

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
BACHARELADO EM ESTATÍSTICA

DANIELLY SANTOS SEVERINO

**PREVISÃO DE SÉRIES EPIDEMIOLÓGICAS
INCORPORANDO ATRASO NA NOTIFICAÇÃO**

BELO HORIZONTE
2023

DANIELLY SANTOS SEVERINO

**PREVISÃO DE SÉRIES EPIDEMIOLÓGICAS
INCORPORANDO ATRASO NA NOTIFICAÇÃO**

Trabalho final de conclusão de curso apresentado ao Departamento de Estatística da Universidade Federal de Minas Gerais (UFMG), como parte dos requisitos necessários à obtenção do título de bacharel em Estatística.

Orientador: Dani Gamerman

Coorientadora: Izabel Nolau

BELO HORIZONTE

2023

Agradecimentos

Algumas épocas da vida a gente trabalha e estuda tanto que o tempo com a família é raro, quando estamos em casa só queremos aproveitar. O engraçado e mais contraditório é que é tudo por eles. Não reclamo porque eu vivo um sonho. Todos os dias posso acessar minha fonte de felicidade, quero ser forte como a Tia Lú (Vera Lúcia) em tempos difíceis.

Agradeço por poder olhar nos olhos do meu irmão Lucas e entender melhor a vida, aprender a me dedicar mais. Sou grata pela minha família e amigos que incentivam minha jornada. Me sinto mais livre sempre que me aproximo mais das minhas raízes. Deus, obrigada por ter me tornado a continuação de um sonho do meu pai e da minha mãe, de todos que vieram antes de mim.

Aos meu orientador Dani Gamerman e minha coorientadora Izabel Nolau, agradeço pelos esforços, dedicação, inspiração, apoio e paciência em me auxiliar a concretizar esse trabalho. Agradeço também aos professores que contribuíram para a minha formação, especialmente Cristina Eterovik e Helinton A. Lopes. Por fim, agradeço a Fundação Mendes Pimental (FUMP) por me proporcionar condições financeiras para cursar o início da graduação.

Resumo

A gestão do risco de doenças infecciosas envolve a notificação e coleta de dados para monitorar a disseminação e identificar tendências das doenças. No entanto, os casos notificados muitas vezes sofrem atraso na notificação, ou seja, casos que são notificados após a incidência da doença. Os atrasos distorcem a relação entre a incidência relatada e a verdadeira incidência da doença. Isso pode ter consequências significativas para a saúde pública e para a eficácia de medidas de medidas de controle. Sem informações adequadas da gravidade da situação atual em relação à doença, as autoridades têm dificuldade em agir de maneira eficaz para identificar e responder adequadamente aos surtos. Sendo assim, é crucial corrigir a informação disponível o mais rápido possível para auxiliar a tomada de decisões como emissões de aviso às autoridades locais e ao público-alvo.

Os modelos hierárquicos Bayesianos são utilizados para lidar com a incerteza em vários níveis dos dados, permitindo combinar informações de diferentes níveis para obter estimativas e previsões mais precisas e robustas. Ao lidar com doenças infecciosas é importante que os modelos sejam capazes de gerar estimativas que recuperem a informação existente com precisão e, especialmente, construir previsões robustas. As previsões podem ser feitas em tempo real (*nowcasting*) com base nos dados disponíveis ou para períodos futuros (*forecasting*) com base no histórico.

Neste trabalho, foram propostos modelos hierárquicos Bayesianos com duas estruturas diferentes para lidar com dados de contagem de incidência de doenças infecciosas e corrigir o atraso nas notificações. Esses modelos são uma forma flexível de corrigir os atrasos nas notificações e permitem gerar previsões e quantificar a incerteza associada. Os modelos propostos foram ilustrados usando dados semanais de incidência de dengue no Rio de Janeiro registrados entre janeiro de 2011 e abril de 2012. Além disso, a técnica de Monte Carlo via Cadeias de Markov (MCMC) foi utilizada para a extração de amostras para aproximar as densidades a posteriori das quantidades de interesse.

Palavras-chaves: Estatística; inferência Bayesiana; séries temporais; dados de contagem; atraso na notificação; *Nowcasting*; *Forecasting*; dengue.

Sumário

Introdução	1
Estrutura típica dos dados com atraso na notificação	3
Metodologia	4
Inferência Bayesiana	4
Estimação e previsão	5
Detalhamento dos modelos propostos.....	6
Operacionalização computacional de modelos Bayesianos.....	10
Comparação de modelos	12
Resultados	14
Descrição das contagens de casos de incidência de dengue	14
Resultados do modelo com estrutura de atraso na notificação independente..	19
Resultados do modelo com estrutura conjunta de atraso na notificação (MC) .	24
Comparação dos resultados dos modelos	30
Conclusões	34
Anexo.....	36
Referências.....	44

Lista de Tabelas

Tabela 1 - Estrutura típica dos dados com atraso na notificação	3
Tabela 2 - Contagem de casos de incidência de dengue notificados no Rio de Janeiro.....	15
Tabela 3 - Estimativas e previsões das médias das distribuições a posteriori – MI.....	19
Tabela 4 - Intervalos de previsão dos casos de dengue com 95% de credibilidade – MI	21
Tabela 5 - Estimativas e previsões das médias das distribuições a posteriori – MC.....	24
Tabela 6 - Intervalos de previsão dos casos de dengue com 95% de credibilidade – MC	26
Tabela 7 - MAE e RMSE por modelo.....	33
Tabela 8 - Intervalos de credibilidade do MI	36
Tabela 9 - Intervalos de credibilidade do MC.....	37

Listas de Figuras

Figura 1 - Total de casos de incidência de dengue ao longo do tempo.....	16
Figura 2 - Distorção entre a incidência de dengue relatada e a verdadeira incidência de dengue	16
Figura 3 - Total de casos de incidência de dengue ao longo do tempo por período de atualização.....	18
Figura 4 - Influência do atraso nas notificações ao longo do tempo.....	18
Figura 5 - Estimativas e previsões para o total de casos de dengue – MI.....	22
Figura 6 - Distribuições a posteriori da média de casos de dengue ao longo do tempo por defasagem – MI	23
Figura 7 - Influência do atraso nas notificações ao longo do tempo – MI	23
Figura 8 - Estimativas e previsões do total de casos de dengue – MC	27
Figura 9 - Estimativas e previsões para os casos de dengue por semana de atraso.....	28
Figura 10 - Médias das distribuições a posteriori para cada semana de atraso por modelo.....	30
Figura 11 - Média da distribuição a posteriori do total de casos de dengue por modelo.....	32
Figura 12 - Traços das cadeias de Markov para os hiper parâmetros do MI.....	38
Figura 13 - Traços das cadeias de Markov para os hiper parâmetros do MC	43

Introdução

O sistema de vigilância de doenças infecciosas possui algumas abordagens comuns que muitos países adotam (Steven and Elliott, 2000). Na maioria dos países a notificação é obrigatória, exigindo que a incidência de doenças infecciosas como tuberculose, hepatite, HIV/AIDS e sarampo seja informada às autoridades de saúde. Os profissionais de saúde, hospitais e laboratórios são responsáveis por relatar os casos suspeitos ou confirmados. Portanto, existe um sistema de notificação e coleta de dados que se caracteriza por estabelecer informações essenciais para monitorar a disseminação e identificar tendências das doenças. Cabe destacar que os sistemas de vigilância epidemiológica ajudam a identificar surtos, avaliar a eficácia das medidas de controle e orientar políticas públicas de saúde. Além disso, a pandemia de COVID-19 trouxe à tona a importância da vigilância de doenças e a necessidade de fortalecer os sistemas de saúde em todo o mundo.

Os casos de doenças infecciosas que são relatados aos profissionais de saúde podem ser analisados através de contagens, sendo que, o número de casos comumente sofre atraso na notificação, ou seja, casos que são relatados após a incidência da doença. Existem diversas situações que podem levar ao atraso na notificação de doenças como tempo até a confirmação do diagnóstico, desafios logísticos, confirmação de exames e barreiras culturais. Dessa forma, os casos notificados distorcem a relação entre a incidência da doença relatada e a verdadeira incidência da doença.

O atraso na notificação de doenças infecciosas pode ter consequências significativas para a saúde pública e para a eficácia de medidas de controle. É extremamente importante corrigir o atraso nas notificações para diminuir o risco de propagação da doença, não retardar o tratamento e cuidados adequados e nem dificultar a detecção de tendências e padrões. Sem informações adequadas da gravidade da situação atual em relação à doença, as autoridades têm dificuldade em agir de forma eficaz para identificar e responder rapidamente aos surtos, permitindo que a infecção entre a população aumente.

