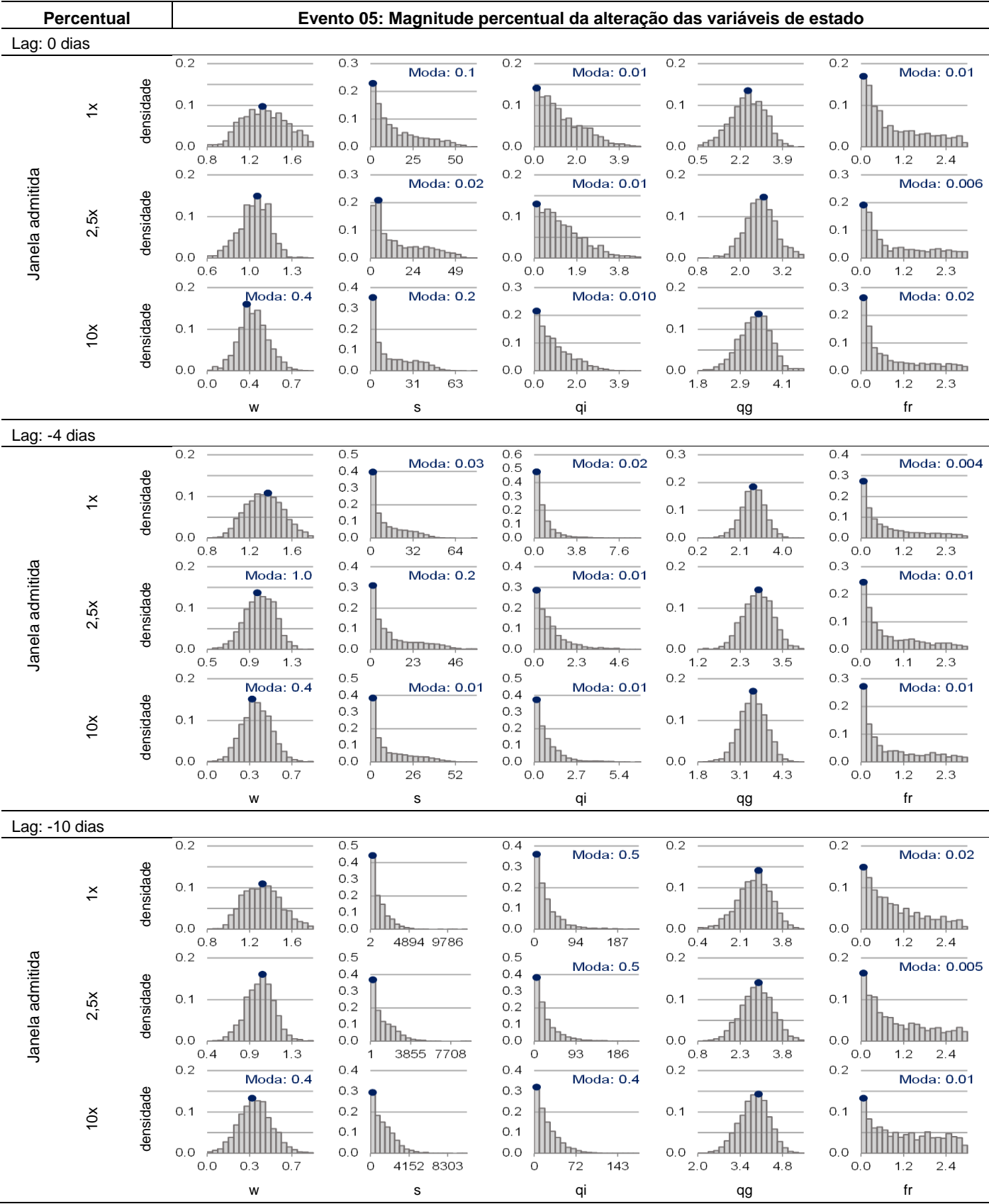
## Evento 04: Comportamento probabilístico dos resíduos da simulação

Janela admitida

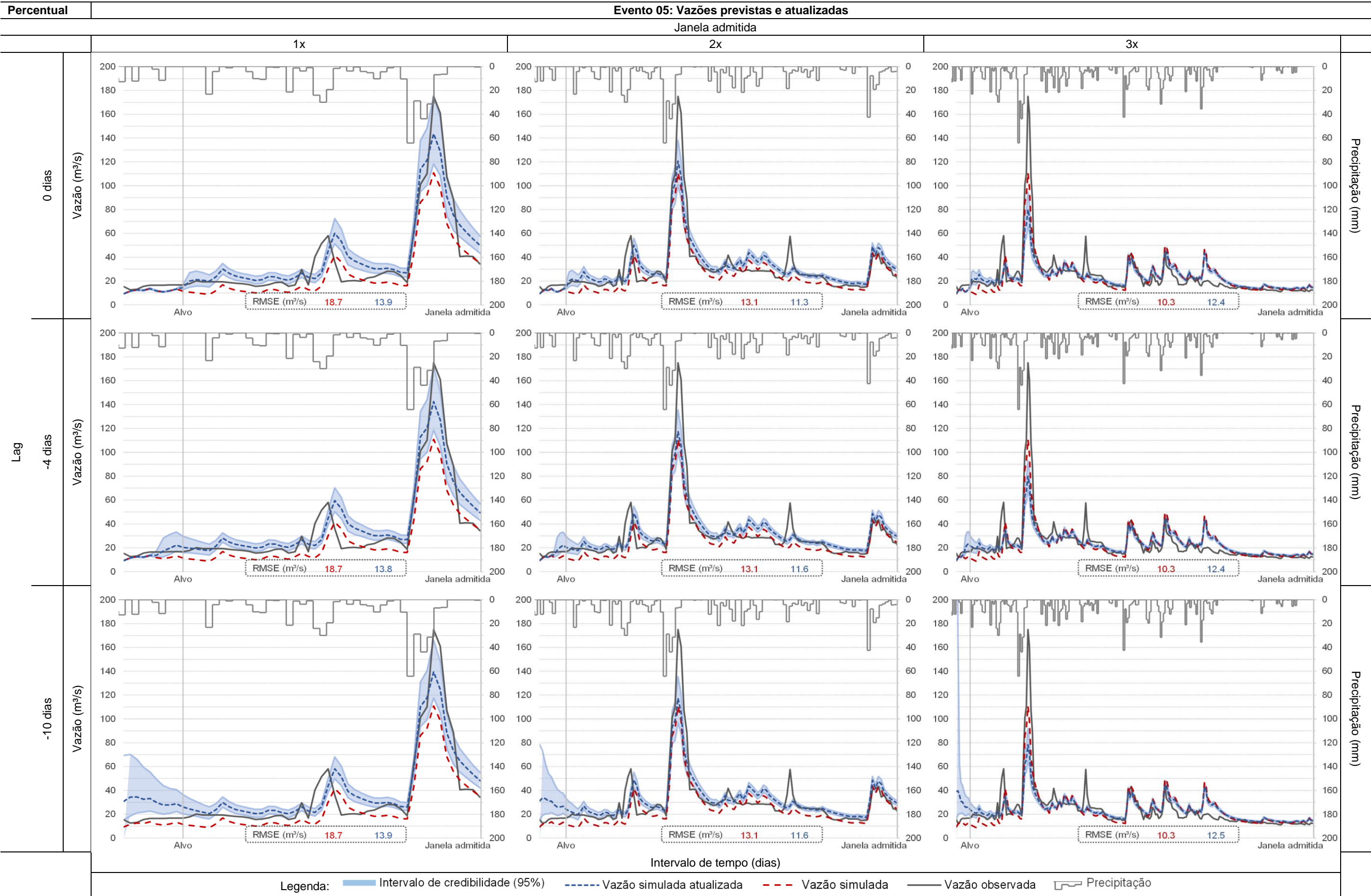


Janela admitida

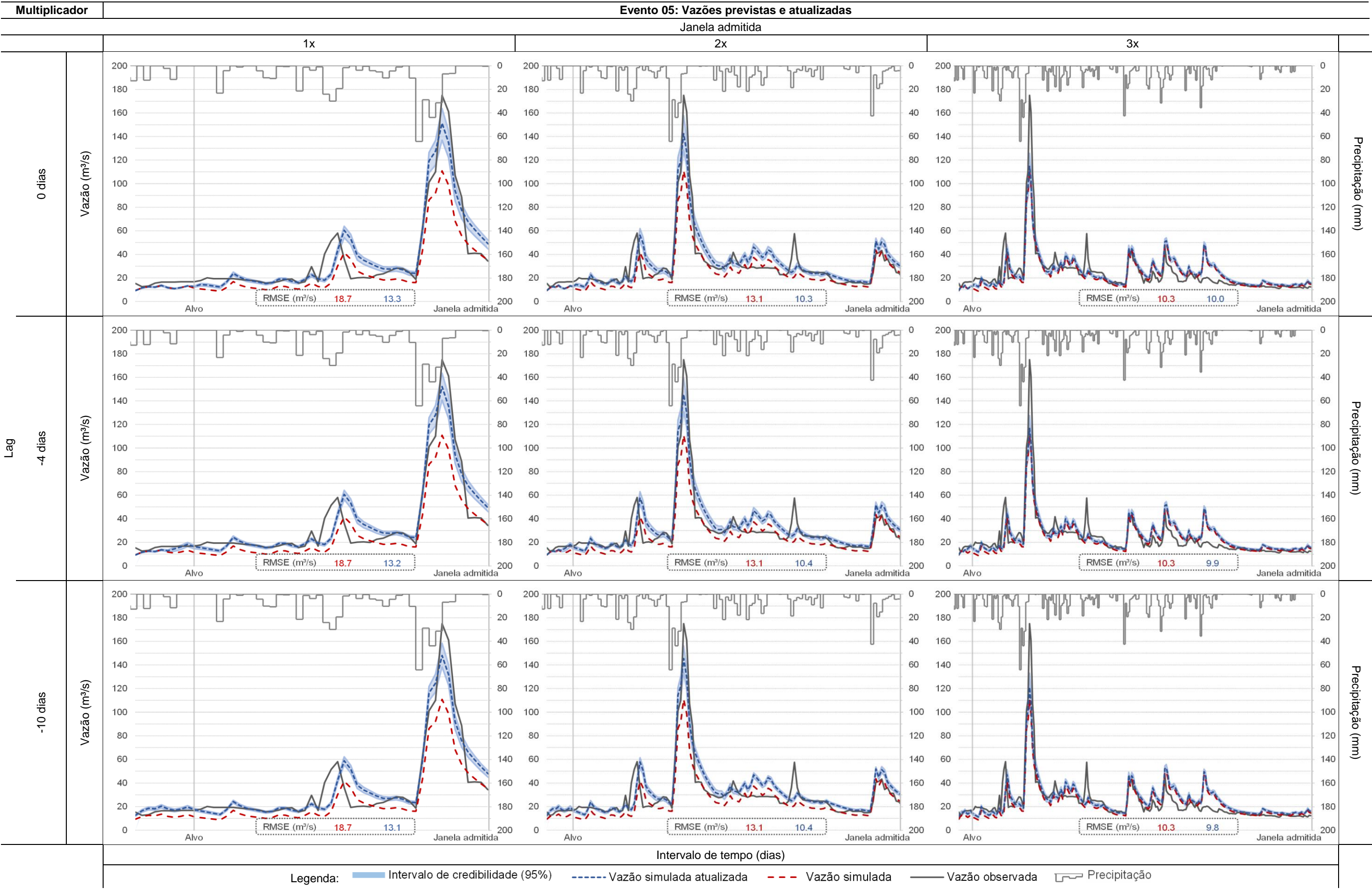


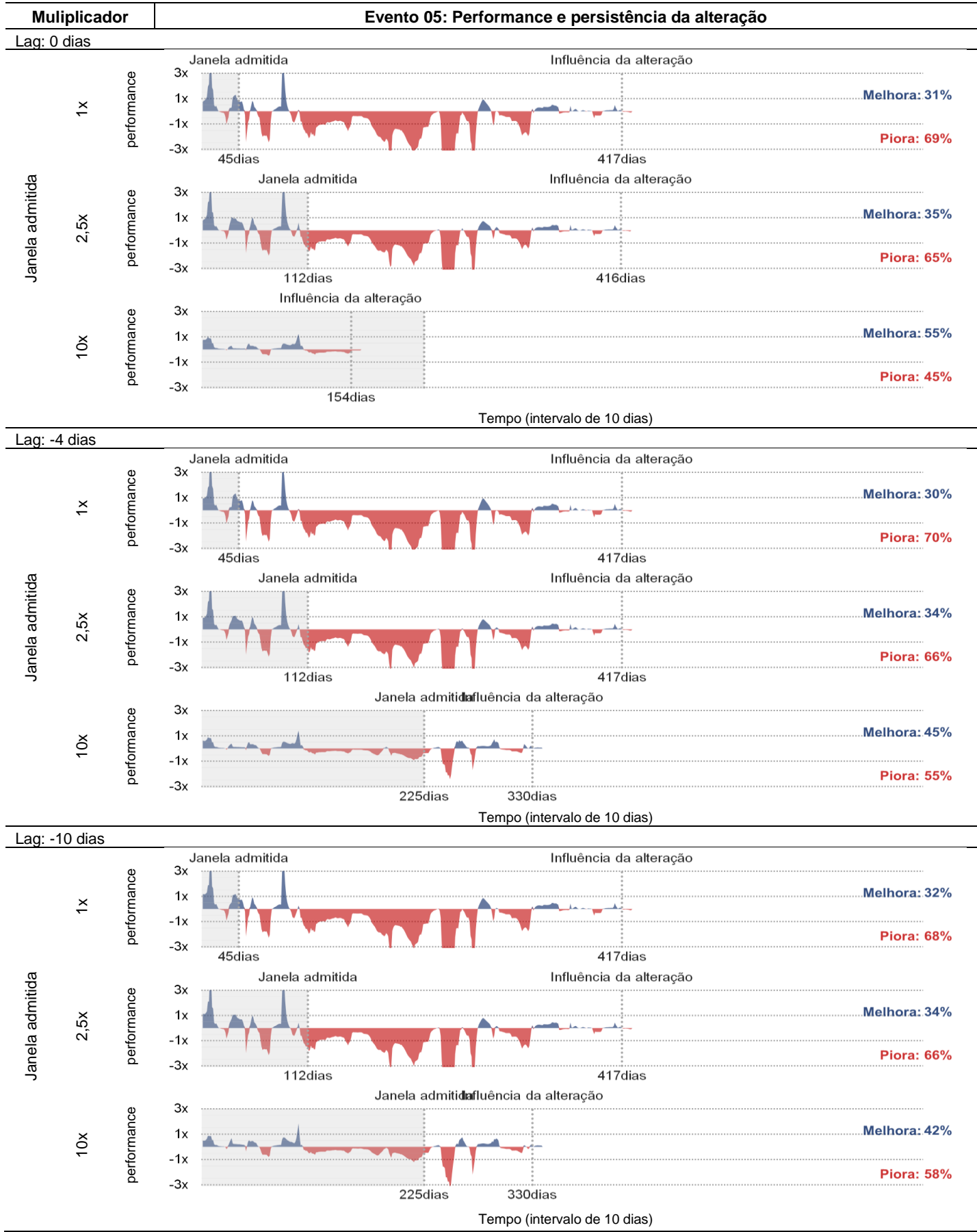
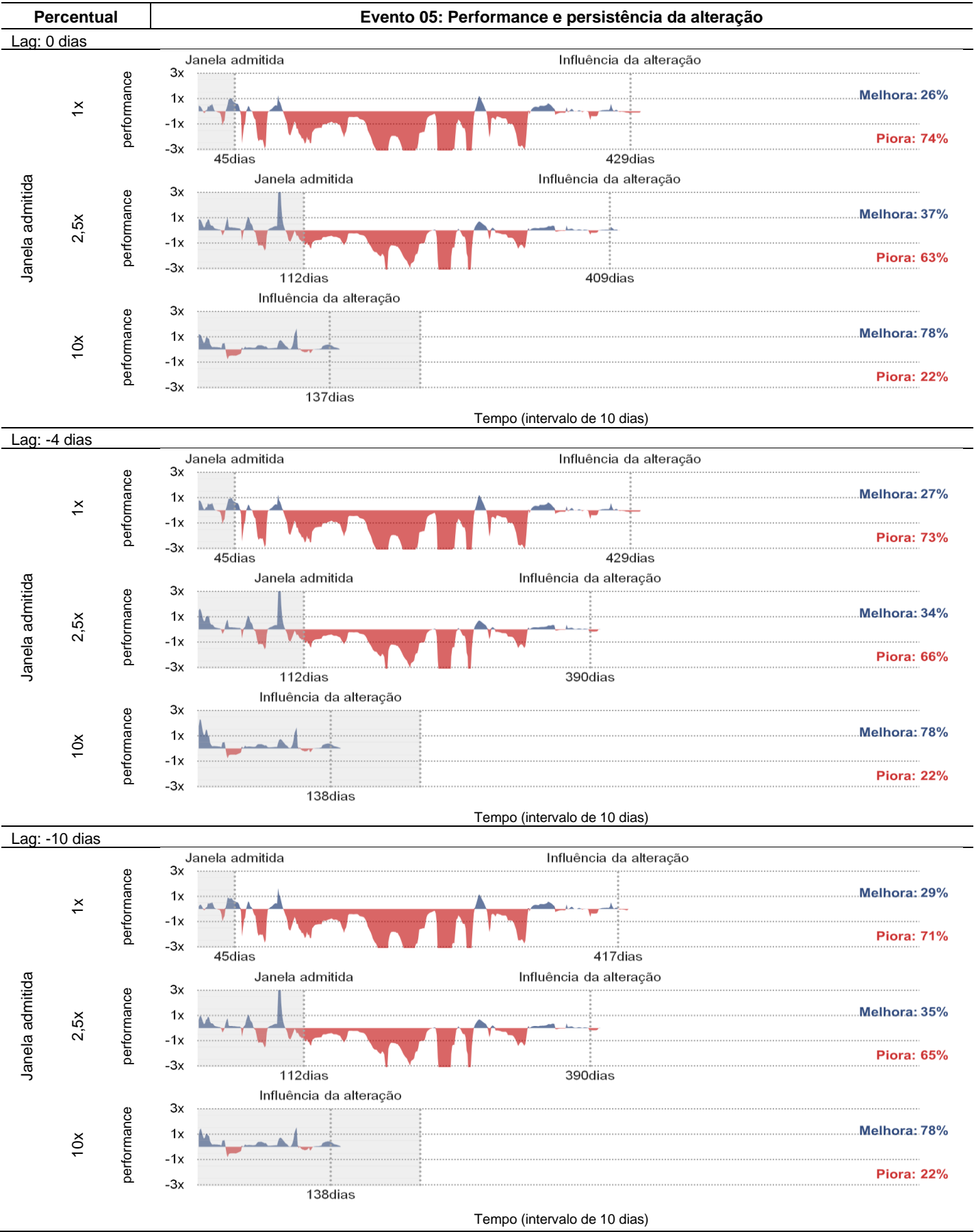






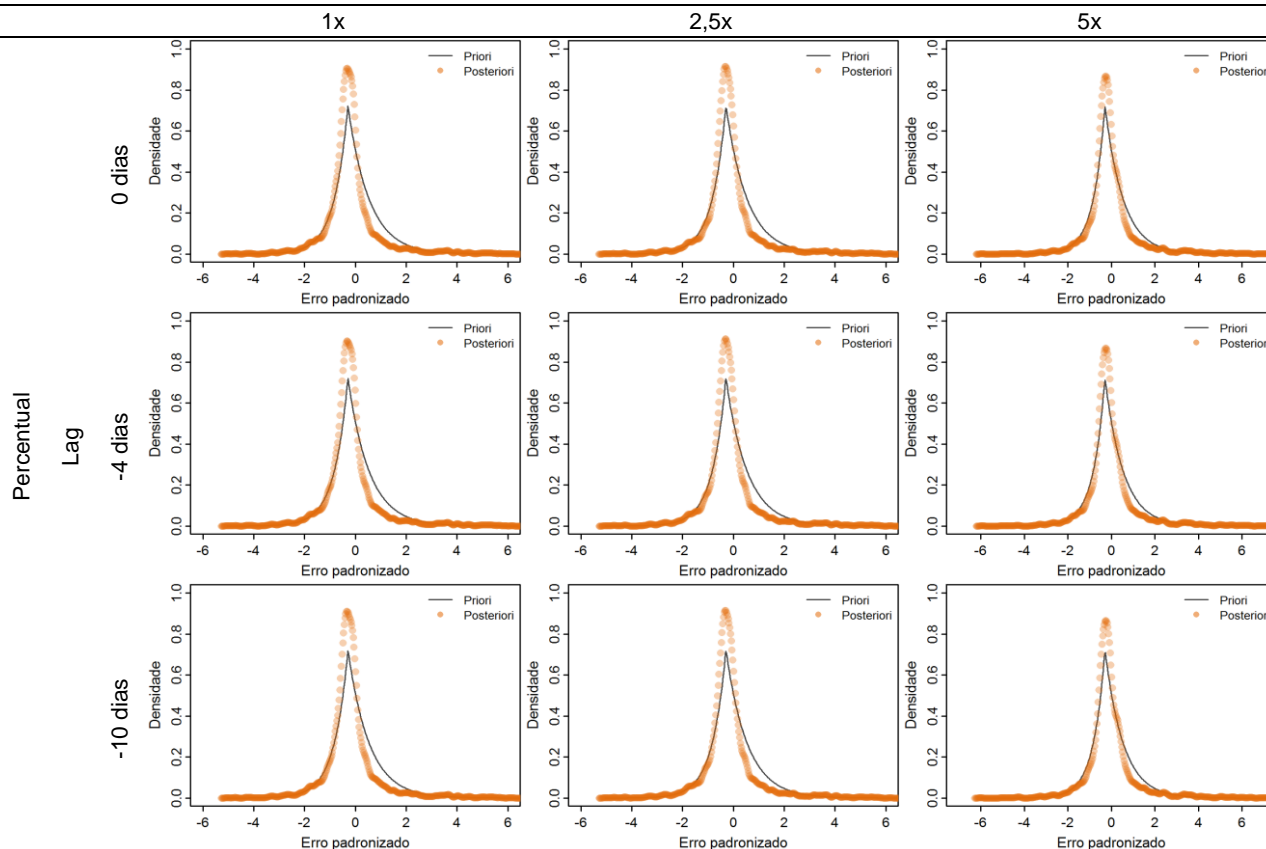




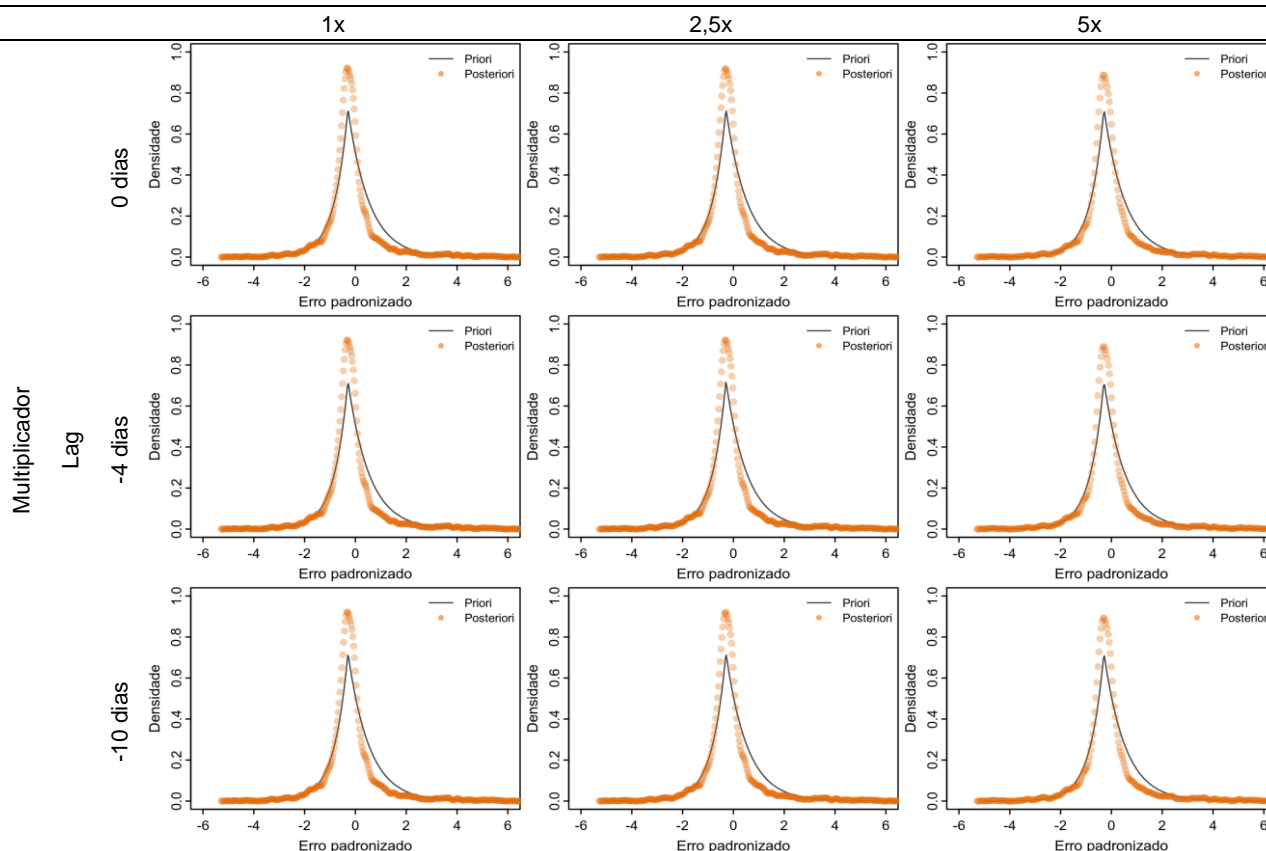


## Evento 05: Comportamento probabilístico dos resíduos da simulação

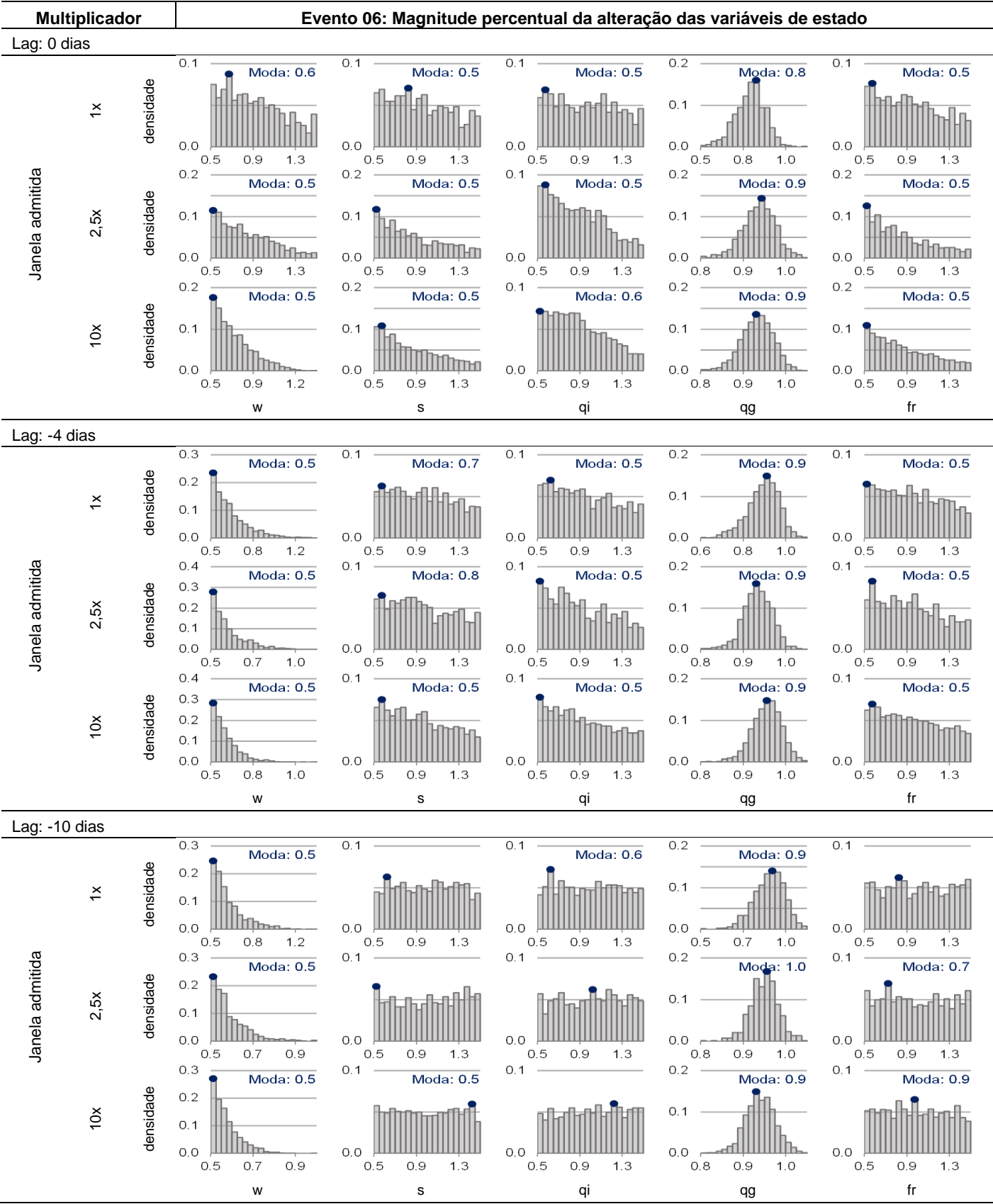
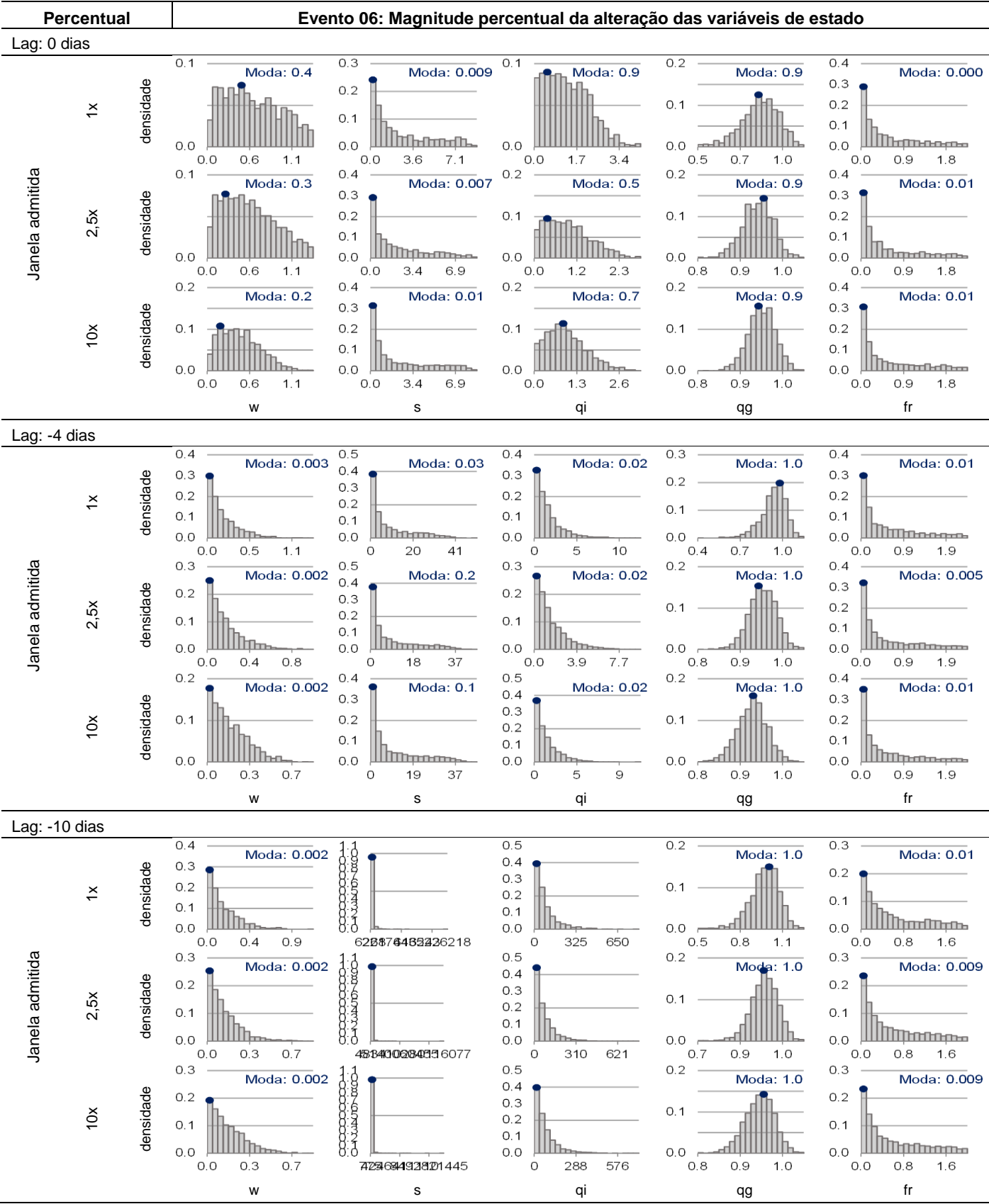
Janela admitida



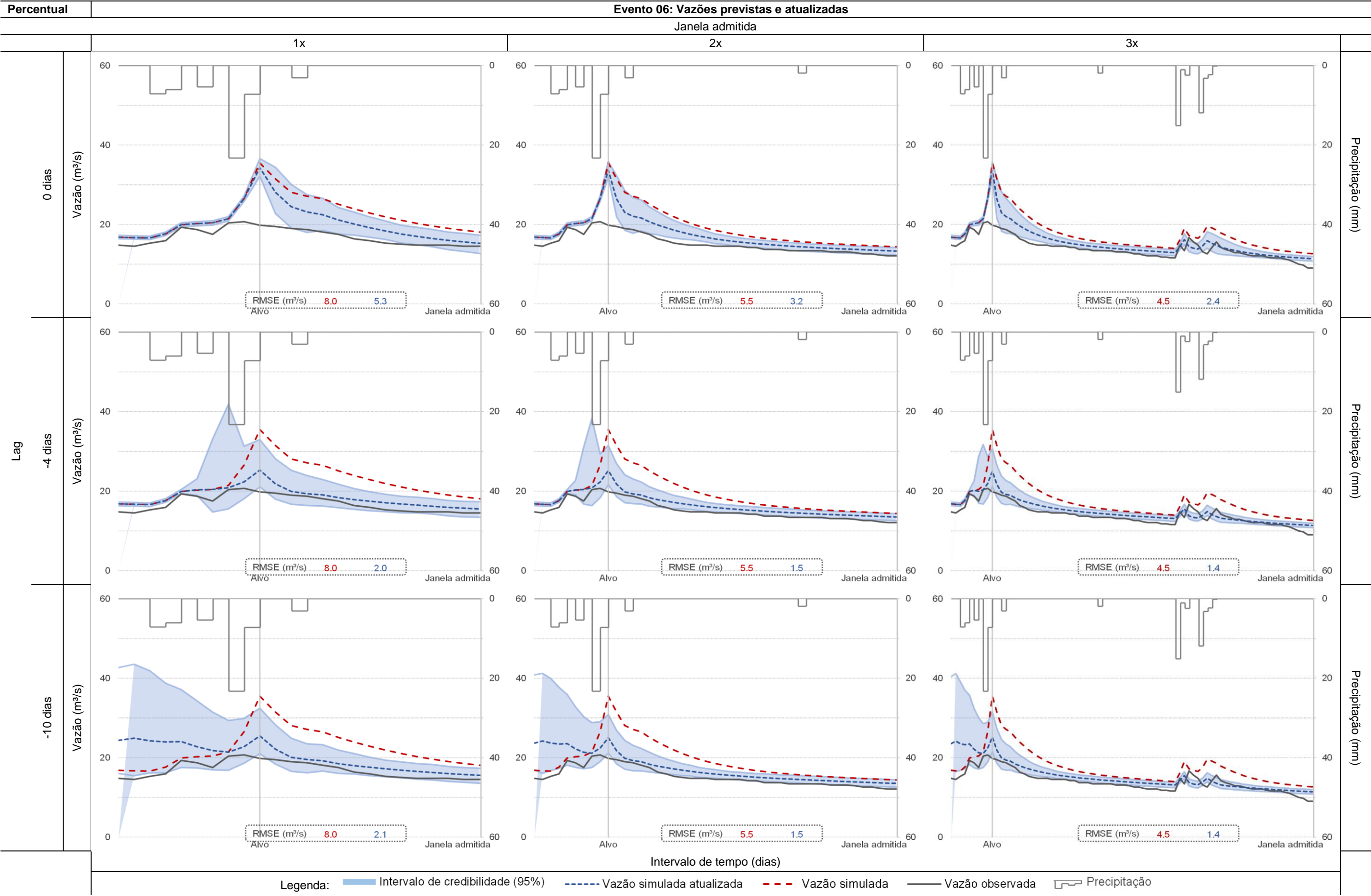
Janela admitida

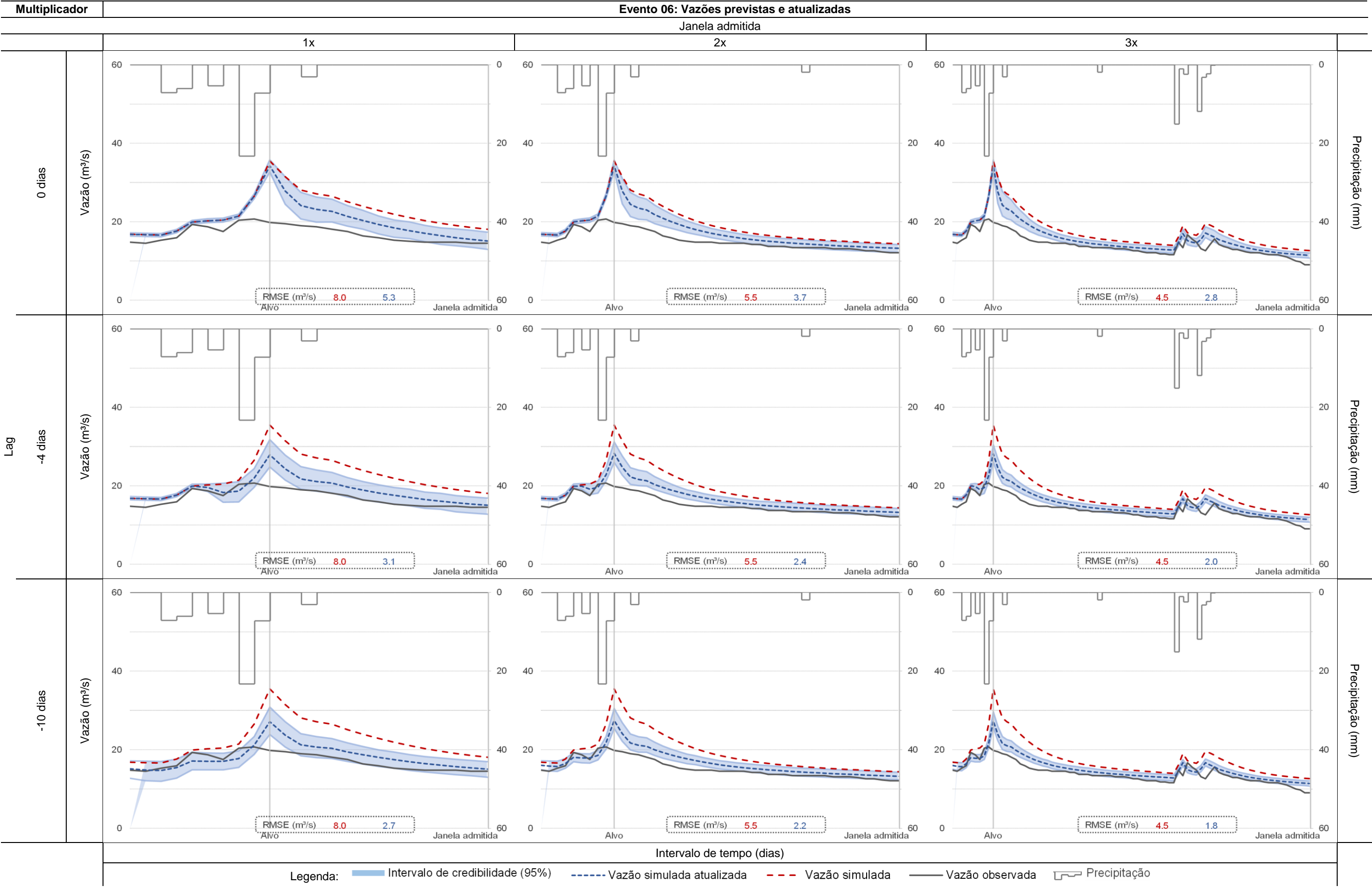


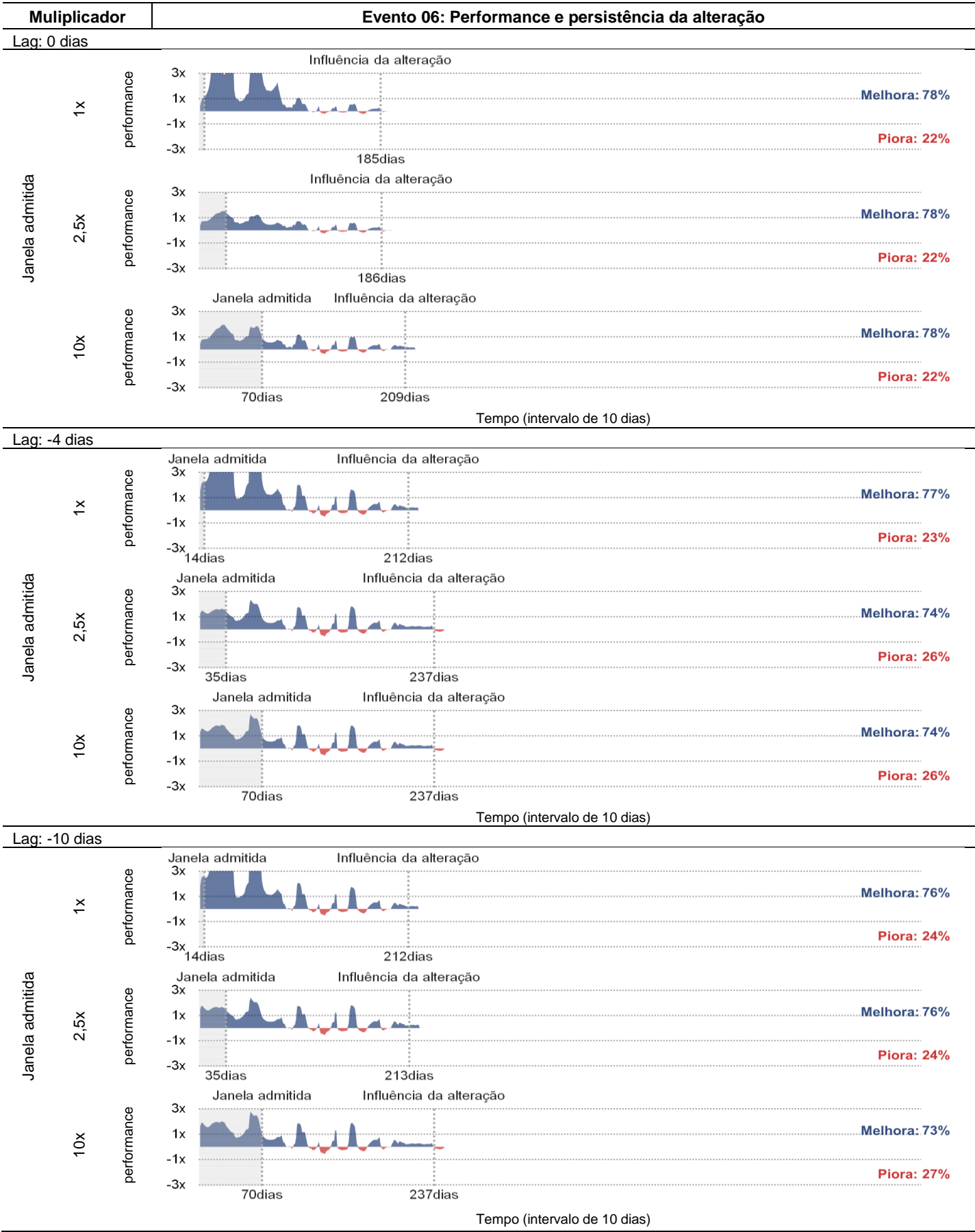
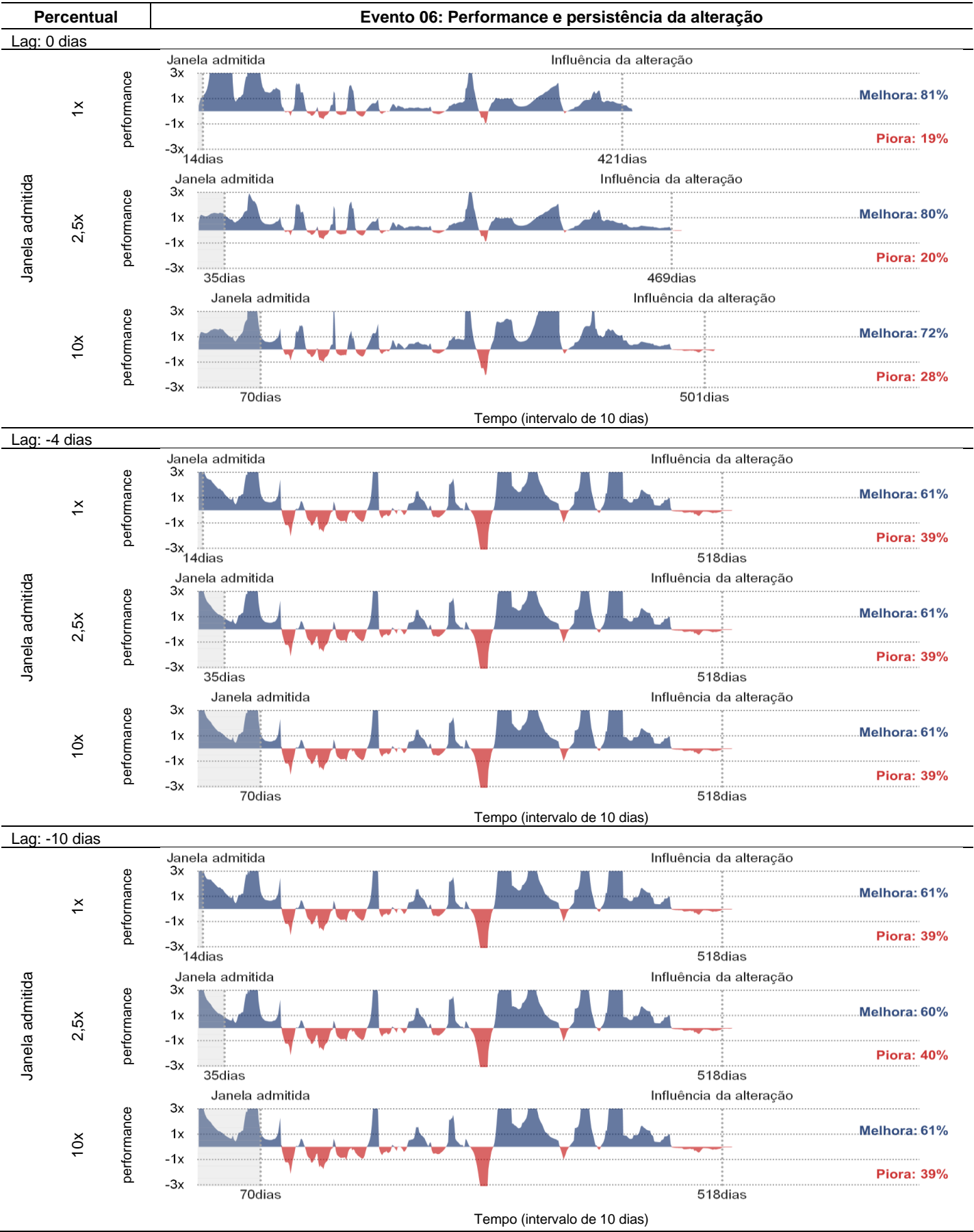






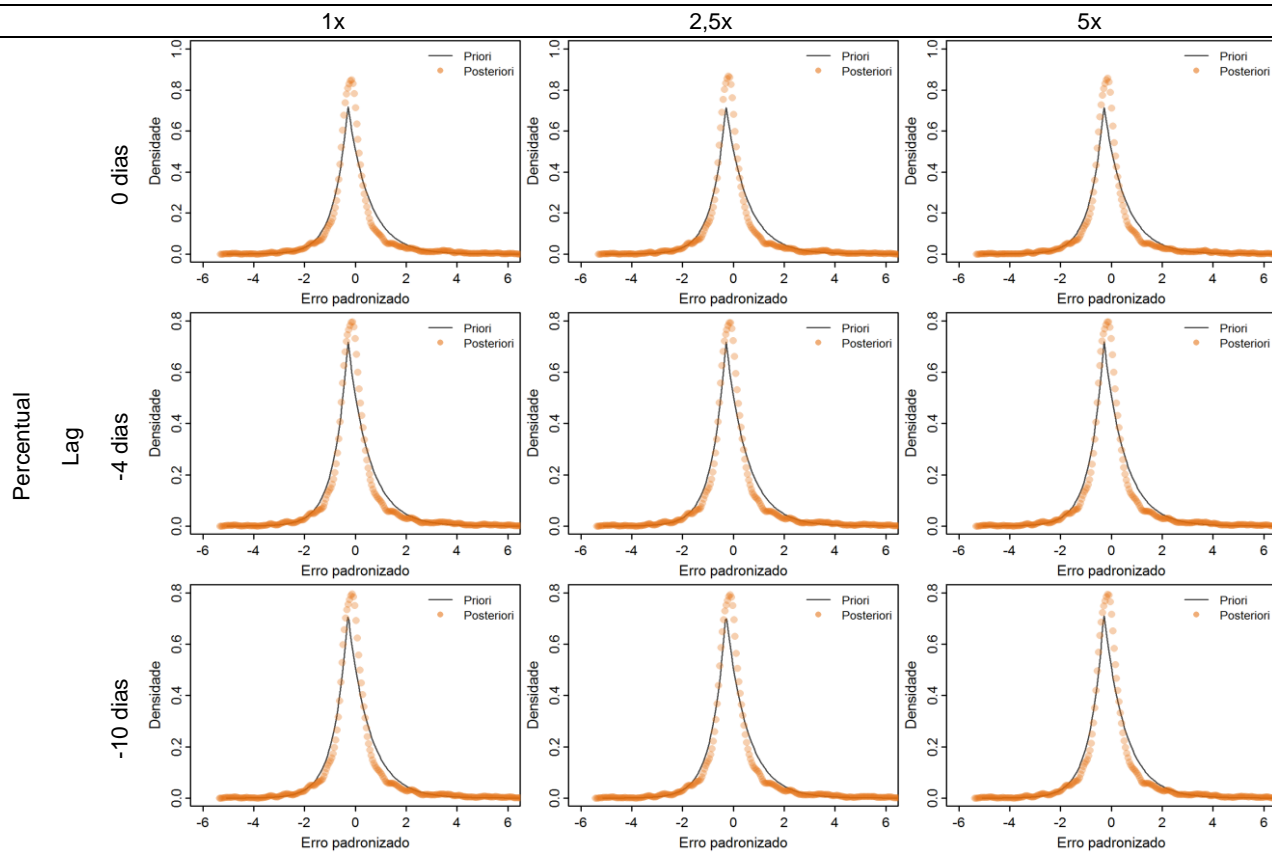




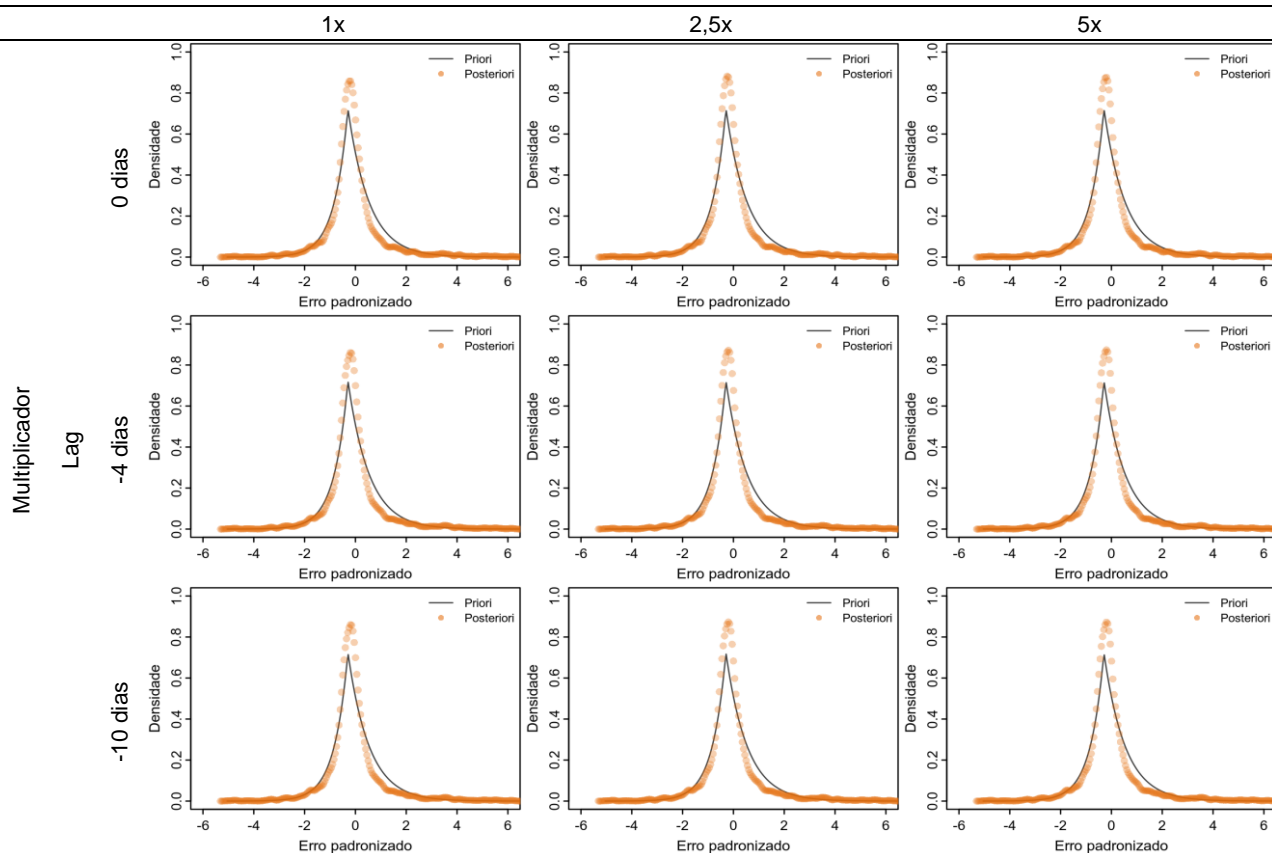


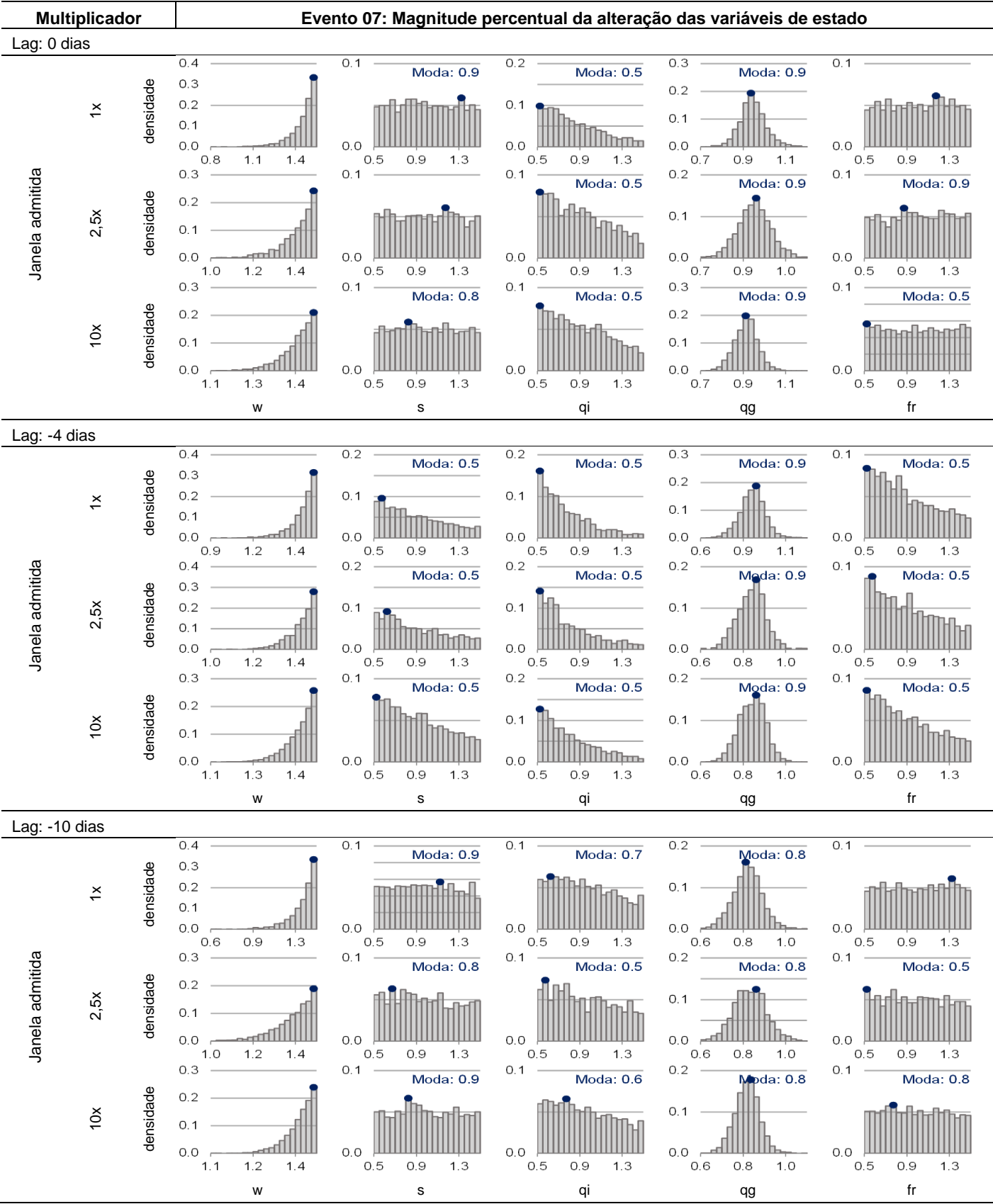
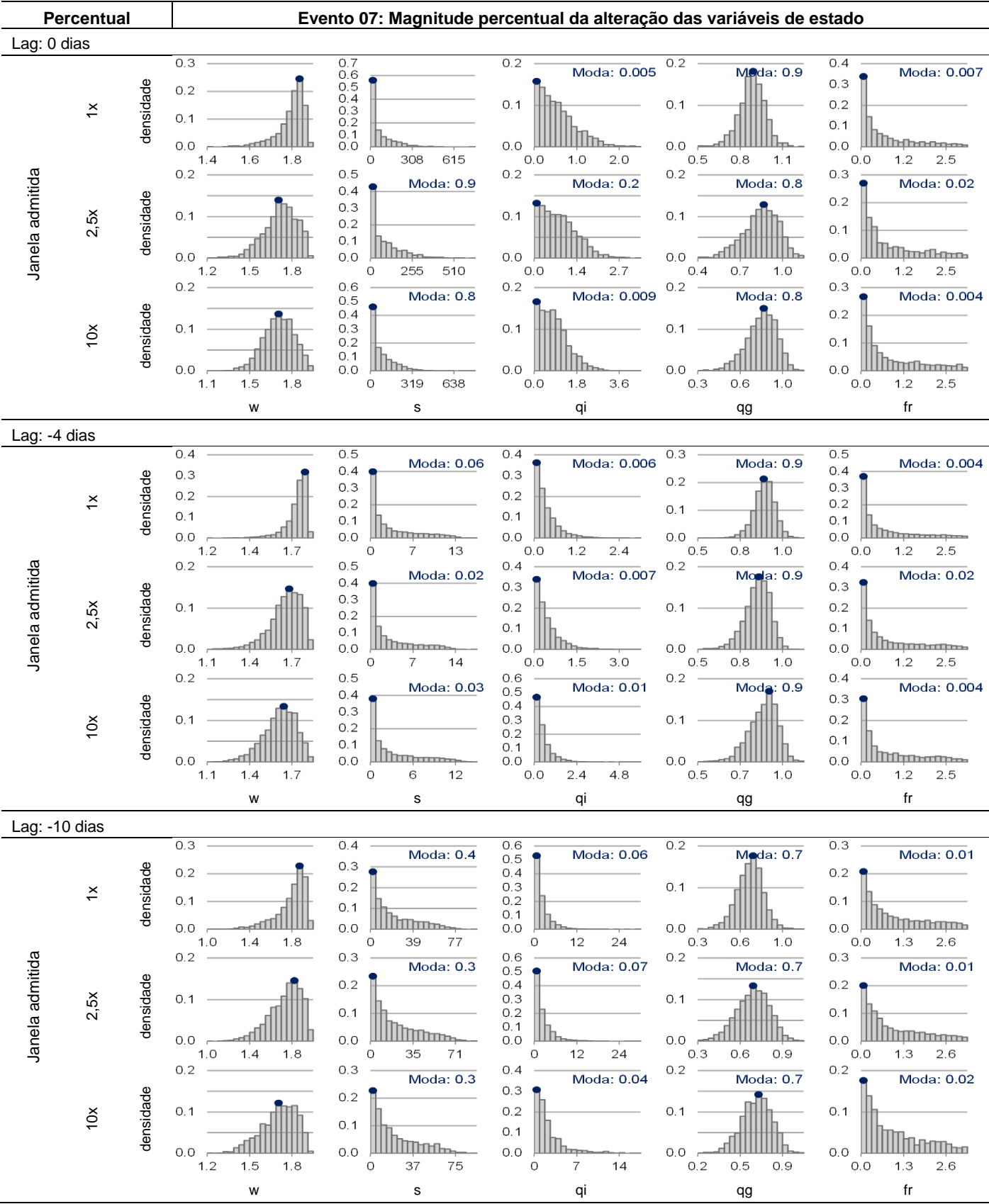
## Evento 06: Comportamento probabilístico dos resíduos da simulação

Janela admitida

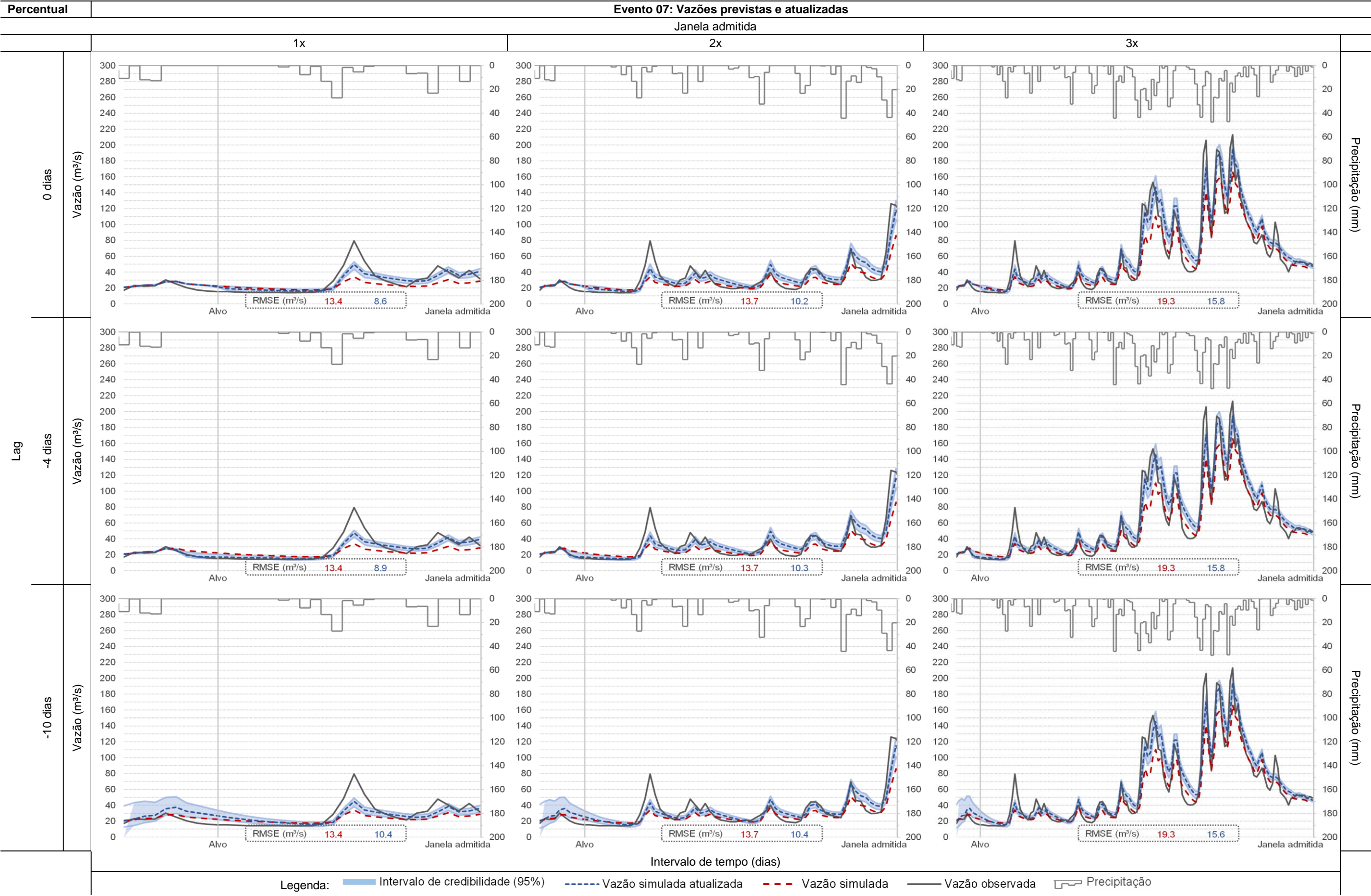


Janela admitida

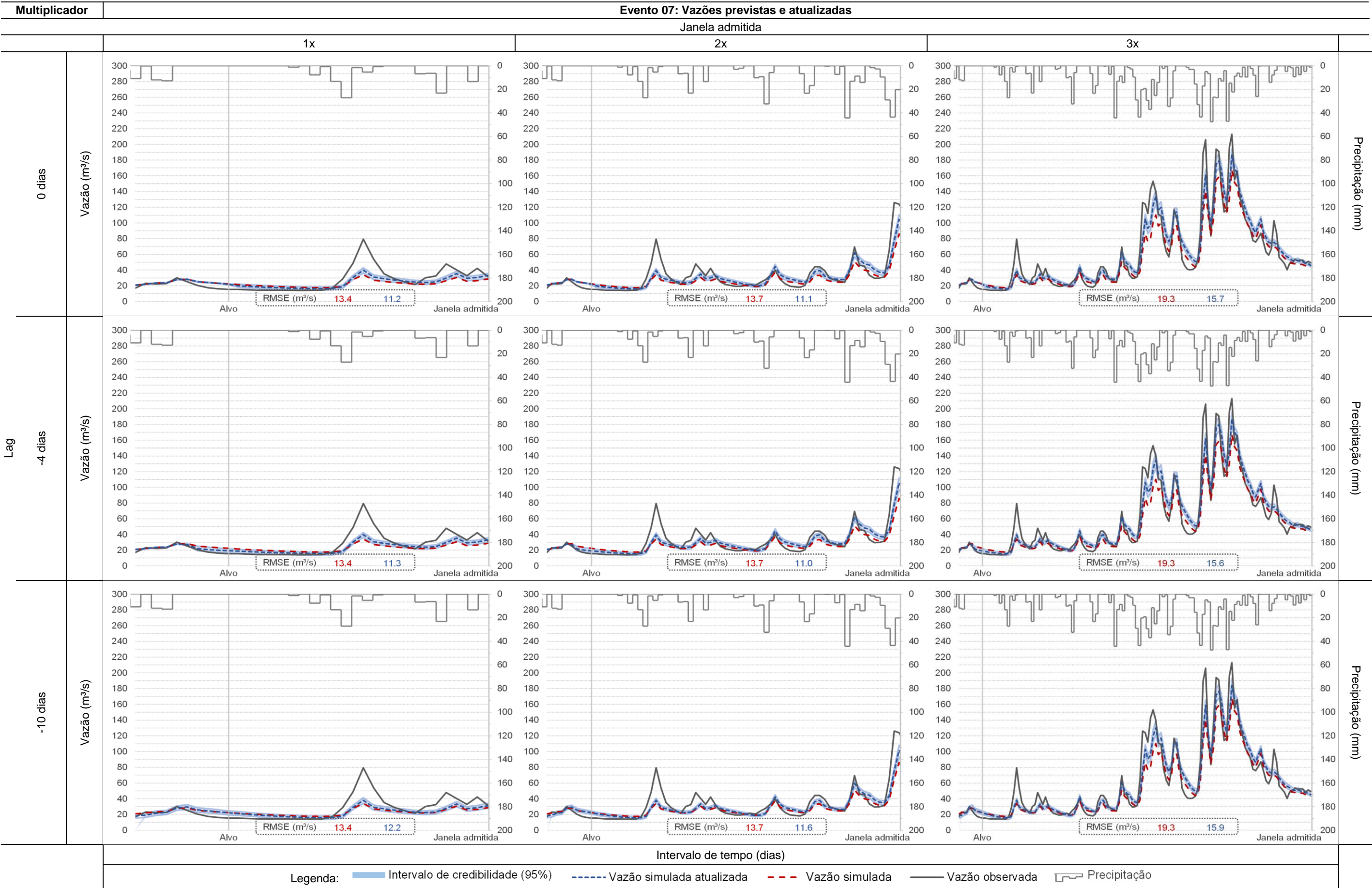


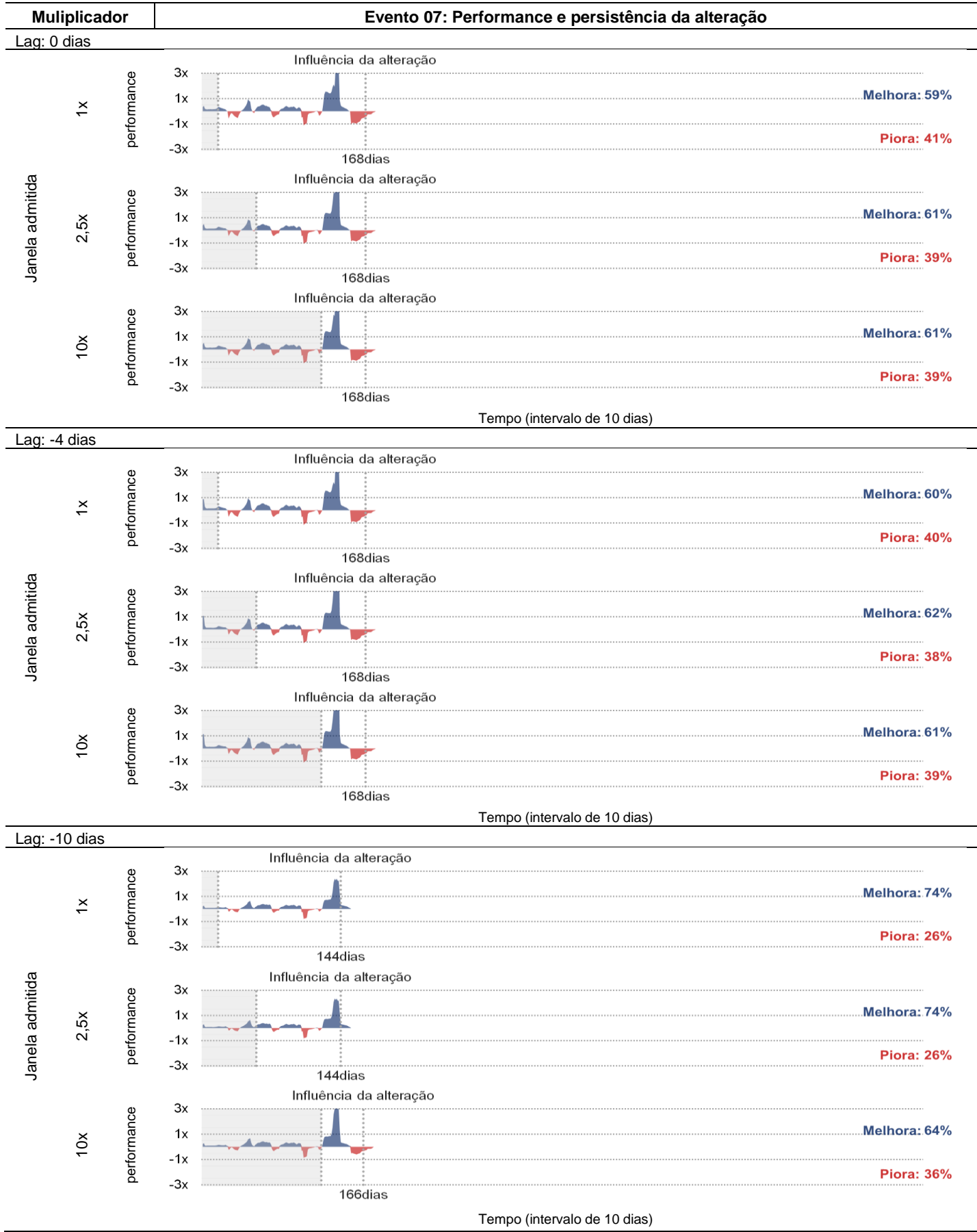
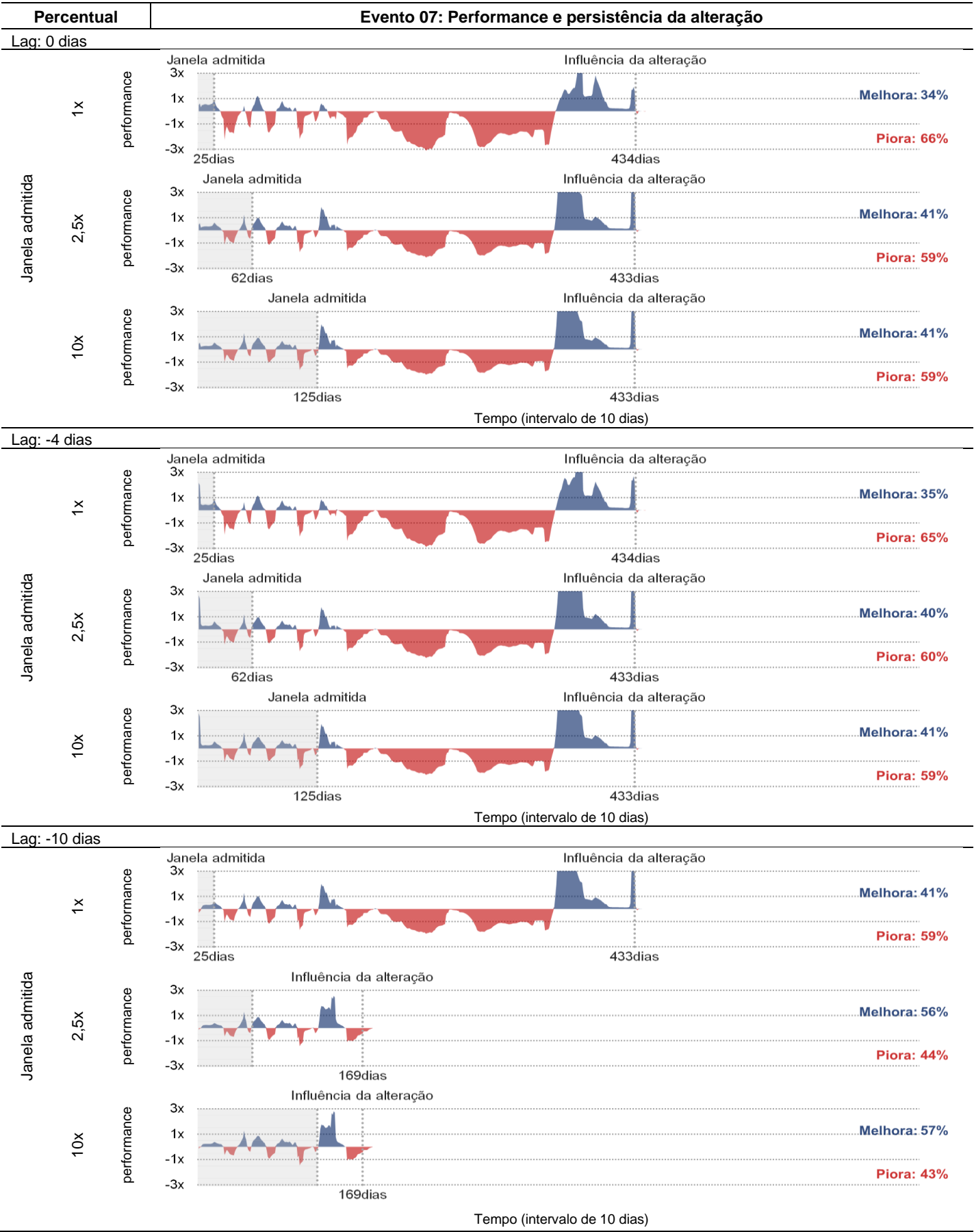