Para diversos casos de doenças é comum que os dados sejam potencialmente subnotificados, ou seja, casos de doenças que não foram detectados ou que foram detectados, mas não notificados. Isso resulta em uma representação imprecisa da verdadeira contagem de doenças. Corrigir a subnotificação de doenças é um desafio complexo que requer uma abordagem abrangente e a implementação de várias estratégias, exigindo fontes adicionais de informação como conhecimento prévio da taxa de subnotificação.

O foco neste trabalho é corrigir o atraso na notificação dos dados. Para isso, foram propostos modelos hierárquicos Bayesianos para dados de contagem de incidência das doenças infecciosas como uma forma flexível de corrigir o atraso nas notificações e quantificar a incerteza associada. A modelagem proposta assume que os dados de contagem seguem uma distribuição de probabilidade Poisson incorporando componentes do efeito do tempo e do efeito das defasagens.

Estrutura típica dos dados com atraso na notificação

Ao lidar com séries epidemiológicas, existe uma estrutura típica que é utilizada para lidar com dados de contagem da incidência de doenças com problema de atraso na notificação. Através dessa estrutura é possível capturar e analisar adequadamente as informações disponíveis (Bastos et al, 2019).

A Tabela 1 exibe a estrutura típica de uma base de dados qualquer para contagens com atraso na notificação. As linhas da tabela correspondem ao tempo, enquanto as colunas indicam o período de atraso na notificação. T é o tempo atual, D é o atraso máximo relevante e H é o horizonte máximo de previsão. Os eventos ocorridos no tempo t com d unidades de atraso são denotados por $n_{t,d}$ e, fixado qualquer intervalo de tempo (linha), a quantidade total de eventos é $N_t = \sum_{d=0}^D n_{t,d}$.

Os eventos ocorridos entre 1 e $T - D$ foram integralmente registrados. Já os eventos ocorridos entre $T - D + 1$ e T são parcialmente registrados, pois existem eventos que ocorreram e não foram registrados devido ao atraso na notificação. Por fim, as contagens de $T + 1$ até $T + H$ são os eventos futuros. Dessa forma, os valores em cinza claro na tabela são as contagens que já ocorreram, mas ainda não foram registradas e os valores em cinza escuro são as contagens futuras.

Tabela 1 - Estrutura típica dos dados com atraso na notificação

$t \setminus d$	0	1	...	$D-1$	D	N
1	$n_{1,0}$	$n_{1,1}$		$n_{1,D-1}$	$n_{1,D}$	N_1
2	$n_{2,0}$	$n_{2,1}$		$n_{2,D-1}$	$n_{2,D}$	N_2
...						
$T - D$	$n_{T-D,0}$	$n_{T-D,1}$		$n_{T-D,D-1}$	$n_{T-D,D}$	N_{T-D}
$T - D + 1$	$n_{T-D+1,0}$	$n_{T-D+1,1}$		$n_{T-D+1,D-1}$	$n_{T-D+1,D}$	N_{T-D+1}
...						
$T - 1$	$n_{T-1,0}$	$n_{T-1,1}$		$n_{T-1,D-1}$	$n_{T-1,D}$	N_{T-1}
T	$n_{T,0}$	$n_{T,1}$		$n_{T,D-1}$	$n_{T,D}$	N_T
$T + 1$	$n_{T+1,0}$	$n_{T+1,1}$		$n_{T+1,D-1}$	$n_{T+1,D}$	N_{T+1}
...						
$T + H$	$n_{T+H,0}$	$n_{T+H,1}$		$n_{T+H,D-1}$	$n_{T+H,D}$	N_{T+H}

Metodologia

Inferência Bayesiana

A inferência Bayesiana (Gelman et al, 2013) é uma abordagem estatística utilizada para fazer inferência sobre parâmetros desconhecidos ou eventos futuros com base em evidências observadas e conhecimento prévio. A inferência Bayesiana parte do princípio de que as probabilidades são representações subjetivas de incerteza. Ela permite a inclusão de informações subjetivas e conhecimento especializado sobre o problema analisado através da especificação de uma distribuição a priori. A inferência Bayesiana envolve três elementos fundamentais para obter estimativas e realizar previsões: uma distribuição/densidade a priori; uma função de verossimilhança; e uma distribuição/densidade a posteriori.

Uma função de probabilidade ou função de densidade de probabilidade a priori $p(\theta)$, é uma expressão probabilística que representa o conhecimento prévio ou falta de conhecimento sobre o problema. A função de verossimilhança $p(y|\theta)$ é uma medida estatística que representa a relação entre os dados observados com os diferentes valores possíveis dos parâmetros desconhecidos. Por sua vez, a distribuição ou densidade a posteriori $p(\theta|y)$ é uma distribuição/densidade de probabilidade atualizada após a observação dos dados, obtida através da combinação da distribuição/densidade a priori e função de verossimilhança usando o teorema de Bayes:

$$p(\theta|y) = \frac{p(y|\theta) p(\theta)}{p(y)}.$$

A inferência Bayesiana é particularmente útil quando existem informações prévias relevantes ou quando a incerteza precisa ser modelada e atualizada à medida que novos dados são disponibilizados. Além disso, essa abordagem é aplicada em diversas áreas como ciências da saúde, ciências sociais, engenharia, economia e aprendizado de máquina.

Estimação e previsão

Sob o contexto da inferência Bayesiana, a estimação pontual é o processo de obter um único valor para estimar o vetor de parâmetros desconhecidos a partir da distribuição ou função de densidade de probabilidade a posteriori. Há diferentes métricas utilizadas para sintetizar a distribuição/densidade a posteriori e obter uma estimativa pontual como calcular o valor esperado (média), moda e mediana.

A estimação intervalar é utilizada para quantificar a incerteza em relação ao(s) parâmetro(s) desconhecido(s) com a construção de intervalos de credibilidade. Os intervalos de credibilidade fornecem uma medida da probabilidade de o parâmetro estar dentro do intervalo, sendo que, o intervalo resultante representa uma faixa de valores mais prováveis para o vetor paramétrico desconhecido considerando as informações disponíveis nos dados observados e conhecimento prévio do problema.

O intervalo HPD (*Highest Posterior Density*) se caracteriza por ser o intervalo de credibilidade mais estreito que contém uma proporção específica baseado na distribuição a posteriori. Esse intervalo (Ming-Hui and Qui-Man, 1988) não depende de pressupostos assintóticos para a sua construção. Além do mais, o intervalo pode ser assimétrico, acomodando funções de probabilidade e funções de densidade de probabilidade a posteriori assimétricas. Em rigor, o procedimento para calcular o HPD envolve obter a aproximação da distribuição/densidade a posteriori, ordenar os valores, calcular a distribuição de probabilidade acumulada, e determinar a faixa mais estreita que contém a proporção desejada da aproximação da distribuição/densidade a posteriori.

No cenário epidemiológico, a previsão de casos de incidência de doenças infecciosas é importante por diversas razões, dentre elas, o planejamento para alocação de recursos e a adoção de medidas de controle. Cabe destacar que a previsão na inferência Bayesiana é flexível pois permite incorporar informações a priori, atualizar o conhecimento a partir dos dados observados e fornecer estimativas probabilísticas completas. Além disso, ela é amplamente aplicada em diversas áreas para previsões do tempo, previsões de eventos raros e previsão de séries temporais.

Dentro do paradigma Bayesiano, a previsão permite gerar estimativas pontuais e intervalares sobre eventos futuros com base nas informações disponíveis da distribuição/densidade a posteriori. Para realizar previsões deve-se gerar amostragens da distribuição ou densidade a posteriori para obter uma amostra de possíveis valores dos parâmetros desconhecidos. A partir da amostra dos valores dos parâmetros aplica-se o modelo probabilístico para gerar valores preditos com base nos parâmetros amostrados. Depois disso, são feitas as análises dos resultados das previsões.

O fato de séries epidemiológicas apresentarem atrasos na notificação torna necessário a utilização de duas abordagens de previsão: *Nowcasting* e *Forecasting*. O *Nowcasting* abrange as previsões em tempo (quase) real, cujo foco principal é obter estimativas precisas e atualizadas para o momento atual baseadas nos dados e informações disponíveis. Esse procedimento é particularmente útil para o problema, uma vez que, permite o monitoramento em tempo real das doenças. Além disso, as informações em tempo real são fundamentais para a tomada de decisões imediatas. Por sua vez, o *Forecasting* refere-se as previsões para os períodos futuros a partir do histórico disponível. Logo, trata-se de uma tentativa de prever o que pode acontecer no futuro considerando padrões e tendências observadas no passado.