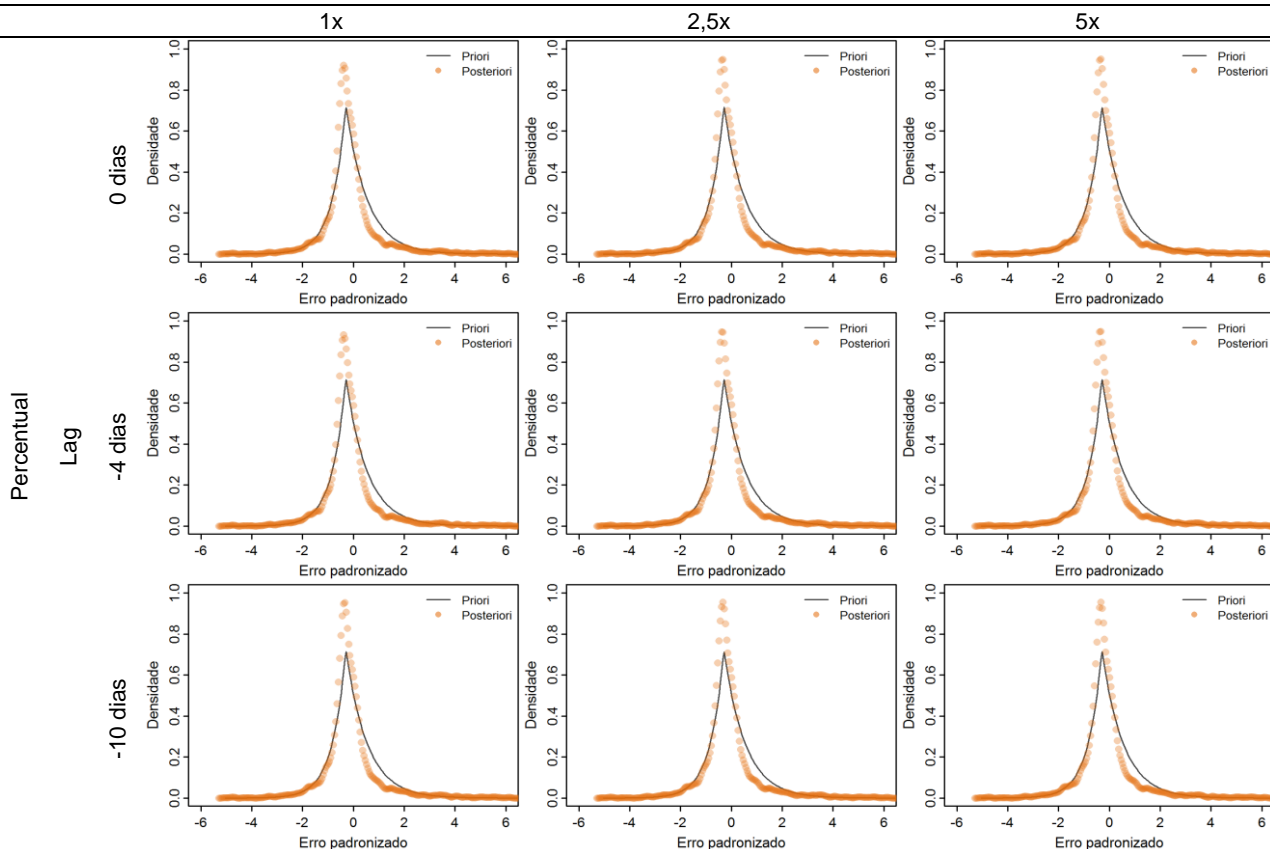




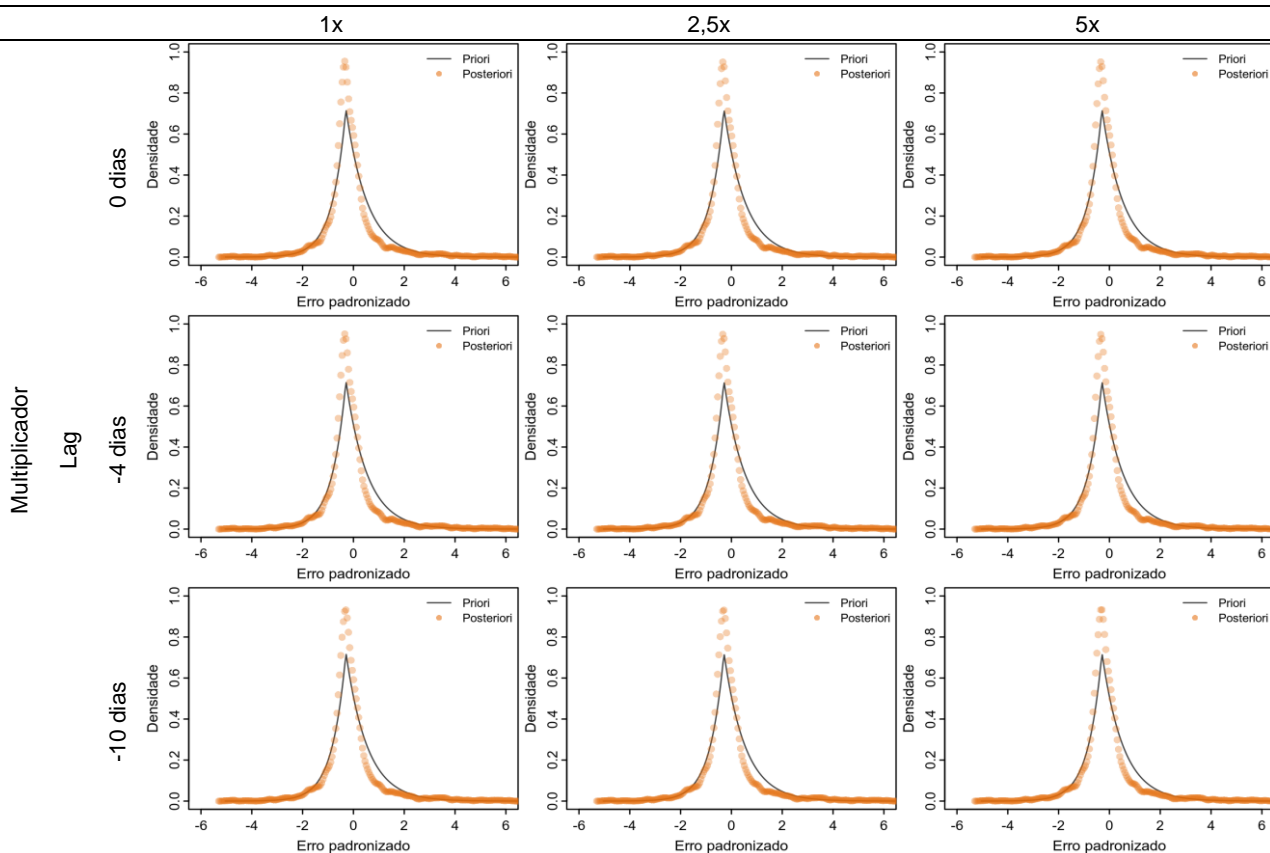


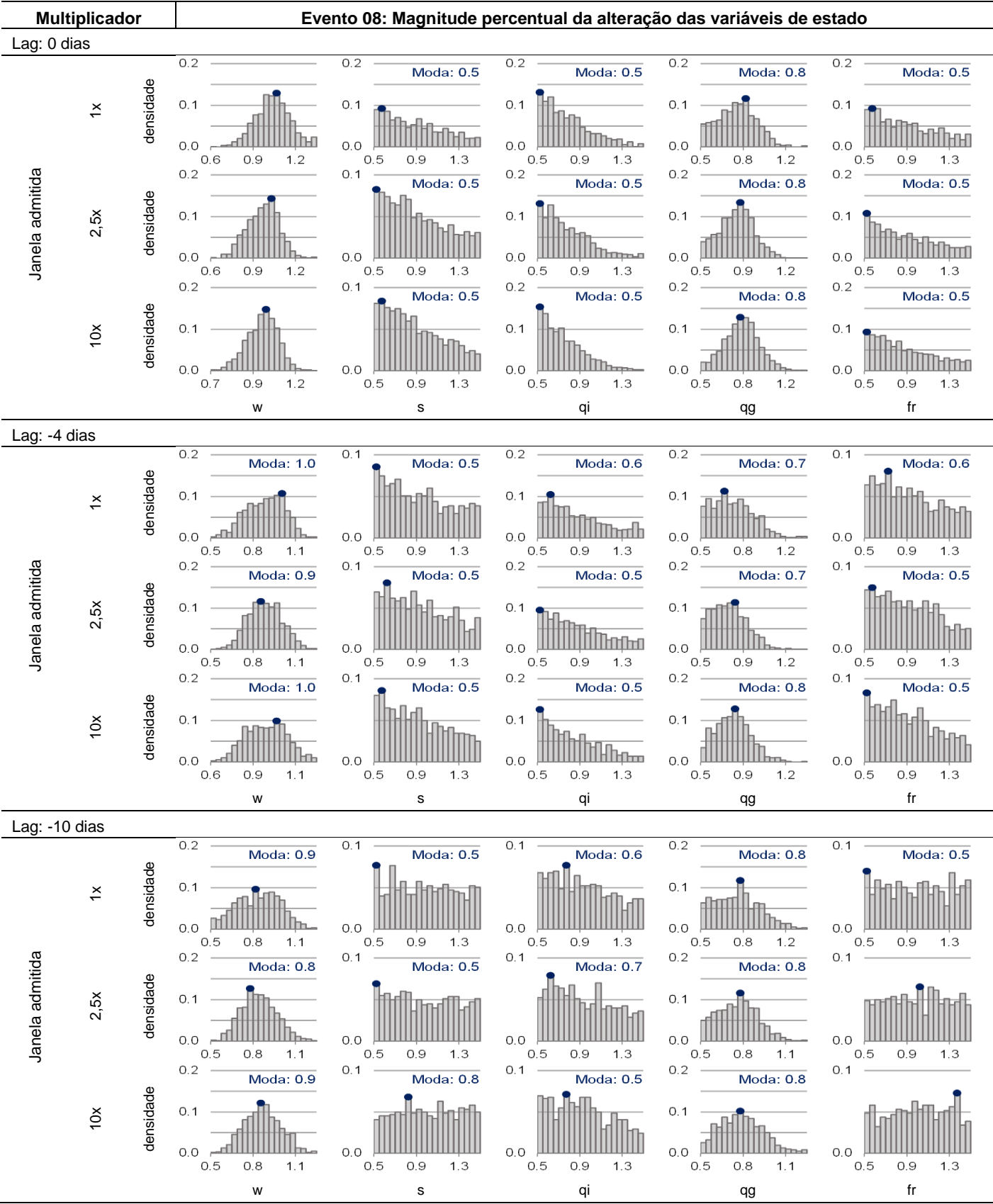
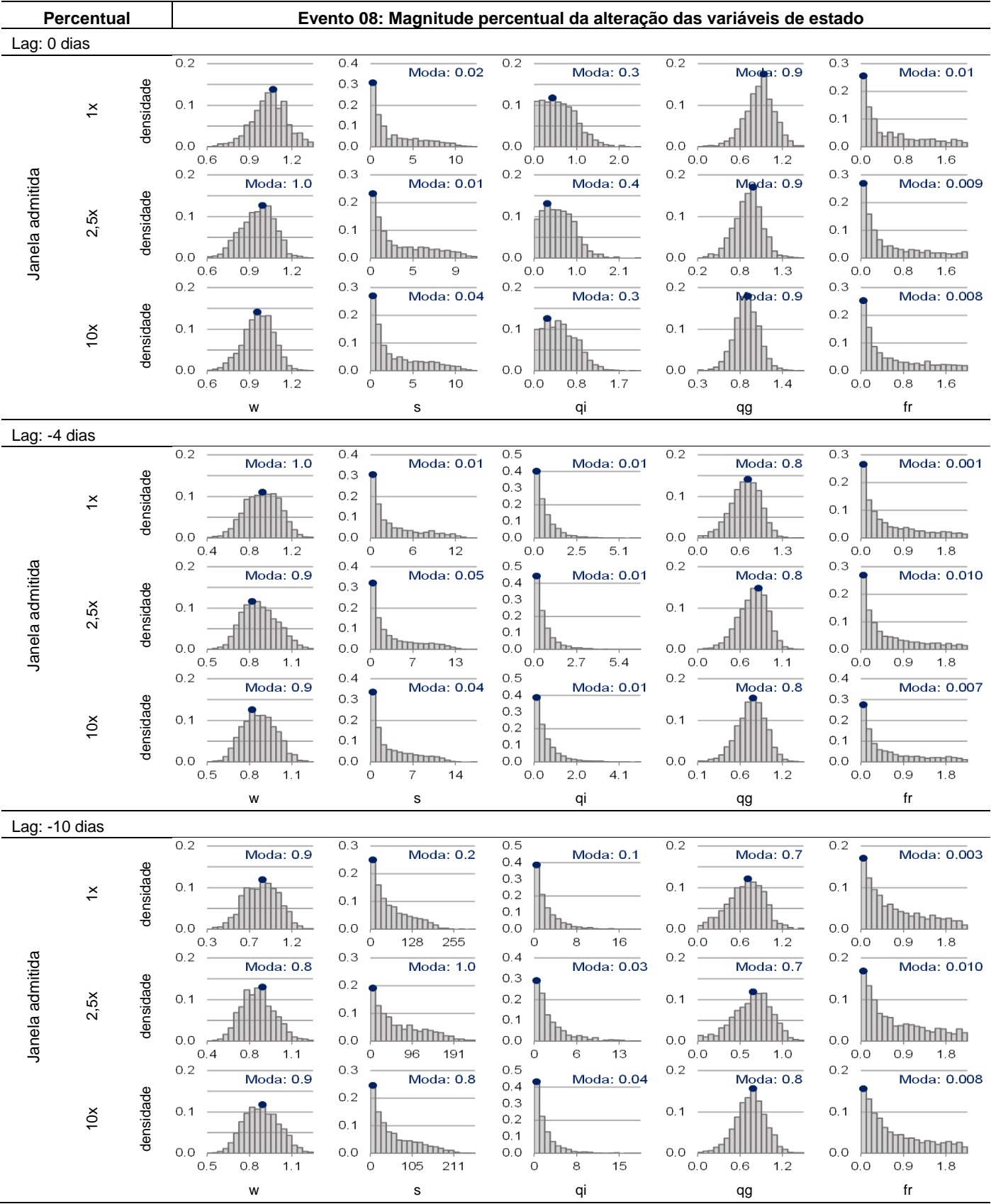
## Evento 07: Comportamento probabilístico dos resíduos da simulação

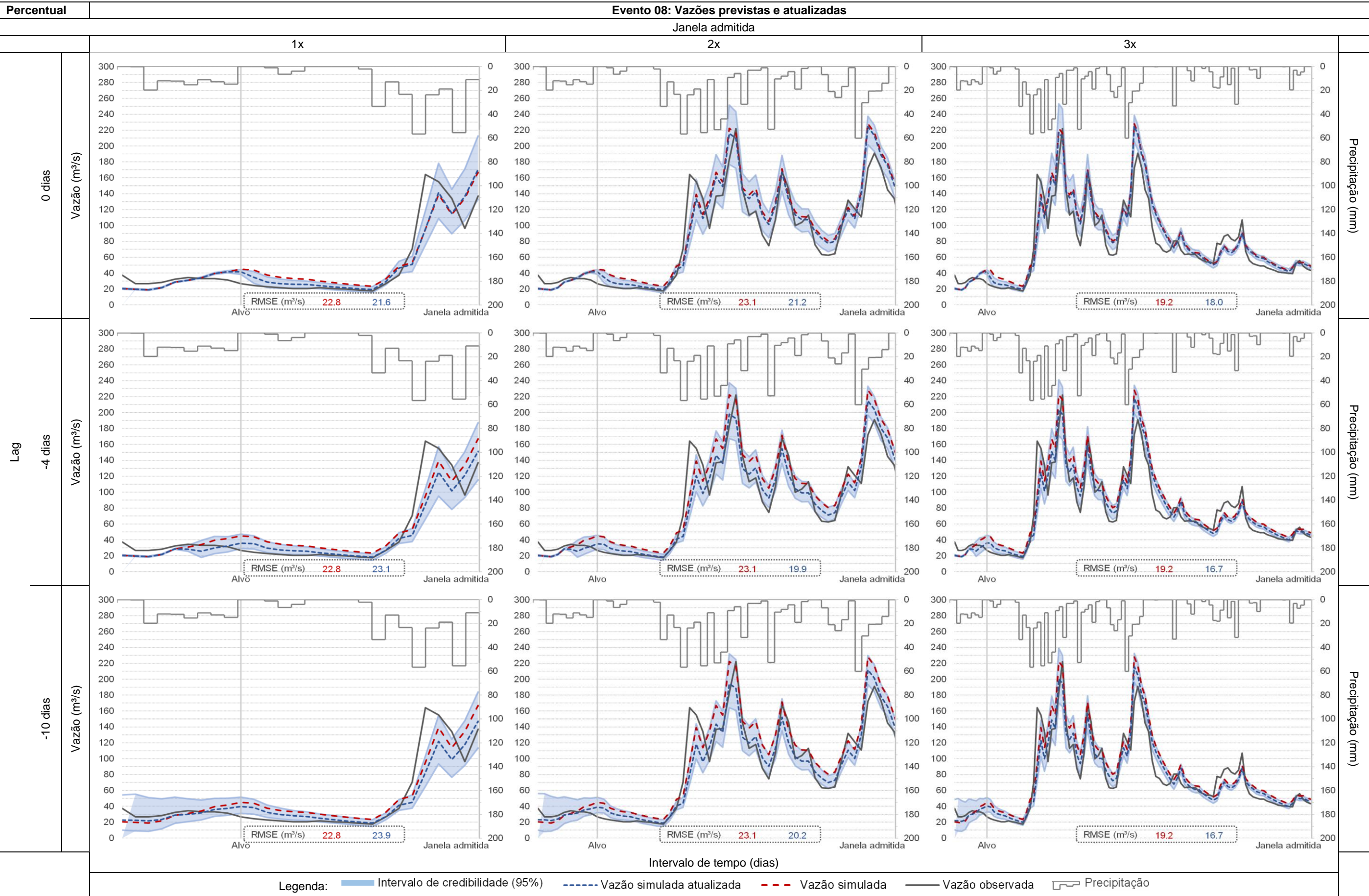
Janela admitida



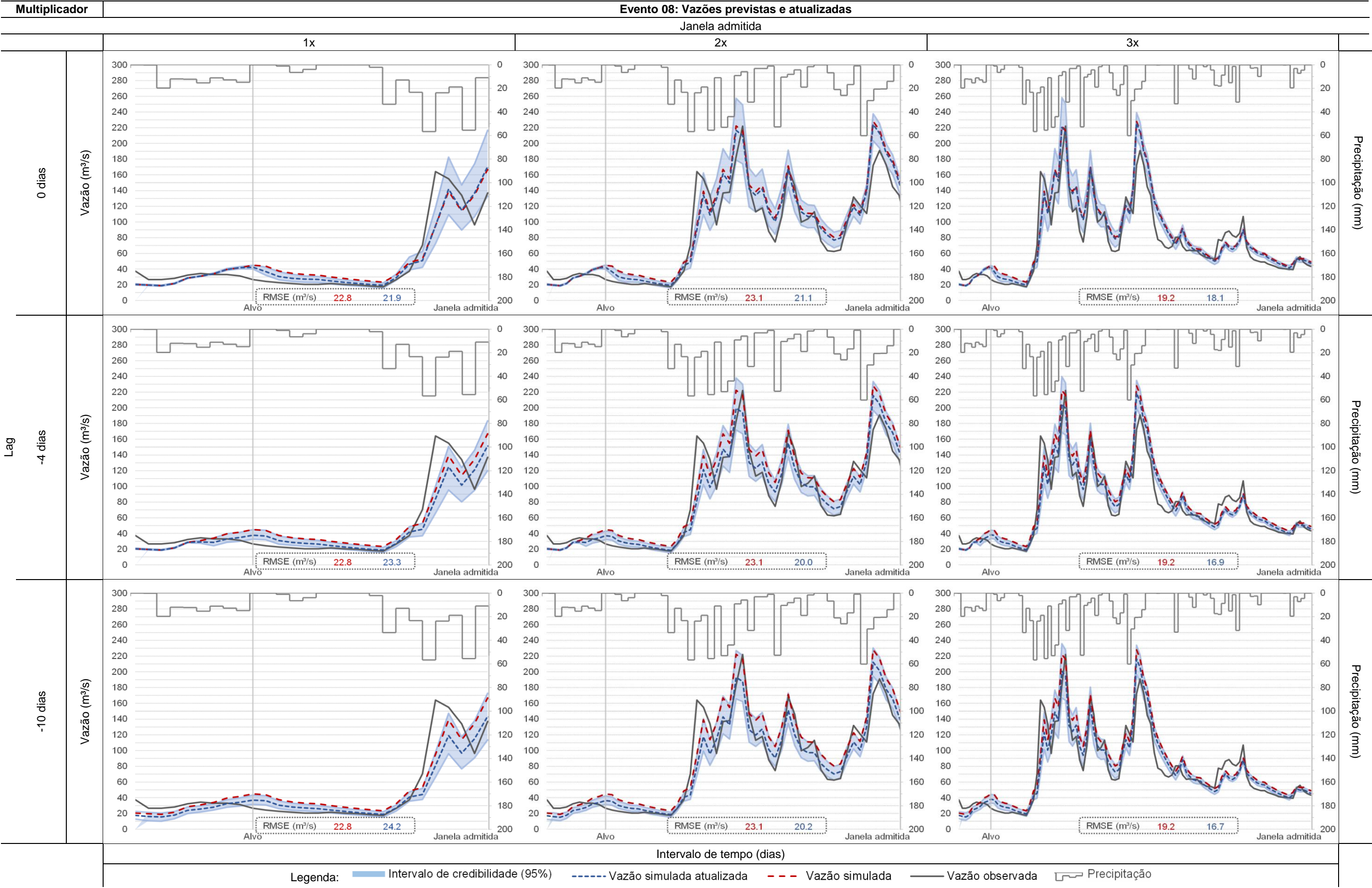
Janela admitida

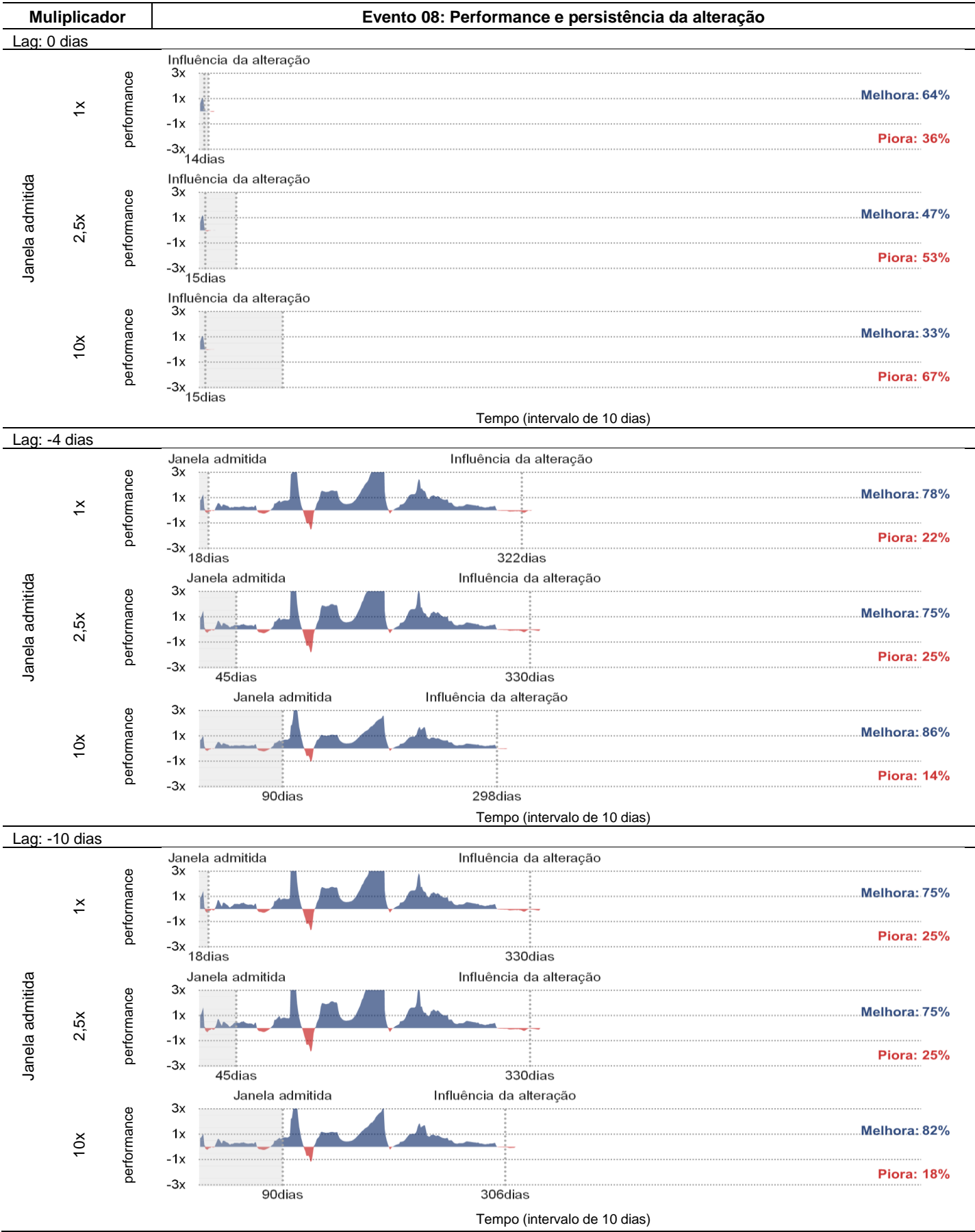
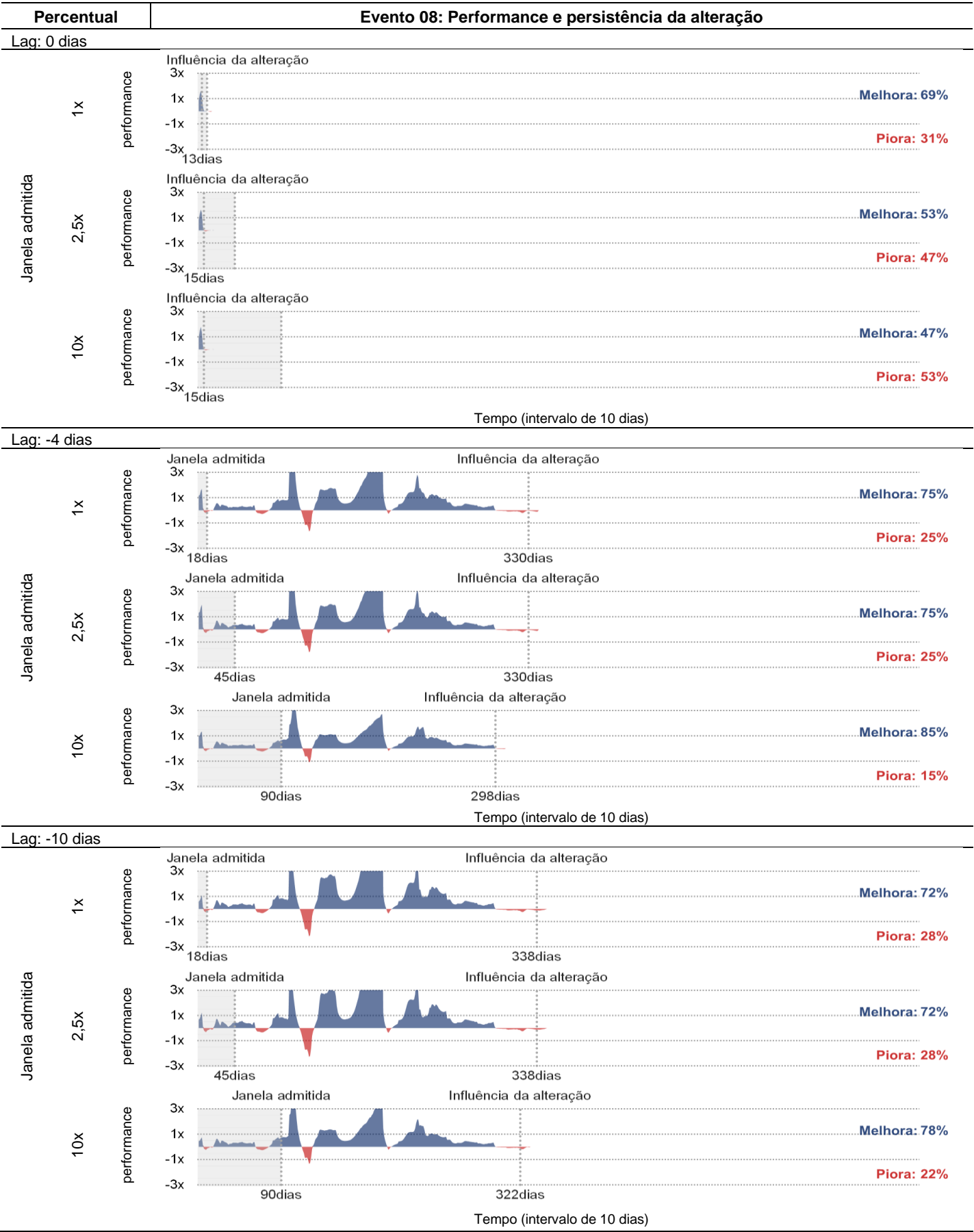






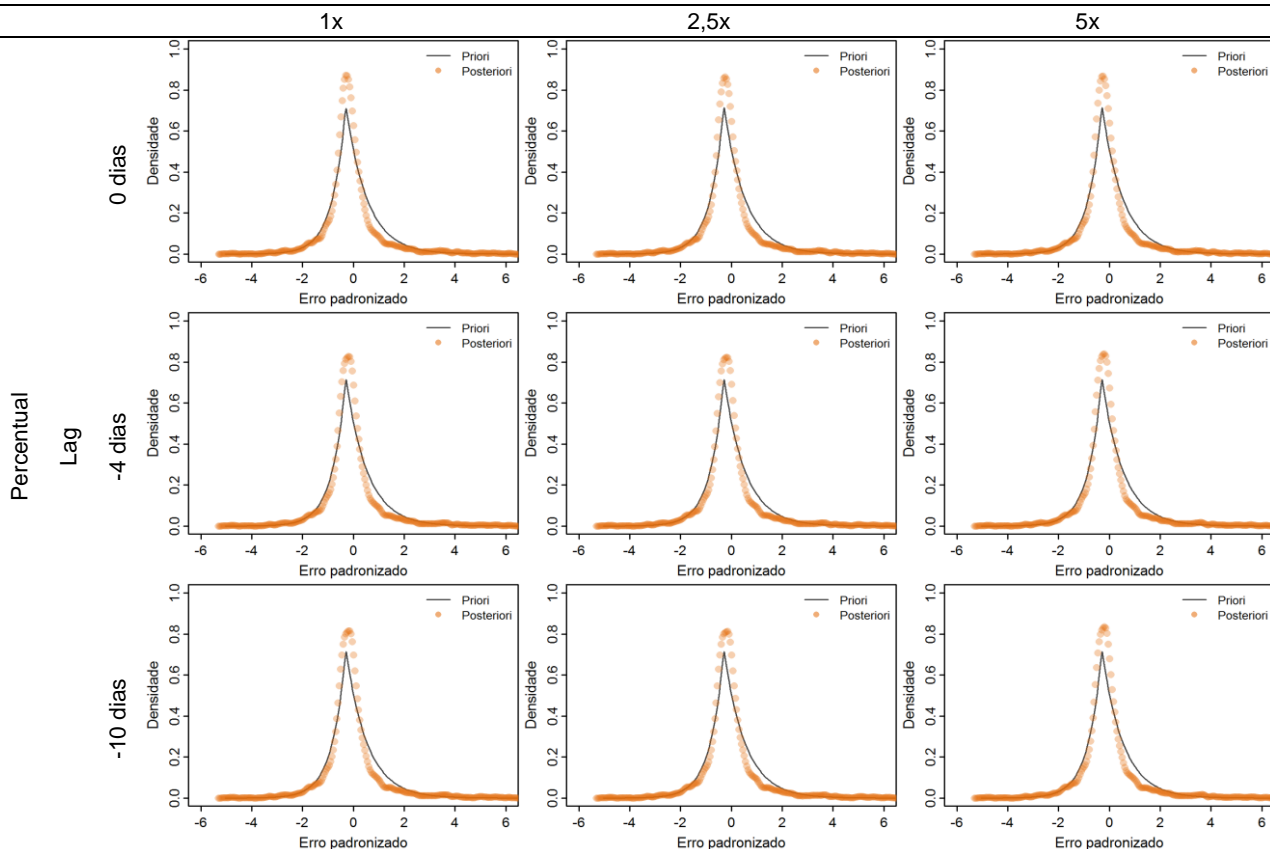




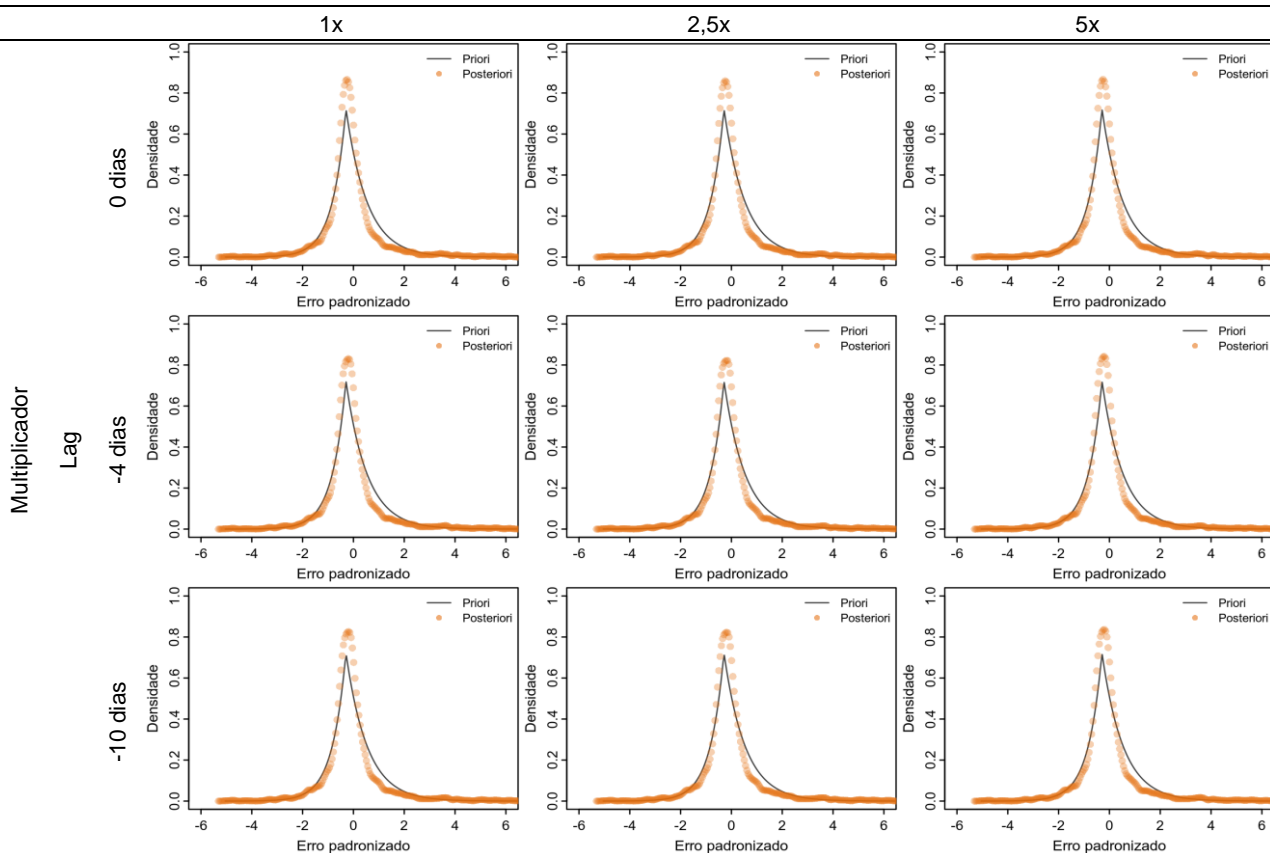


## Evento 08: Comportamento probabilístico dos resíduos da simulação

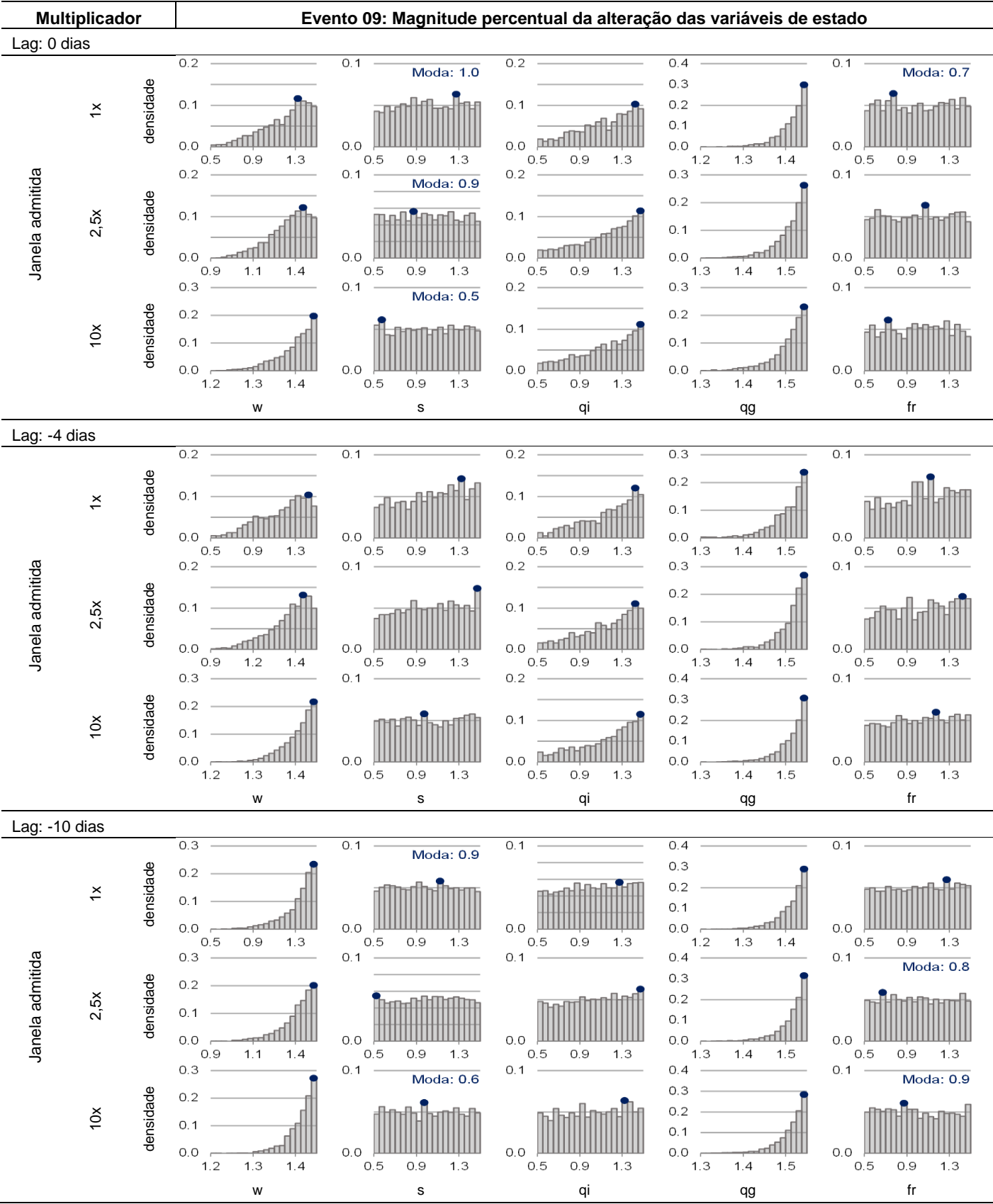
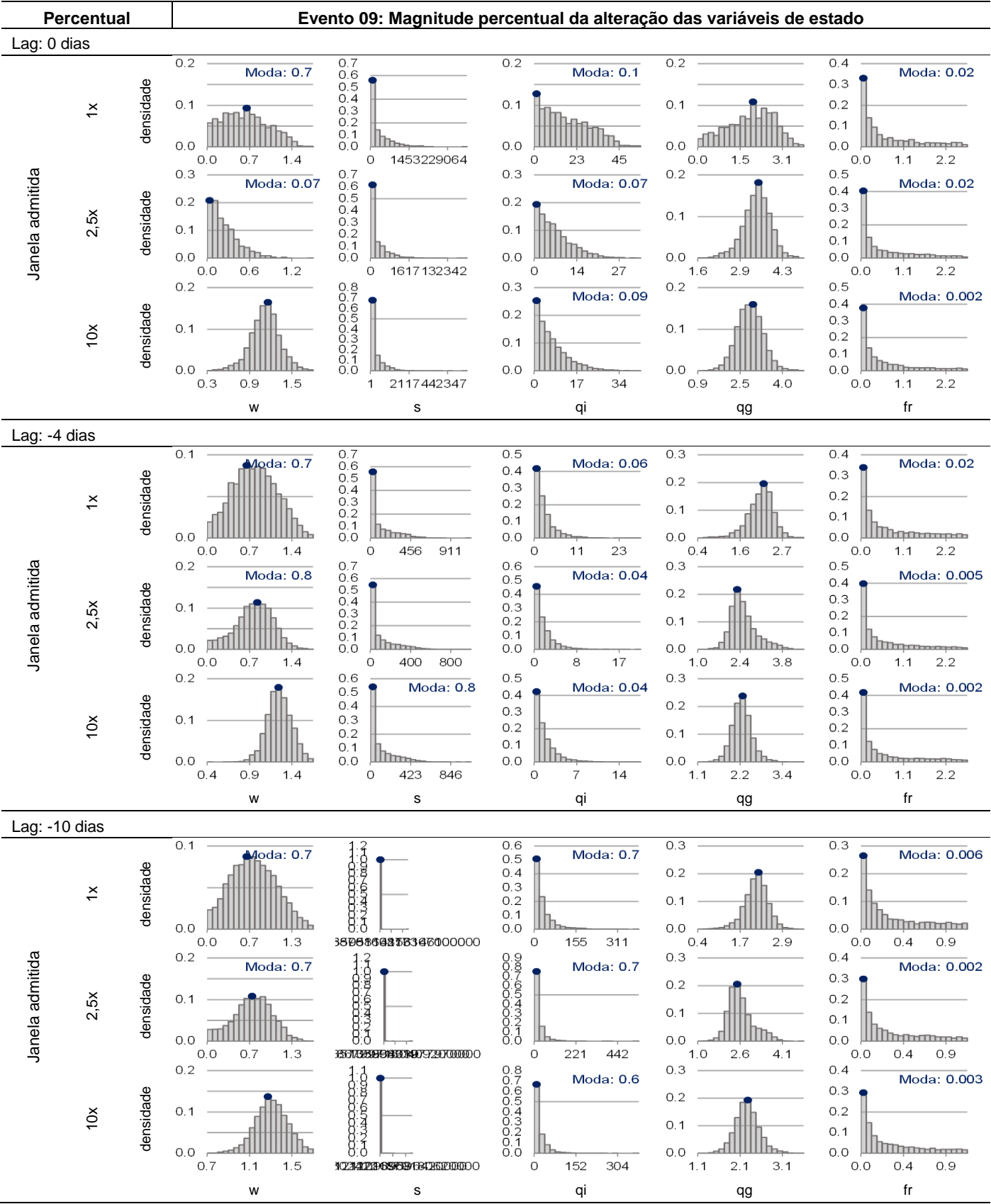
Janela admitida