Detalhamento dos modelos propostos

Sob o paradigma Bayesiano, os modelos hierárquicos são uma abordagem utilizada para lidar com a incerteza e a variabilidade em vários níveis dos dados. Esses modelos podem ser especialmente úteis quando há uma estrutura hierárquica ou agrupamento dos dados, permitindo que as informações de cada nível sejam compartilhadas e combinadas para obter estimativas mais precisas dos parâmetros desconhecidos e quantidades de interesse.

Em um modelo hierárquico, os dados são organizados em vários níveis, sendo que cada nível do modelo possui seus próprios parâmetros e a incerteza sobre esses parâmetros é modelada usando distribuições/densidades a priori. A principal ideia de um modelo hierárquico é que as distribuições ou densidades a priori de cada nível são influenciadas pelas informações dos níveis superiores.

Os modelos hierárquicos podem fornecer estimativas mais precisas e robustas, principalmente quando os dados em cada nível são limitados ou possuem ruídos. Além do mais, a estrutura hierárquica possibilita que os níveis mais baixos sejam regularizados pelos dados nos níveis superiores. Portanto, modelos hierárquicos aderem uma modelagem mais realista em relação à incerteza e variabilidade em dados complexos.

Neste trabalho foram propostos modelos hierárquicos Bayesianos com duas estruturas diferentes para lidar com dados de contagem de incidência de doenças incorporando atrasos nas notificações. Ambos os modelos permitem mudanças suaves na variação temporal do número total de eventos e no mecanismo de atraso. Cabe destacar que as contagens de incidência de doenças podem ser modeladas através de múltiplas distribuições de probabilidade discreta, como as distribuições Poisson e Binomial Negativa.

Existe uma relação de dependência entre as contagens de incidência de doenças infecciosas com atraso na notificação, isso ocorre porque $N_t = \sum_{d=0}^D n_{t,d}$, ou seja, o número total de eventos ocorridos no tempo t é exatamente a soma dos eventos ocorridos no tempo t com d unidades de atraso para todas as defasagens. Portanto, não é possível especificar um modelo para N_t e $n_{t,d}$ acatando todos os atrasos ao mesmo tempo e respeitando a disposição das contagens. Acatando a disposição típica dos dados de contagens com atraso na notificação, os modelos podem ser especificados para todos os eventos $n_{t,d}$ com $t = 1, \dots, T$ e $d = 0, 1, \dots, D$, ou podem ser especificados para N_t e $n_{t,d}$ com $t = 1, \dots, T$ e $d = 1, \dots, D$. Salienta-se que $n_{t,0}$ são os eventos notificados sem nenhum atraso e, por isso, pode ser mais interessante especificar modelos para N_t e $n_{t,d}$ com $t = 1, \dots, T$ e $d = 1, \dots, D$.

Em ambos os modelos propostos se assume que as contagens de incidência de doenças infecciosas seguem uma distribuição de probabilidade Poisson com médias que correspondem a uma expressão que representa o comportamento típico de epidemias, sendo conhecida como curva logística generalizada (Gamerman et al., 2022). Além disso os modelos apresentam diferentes especificações para N_t e $n_{t,d}$ com $t = 1, \dots, T$ e $d = 1, \dots, D$.

Um dos modelos propostos incorpora uma estrutura de atraso na notificação independente e o outro abrange uma estrutura conjunta de atraso na notificação. Esses modelos apresentam as descrições apresentadas abaixo.

Modelo com estrutura de atraso na notificação independente (MI)

As contagens de incidências de doenças N_t e $n_{t,d}$ seguem uma distribuição de probabilidade Poisson, sendo essa uma distribuição base para dados de contagem, com médias θ_t e $\lambda_{t,d}$ respectivamente. Cabe ressaltar que as médias são curvas logísticas generalizadas (Gamerman et al., 2022) que representam o comportamento típico de doenças infecciosas. As especificações utilizadas foram:

$$n_{t,d} \sim Poisson(\lambda_{t,d}) ,$$

$$\lambda_{t,d} = \exp(\alpha_{t,d}) = \frac{a_d c_d f_d \exp(-c_d t)}{[b_d + \exp(-c_d t)]^{f_d+1}} ,$$

$$N_t \sim Poisson(\theta_t) \text{ e}$$

$$\theta_t = \frac{a c f \exp(-c t)}{[b + \exp(-c t)]^{f+1}} ,$$

para $t = 1, 2, \dots, T - 1, T, T + 1, \dots, T + H$ e $d = 0, 1, \dots, D$.

Além disso, $N_t = \sum_{d=0}^D n_{t,d}$ e, portanto $\theta_t = \sum_{d=0}^D \lambda_{t,d}$. Logo, tem-se a restrição que

$$\theta_t > \sum_{d=1}^D \lambda_{t,d} .$$

Em modelos Bayesianos é necessário especificar as distribuições ou densidades de probabilidade a priori. Para esse modelo proposto, foram utilizadas as densidades de probabilidade a priori vagas apresentadas em Bastos et al. (2019) para os parâmetros desconhecidos conforme segue abaixo:

$$a_d \sim Gama(0,1; 0,1) ,$$

$$\exp(b_d) \sim Normal(0; \sqrt{20}) ,$$

$$c_d \sim Gama(2; 9) ,$$

$$f_d \sim Gama(0,01; 0,01) ,$$

$$a \sim Gama(0,1; 0,1) ,$$

$$\exp(b) \sim Normal(0; \sqrt{20}) ,$$

$$c \sim Gama(2; 9) \text{ e}$$

$$f \sim Gama(0,01; 0,01) .$$

Modelo com estrutura conjunta de atraso na notificação (MC)

As contagens de incidências de doenças N_t e $n_{t,d}$ seguem uma distribuição de probabilidade Poisson com médias θ_t e $\lambda_{t,d}$, respectivamente. Salienta-se que Bastos et al. (2019) propõe a separação dos efeitos do tempo (α_t) e efeitos do atraso (β_d) na notificação no logaritmo da média das contagens ($\lambda_{t,d}$). Neste trabalho, as especificações do efeito do tempo e efeito da defasagem são diferentes das apresentadas em Bastos et al. (2019). O efeito do tempo foi caracterizado por uma curva logística generalizada (Gamerman et al., 2022) para representar o comportamento de epidemias. Por sua vez, o efeito das defasagens apresenta uma relação linear nos atrasos das notificações de incidência das doenças infecciosas e esse detalhamento foi incentivado por características das contagens de casos de dengue utilizadas na aplicação dos modelos. Além disso, as especificações utilizadas foram:

$$n_{t,d} \sim Poisson(\lambda_{t,d}) ,$$

$$\log(\lambda_{t,d}) = \alpha_t + \beta_d ,$$

$$\exp(\alpha_t) = \frac{a_\alpha c_\alpha f_\alpha \exp(-c_\alpha t)}{[b_\alpha + \exp(-c_\alpha t)]^{f_\alpha+1}} ,$$

$$\beta_d = \gamma d ,$$

$$N_t \sim Poisson(\theta_t) \text{ e}$$

$$\theta_t = \frac{a_\theta c_\theta f_\theta \exp(-c_\theta t)}{[b_\theta + \exp(-c_\theta t)]^{f_\theta+1}},$$

para $t = 1, 2, \dots, T - 1, T, T + 1, \dots, T + H$ e $d = 0, 1, \dots, D$.

Além disso, $N_t = \sum_{d=0}^D n_{t,d}$ e, portanto $\theta_t = \sum_{d=0}^D \lambda_{t,d}$. Logo, tem-se a restrição que

$$\theta_t > \sum_{d=1}^D \lambda_{t,d}.$$

Para esse modelo proposto, foram utilizadas as densidades de probabilidade a priori vagas apresentadas em Bastos et al. (2019) para os parâmetros desconhecidos conforme segue abaixo:

$$a_\alpha \sim Gama(0,1; 0,1),$$

$$\exp(b_\alpha) \sim Normal(0; \sqrt{20}),$$

$$c_\alpha \sim Gama(2; 9),$$

$$f_\alpha \sim Gama(0,01; 0,01),$$

$$\gamma \sim Normal(0; 100),$$

$$a_\theta \sim Gama(0,1; 0,1),$$

$$\exp(b_\theta) \sim Normal(0; \sqrt{20}),$$

$$c_\theta \sim Gama(2; 9) \text{ e}$$

$$f_\theta \sim Gama(0,01; 0,01).$$

Operacionalização computacional de modelos Bayesianos

A forma analítica exata das distribuições a posteriori dos modelos propostos é desconhecida. Isso comumente ocorre quando o modelo estatístico é complexo e

envolve múltiplos parâmetros e/ou possuí uma estrutura hierárquica. Portanto, é necessário aproximar as distribuições a posteriori das quantidades desconhecidas.