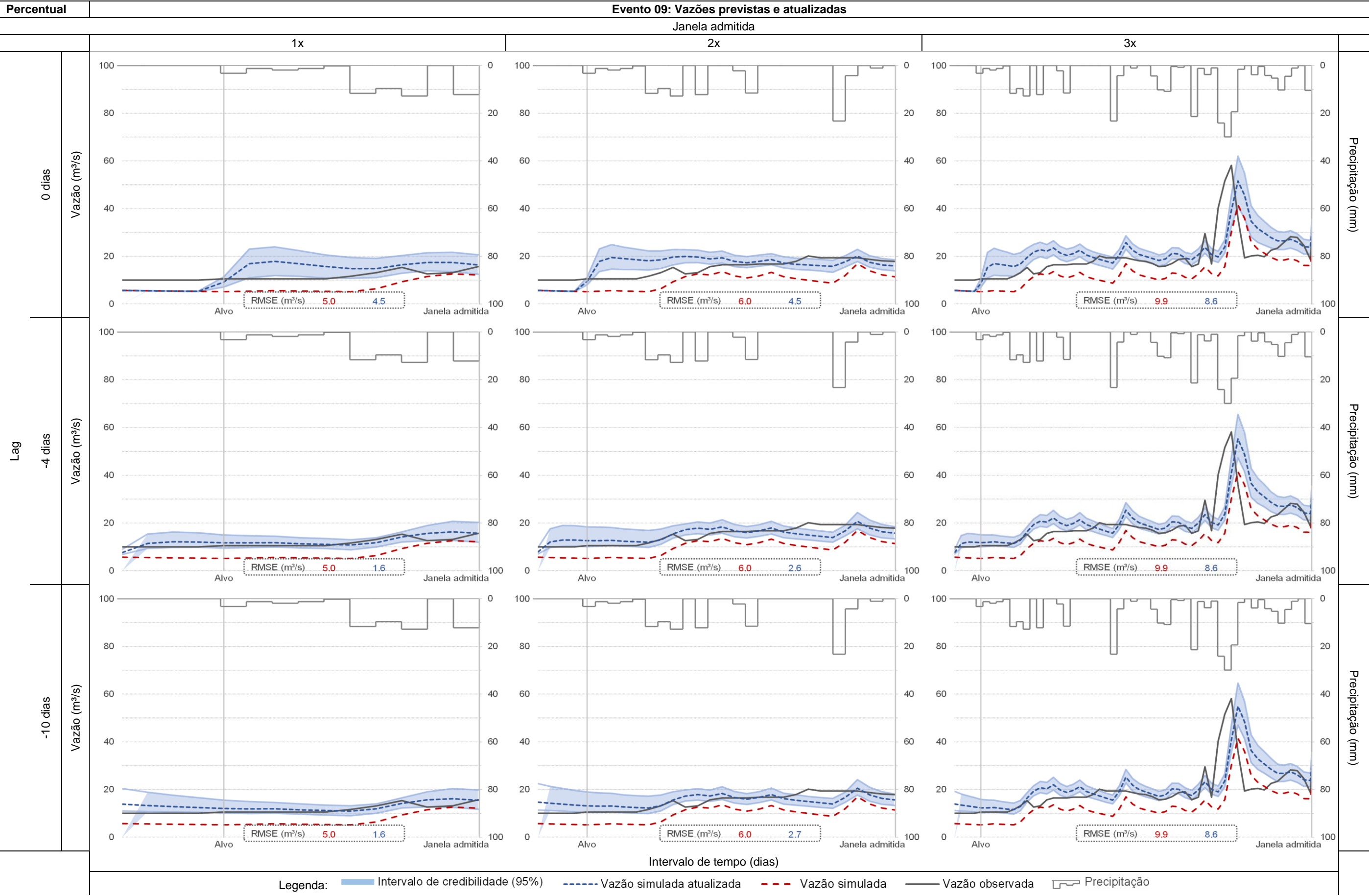


Janela admitida

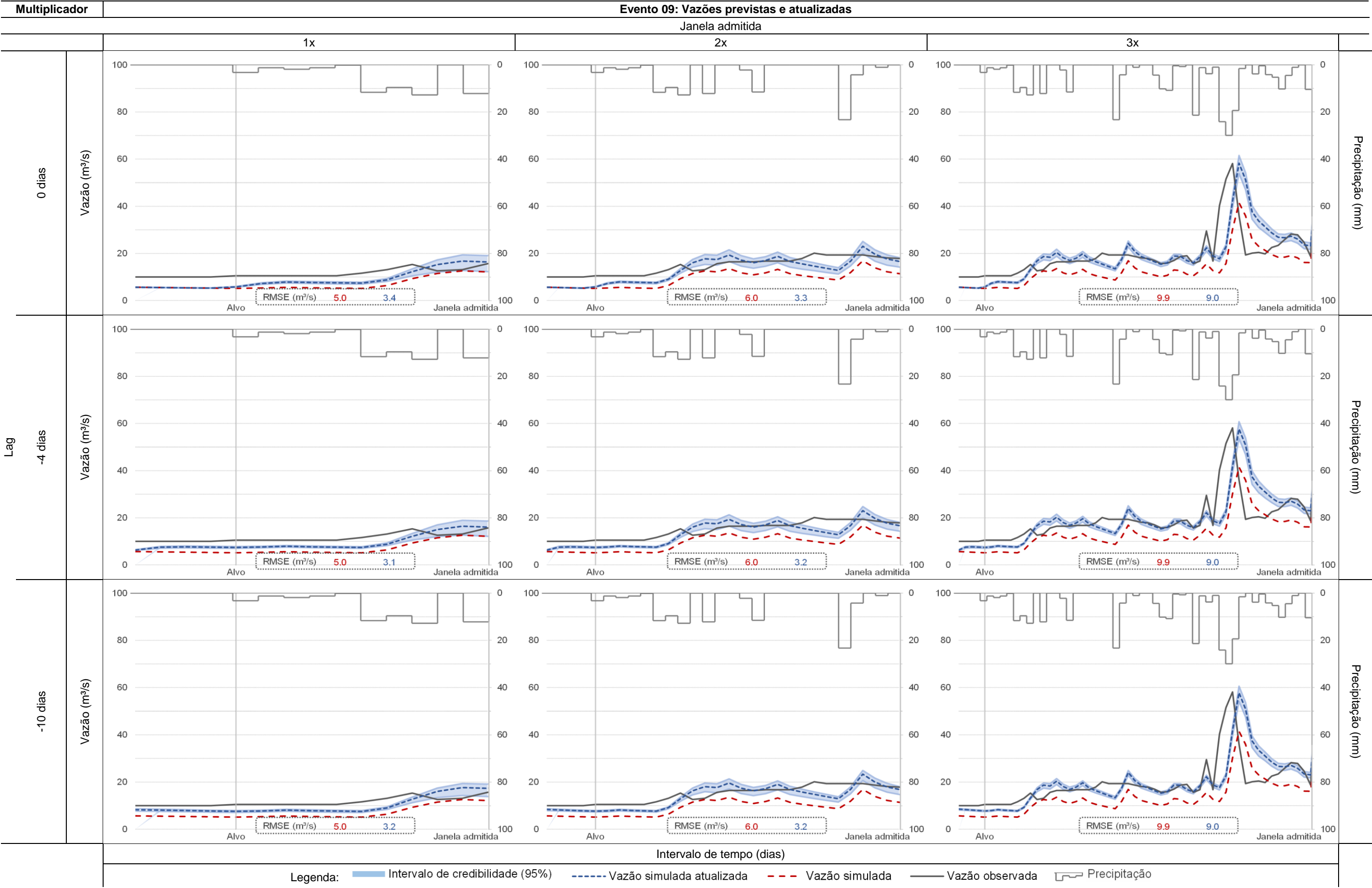


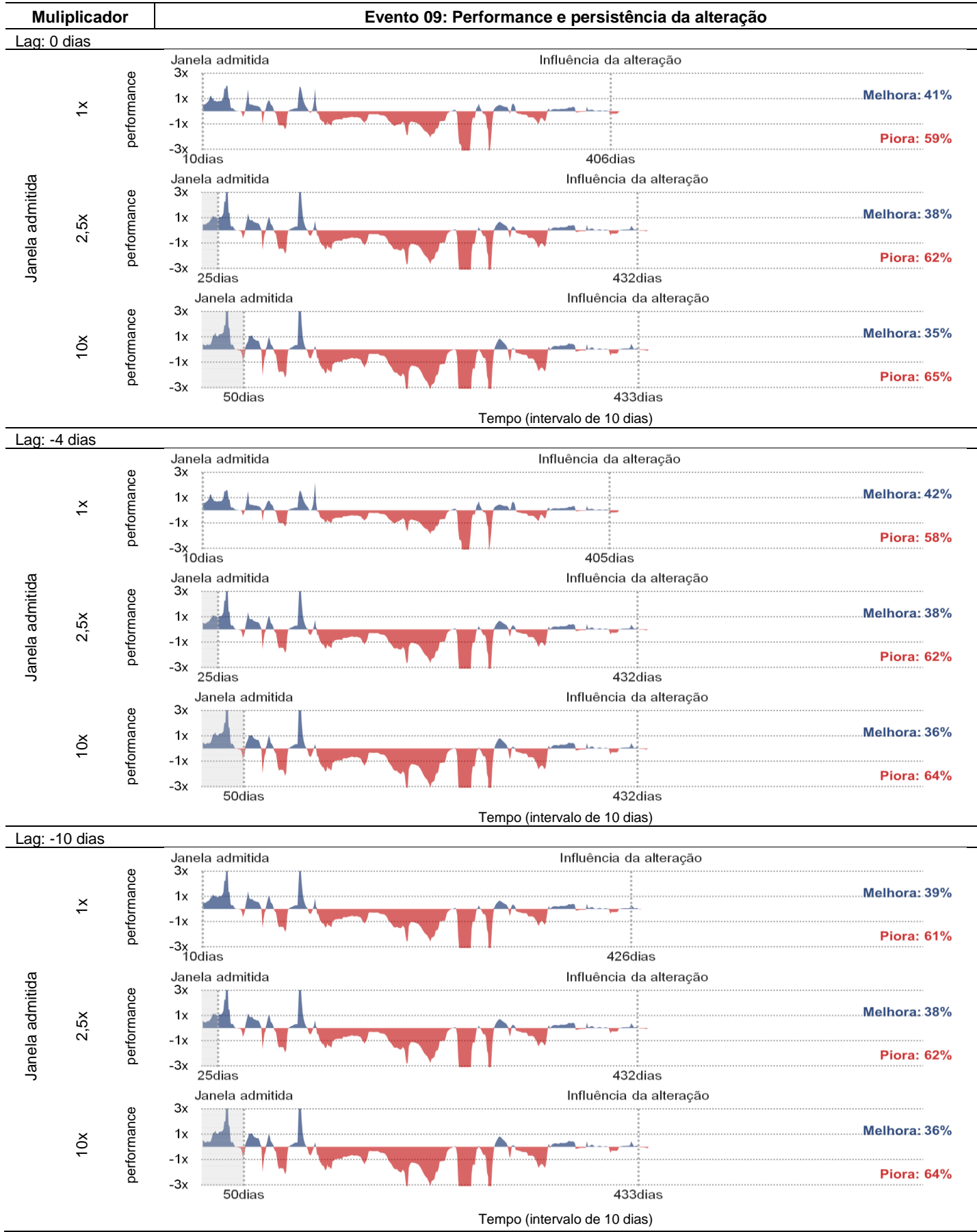
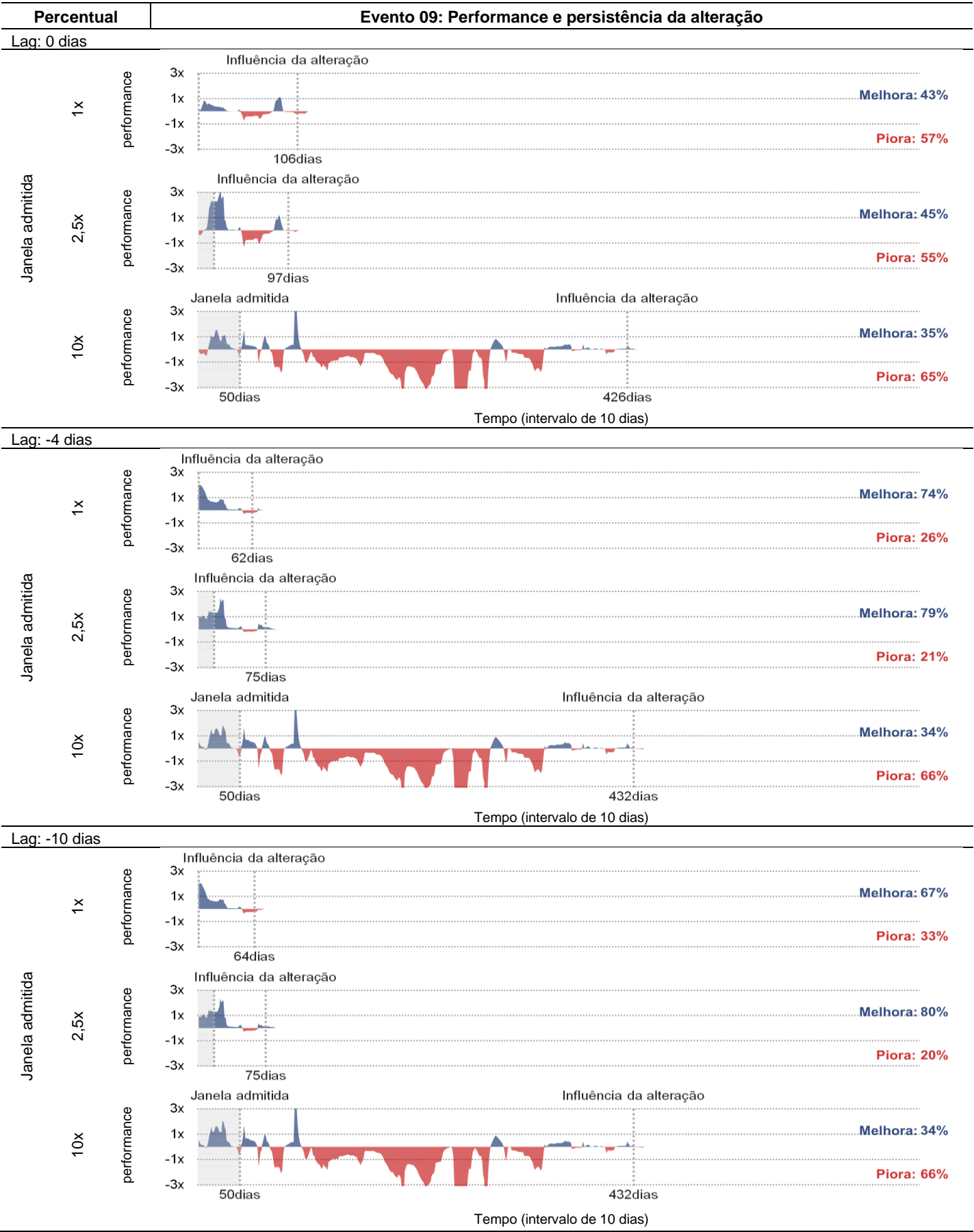






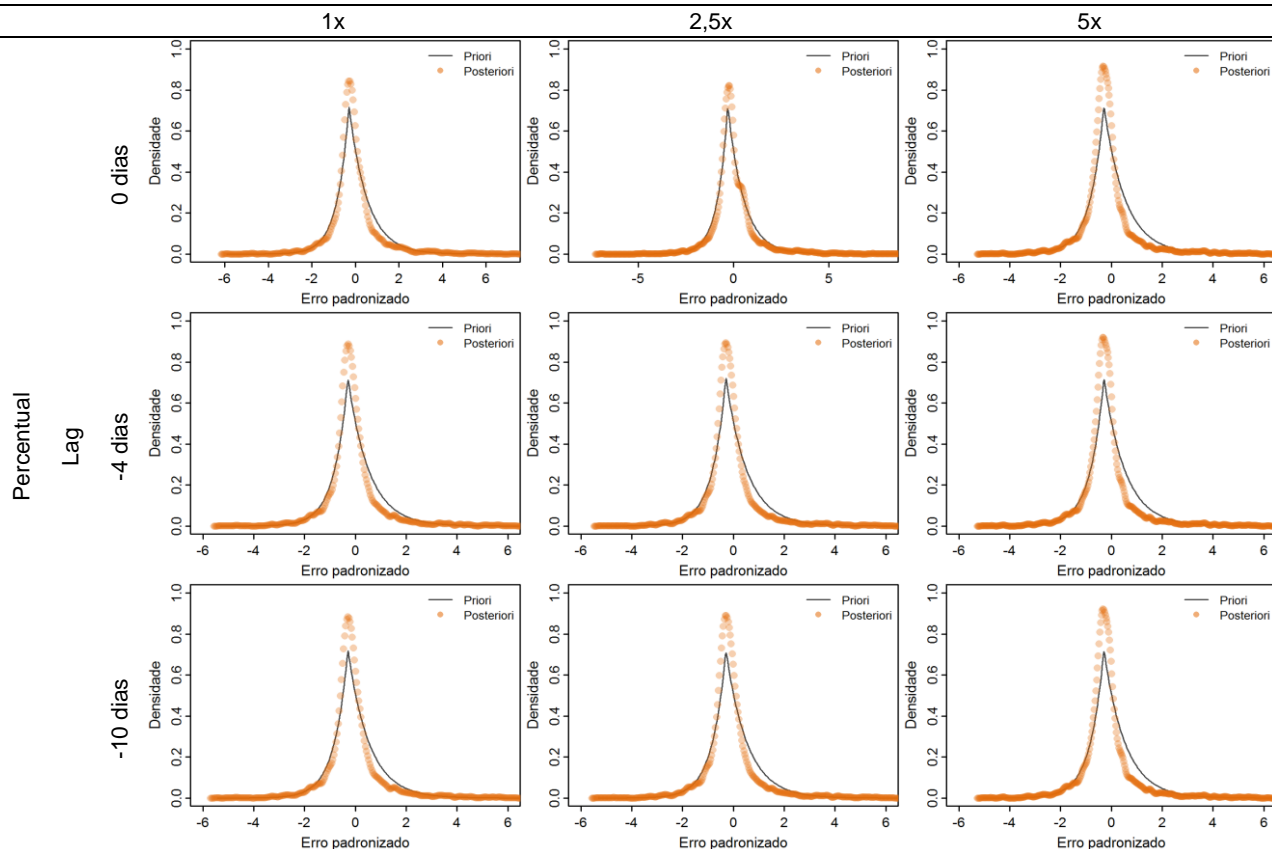




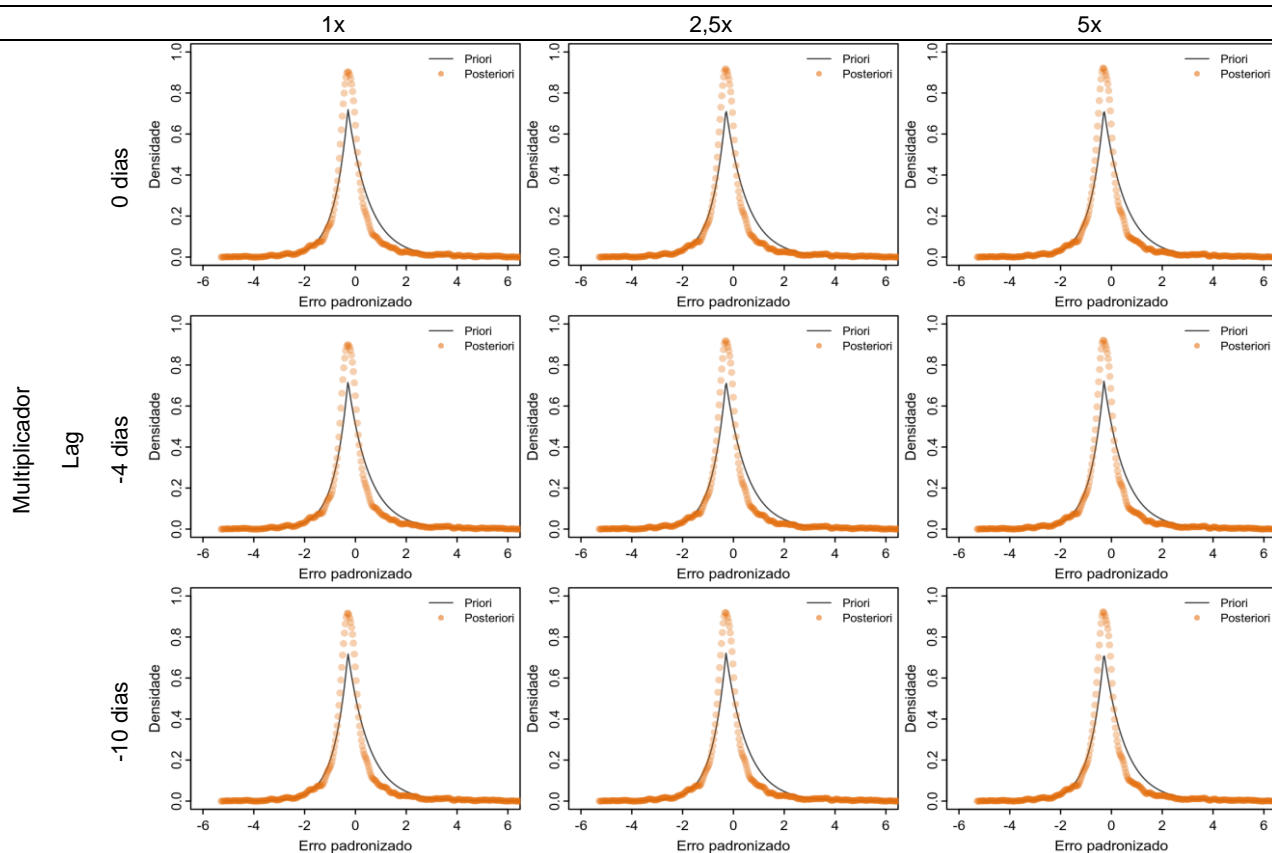


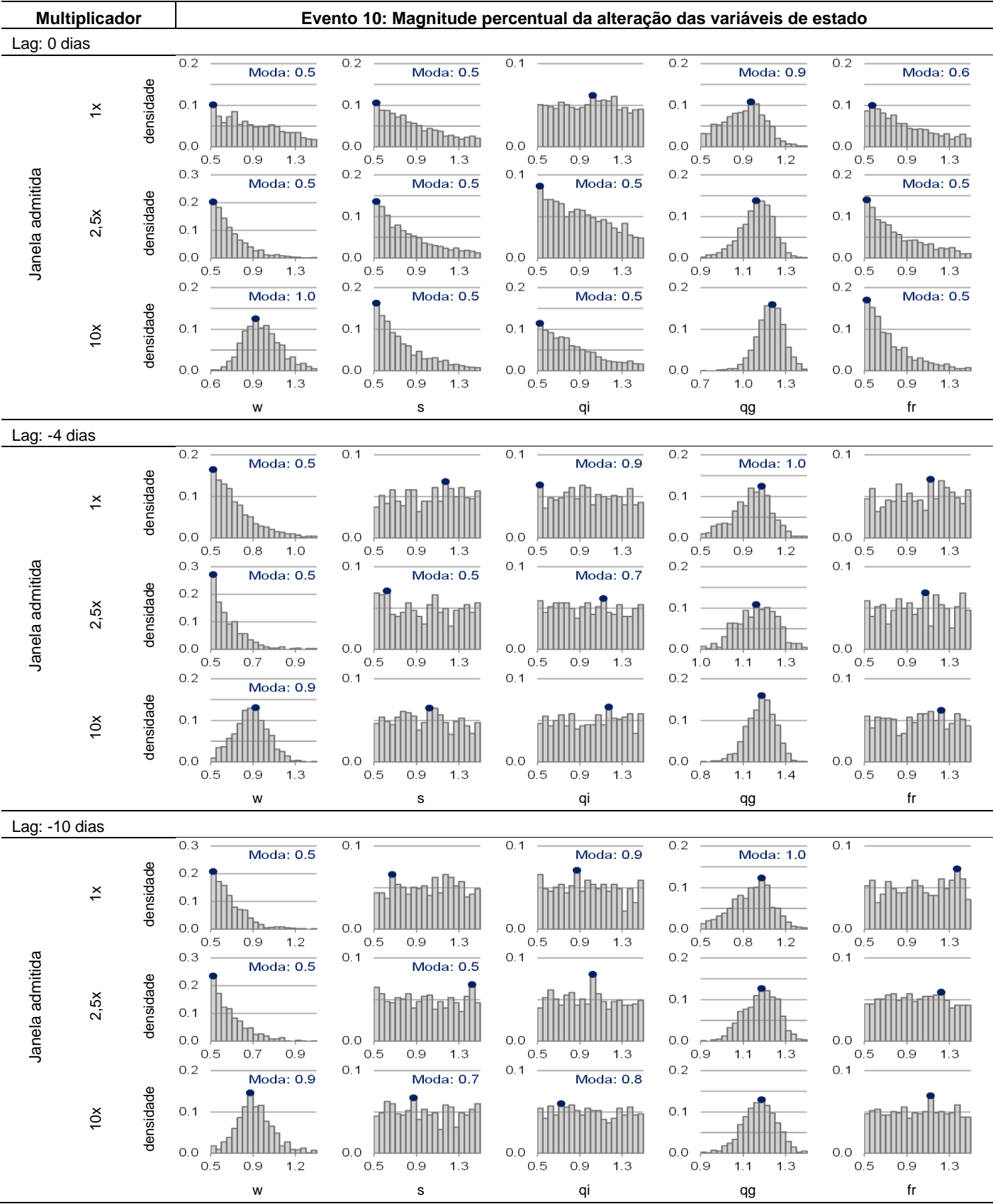
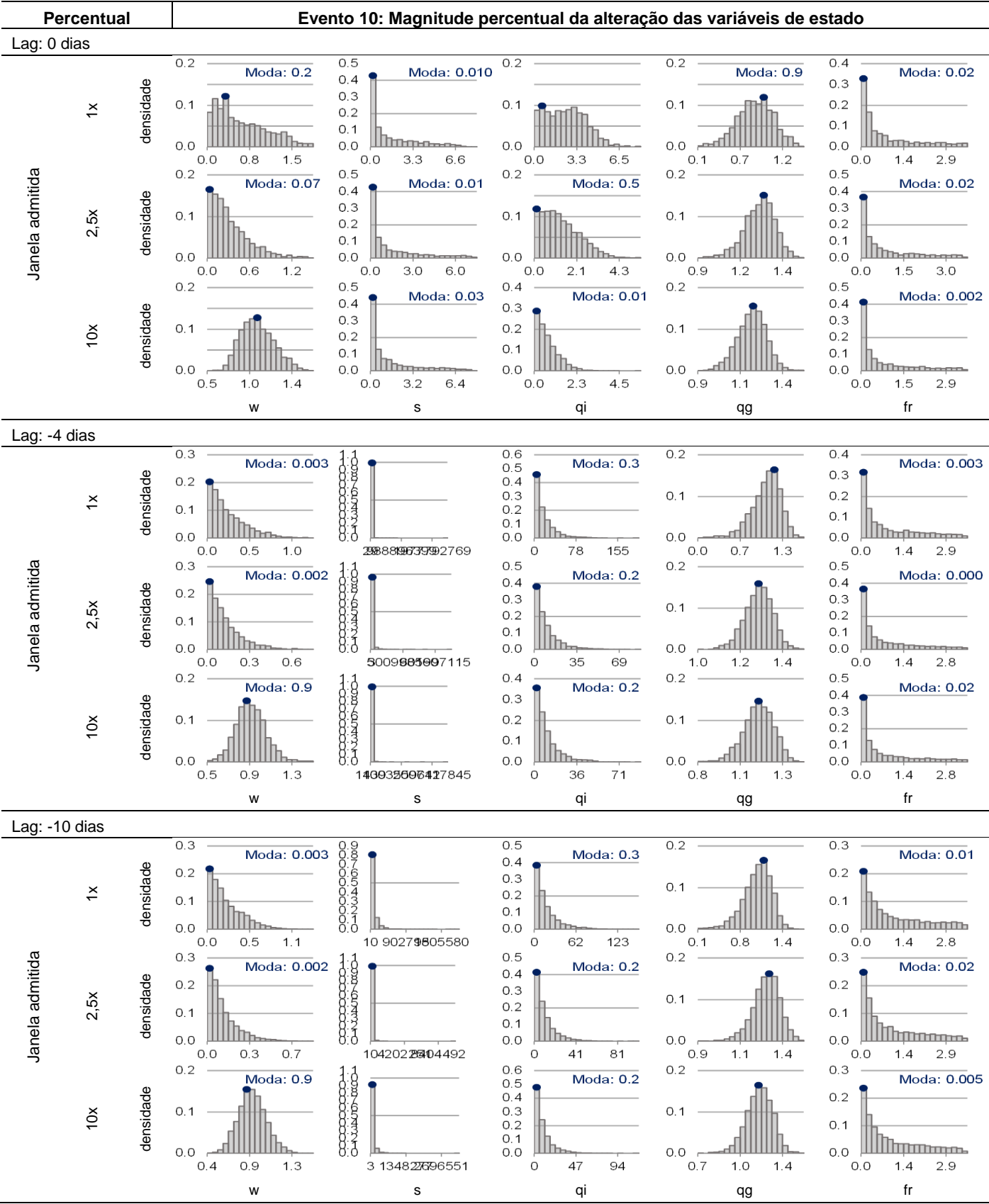
## Evento 09: Comportamento probabilístico dos resíduos da simulação