A técnica de amostragem de Markov Chain Monte Carlo (MCMC) é utilizada para obter uma amostra de possíveis valores dos parâmetros. Essa é uma técnica computacional comumente usada em inferência Bayesiana para realizar amostragem de distribuições ou densidades a posteriori que podem ser muito complexas. Além do mais, o MCMC é muito útil quando a distribuição/densidade a posteriori não pode ser calculada diretamente ou quando não há uma solução analítica disponível.

O MCMC (Gamerman and Lopes, 2006) permite gerar uma sequência de amostras que, em equilíbrio, representa uma aproximação da distribuição desejada. O processo de amostragem da distribuição ou densidade a posteriori usando MCMC envolve a construção de uma cadeia de Markov irredutível e aperiódica que tem a distribuição estacionária equivalente a distribuição/densidade a posteriori. A cadeia de Markov é construída por meio de uma série de iterações que consiste em propor um novo estado, geralmente por meio de amostragem aleatória baseada no estado atual, e decidir sobre a aceitação ou rejeição desse novo estado. Em rigor, um novo estado proposto pode ser aceito ou rejeitado com base em uma regra de aceitação, sendo que, essa regra geralmente envolve calcular a razão de verossimilhança e a probabilidade de aceitar o novo estado é determinada com base nessa razão.

As técnicas de MCMC geralmente usadas incluem o Algoritmo de Metropolis-Hastings e a Amostragem de Gibbs. Ambas as técnicas fornecem uma forma de gerar amostras da distribuição/densidade a posteriori e obter estimativas robustas para os parâmetros desconhecidos. Esses algoritmos são apresentados com mais detalhes em Gamerman e Lopes (2006).

A convergência no MCMC é atingida quando o algoritmo é capaz de gerar amostras que se aproximam da distribuição/densidade alvo desejada. É essencial garantir que as amostras geradas sejam representações da distribuição ou densidade de probabilidade alvo, sejam não viesadas e não apresentem um erro amostral relevante. O método utilizado para verificar a convergência no MCMC foi o monitoramento dos traços das cadeias. Os traços das cadeias MCMC são gráficos

que ilustram os valores dos parâmetros amostrados em função do tempo ou iteração. Os gráficos indicam convergência quando as cadeias parecem estar “flutuando” dentro de uma mesma região, sem tendências aparentes.

Comparação de modelos

Comparar os resultados de diferentes modelos é uma etapa importante para selecionar o modelo que melhor se adequou aos dados e ao problema determinado. O desempenho dos modelos pode ser avaliado através de múltiplas técnicas e métricas. A comparação dos modelos propostos foi feita de maneira visual através de gráficos das aproximações das distribuições a posteriori para as médias das quantidades de interesse, equiparação da porcentagem de cobertura dos intervalos de credibilidade para as previsões, e comparação das métricas MAE (*Mean Absolute Error*) e RMSE (*Root Mean Squared Error*).

A porcentagem de cobertura de um intervalo de credibilidade é uma medida que avalia a precisão do intervalo obtido. Essa medida indica a frequência com que o intervalo de credibilidade capturou o verdadeiro valor da quantidade desconhecida em repetidas amostragens, sendo que, quanto maior a porcentagem de cobertura, maior a segurança que o intervalo de credibilidade seja informativo e confiável. No entanto, quando a cobertura é muito alta, o intervalo será mais amplo e poderá fornecer pouca informação. Por outro lado, quando a cobertura é muito baixa, o intervalo será muito estreito e poderá não capturar adequadamente a variabilidade dos dados.

O MAE e o RMSE (Chai and Draxler, 2014) medem a magnitude média dos erros em um conjunto de estimativas ou previsões. O MAE é uma métrica calculada como a média das diferenças absolutas entre as previsões ou estimativas e os valores reais observados. Ele fornece uma medida média do erro absoluto sem considerar a direção do erro. Já o RMSE calcula a raiz quadrada da média dos erros quadrados dos valores previstos e quantidades observadas. Ele fornece uma medida da raiz quadrada do erro quadrático e é sensível a erros maiores.

O MAE e o RMSE são calculados através das seguintes fórmulas:

$$MAE = \sum_t \frac{|y_t - \hat{y}_t|}{m}$$

$$RMSE = \sqrt[2]{\frac{(y_t - \hat{y}_t)^2}{m}}$$

Onde, m corresponde ao número de parcelas da soma. Os índices do somatório dependem do tempo t para N_t , enquanto para $n_{t,d}$ os índices dependem do tempo t e do atraso d . Quando os erros de estimativa ou previsão são calculados para $n_{t,d}$, $y_t = n_{t,d}$ e $\hat{y}_t = E(\lambda_{t,d}|\underline{n})$, no cálculo dos erros de estimativa ou previsão para N_t , $y_t = N_t$ e $\hat{y}_t = E(\theta_t|\underline{n})$, onde $\underline{n} = \{n_{t,d}\}$ é a coleção dos dados observados.

Resultados

Descrição das contagens de casos de incidência de dengue

Os modelos propostos foram ajustados para dados reais de incidência de dengue notificados semanalmente no Rio de Janeiro entre janeiro de 2011 e abril de 2012. Parte das contagens foi omitida para que fosse possível verificar a qualidade das previsões geradas pelos modelos, portanto, a omissão parcial das contagens permite avaliar a capacidade dos modelos de construir previsões para as contagens não utilizadas durante o processo de ajuste dos modelos. Neste trabalho, foi feito o exercício de simular uma situação em que parte das contagens de incidência de dengue ainda não foram observadas. Como os dados de contagem da incidência de dengue disponíveis possuíam no máximo 10 semanas de atraso, definiu-se o atraso máximo relevante de $D = 10$ semanas. Ao avaliar a série epidemiológica da incidência de dengue, observa-se que boa parte das contagens haviam sido notificadas no tempo 25 e, portanto, foi especificado o tempo atual $T = 25$ semanas. A série foi observada para 35 semanas no total, dessa forma o horizonte de previsão equivale a $H = 10$ semanas. Cabe ressaltar que outros valores poderiam ter sido especificados para T , D e H , mas esses casos podem ser avaliados em aplicações de trabalhos futuros.

A Tabela 2 apresenta a contagem dos casos de incidência de dengue notificados. Parte dos valores (contagens em vermelho) foram omitidas para ajustar os modelos e ser possível verificar a qualidade das previsões obtidas. Dessa forma, foi realizada uma simulação onde as contagens até $t = 15$ são integralmente observadas, as contagens para $t = 16, \dots, 25$ são parcialmente observadas, e as contagens para $t = 26, \dots, 35$ equivalem aos valores futuros. Destaca-se que as previsões *Nowcasting* devem ser feitas para as contagens em cinza claro na tabela, uma vez que essas são as contagens que ocorreram, mas não foram notificadas. Já as previsões *Forecasting* devem ser calculadas para os valores em cinza escuro, pois correspondem as contagens que ainda vão ocorrer na simulação.

A Figura 1 exibe de maneira visual o total de casos de dengue ao logo do tempo. Através dessa ilustração percebe-se que a incidência de dengue começa com

contagens baixas, atinge seu pico (com 15 semanas) e depois, volta a diminuir. Esse é um comportamento típico de doenças infecciosas.

Através da ilustração na Figura 2 pode-se avaliar visualmente a distorção entre a incidência da doença relatada e a verdadeira incidência da doença. Observa-se que existe uma tendência de aumento na distorção das contagens à medida que o tempo passa, isso acontece porque a quantidade de casos ocorridos, mas não notificados aumenta.