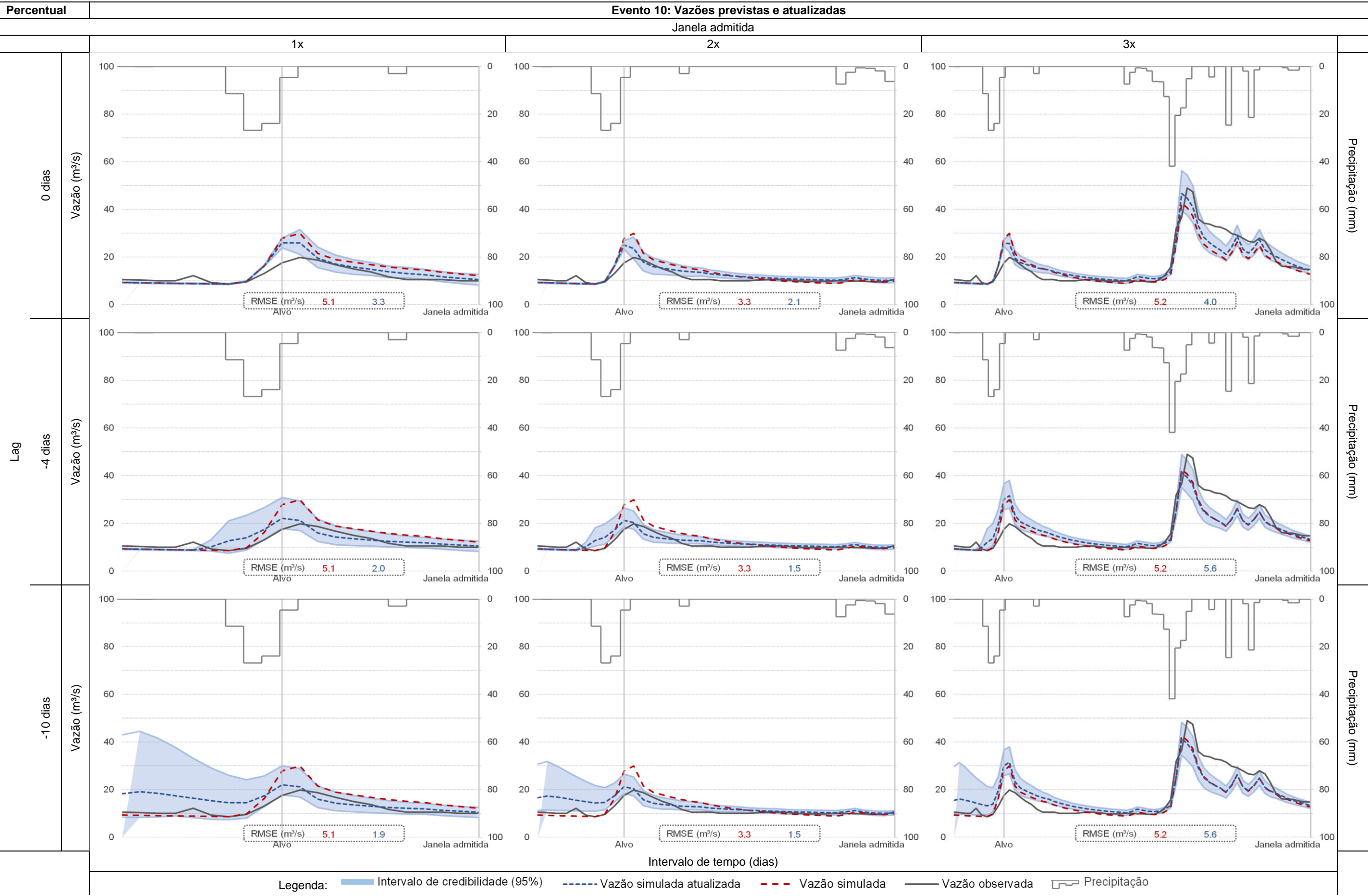
Janela admitida



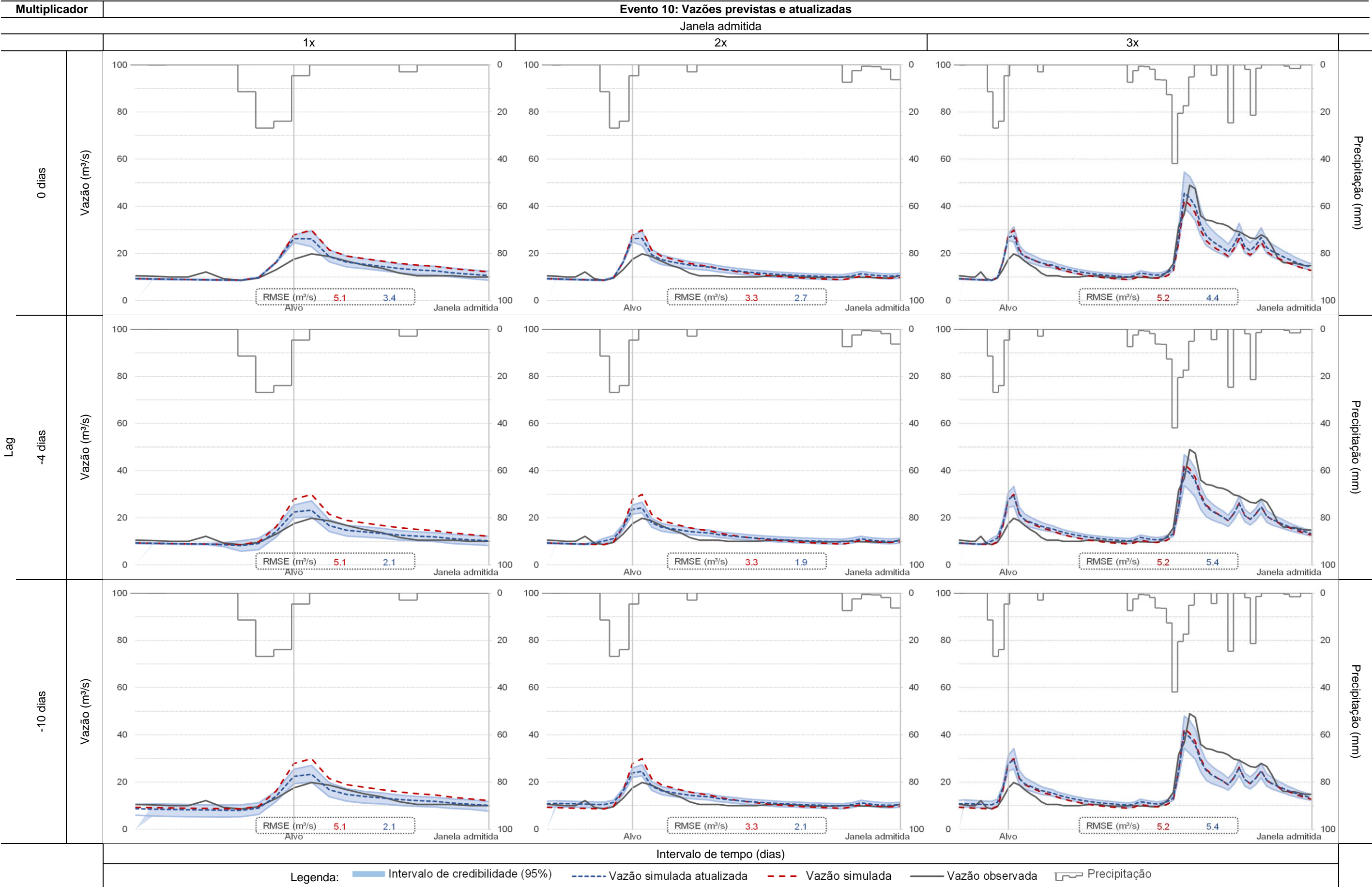
Janela admitida

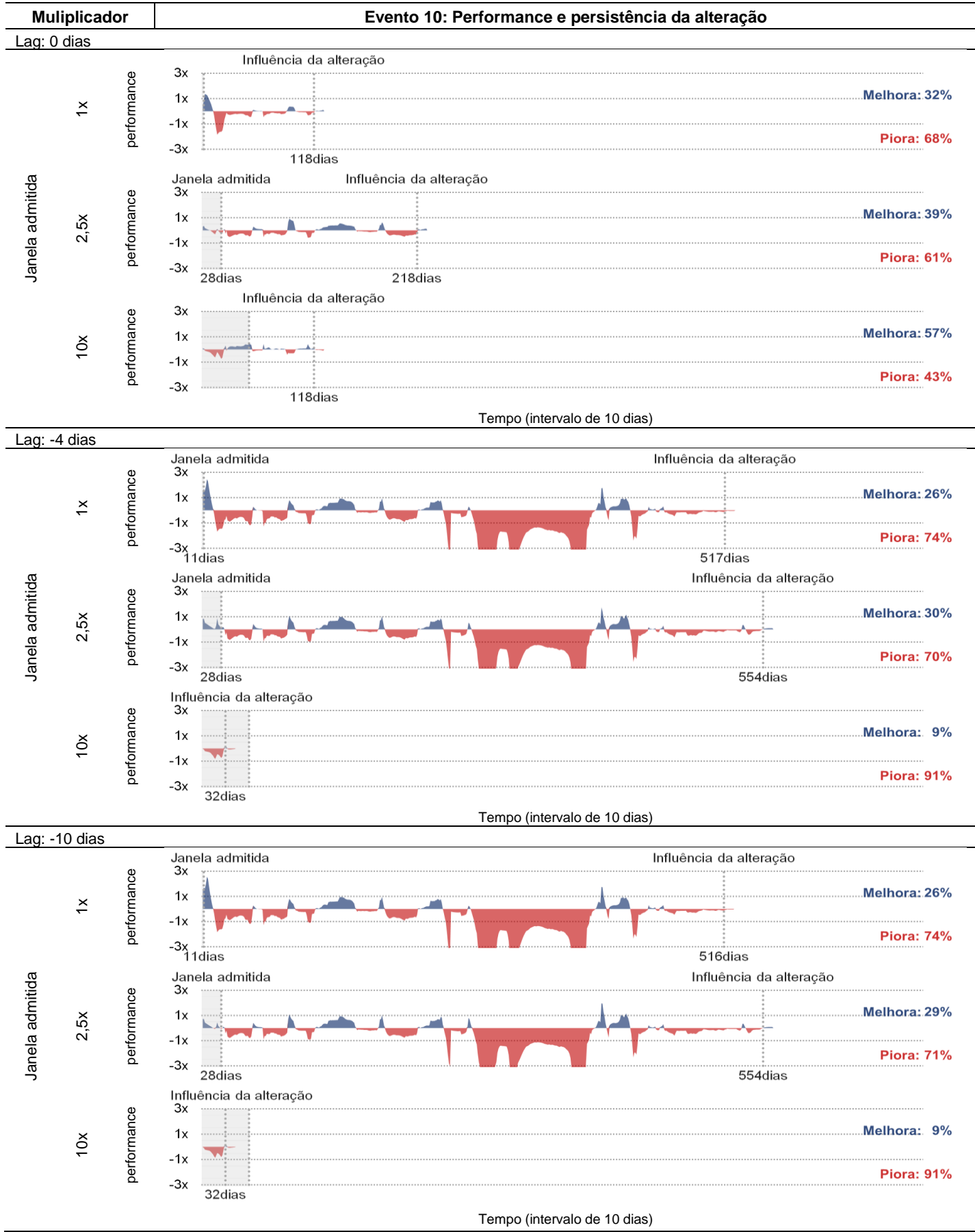
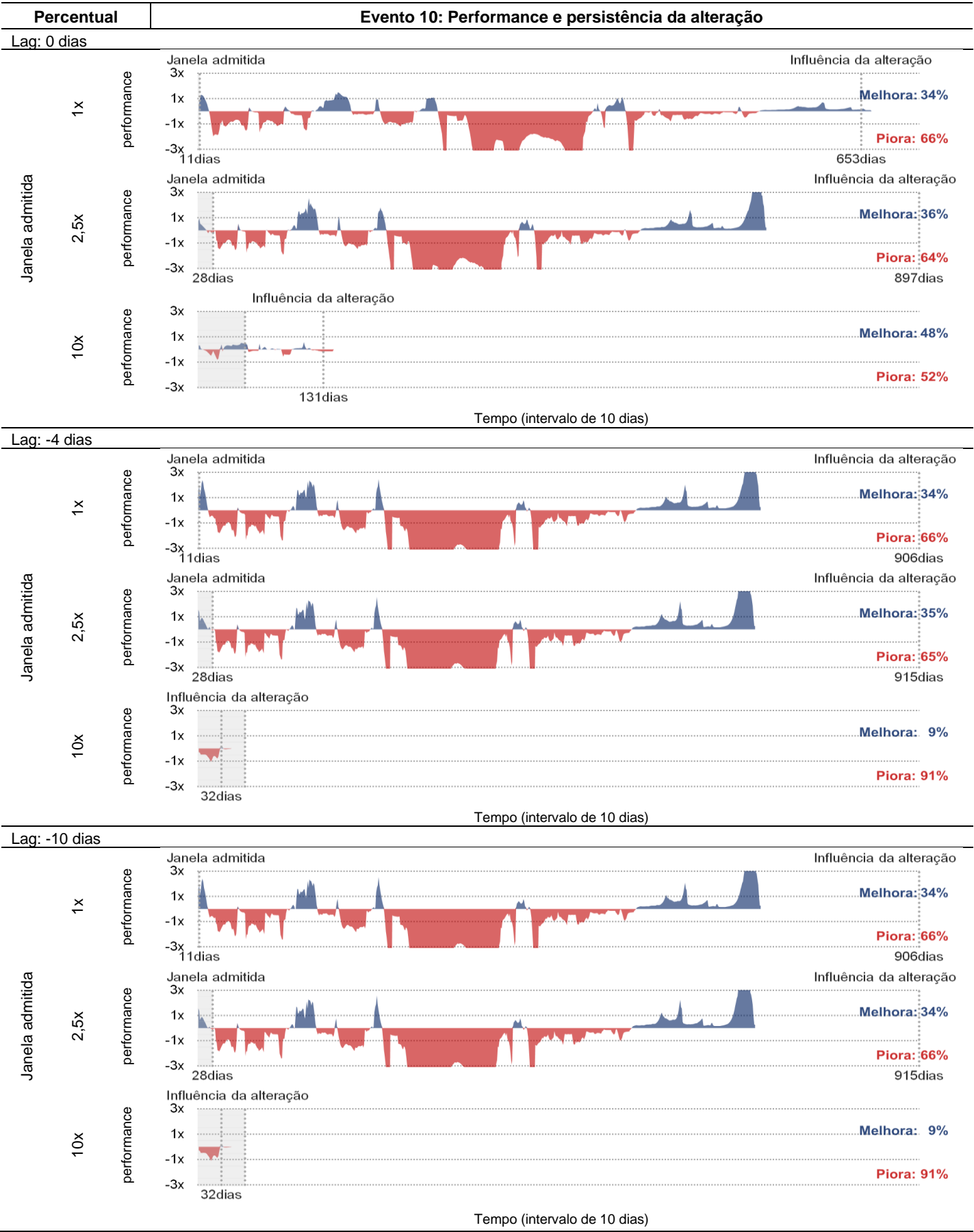






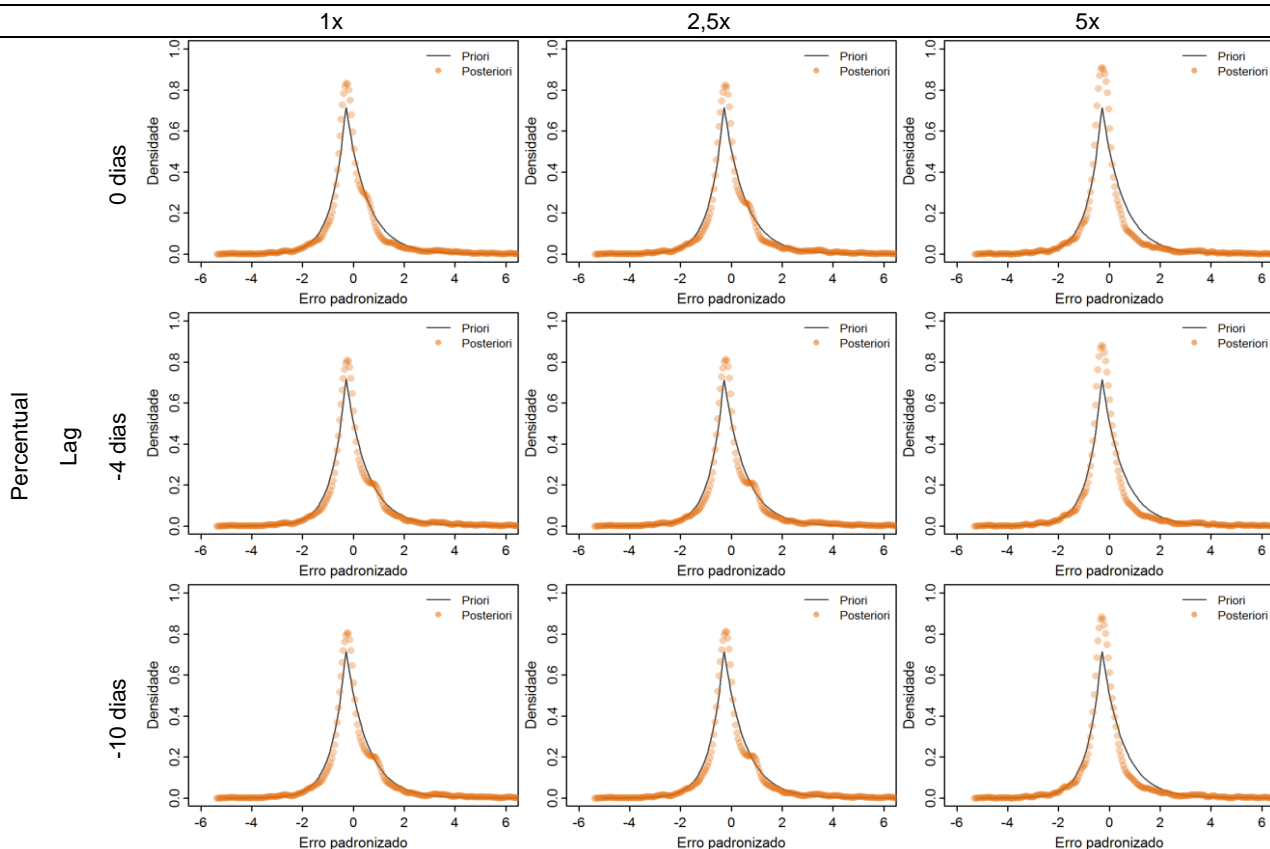




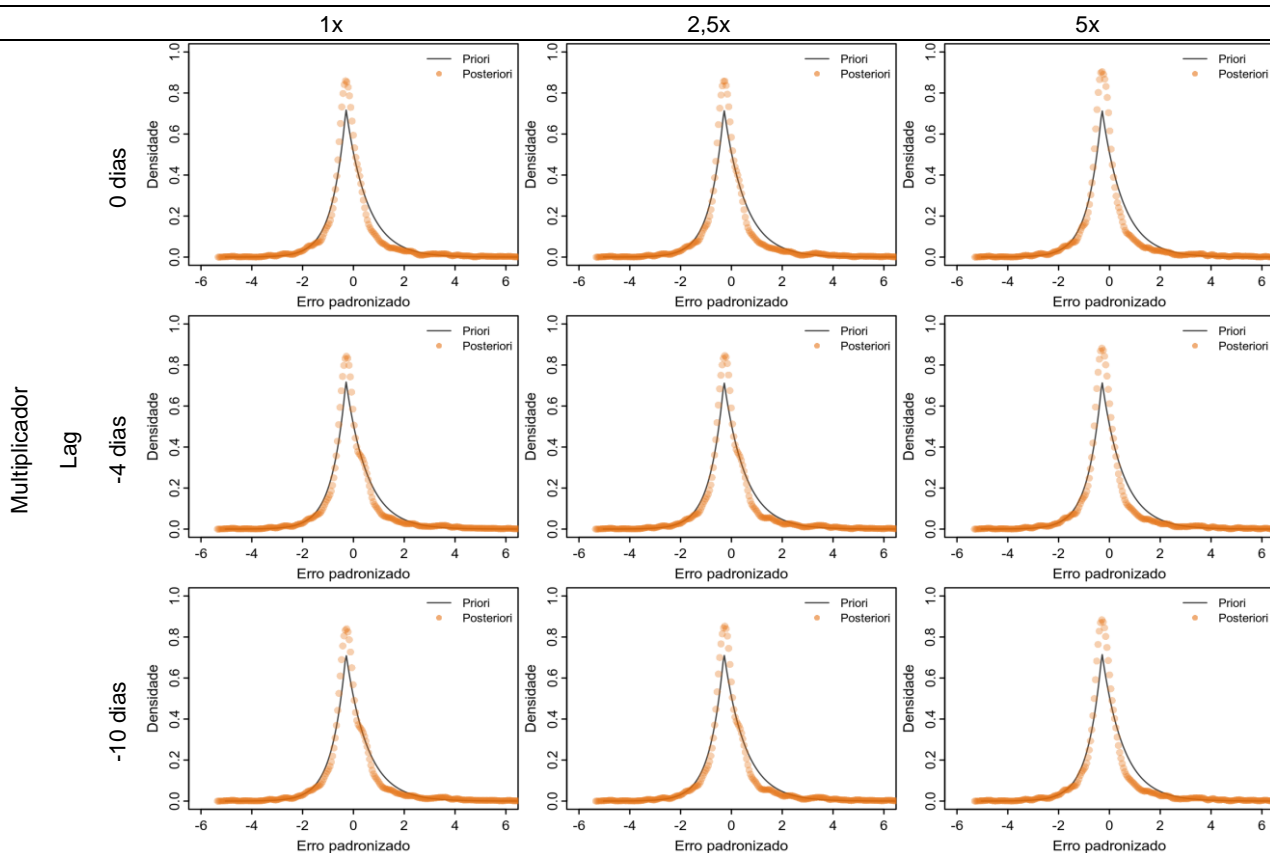


## Evento 10: Comportamento probabilístico dos resíduos da simulação

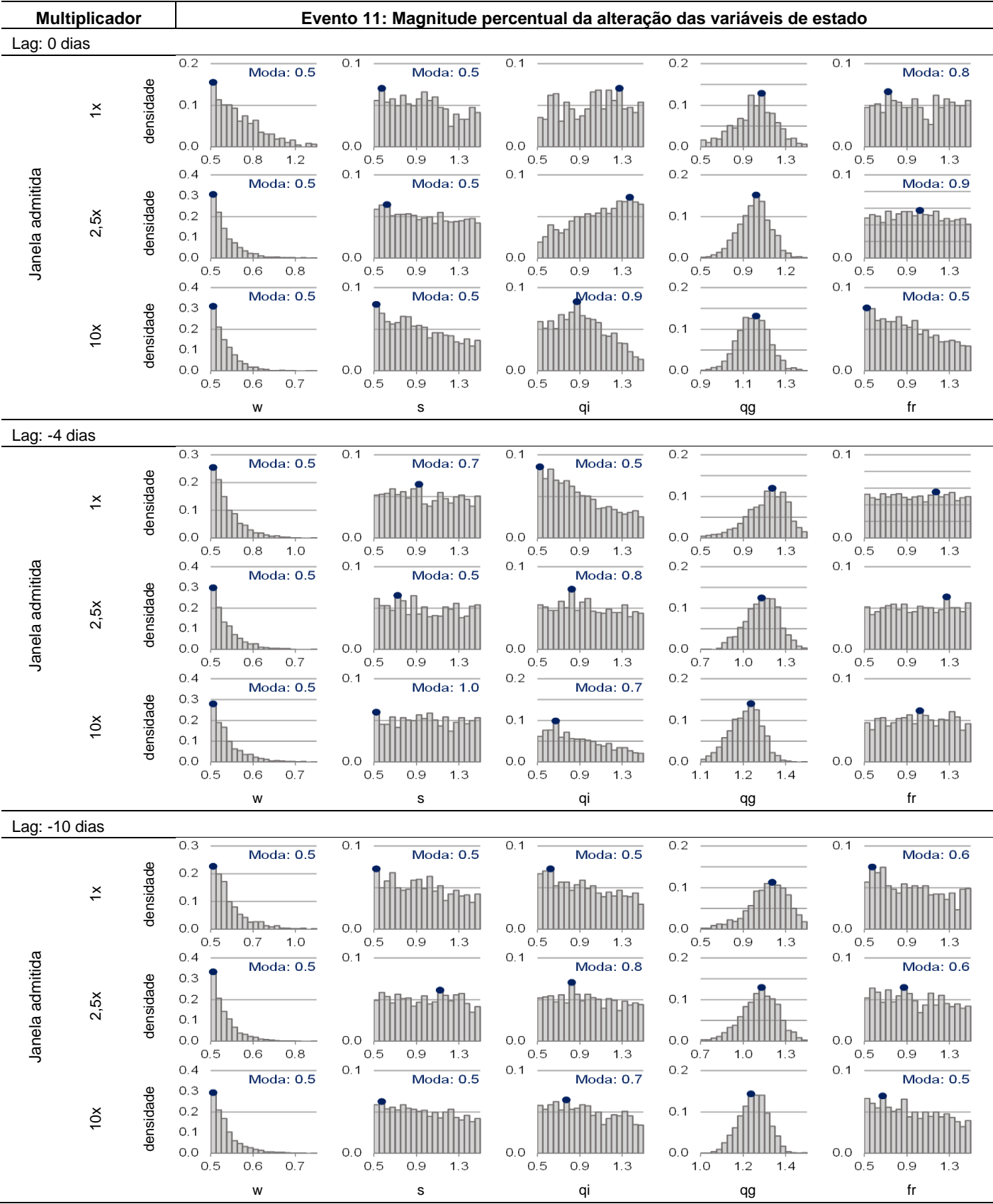
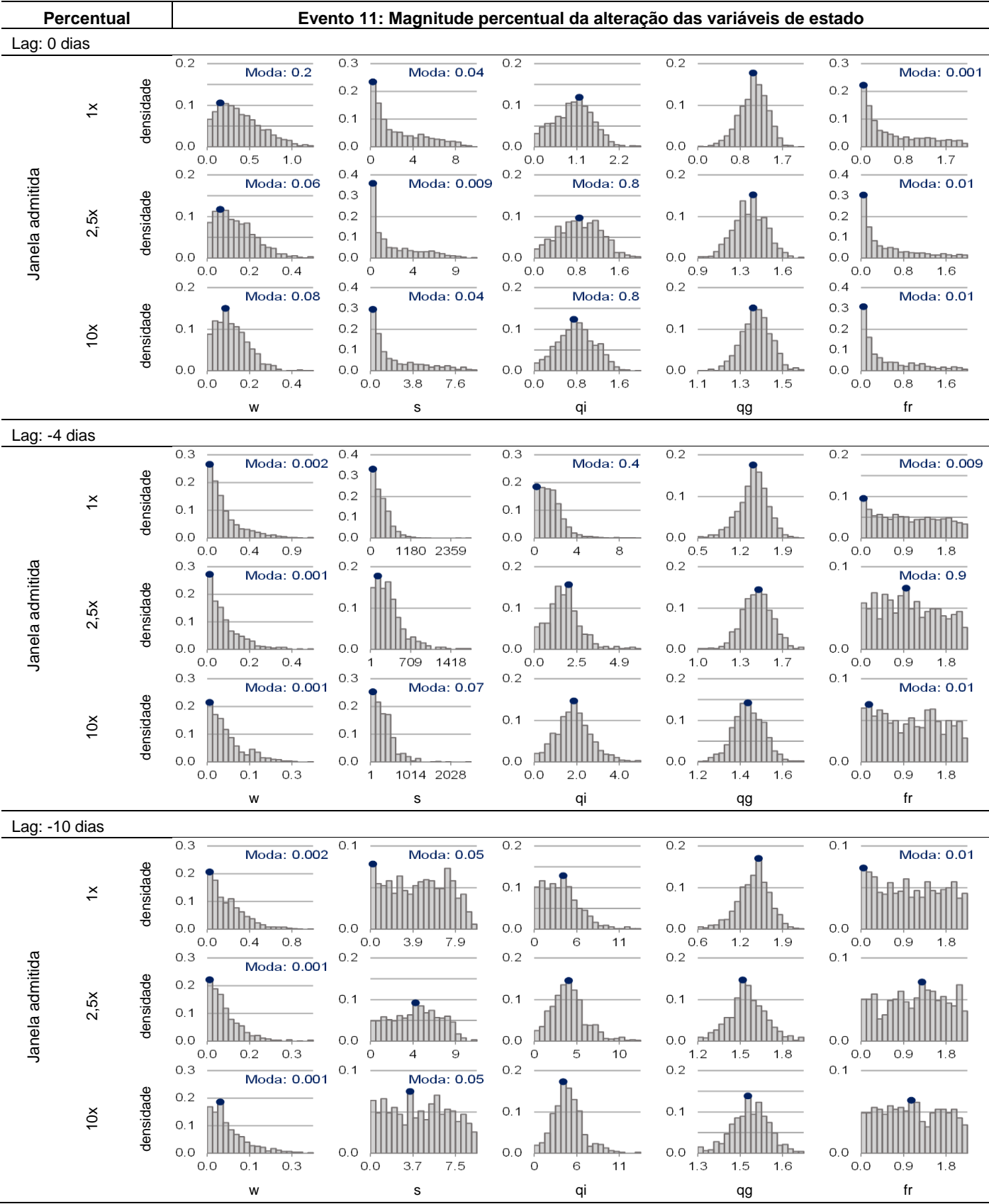
Janela admitida

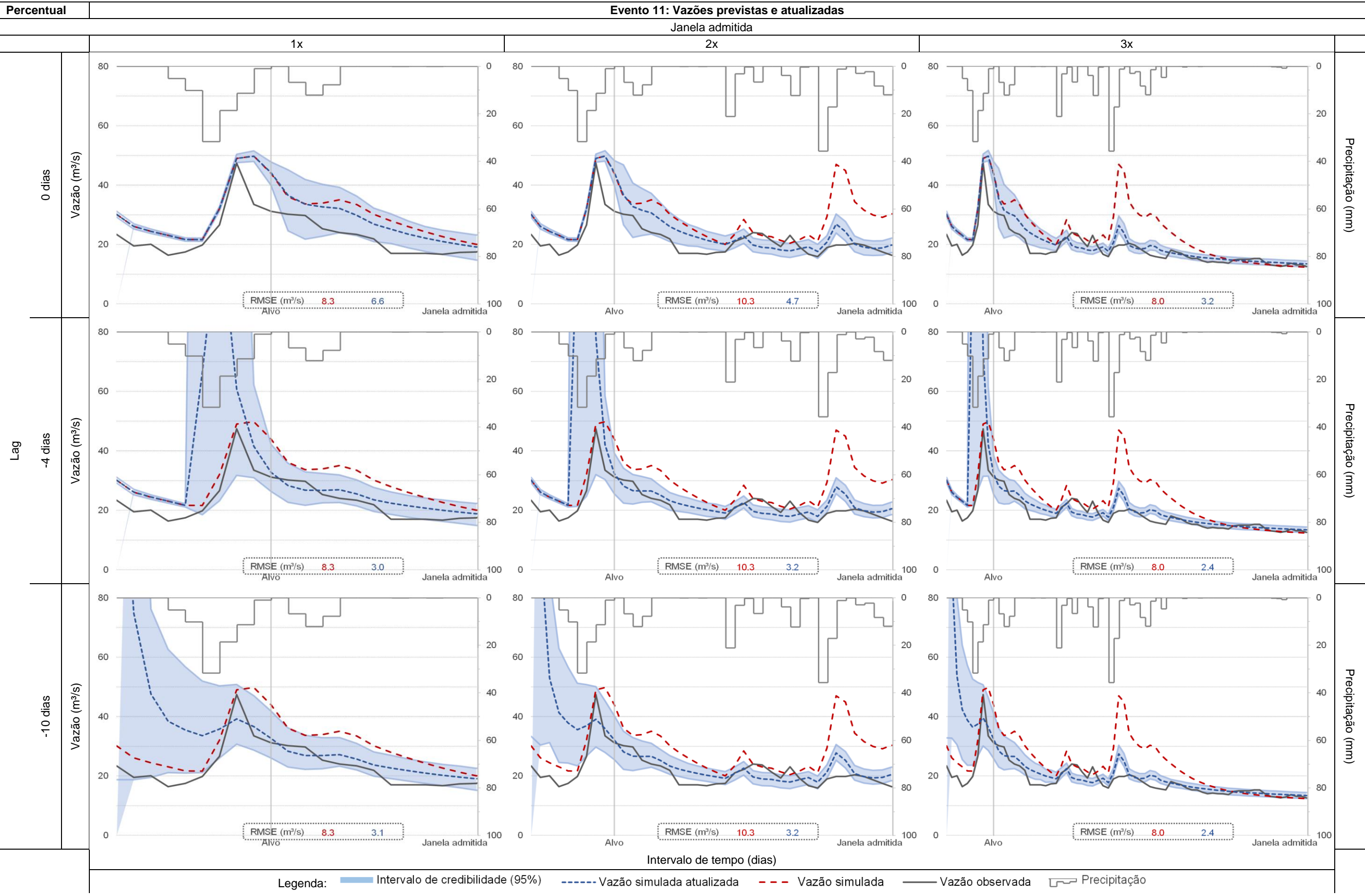


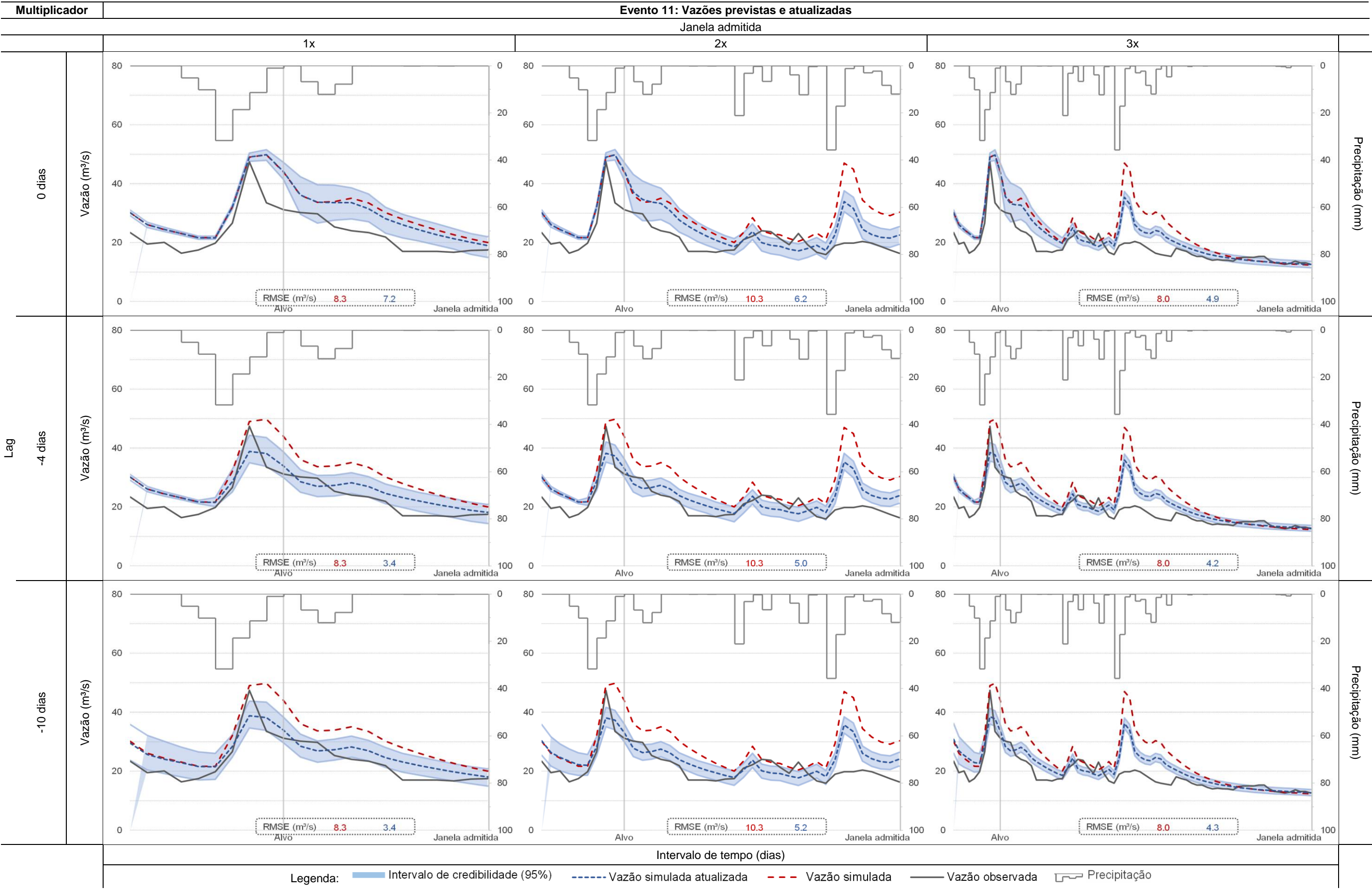
Janela admitida



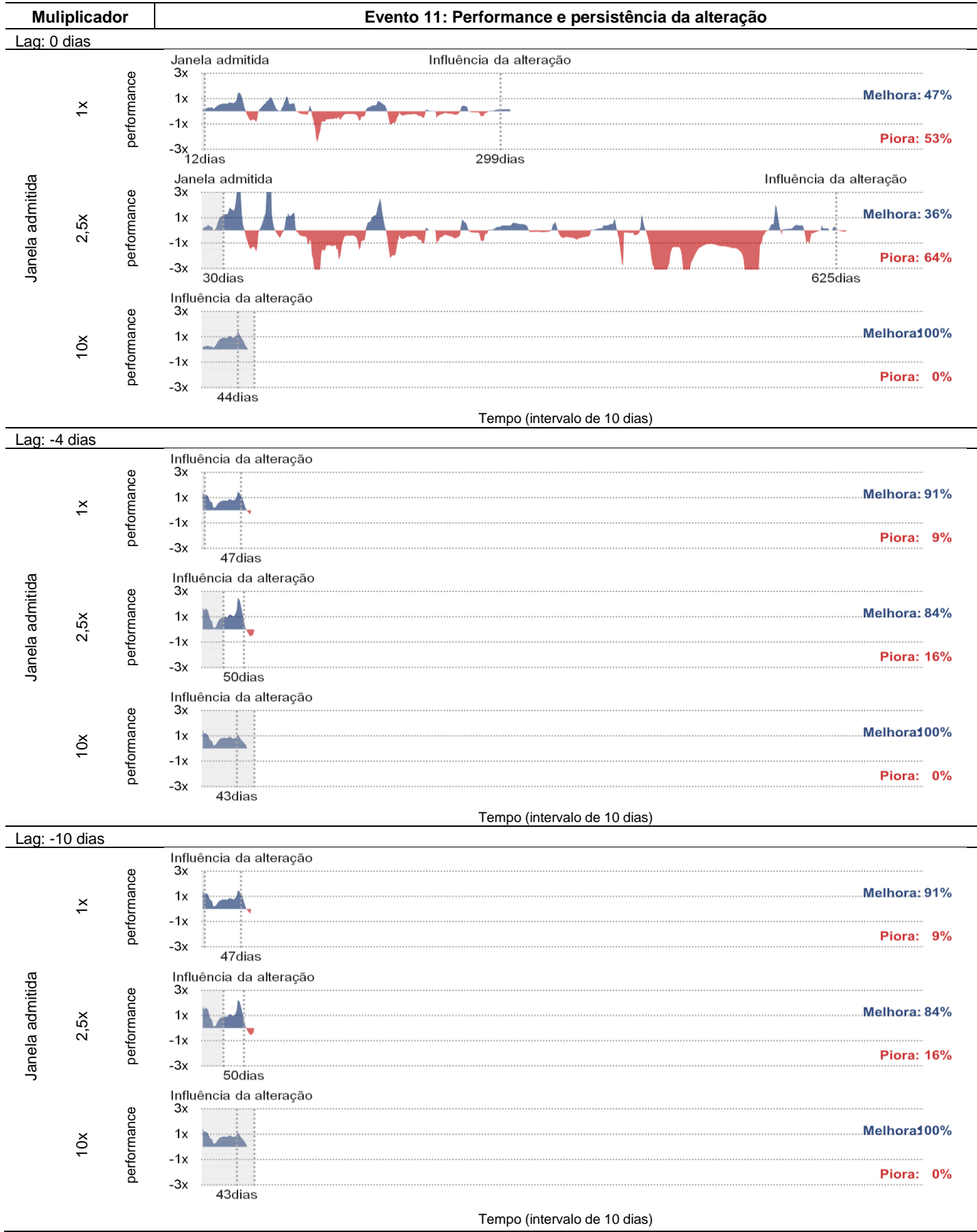
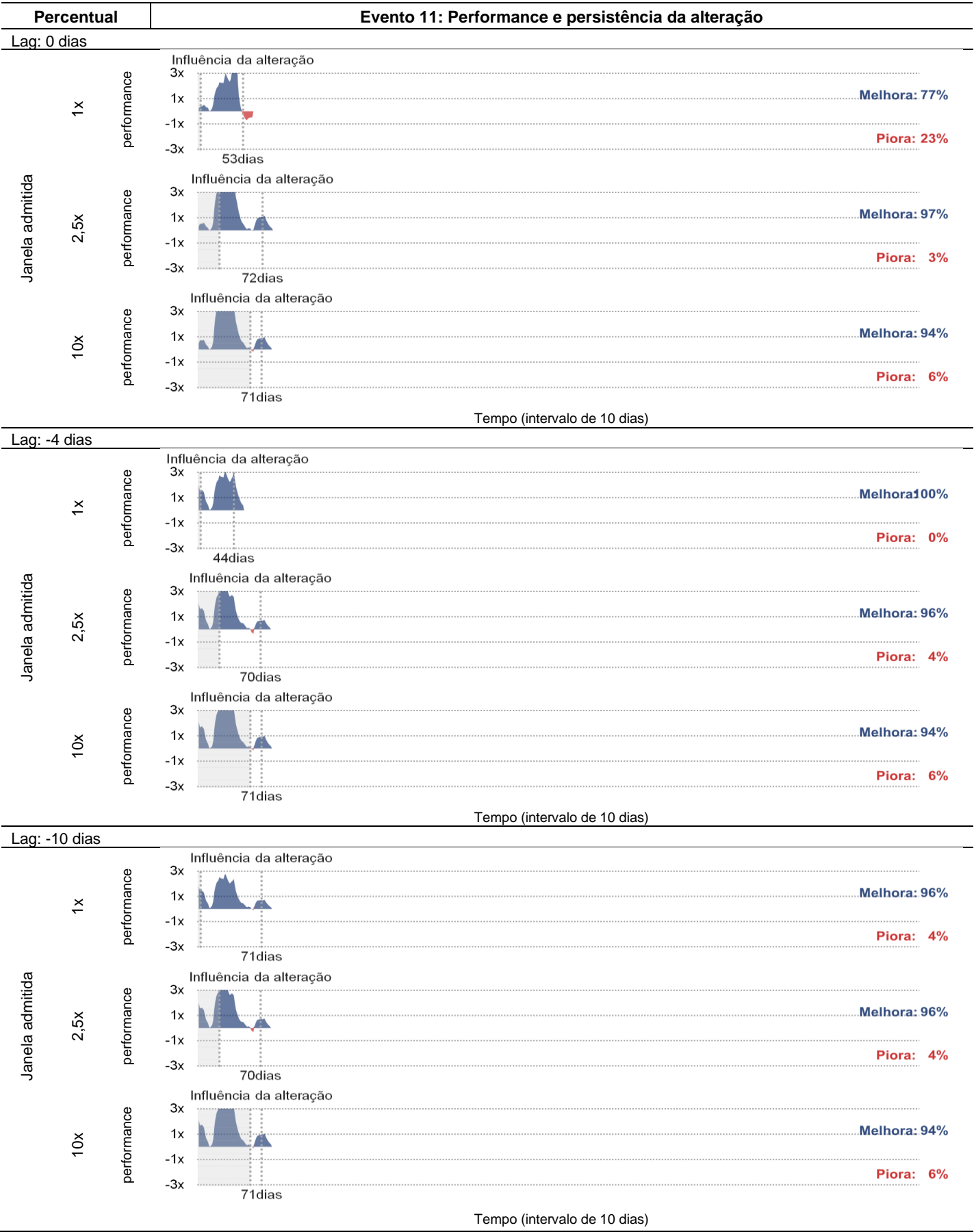






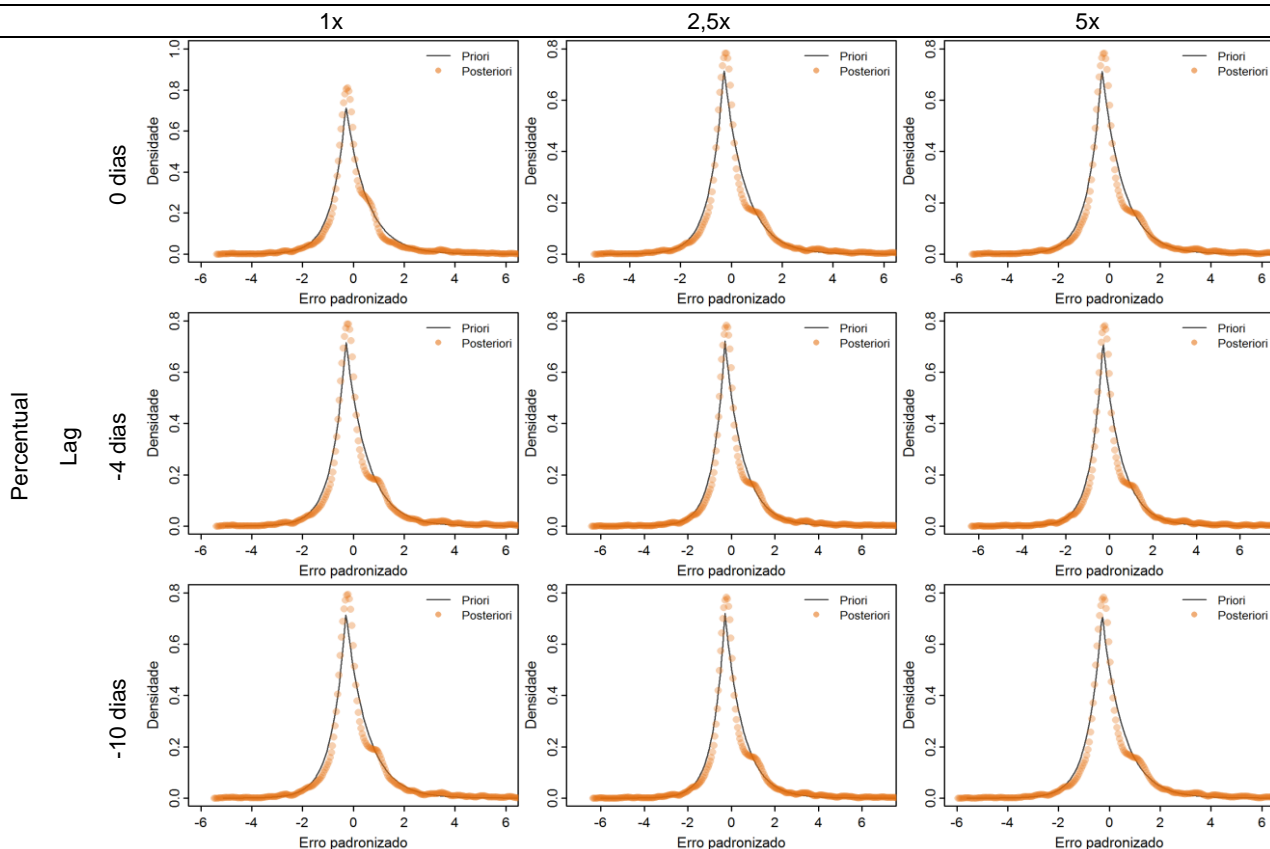




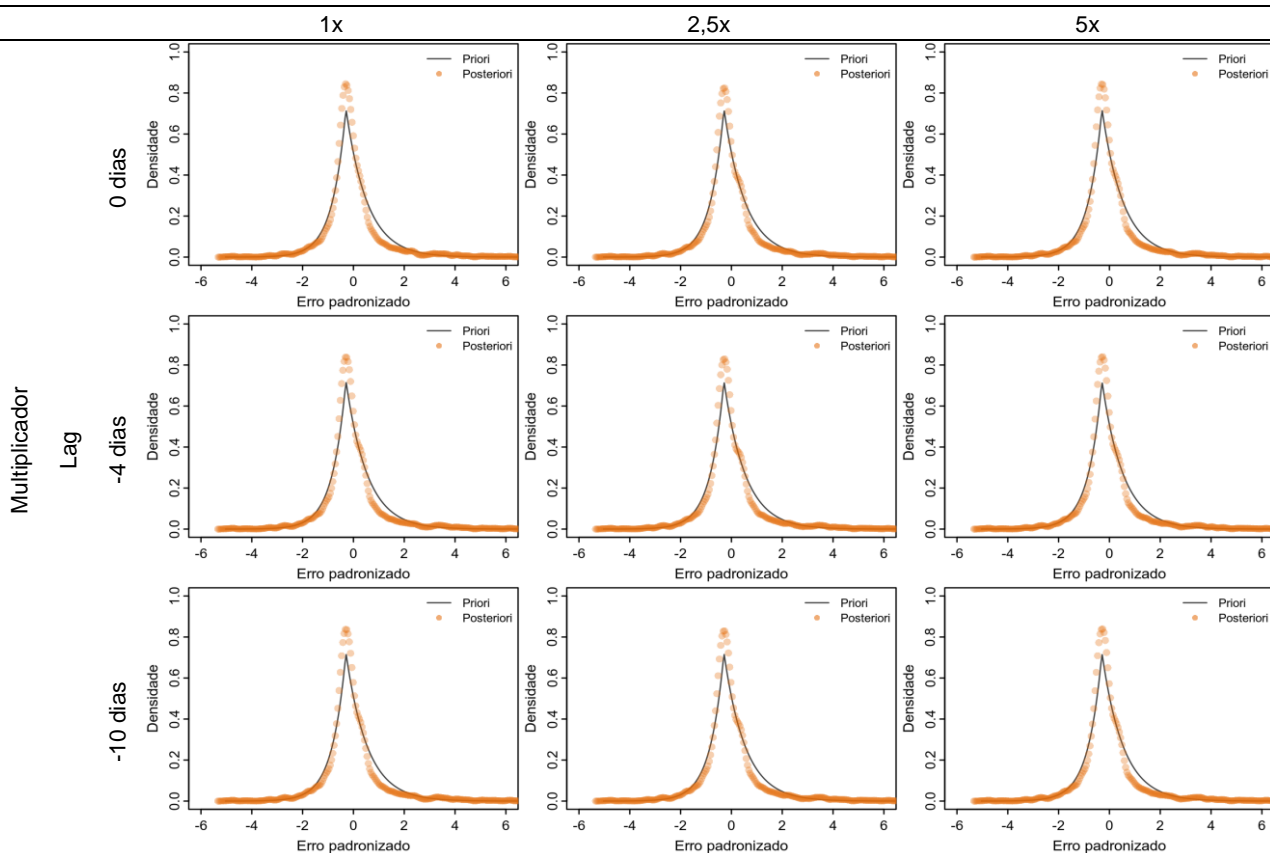


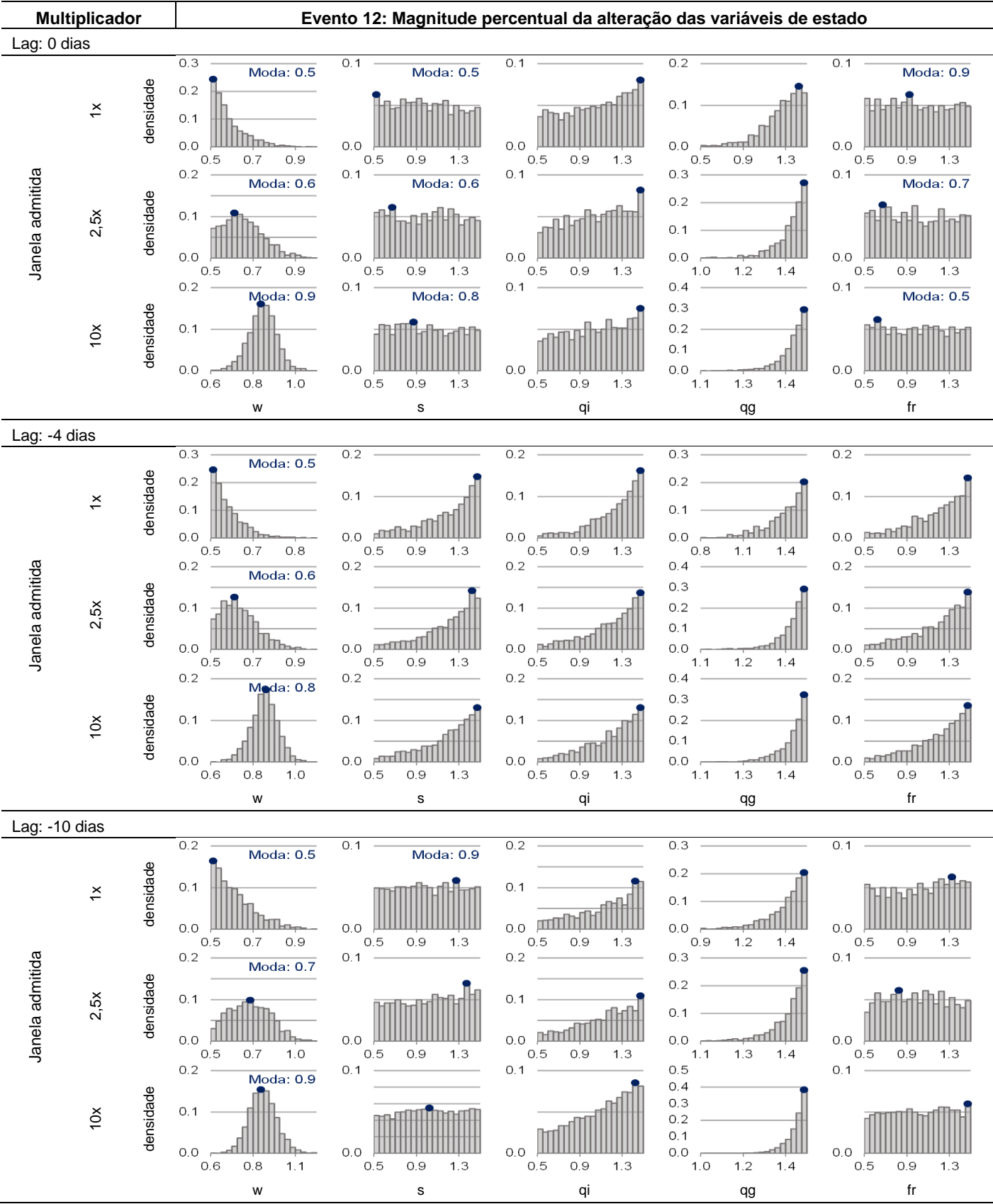
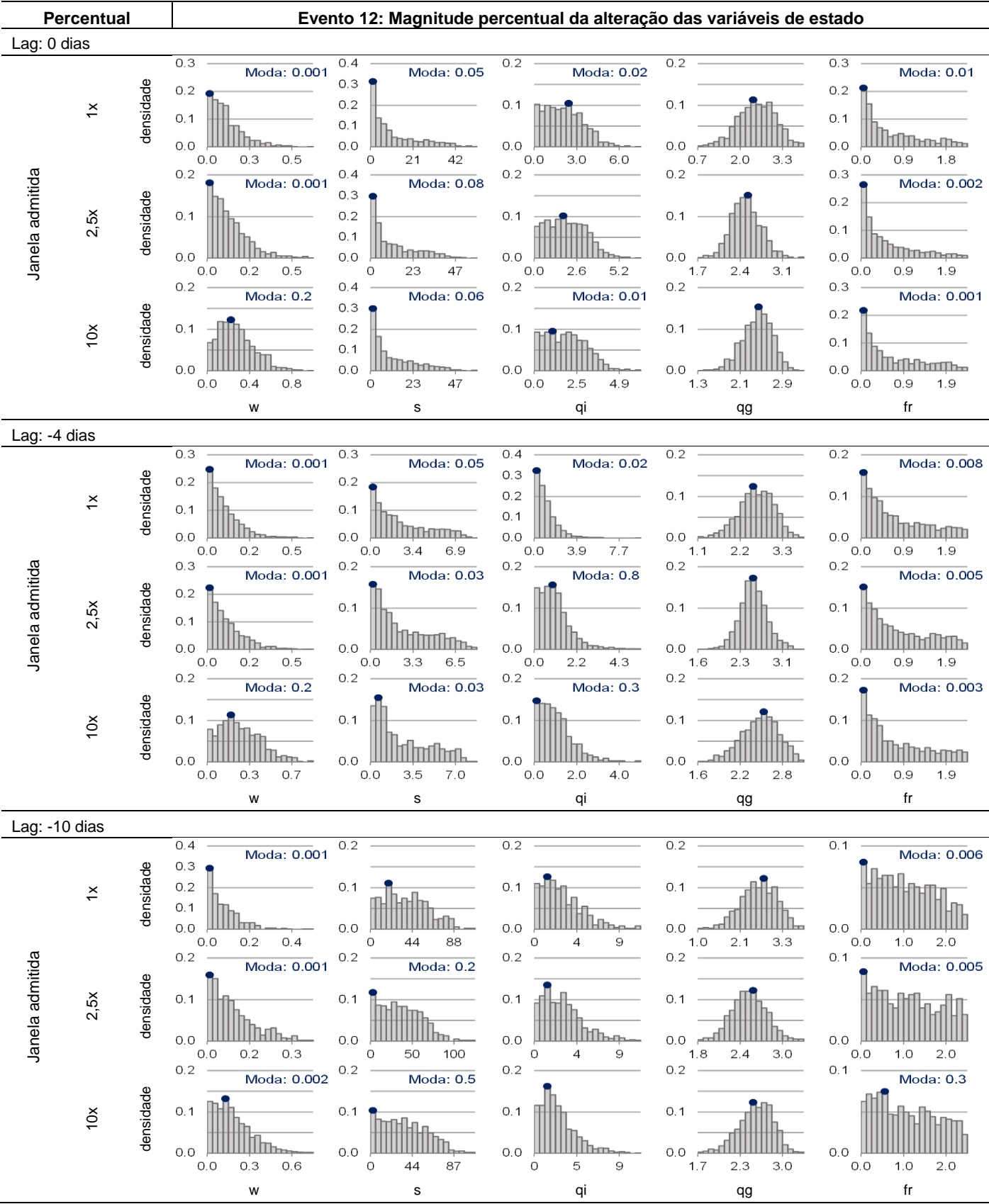
## Evento 11: Comportamento probabilístico dos resíduos da simulação

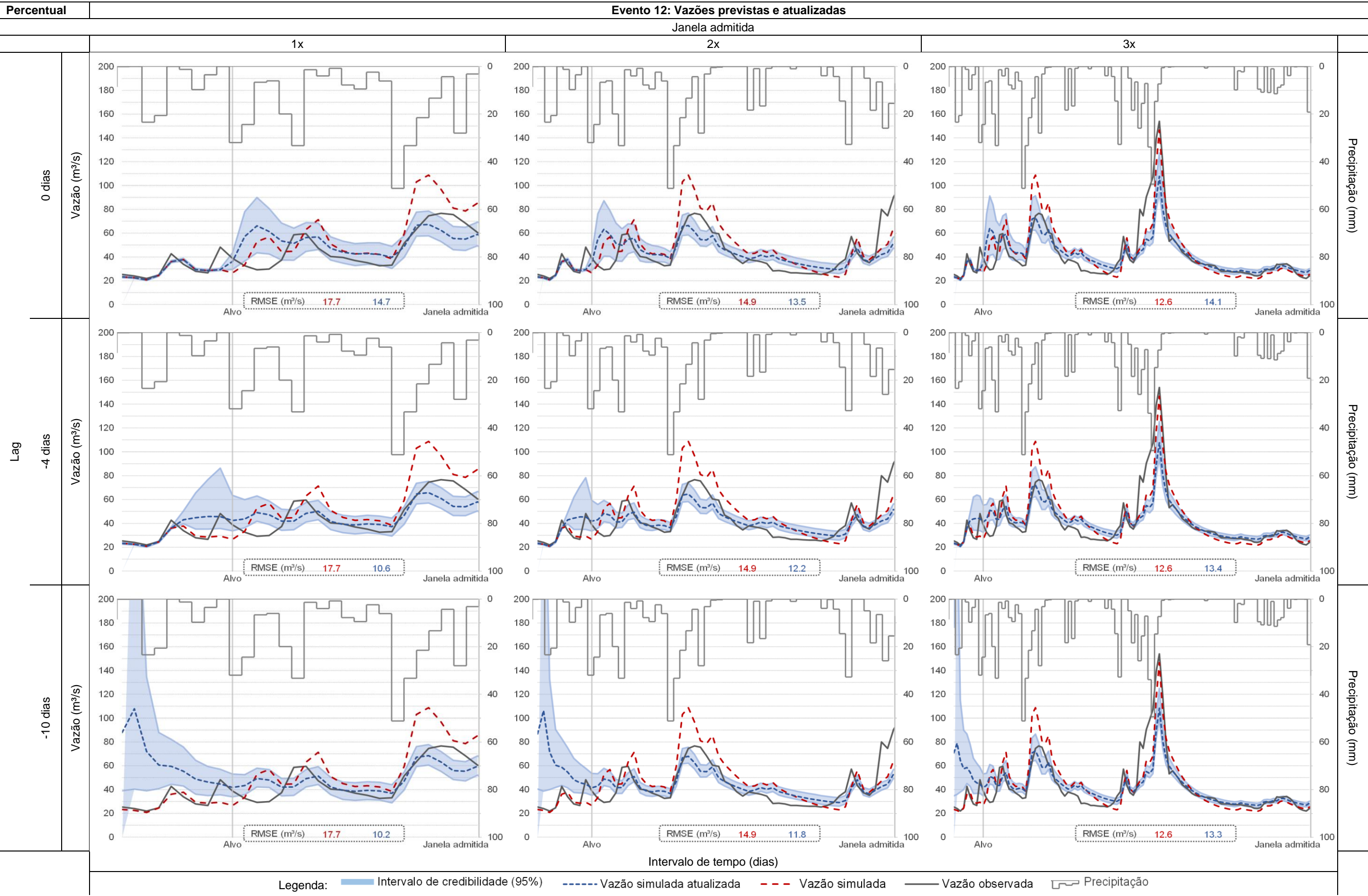
Janela admitida



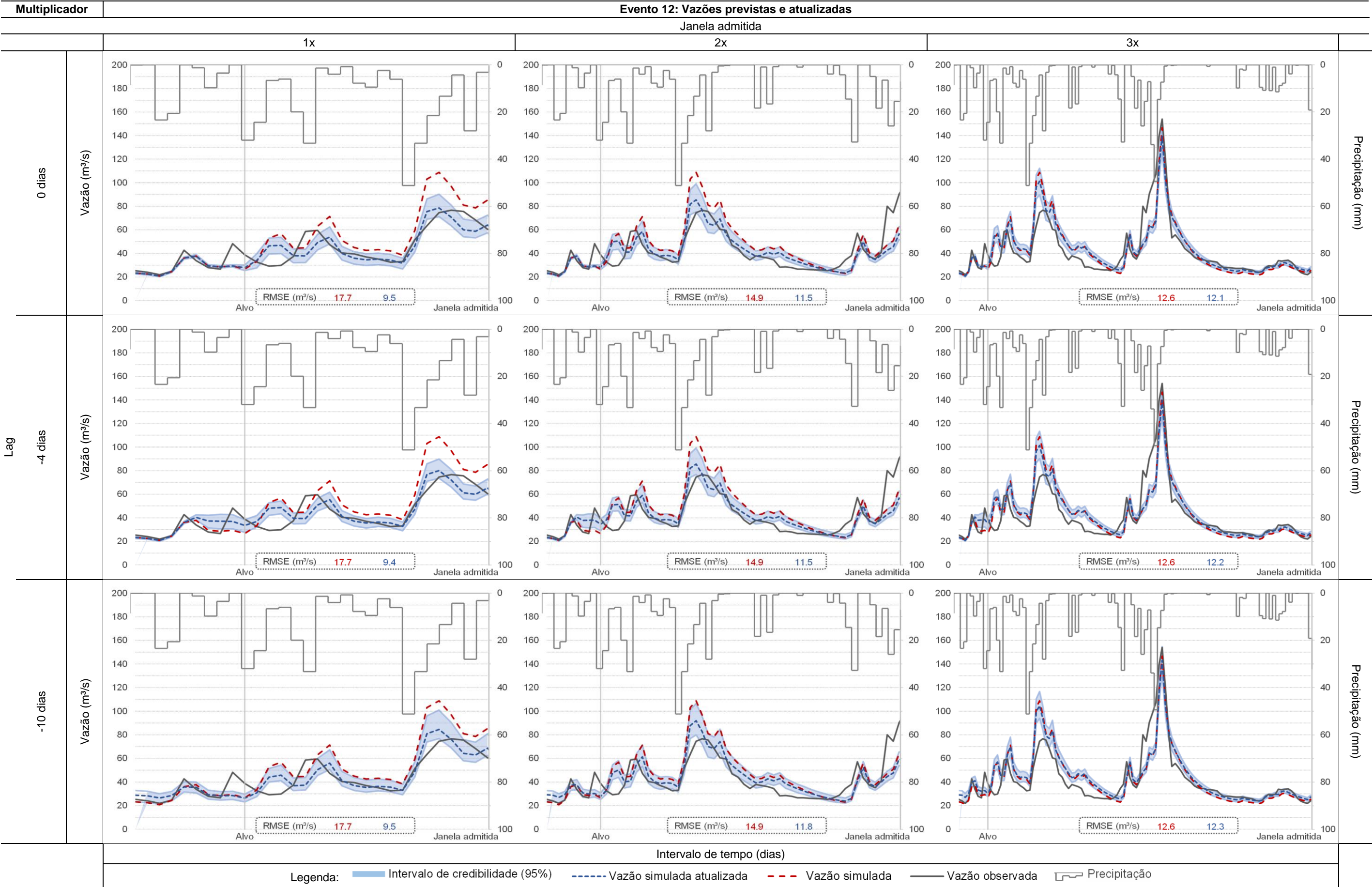
Janela admitida



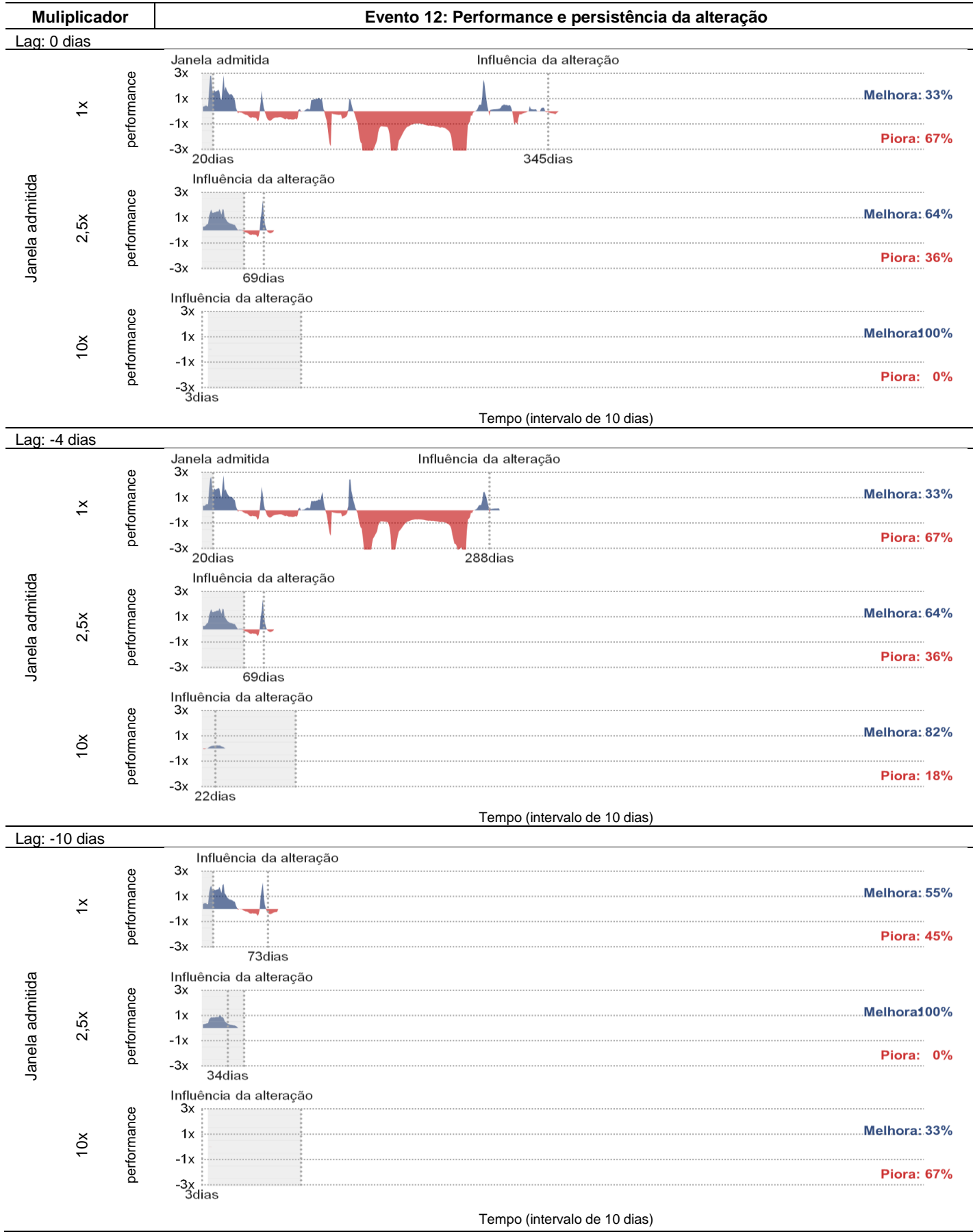
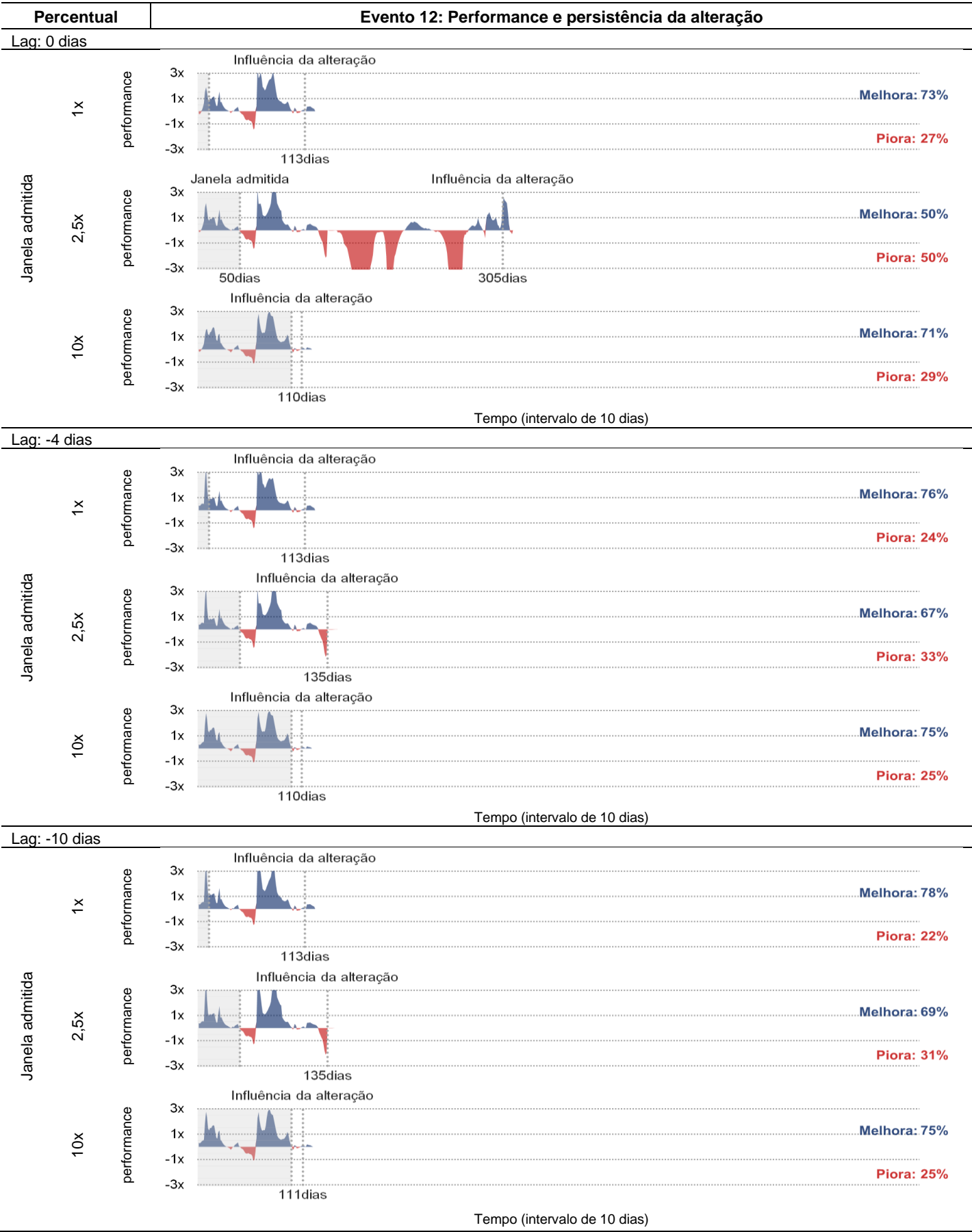






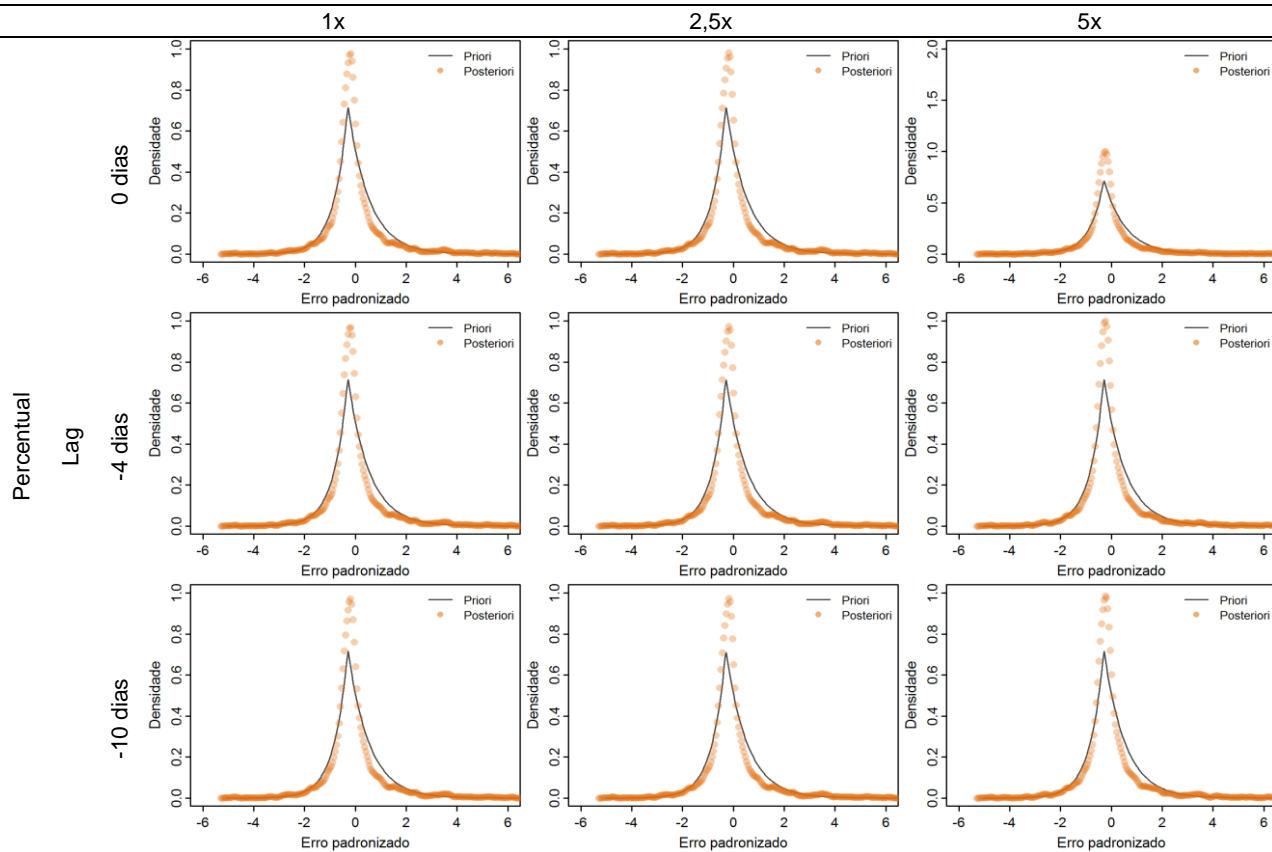




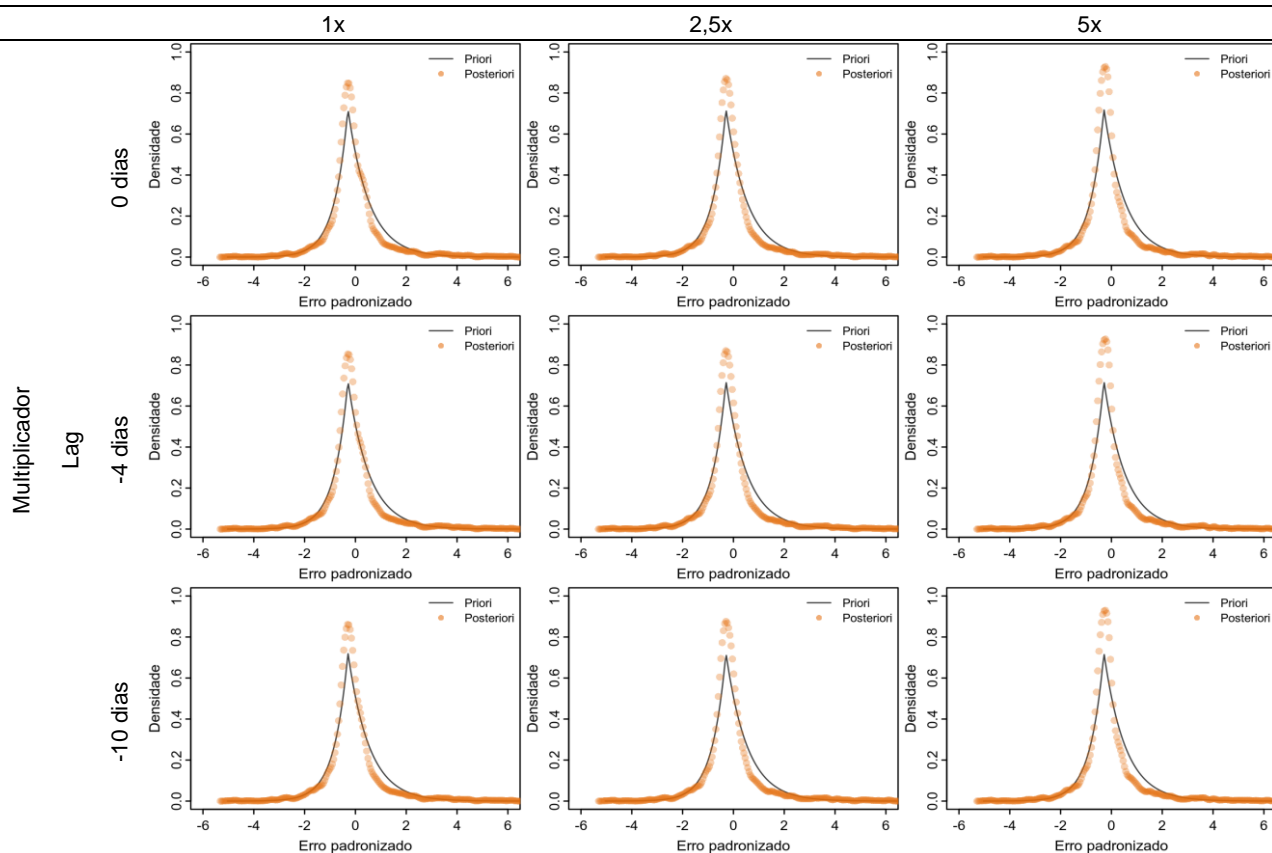


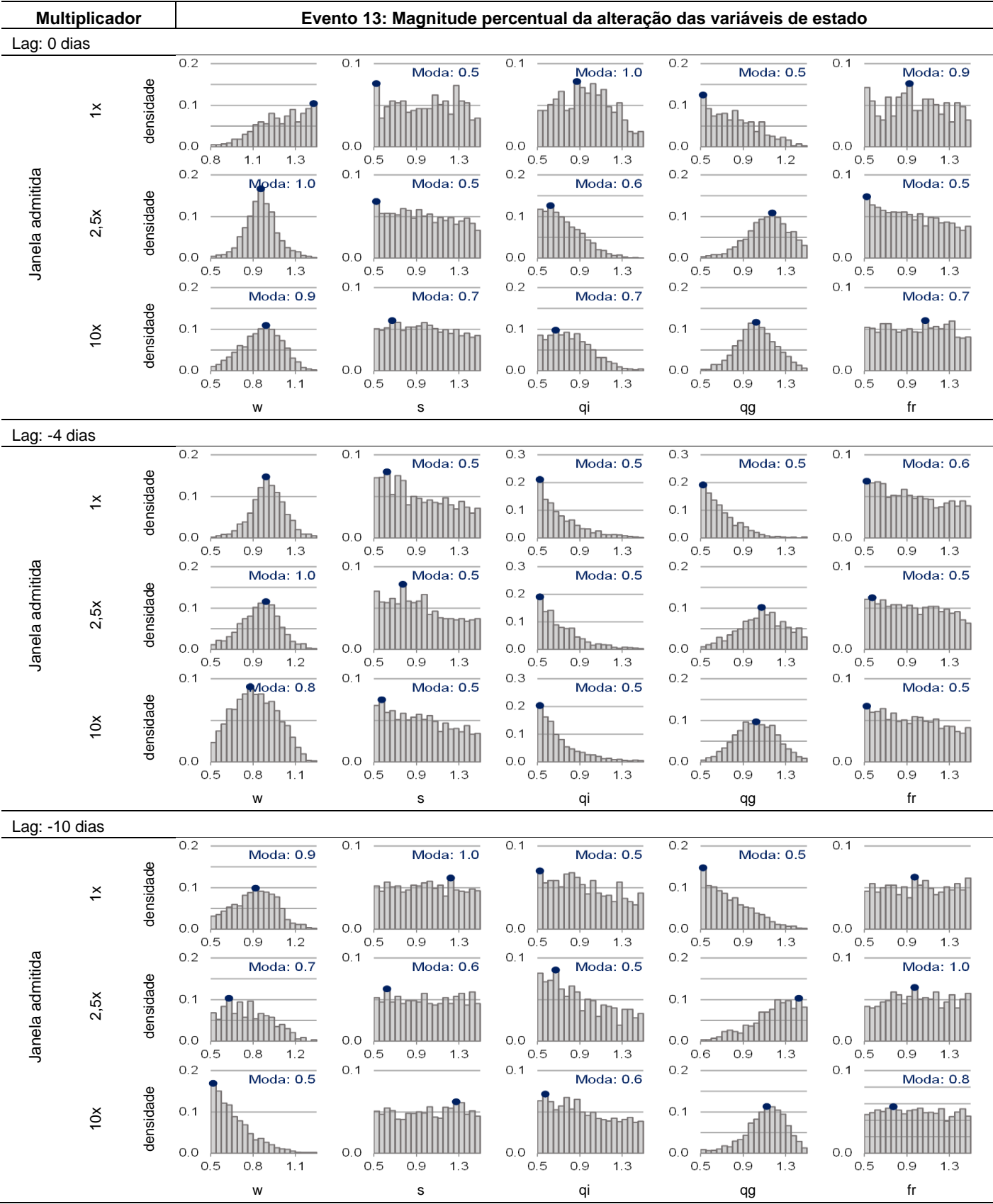
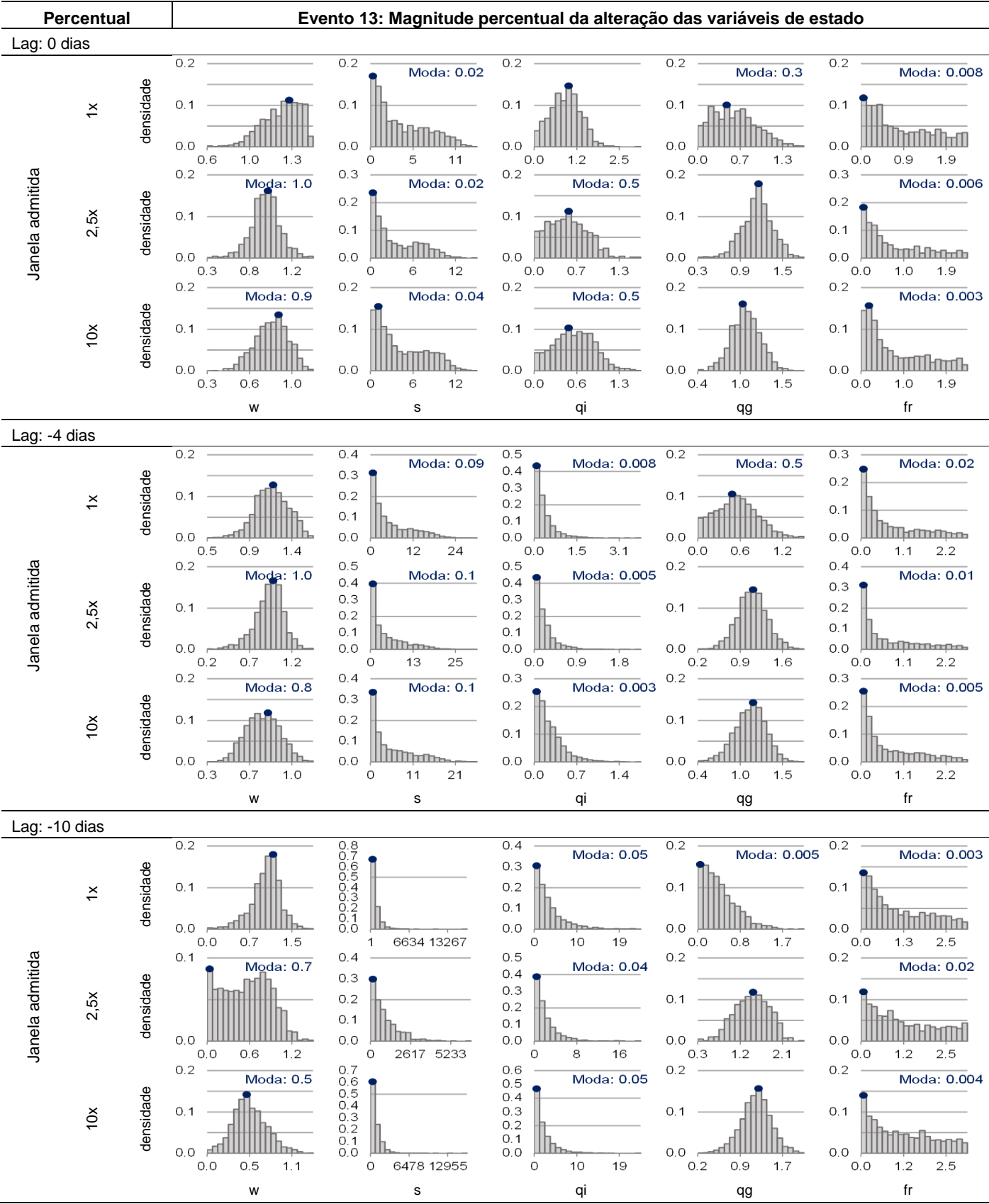
## Evento 12: Comportamento probabilístico dos resíduos da simulação