Tabela 2 - Contagem de casos de incidência de dengue notificados no Rio de Janeiro

t/d	0	1	2	3	4	5	6	7	8	9	10	N parcial	N real
1	0	2	1	0	2	0	1	1	0	0	0	-	7
2	77	24	11	2	19	8	7	5	2	1	3	-	159
3	109	13	11	31	13	2	3	2	0	2	0	-	186
4	89	23	31	24	10	8	4	1	0	1	1	-	192
5	154	84	68	30	14	4	1	4	7	5	3	-	374
6	291	33	56	17	24	22	3	13	5	2	10	-	476
7	460	133	68	38	27	29	37	23	6	2	1	-	824
8	693	189	89	157	48	71	40	13	9	9	8	-	1326
9	779	187	189	107	91	48	17	12	14	8	9	-	1461
10	577	371	179	115	67	35	16	22	5	8	4	-	1399
11	533	242	168	58	20	26	8	4	12	6	0	-	1077
12	1137	325	268	118	60	9	14	9	9	3	2	-	1954
13	1716	607	234	147	64	42	26	20	24	26	1	-	2907
14	2240	798	375	194	85	78	67	63	83	52	6	-	4041
15	2410	950	324	222	97	85	227	205	198	249	6	-	4973
16	1629	989	435	197	243	180	268	302	116	34	11	4393	4404
17	501	1369	594	332	131	170	144	122	19	29	28	3382	3439
18	1032	1082	664	370	134	239	140	31	31	44	3	3692	3770
19	1079	1093	691	262	132	85	17	4	103	79	16	3359	3561
20	878	778	397	243	94	64	6	104	121	36	5	2454	2726
21	859	807	381	142	25	8	129	71	11	2	3	2214	2438
22	853	537	195	29	26	59	73	45	10	9	2	1614	1838
23	672	536	56	22	64	151	33	10	8	22	5	1264	1579
24	513	171	37	100	55	21	19	1	1	1	1	684	920
25	273	77	163	156	22	9	4	2	2	4	6	273	718
26	124	172	92	31	17	2	2	31	3	4	4	-	482
27	142	77	95	53	4	27	35	17	3	1	0	-	454
28	73	56	60	19	19	27	1	1	0	2	1	-	259
29	72	58	33	3	8	5	1	1	1	1	0	-	183
30	70	39	22	10	1	6	5	2	3	0	0	-	158
31	70	60	44	1	5	3	2	1	0	3	0	-	189
32	32	47	3	3	27	0	1	1	5	1	1	-	121

Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO

33	39	29	4	21	3	5	2	12	1	0	0	-	116
34	41	25	31	12	5	2	7	0	1	2	0	-	126
35	30	41	11	10	6	8	3	3	0	0	0	-	112

Figura 1 - Total de casos de incidência de dengue ao longo do tempo

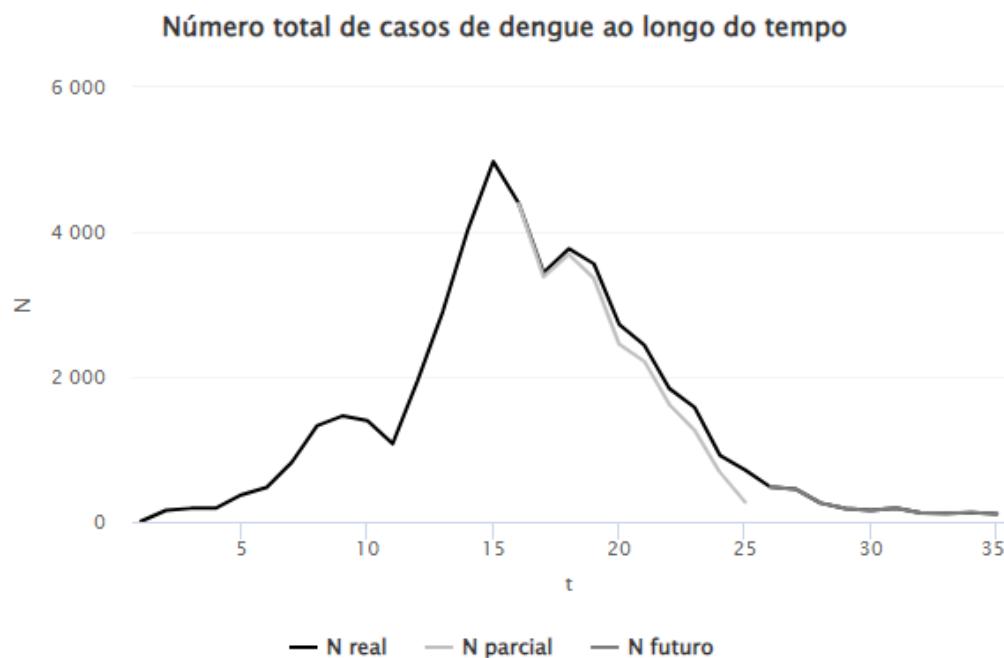
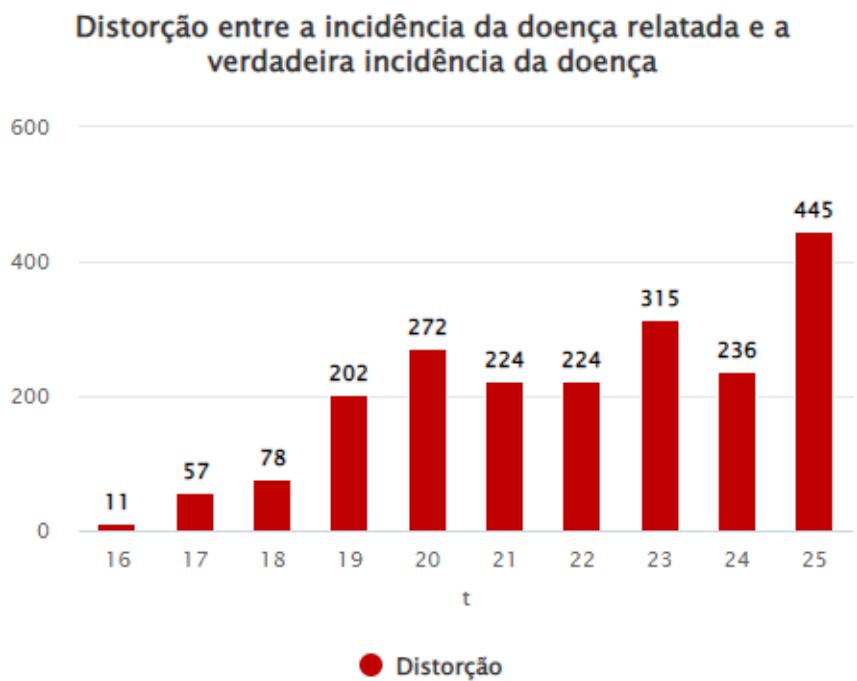


Figura 2 - Distorção entre a incidência de dengue relatada e a verdadeira incidência de dengue



É importante lembrar que a especificação do efeito de defasagem β_d no modelo com estrutura conjunta de atraso na notificação foi feita de modo a assumir uma relação linear nos atrasos, como mostra abaixo:

$$n_{t,d} \sim Poisson(\lambda_{t,d}) ,$$

$$\log(\lambda_{t,d}) = \alpha_t + \beta_d ,$$

$$\exp(\alpha_t) = \frac{a_\alpha c_\alpha f_\alpha \exp(-c_\alpha t)}{[b_\alpha + \exp(-c_\alpha t)]^{f_\alpha+1}} e$$

$$\beta_d = \gamma d ,$$

para $t = 1, 2, \dots, T-1, T, T+1, \dots, T+H$ e $d = 0, 1, \dots, D$.

Na Figura 3 são exibidas as contagens do total de casos de incidência de dengue conforme as atualizações foram realizadas. Já na Figura 4 foram ilustrados o logaritmo das contagens ao longo do tempo para cada atraso na notificação. Desse modo, observa-se que a grande maioria as atualizações apresentam um padrão de comportamento no crescimento do total de casos ao longo do tempo. Além do mais, para a maioria dos dados, há um decaimento no logaritmo das contagens com as defasagens, reforçando que a condição de linearidade nos atrasos para o modelo com estrutura conjunta de atraso na notificação parece razoável para esses dados.

Figura 3 - Total de casos de incidência de dengue ao longo do tempo por período de atualização