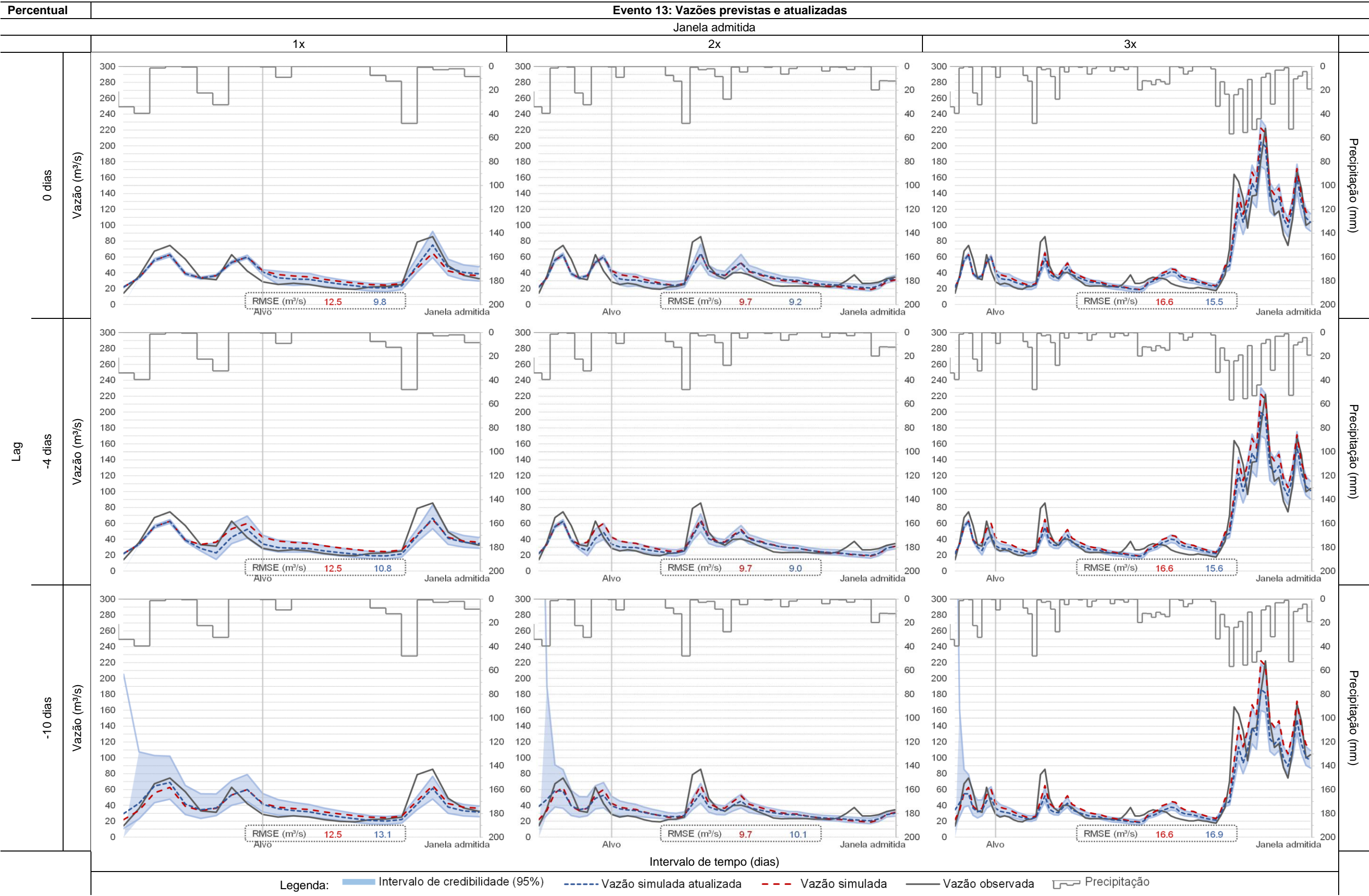
Janela admitida



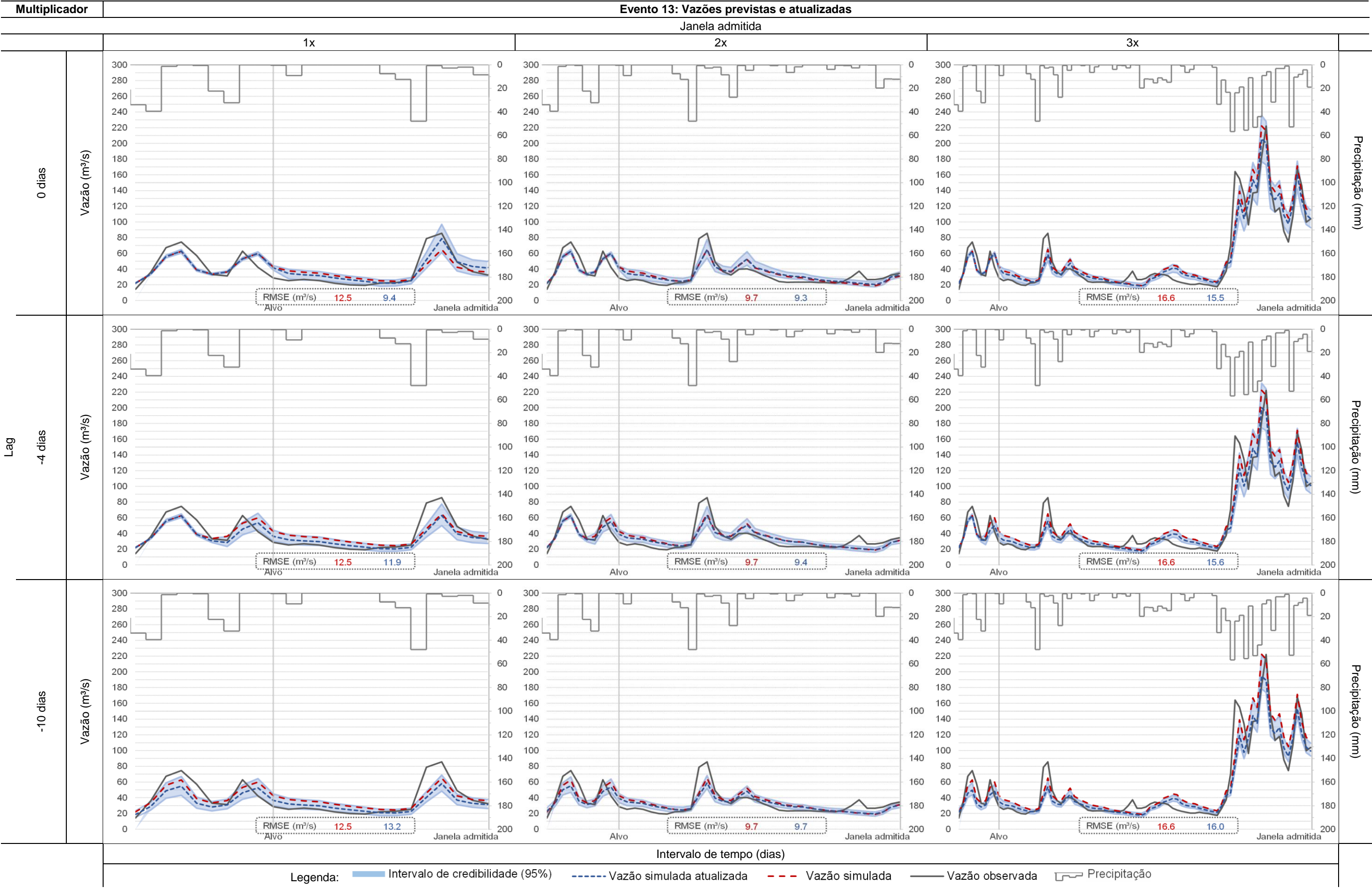
Janela admitida

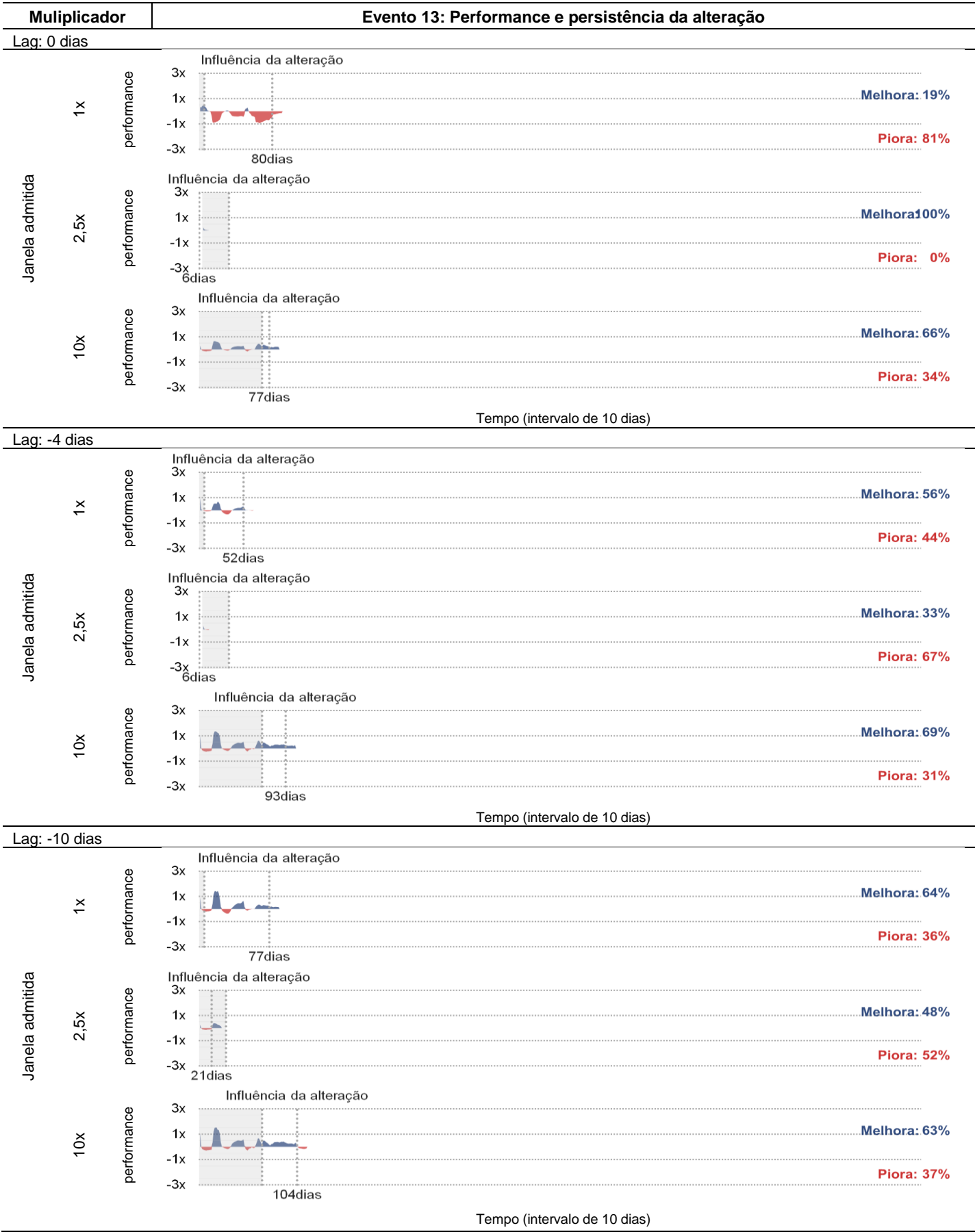
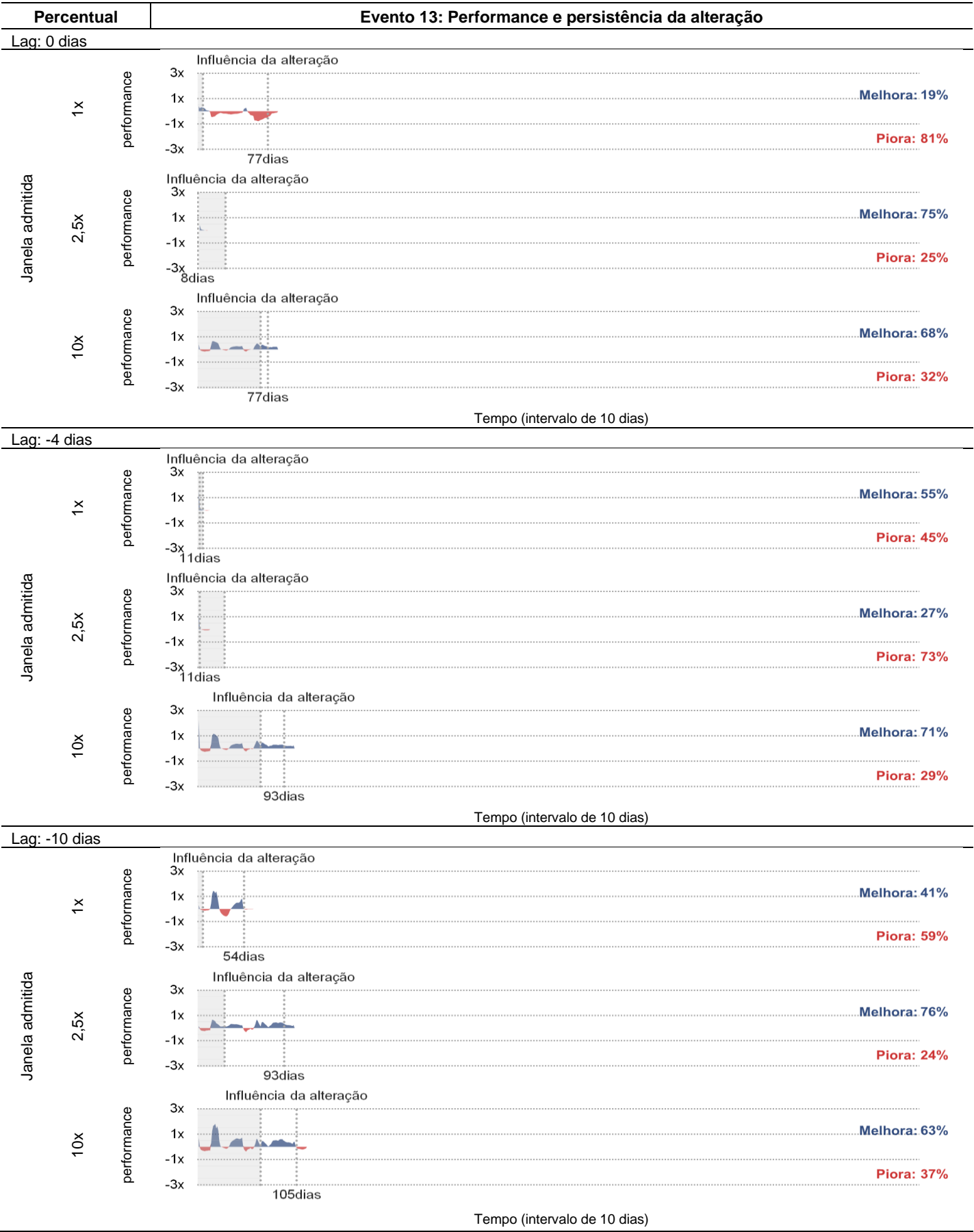






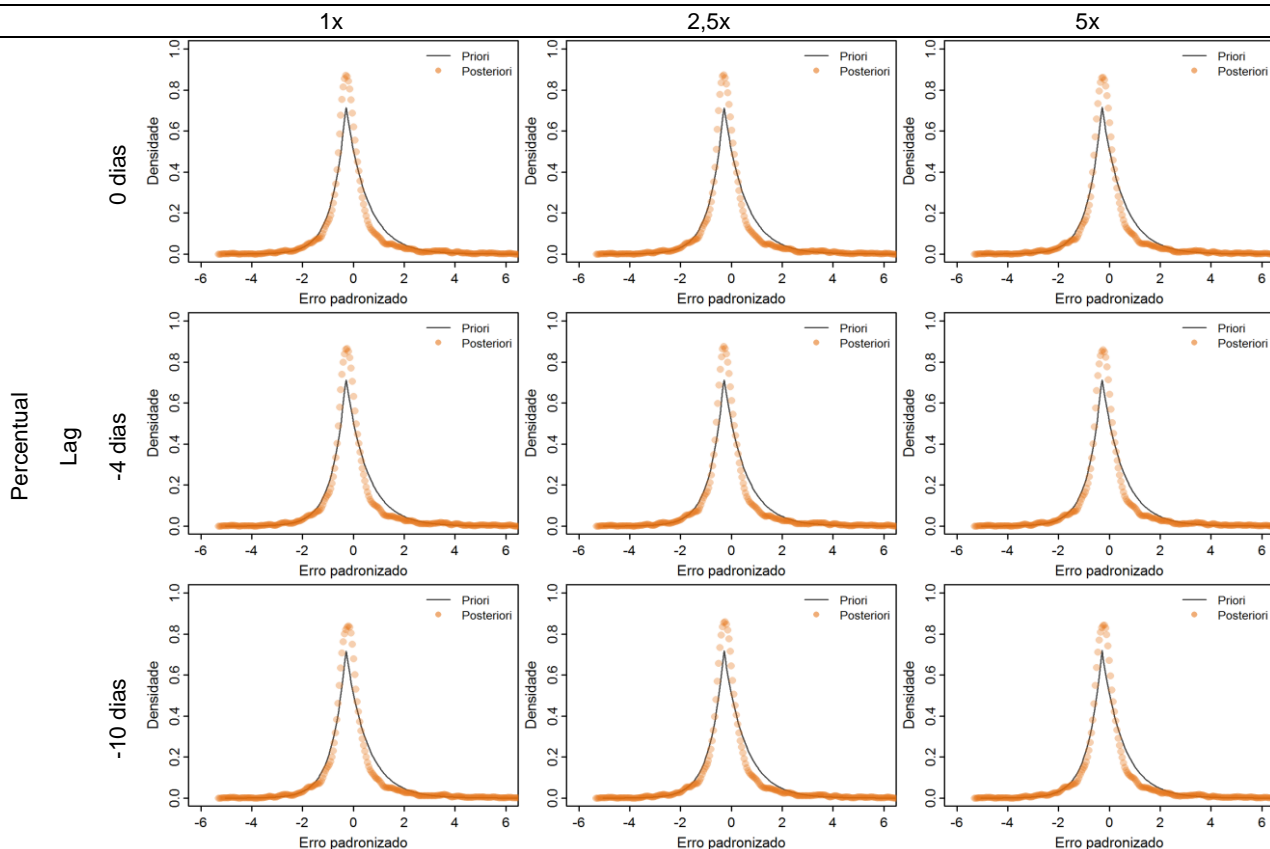




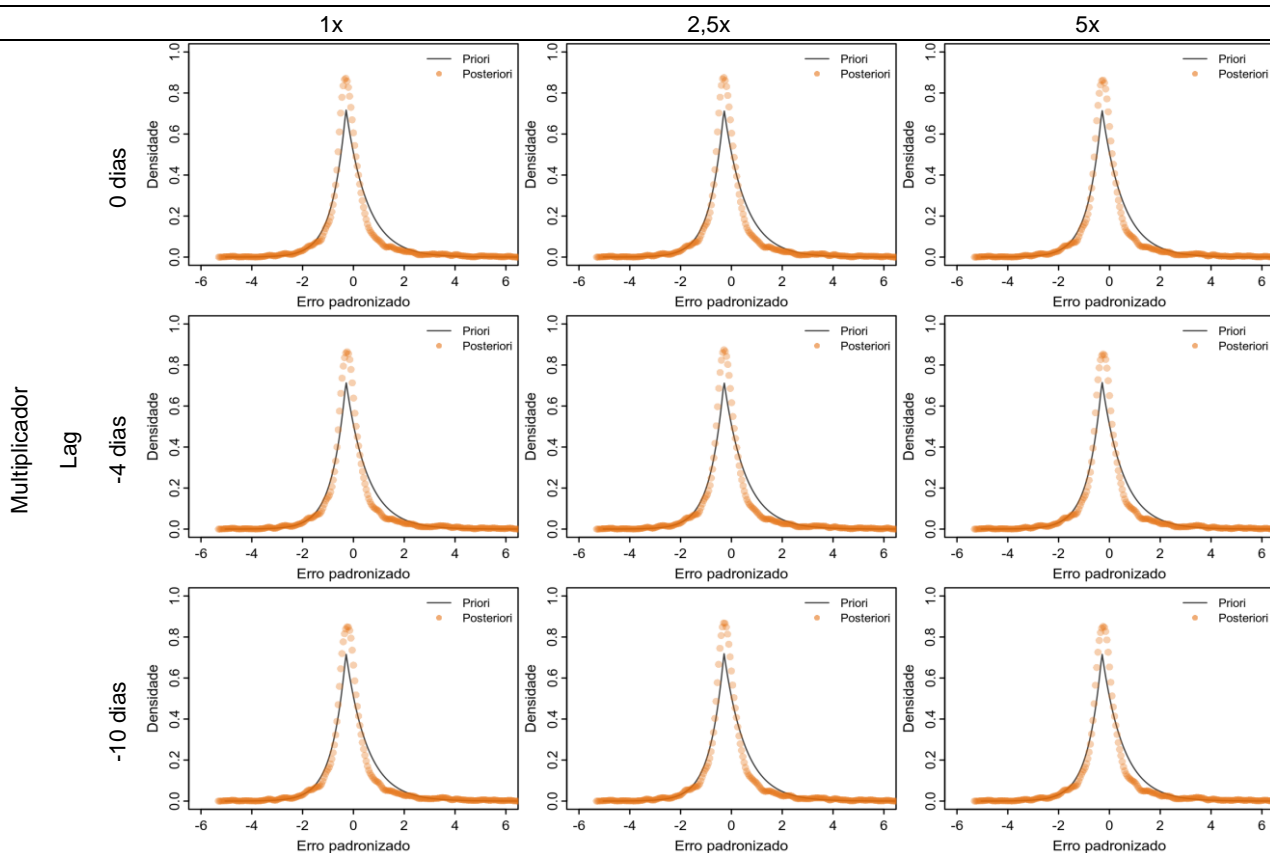


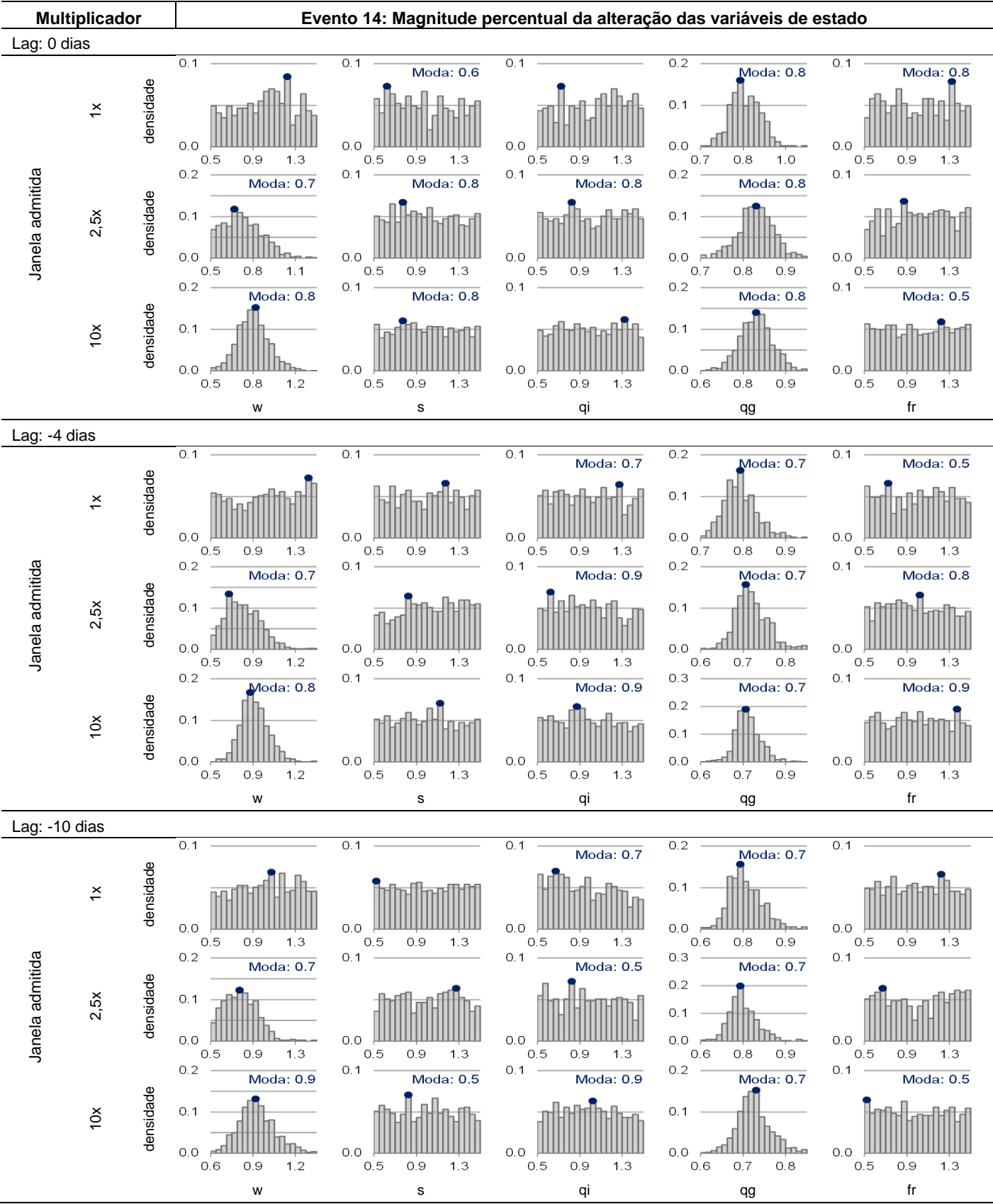
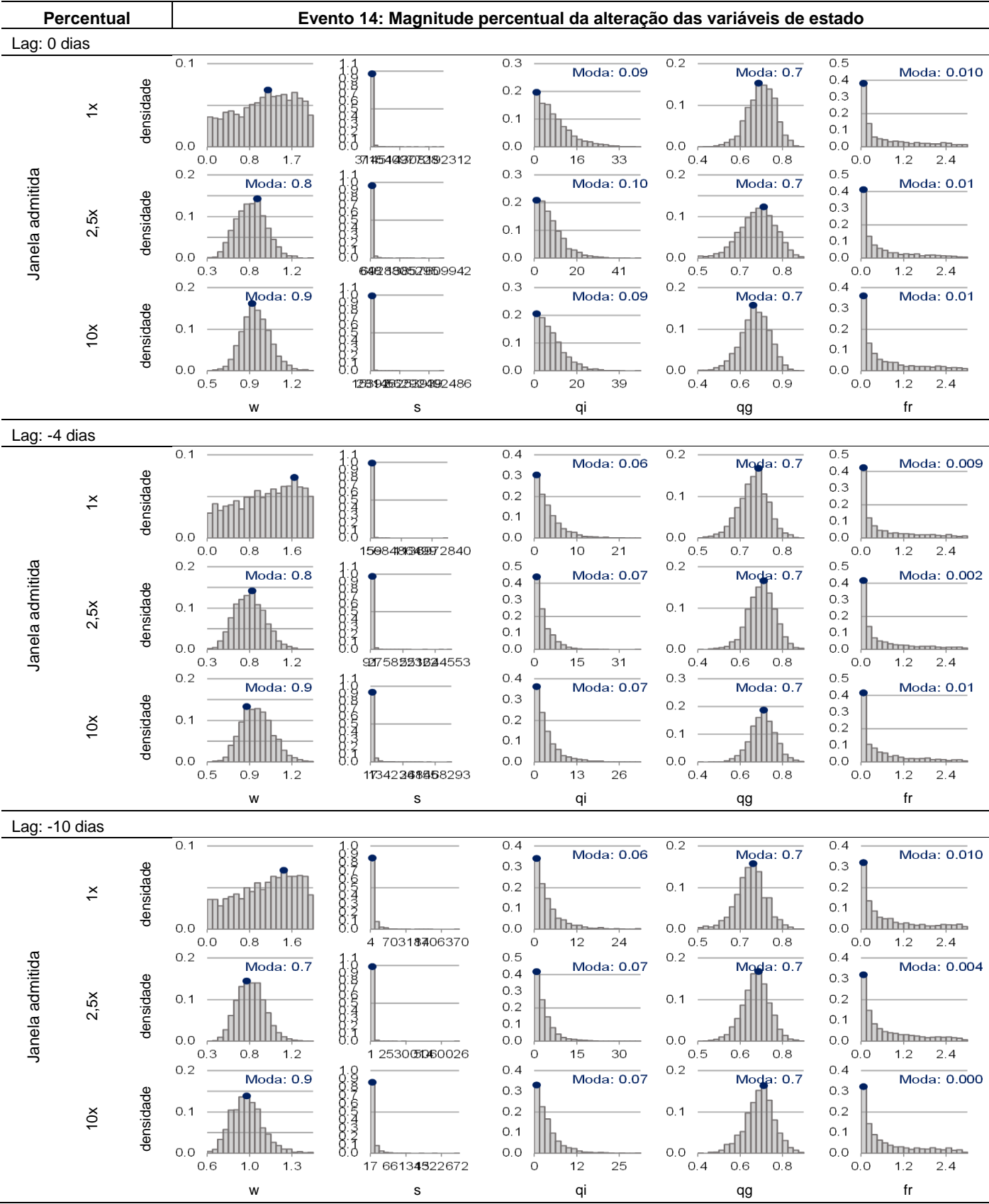
## Evento 13: Comportamento probabilístico dos resíduos da simulação

Janela admitida

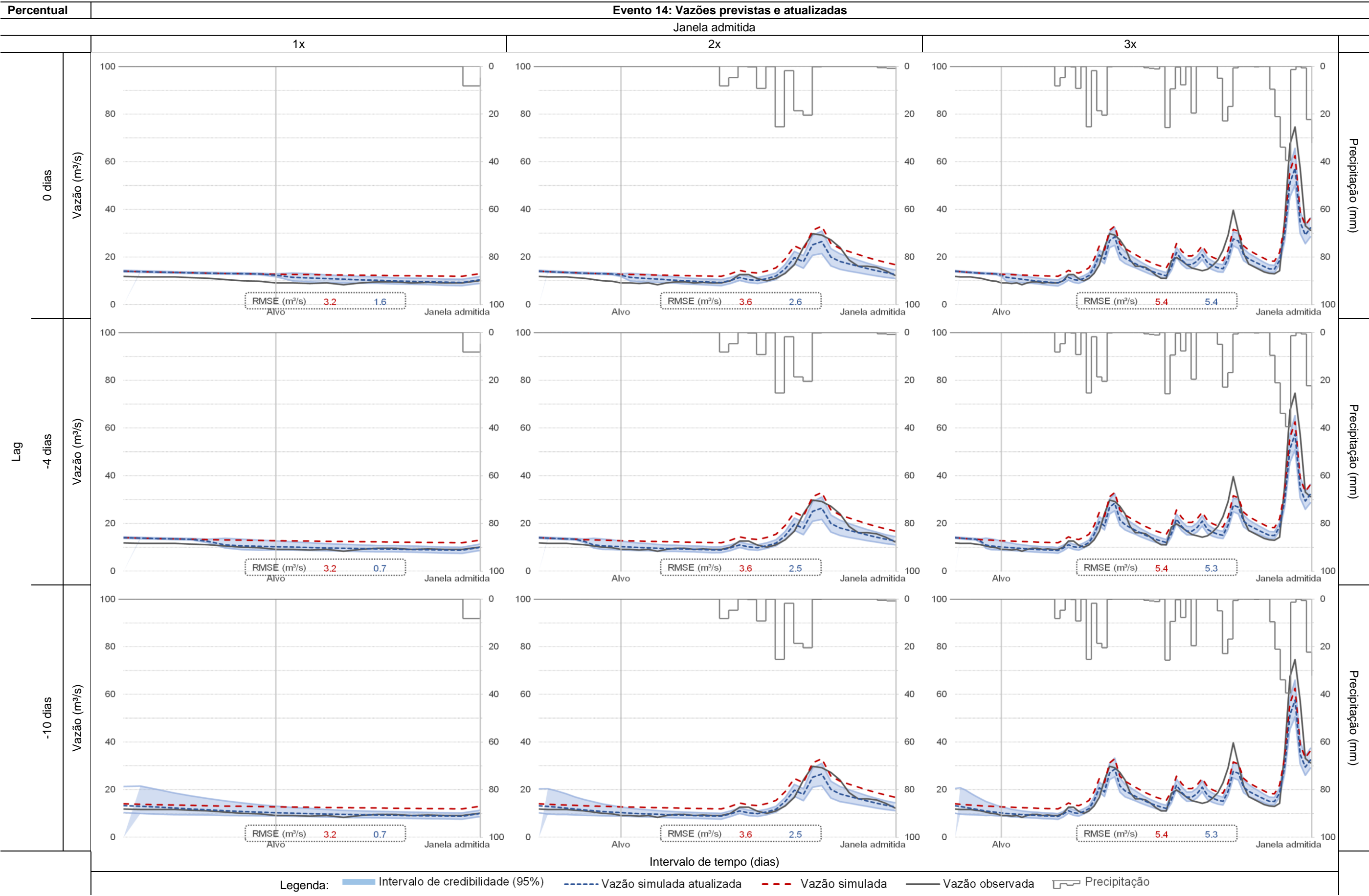


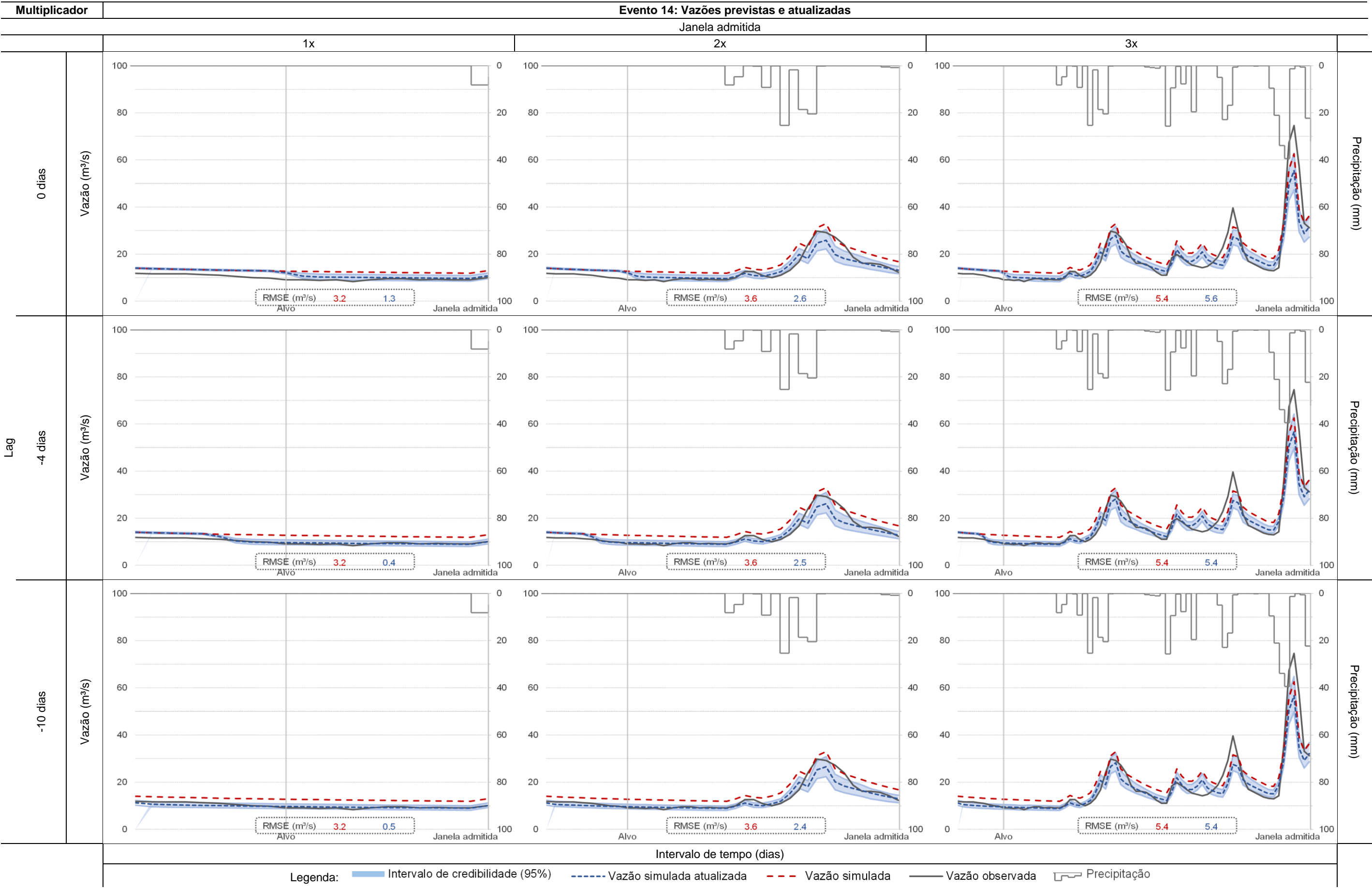
Janela admitida

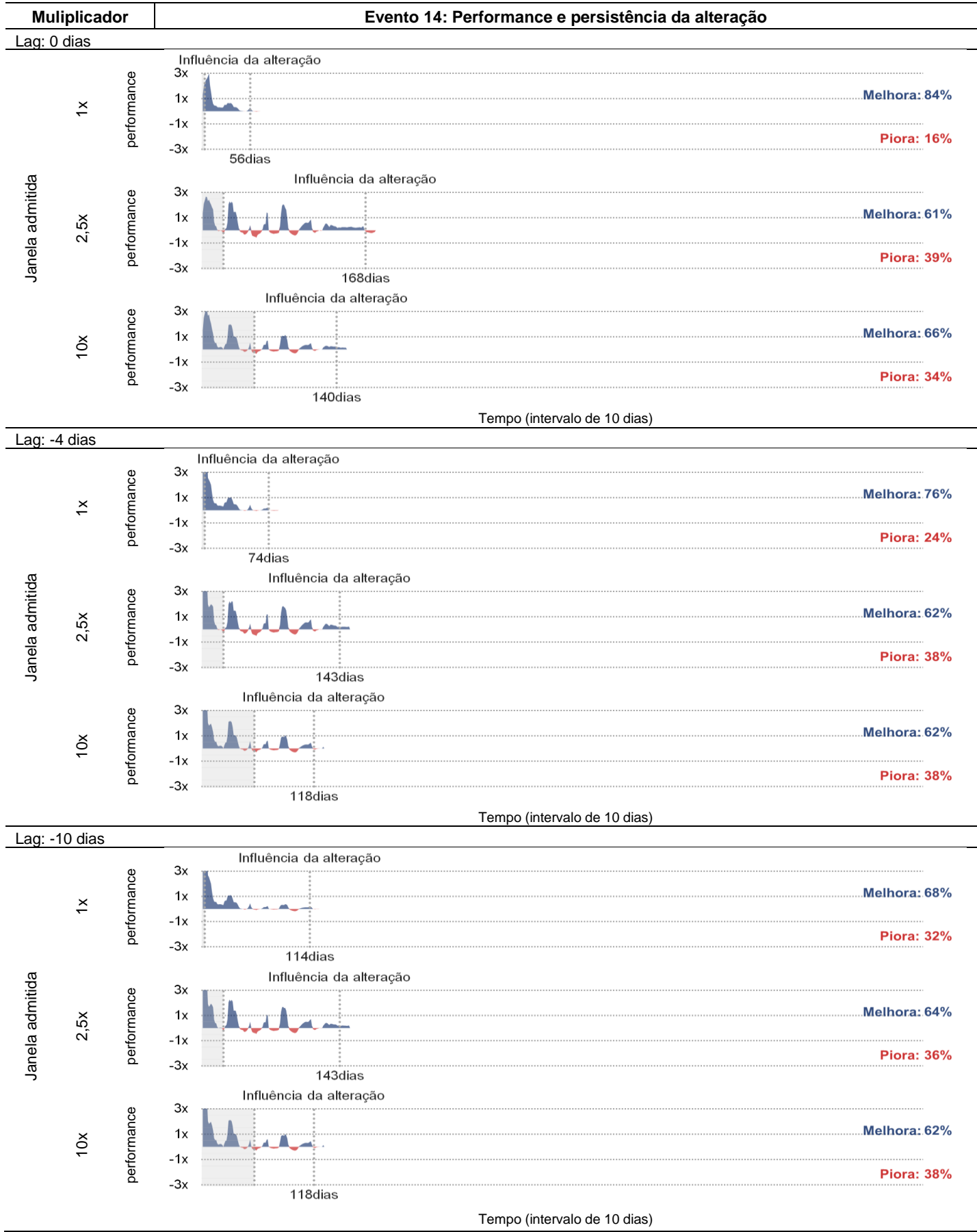
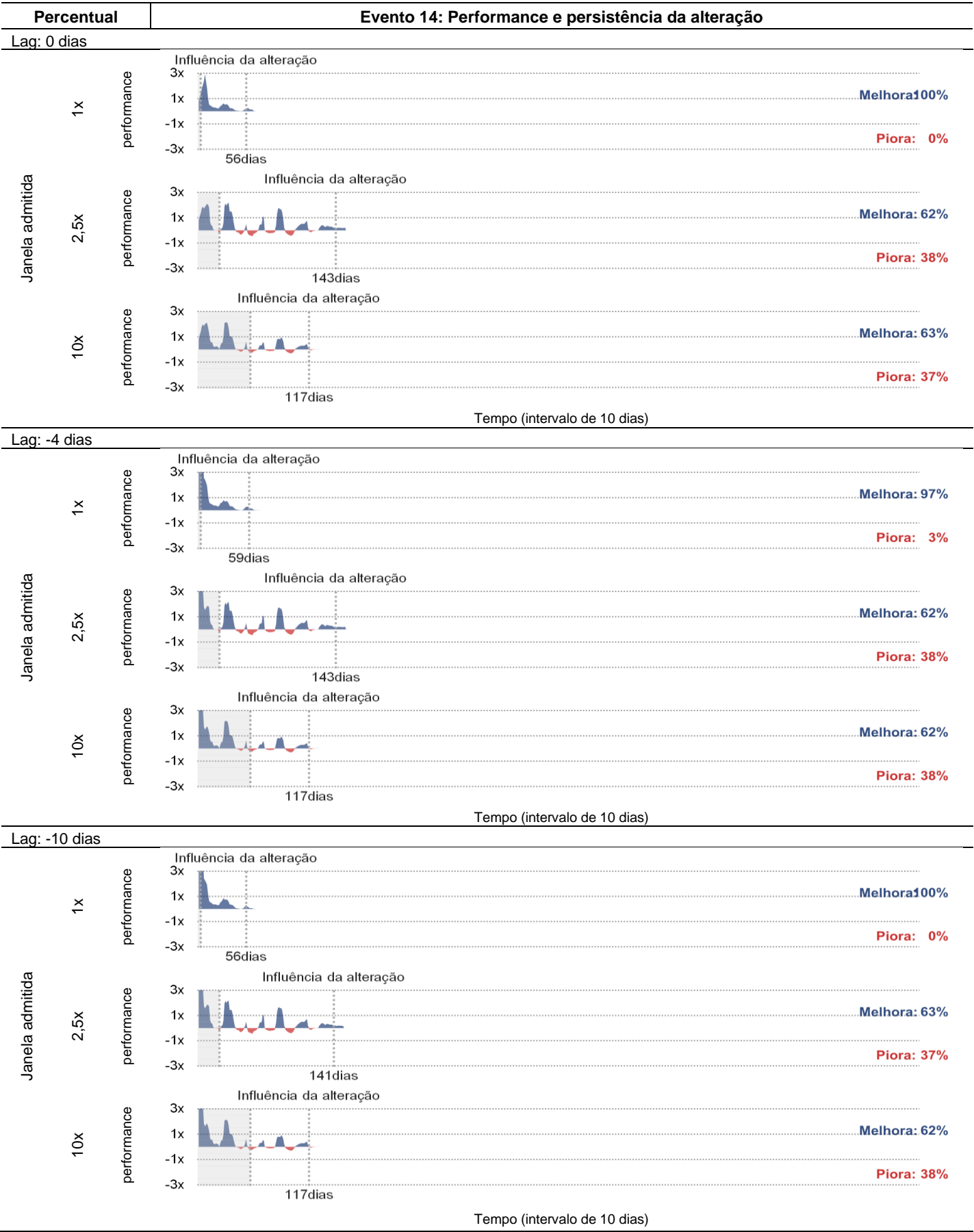






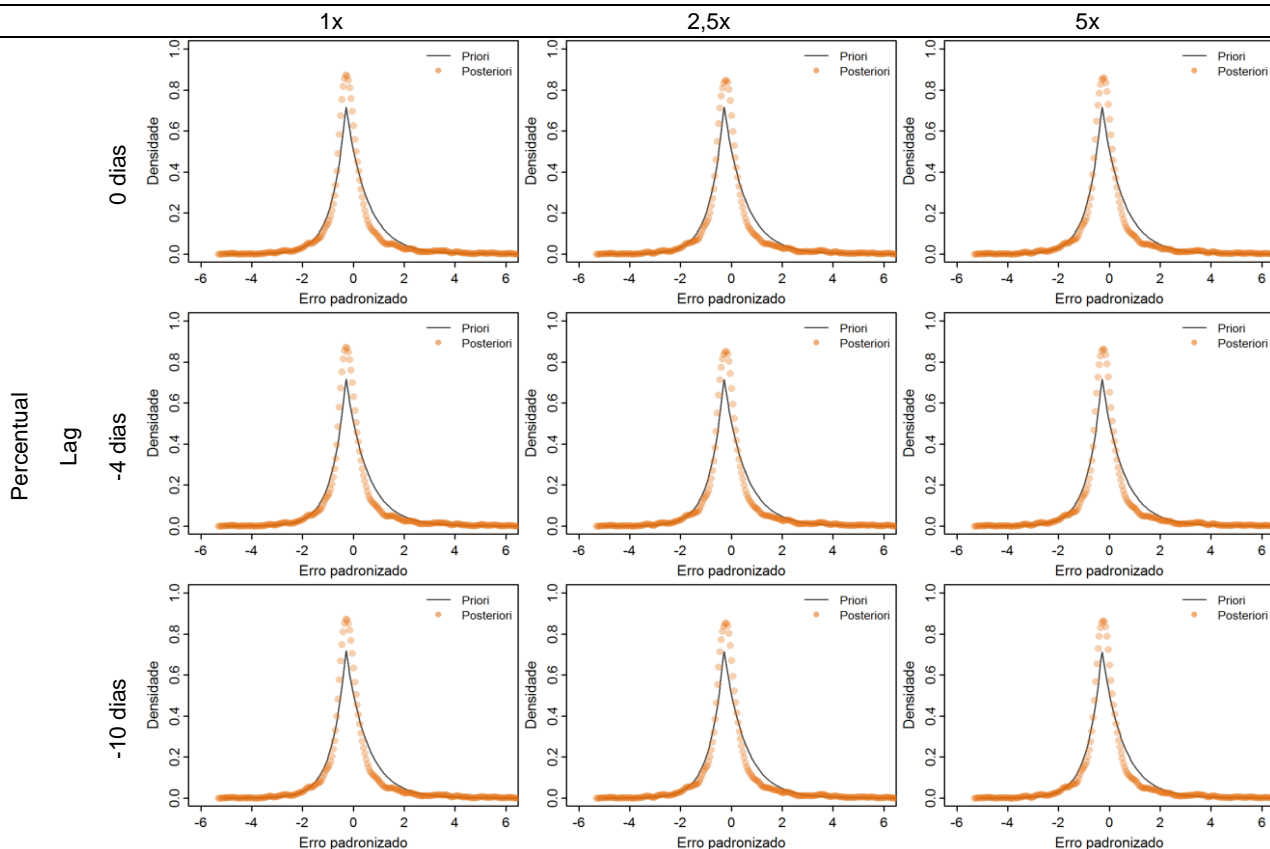




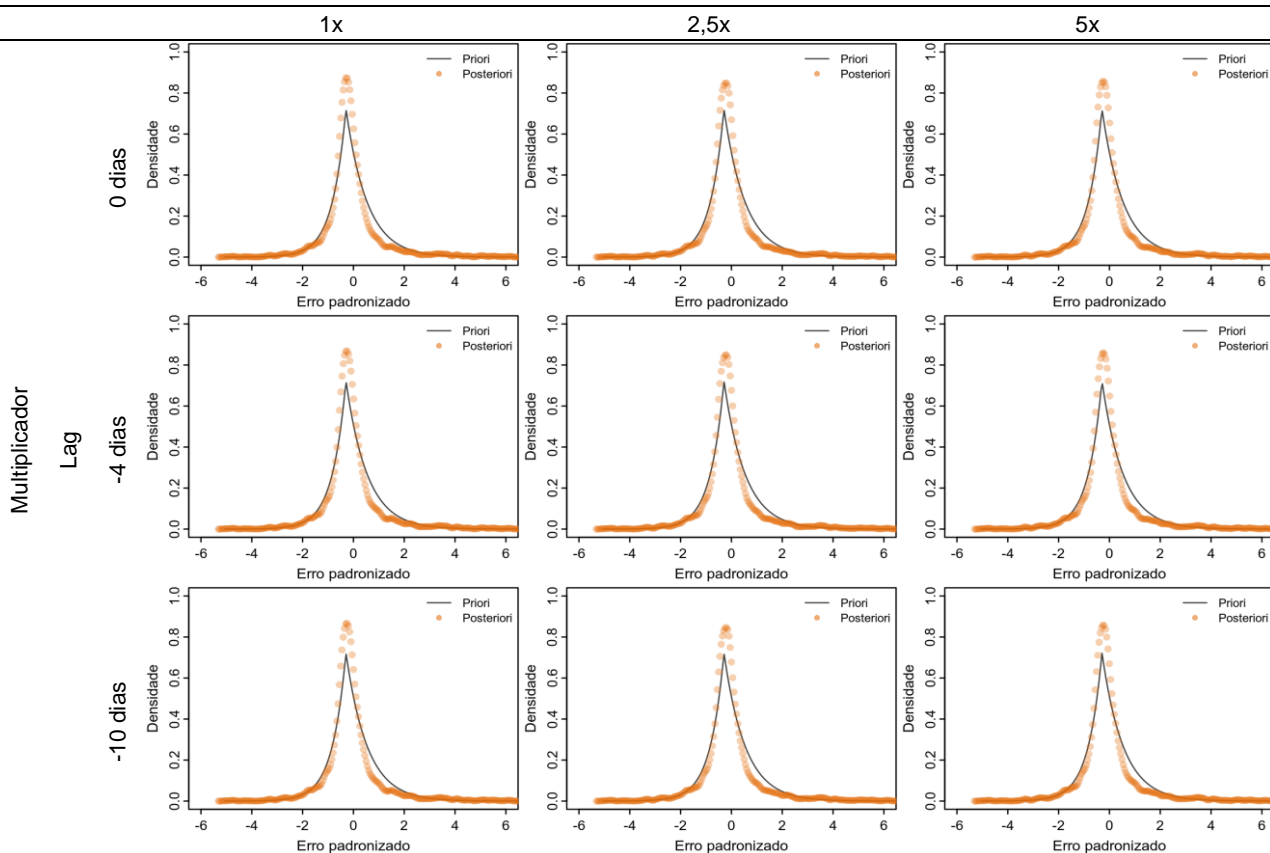


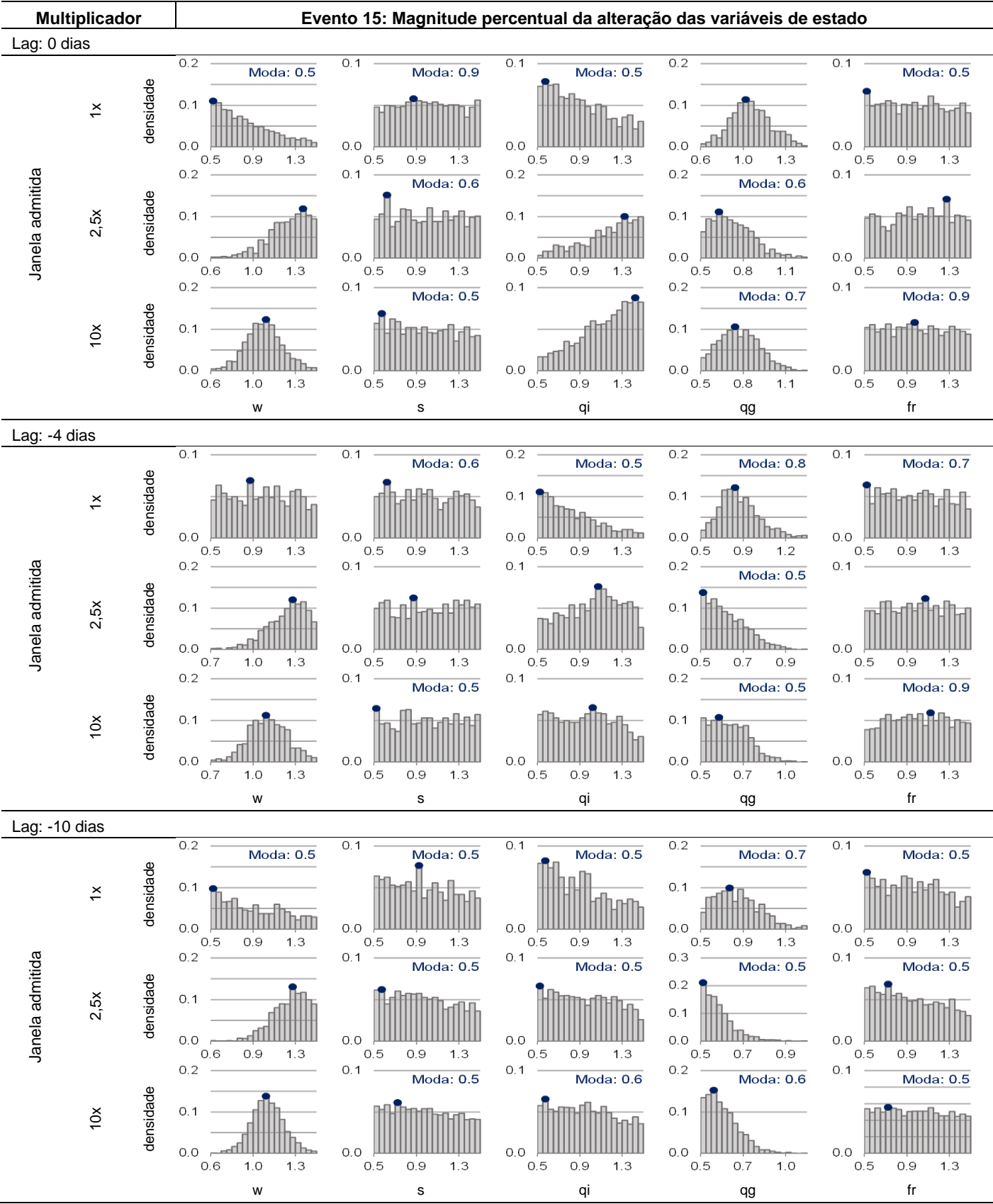
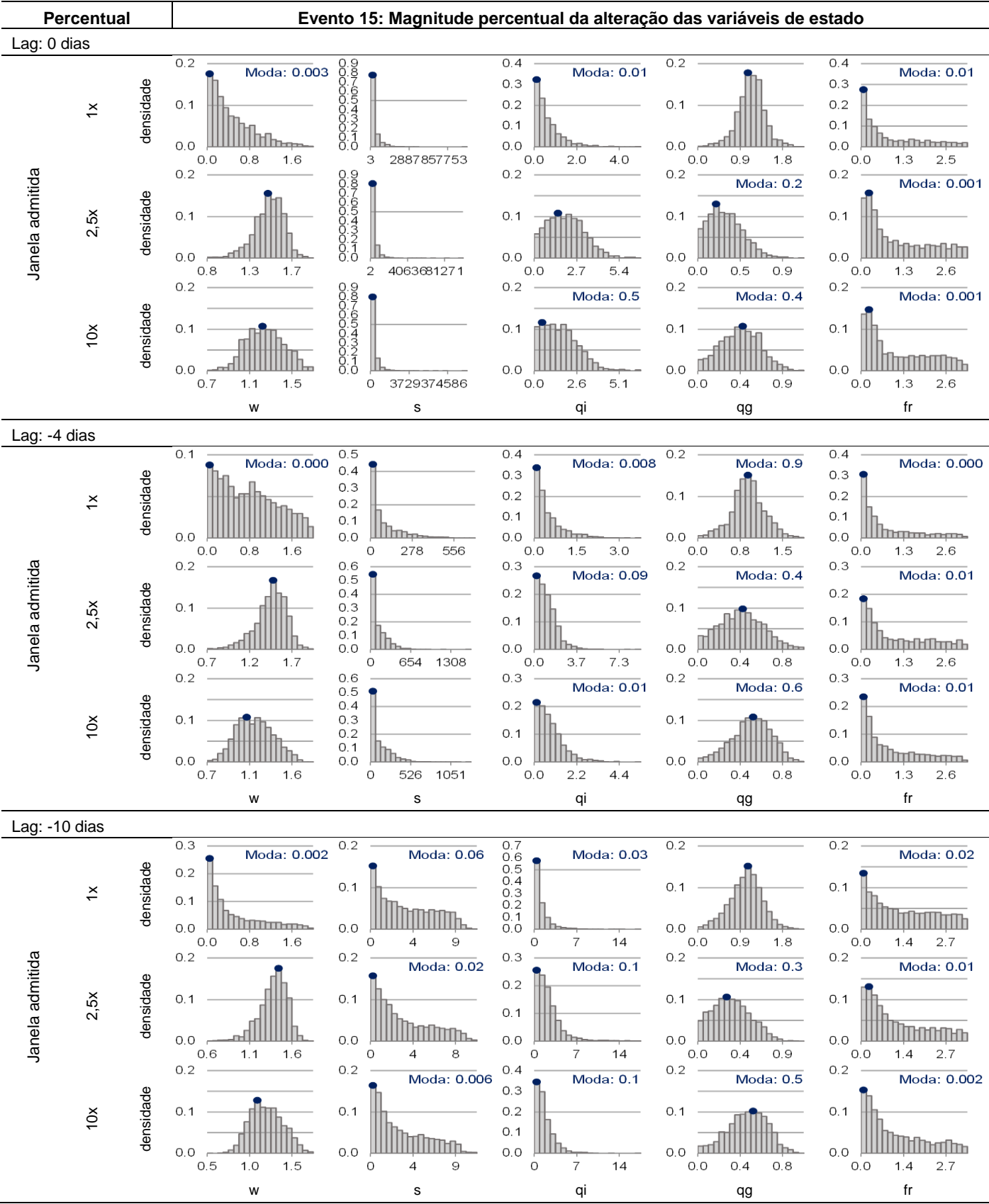
## Evento 14: Comportamento probabilístico dos resíduos da simulação

Janela admitida

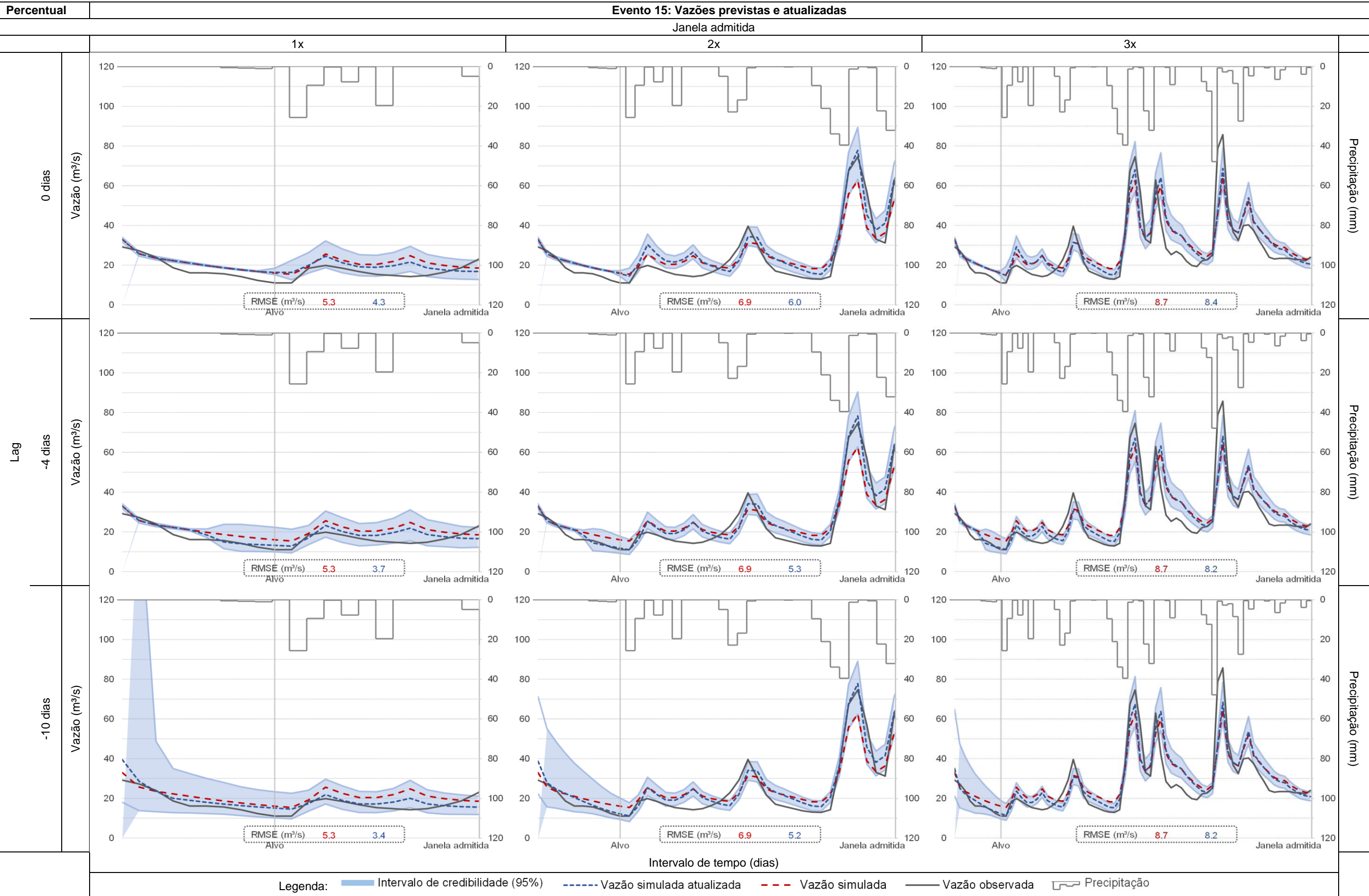


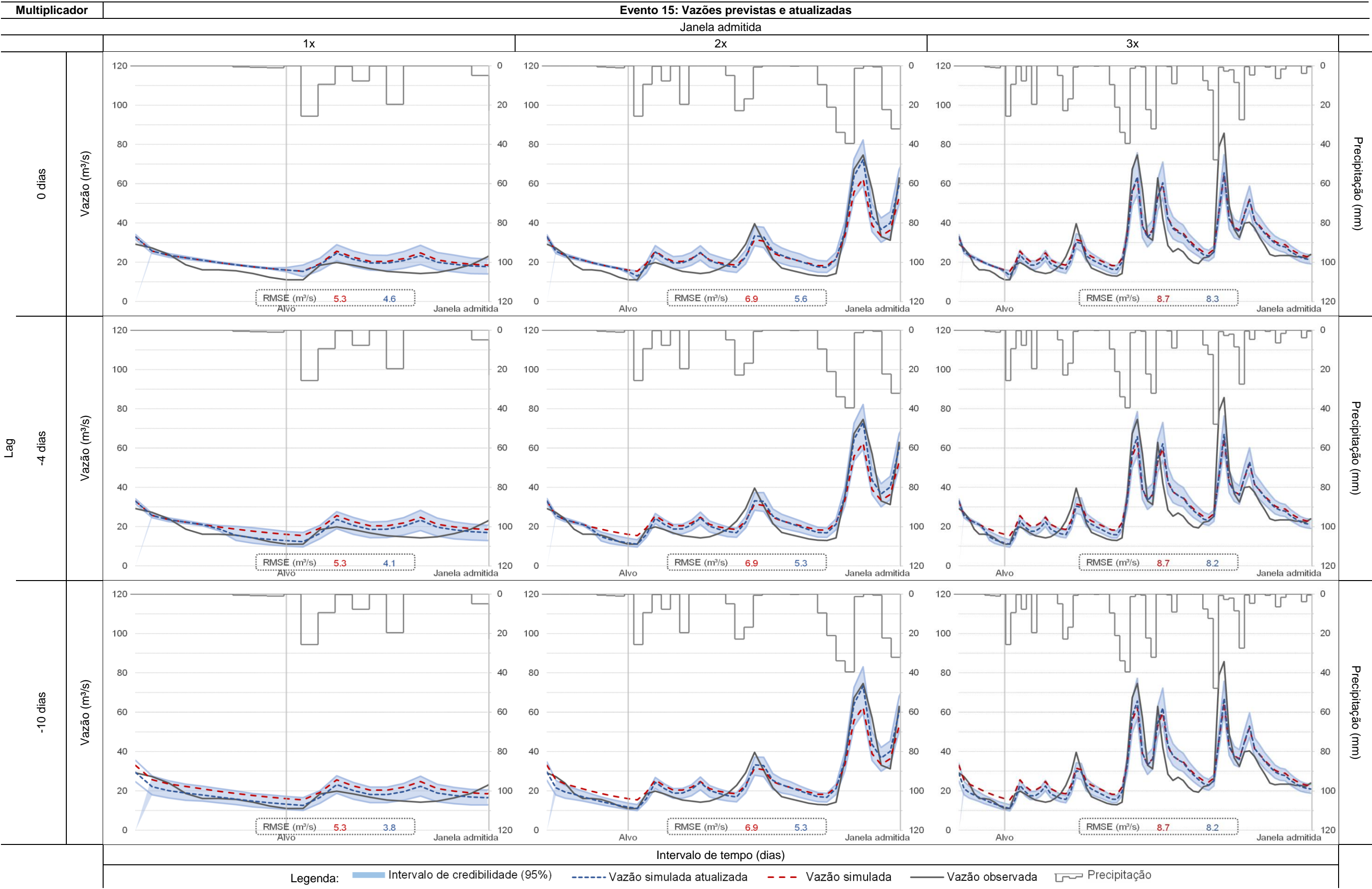
Janela admitida

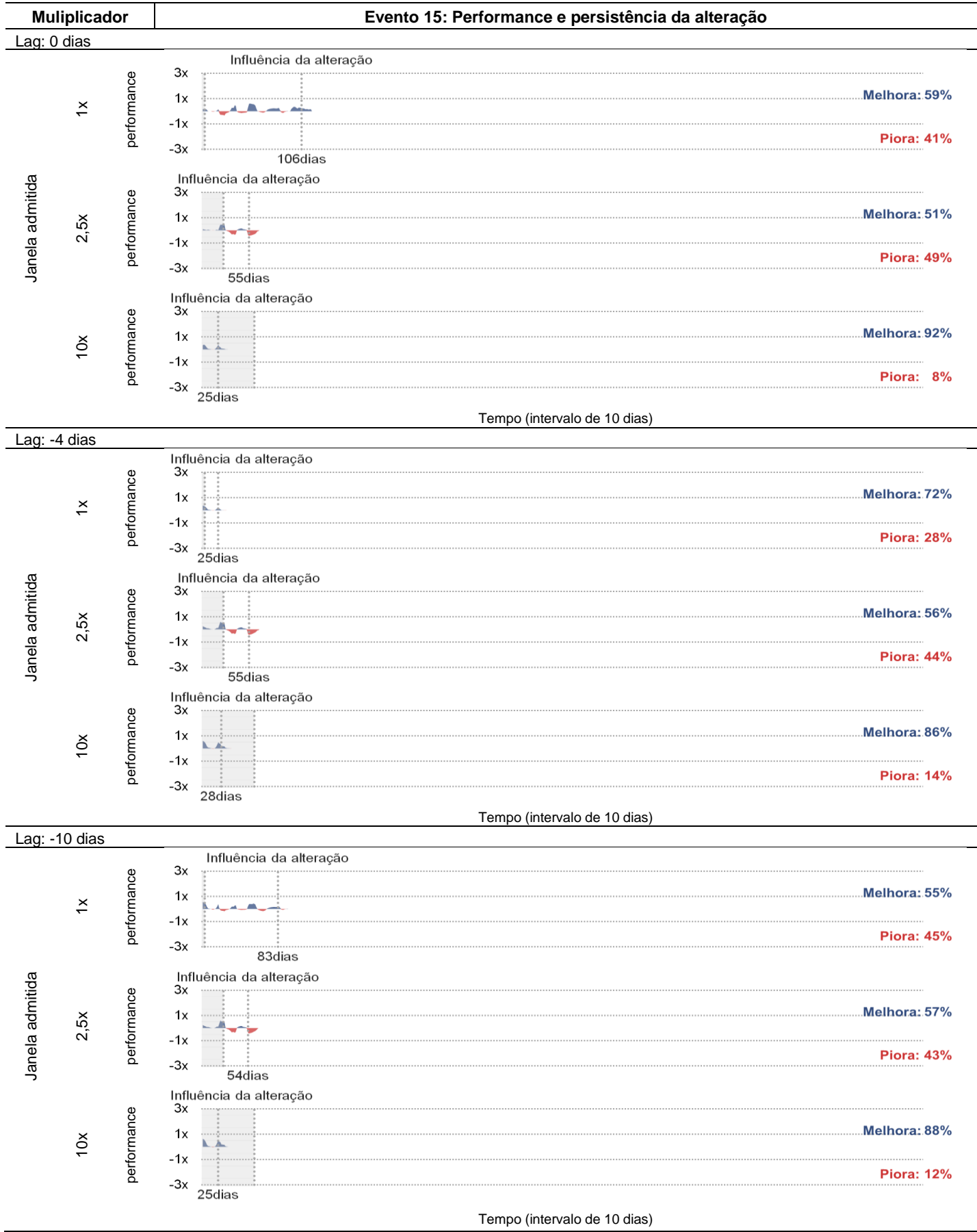
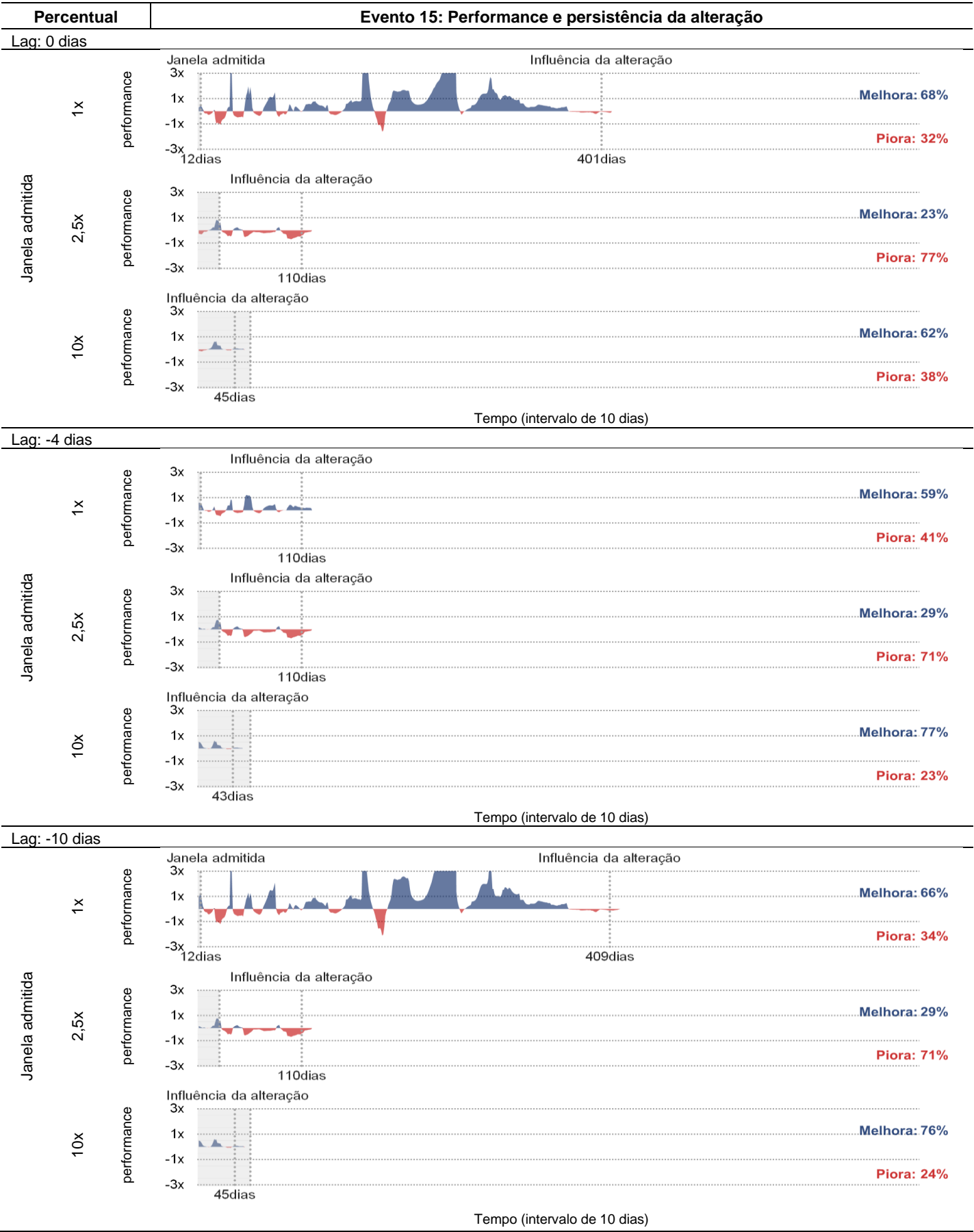








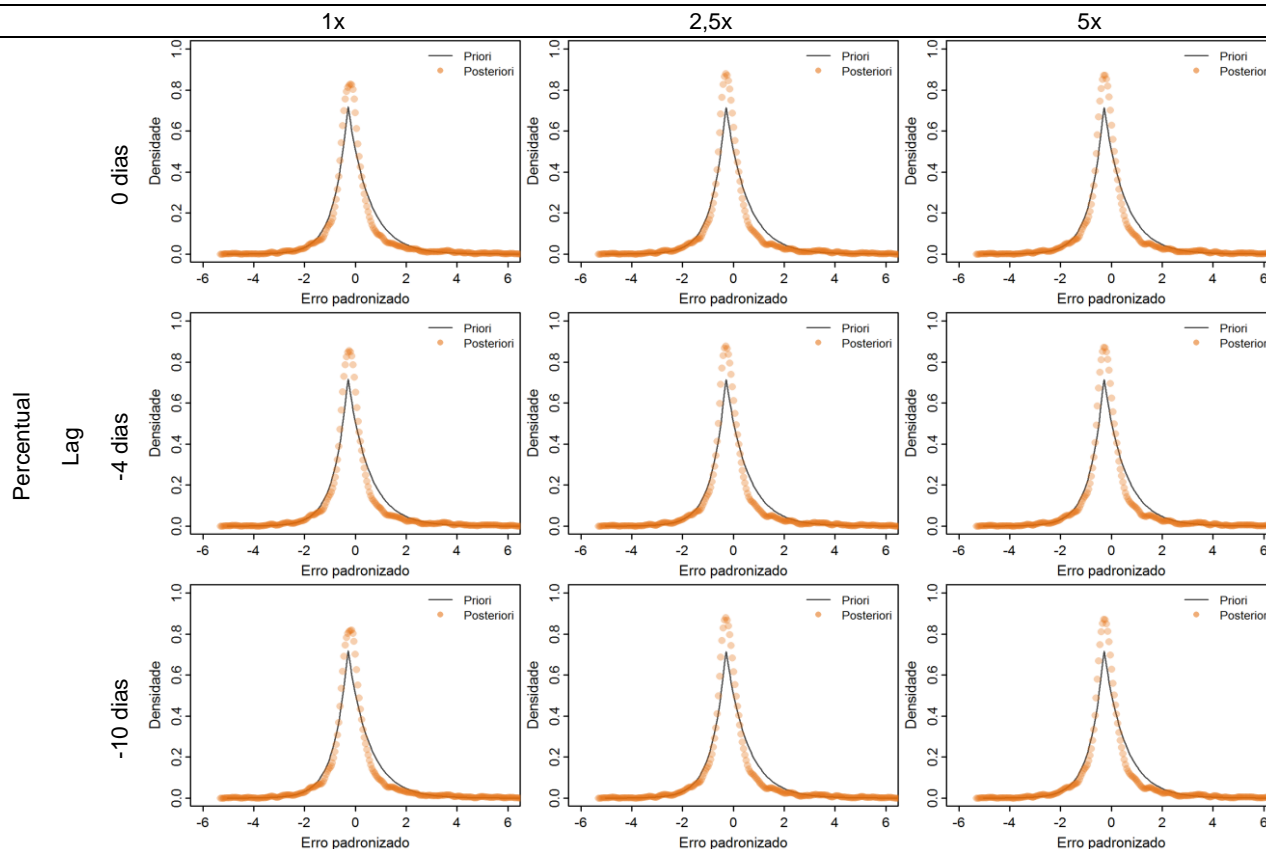






## Evento 15: Comportamento probabilístico dos resíduos da simulação

Janela admitida



Janela admitida