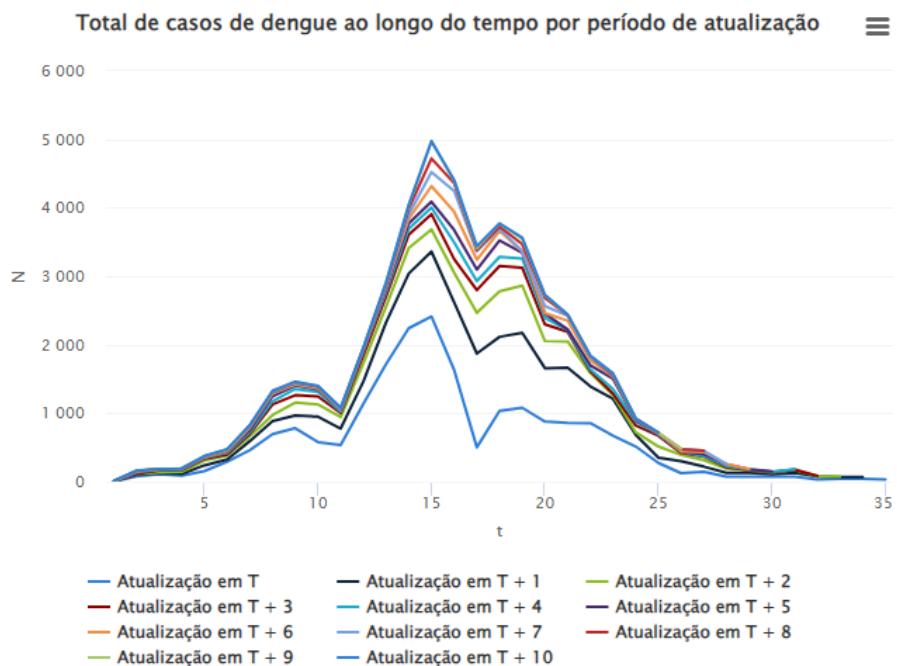
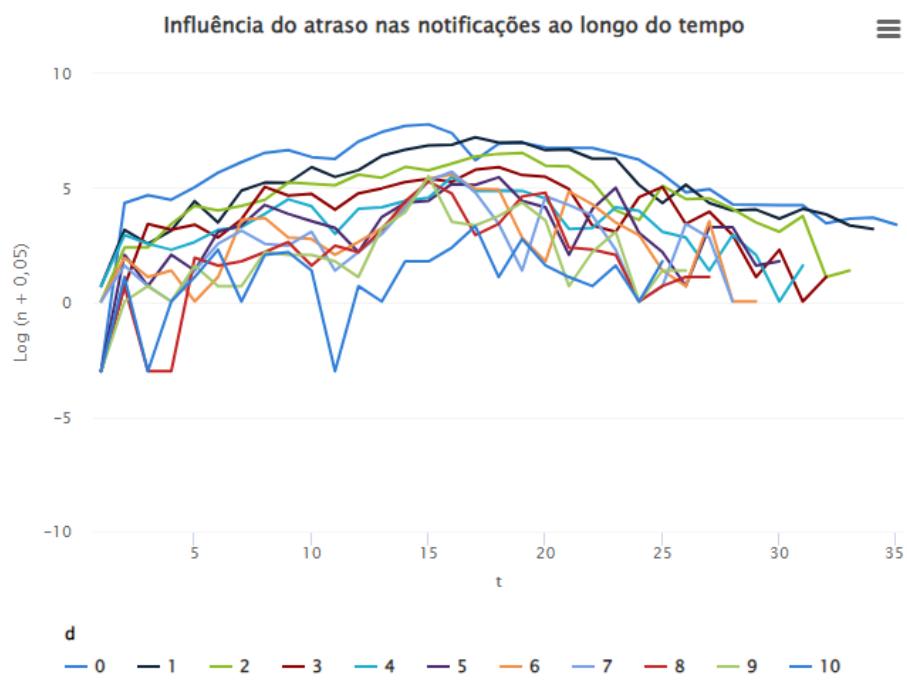


Figura 4 - Influência do atraso nas notificações ao longo do tempo



Resultados do modelo com estrutura de atraso na notificação independente

Ressalta-se que parte das contagens (valores em cinza claro e cinza escuro) foram omitidas para que fosse possível verificar a capacidade dos modelos de gerar previsões para os casos de dengue não utilizados no processo de ajuste dos modelos. Portanto, foi feito o exercício de simular uma situação em que parte das contagens não haviam sido observadas. Ademais, para as contagens não omitidas, os intervalos de credibilidade foram calculados para as médias das distribuições a posteriori. Por outro lado, para os valores omitidos no ajuste dos modelos, os intervalos foram feitos para os casos de incidência de dengue, já que existe o interesse de prever essas quantidades.

A Tabela 3 exibe os resultados das estimativas e previsões das médias das distribuições a posteriori do modelo com estrutura de atraso na notificação independente. Os valores reais dos casos de incidência de dengue são os valores acima, abaixo são apresentadas as médias das aproximações das distribuições a posteriori. Na Tabela 4 são apresentados os intervalos HPD de previsão com 95% de credibilidade das contagens de incidência de dengue, ou seja, $N_t|\underline{n}$ ou $n_{t,d}|\underline{n}$, onde $\underline{n} = \{n_{t,d}\}$ é a coleção dos dados observados. Além disso, a Figura 5 ilustra as estimativas, as previsões e os intervalos HPD com 95% de credibilidade para a média da distribuição a posteriori ou para o total de casos de dengue ao longo do tempo.

Tabela 3 - Estimativas e previsões das médias das distribuições a posteriori – MI

t/d	1	2	3	4	5	6	7	8	9	10	N
1	2	1	0	2	0	1	1	0	0	0	7
	10,43	11,01	10,01	6,89	3,65	1,05	0,53	0,34	0,13	1,07	82,51
2	24	11	2	19	8	7	5	2	1	3	159
	15,07	14,93	13,02	8,77	4,81	1,53	0,81	0,51	0,22	1,36	117,00
3	13	11	31	13	2	3	2	0	2	0	186
	21,75	20,22	16,91	11,16	6,33	2,23	1,24	0,78	0,36	1,72	165,55
4	23	31	24	10	8	4	1	0	1	1	192
	31,37	27,36	21,95	14,19	8,34	3,25	1,90	1,18	0,59	2,15	233,59
5	84	68	30	14	4	1	4	7	5	3	374
	45,17	36,97	28,46	18,02	10,96	4,74	2,90	1,79	0,98	2,65	328,27
6	33	56	17	24	22	3	13	5	2	10	476
	64,88	49,85	36,81	22,84	14,38	6,90	4,44	2,72	1,62	3,21	458,81
7	133	68	38	27	29	37	23	6	2	1	824
	92,86	67,01	47,48	28,86	18,83	10,02	6,78	4,14	2,68	3,84	636,44

Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO

8	189 132,20	89 89,67	157 60,98	48 36,30	71 24,56	40 14,53	13 10,32	9 6,28	9 4,42	8 4,51	1326 873,86
9	187 186,80	189 119,23	107 77,84	91 45,40	48 31,86	17 20,96	12 15,63	14 9,53	8 7,26	9 5,19	1461 1.183,41
10	371 261,08	179 157,11	115 98,52	67 56,29	35 41,02	16 30,01	22 23,47	5 14,43	8 11,83	4 5,86	1399 1.573,56
11	242 359,33	168 204,40	58 123,22	20 68,95	26 52,22	8 42,46	4 34,76	12 21,73	6 19,00	0 6,48	1077 2.043,02
12	325 484,04	268 261,28	118 151,62	60 83,07	9 65,49	14 58,97	9 50,36	9 32,38	3 29,74	2 7,03	1954 2.573,16
13	607 633,24	234 326,08	147 182,52	64 97,89	42 80,46	26 79,62	20 70,52	24 47,32	26 44,59	1 7,48	2907 3.121,42
14	798 797,03	375 394,33	194 213,51	85 112,04	78 96,17	67 103,18	63 93,94	83 66,75	52 62,49	6 7,81	4041 3.620,74
15	950 955,15	324 458,07	222 240,86	97 123,61	85 110,98	227 126,28	205 116,68	198 88,67	249 79,34	6 7,99	4973 3.990,72
16	989 1.078,78	435 506,69	197 260,03	243 130,45	180 122,63	268 143,40	302 132,40	116 107,14	34 88,51	11 8,04	4404 4.160,89
17	1369 1.138,96	594 529,61	332 266,82	131 130,82	170 128,74	144 148,78	122 135,01	19 113,26	29 85,02	28 7,94	3439 4.096,97
18	1082 1.119,26	664 520,52	370 259,03	134 124,15	239 127,69	140 139,85	31 123,01	31 101,87	44 70,40	3 7,73	3770 3.814,93
19	1093 1.024,51	691 480,50	262 237,60	132 111,40	85 119,34	17 119,28	4 100,81	103 77,86	79 51,36	16 7,40	3561 3.373,47
20	778 878,35	397 417,90	243 206,39	94 94,79	64 105,24	6 93,36	104 75,60	121 51,96	36 34,10	5 7,00	2726 2.850,53
21	807 711,67	381 344,64	142 170,69	25 76,93	8 88,01	129 68,18	71 52,97	11 31,51	2 21,26	3 6,53	2438 2.318,18
22	537 550,66	195 271,80	29 135,39	26 60,00	59 70,32	73 47,30	45 35,38	10 18,00	36 12,74	2 6,03	1838 1.827,56
23	536 411,10	56 206,83	22 103,80	64 45,34	151 54,13	33 31,65	10 22,89	8 9,94	22 7,46	5 5,52	1579 1.405,95
24	171 298,78	37 153,11	100 77,52	55 33,44	21 40,48	19 20,67	1 14,50	1 5,39	1 4,32	1 5,01	920 1.061,46
25	77 212,95	163 111,04	156 56,75	22 24,23	9 29,61	4 13,29	2 9,07	2 2,89	4 2,48	6 4,51	718 790,09
26	172 149,67	92 79,34	31 40,94	17 17,32	2 21,32	2 8,46	3 5,62	3 1,55	4 1,42	4 4,04	482 581,92
27	77 104,17	95 56,08	53 29,22	4 12,27	27 15,18	35 5,35	17 3,47	3 0,83	1 0,81	0 3,60	454 425,28
28	56 72,01	60 39,34	19 20,70	19 8,63	27 10,72	1 3,37	1 2,14	0 0,44	2 0,46	1 3,19	259 309,05
29	58 49,54	33 27,45	3 14,59	8 6,05	5 7,52	1 2,12	1 1,31	1 0,24	1 0,26	0 2,82	183 223,66
30	39 33,98	22 19,09	10 10,24	1 4,22	6 5,26	5 1,33	2 0,81	3 0,13	0 0,15	0 2,48	158 161,38
31	60 23,25	44 13,24	1 7,17	5 2,94	3 3,67	2 0,84	1 0,50	0 0,07	3 0,09	0 2,18	189 116,19
32	47 15,88	3 9,16	3 5,02	27 2,05	0 2,56	1 0,52	1 0,30	5 0,04	1 0,05	1 1,91	121 83,53
33	29 10,84	4 6,34	21 3,50	3 1,43	5 1,78	2 0,33	12 0,19	1 0,02	0 0,03	0 1,68	116 59,98
34	25	31	12	5	2	7	0	1	2	0	126

Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO

	7,40	4,38	2,44	0,99	1,24	0,21	0,11	0,01	0,02	1,46	43,04
35	41	11	10	6	8	3	3	0	0	0	112
	5,04	3,02	1,70	0,69	0,86	0,13	0,07	0,01	0,01	1,28	30,86

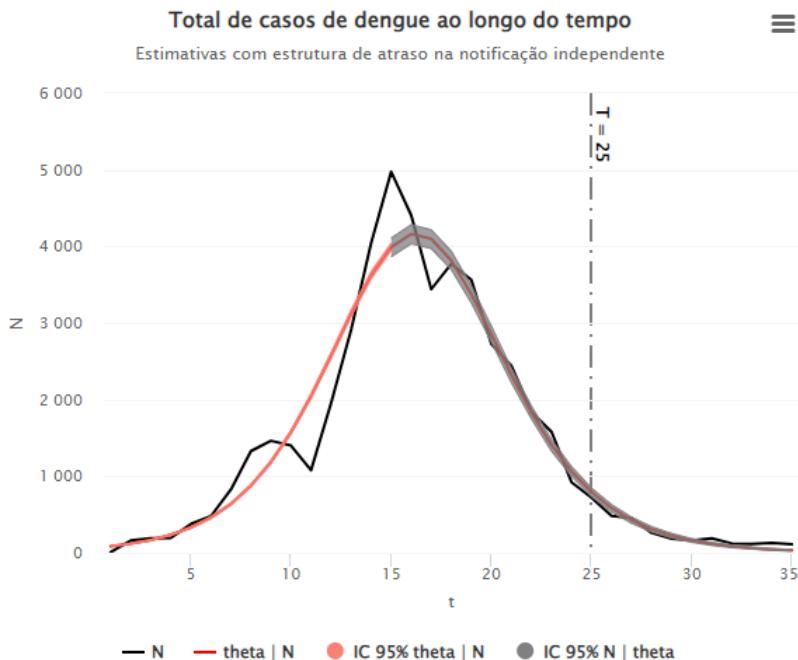
Valor real acima, média a posteriori abaixo.

Tabela 4 - Intervalos de previsão dos casos de dengue com 95% de credibilidade – MI

t/d	1	2	3	4	5	6	7	8	9	10	N
16										11	4404
										[3; 13]	[4032; 4282]
17									29	28	3439
									[66; 102]	[3; 13]	[3970; 4216]
18							31	44	3		3770
							[82; 121]	[55; 87]	[3; 13]		[3695; 3936]
19						4	103	79	16		3561
						[80; 119]	[61; 94]	[37; 64]	[2; 12]		[3260; 3484]
20					6	104	121	36	5		2726
					[75; 111]	[59; 92]	[37; 65]	[23; 45]	[2; 12]		[2750; 2957]
21				8	129	71	11	2	3		2438
				[69; 105]	[51; 83]	[38; 66]	[21; 42]	[13; 30]	[2; 11]		[2220; 2406]
22			26	59	73	45	10	9	2		1838
			[44; 74]	[53; 85]	[34; 60]	[23; 46]	[9; 25]	[6; 19]	[1; 10]		[1747; 1915]
23		22	64	151	33	10	8	22	5		1579
		[84; 123]	[32; 58]	[40; 68]	[21; 42]	[13; 31]	[3; 15]	[2; 12]	[1; 10]		[1328; 1475]
24	37	100	55	21	19	1	1	1	1		920
	[129; 176]	[59; 93]	[22; 44]	[29; 53]	[12; 29]	[7; 21]	[2; 10]	[1; 8]	[1; 9]		[998; 1123]
25	77	163	156	22	9	4	2	2	4	6	718
	[183; 240]	[90; 131]	[42; 71]	[15; 33]	[18; 39]	[6; 20]	[3; 14]	[0; 6]	[0; 5]	[1; 8]	[733; 843]
26	172	92	31	17	2	2	31	3	4	4	482
	[124; 172]	[61; 96]	[27; 52]	[9; 25]	[13; 30]	[2; 13]	[1; 10]	[0; 4]	[0; 4]	[1; 8]	[533; 628]
27	77	95	53	4	27	35	17	3	1	0	454
	[83; 123]	[40; 69]	[18; 39]	[5; 18]	[7; 22]	[2; 10]	[0; 7]	[0; 3]	[0; 2]	[0; 7]	[382; 463]
28	56	60	19	19	27	1	1	0	2	1	259
	[56; 88]	[26; 50]	[12; 29]	[3; 14]	[5; 17]	[0; 7]	[0; 5]	[0; 2]	[0; 2]	[0; 6]	[272; 340]
29	58	33	3	8	5	1	1	1	1	0	183
	[35; 62]	[17; 37]	[6; 21]	[1; 10]	[3; 13]	[0; 5]	[0; 3]	[0; 1]	[0; 1]	[0; 6]	[192; 251]
30	39	22	10	1	6	5	2	3	0	0	158
	[22; 44]	[11; 27]	[4; 16]	[1; 8]	[1; 9]	[0; 3]	[0; 2]	[0; 1]	[0; 1]	[0; 5]	[135; 184]
31	60	44	1	5	3	2	1	0	3	0	189
	[13; 32]	[6; 20]	[2; 12]	[0; 6]	[0; 7]	[0; 3]	[0; 2]	[0; 1]	[0; 1]	[0; 5]	[94; 136]
32	47	3	3	27	0	1	1	5	1	1	121
	[8; 23]	[4; 15]	[1; 9]	[0; 5]	[0; 5]	[0; 2]	[0; 1]	[0; 0]	[0; 0]	[0; 4]	[65; 100]
33	29	4	21	3	5	2	12	1	0	0	116
	[5; 17]	[2; 11]	[0; 7]	[0; 4]	[0; 4]	[0; 1]	[0; 1]	[0; 0]	[0; 0]	[0; 4]	[46; 75]
34	25	31	12	5	2	7	0	1	2	0	126
	[2; 12]	[1; 8]	[0; 5]	[0; 3]	[0; 3]	[0; 1]	[0; 1]	[0; 0]	[0; 0]	[0; 4]	[30; 55]
35	41	11	10	6	8	3	3	0	0	0	112
	[1; 9]	[0; 6]	[0; 4]	[0; 2]	[0; 3]	[0; 1]	[0; 1]	[0; 0]	[0; 0]	[0; 3]	[20; 41]

Valor real acima, intervalo de credibilidade abaixo.

Figura 5 - Estimativas e previsões para o total de casos de dengue – MI



A Figura 6 ilustra as médias das aproximações das distribuições a posteriori dos casos de incidência de dengue ao longo do tempo para cada atraso na notificação. Já na Figura 7 observa-se o logaritmo das médias das aproximações das distribuições a posteriori dos casos de dengue ao longo do tempo por defasagem.

Com base no modelo, obtêm-se uma suavização das contagens dos casos de dengue. Portanto, é possível reproduzir as Figuras 3 e 4 de forma mais sistematizada e livre dos ruídos que as contagens estavam apresentando. Através dessas imagens suavizadas pode-se concluir que, em sua maioria, há um padrão de decaimento no logaritmo da média das distribuições a posteriori das contagens com os atrasos, reforçando a especificação de linearidade nos atrasos ($\beta_d = \gamma d$) para o modelo com estrutura conjunta de atraso na notificação.

Figura 6 - Distribuições a posteriori da média de casos de dengue ao longo do tempo por defasagem – MI

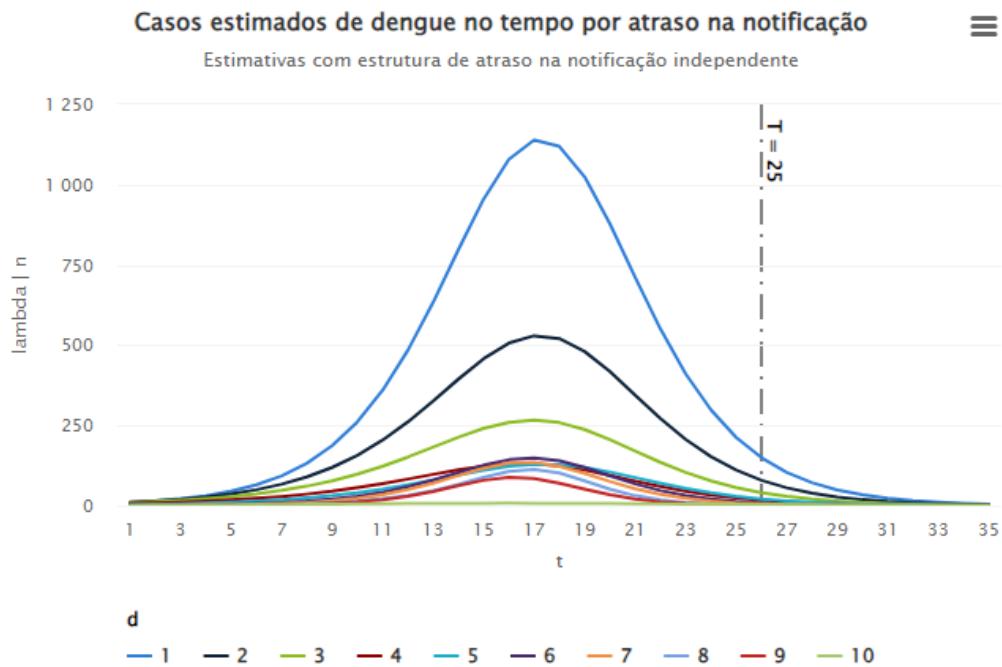
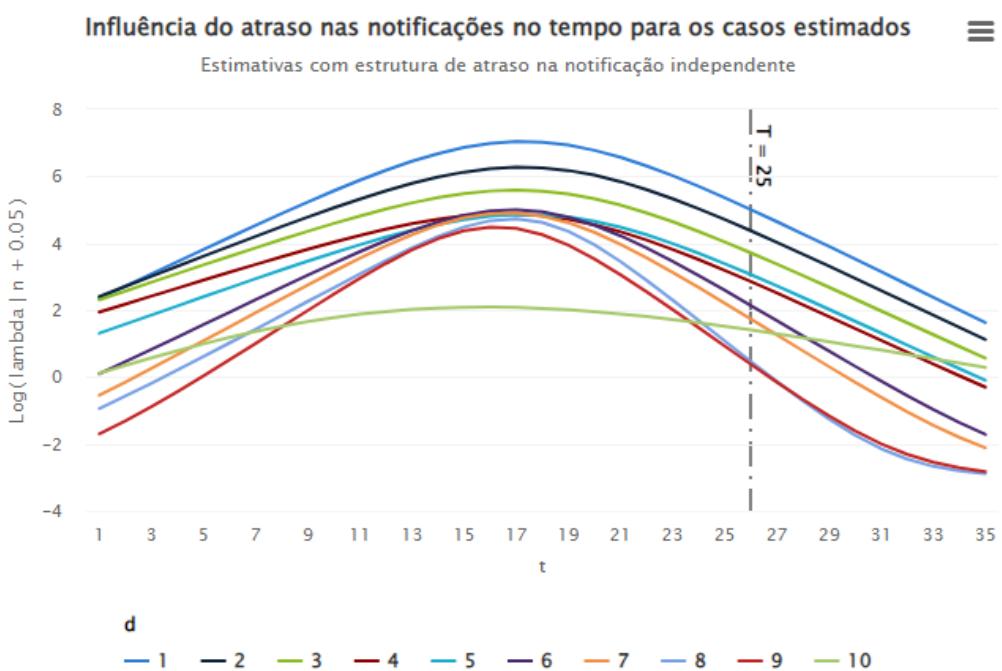


Figura 7 - Influência do atraso nas notificações ao longo do tempo – MI



Resultados do modelo com estrutura conjunta de atraso na notificação (MC)

A Tabela 5 expõe os resultados das estimativas e previsões das médias das distribuições a posteriori do modelo com estrutura conjunta de atraso na notificação. Os valores reais dos casos de incidência de dengue são os valores acima, abaixo são apresentadas as médias das aproximações das distribuições a posteriori. Na Tabela 6 são exibidos os intervalos HPD de previsão com 95% de credibilidade das contagens de incidência de dengue, ou seja, $N_t|\underline{n}$ ou $n_{t,d}|\underline{n}$, onde $\underline{n} = \{n_{t,d}\}$ é a coleção dos dados observados. Ademais, a Figura 8 exibe visualmente as estimativas, as previsões e os intervalos HPD com 95% de credibilidade para a média da distribuição a posteriori ou para o total de casos de dengue ao longo do tempo.

Tabela 5 - Estimativas e previsões das médias das distribuições a posteriori – MC

t/d	1	2	3	4	5	6	7	8	9	10	N
1	2	1	0	2	0	1	1	0	0	0	7
	18,49	12,01	7,80	5,07	3,29	2,14	1,39	0,90	0,59	0,38	91,73
2	24	11	2	19	8	7	5	2	1	3	159
	24,62	15,99	10,39	6,75	4,38	2,85	1,85	1,20	0,78	0,51	127,34
3	13	11	31	13	2	3	2	0	2	0	186
	32,76	21,28	13,83	8,98	5,83	3,79	2,46	1,60	1,04	0,68	176,56
4	23	31	24	10	8	4	1	0	1	1	192
	43,61	28,33	18,40	11,95	7,76	5,04	3,28	2,13	1,38	0,90	244,34
5	84	68	30	14	4	1	4	7	5	3	374
	58,02	37,69	24,48	15,90	10,33	6,71	4,36	2,83	1,84	1,20	337,26
6	33	56	17	24	22	3	13	5	2	10	476
	77,18	50,13	32,57	21,15	13,74	8,93	5,80	3,77	2,45	1,59	463,77
7	133	68	38	27	29	37	23	6	2	1	824
	102,59	66,64	43,29	28,12	18,27	11,87	7,71	5,01	3,25	2,11	634,35
8	189	89	157	48	71	40	13	9	9	8	1326
	136,23	88,49	57,48	37,34	24,26	15,76	10,24	6,65	4,32	2,81	861,21
9	187	189	107	91	48	17	12	14	8	9	1461
	180,57	117,30	76,19	49,50	32,15	20,89	13,57	8,82	5,73	3,72	1157,03
10	371	179	115	67	35	16	22	5	8	4	1399
	238,65	155,02	100,70	65,42	42,49	27,61	17,93	11,65	7,57	4,92	1532,23
11	242	168	58	20	26	8	4	12	6	0	1077
	313,84	203,87	132,43	86,03	55,88	36,30	23,58	15,32	9,95	6,47	1989,71
12	325	268	118	60	9	14	9	9	3	2	1954
	409,38	265,93	172,75	112,22	72,90	47,36	30,76	19,99	12,98	8,43	2517,27
13	607	234	147	64	42	26	20	24	26	1	2907
	526,93	342,28	222,35	144,44	93,83	60,95	39,60	25,72	16,71	10,86	3078,99
14	798	375	194	85	78	67	63	83	52	6	4041
	663,86	431,23	280,13	181,97	118,21	76,79	49,89	32,41	21,05	13,68	3610,49
15	950	324	222	97	85	227	205	198	249	6	4973
	808,93	525,47	341,34	221,74	144,04	93,57	60,79	39,49	25,65	16,67	4025,40

Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO

16	989 937,85	435 609,21	197 395,74	243 257,07	180 167,00	268 108,49	302 70,48	116 45,78	34 29,74	11 19,32	4404 4238,14
17	1369 1014,29	594 658,87	332 428,00	131 278,03	170 180,61	144 117,33	122 76,22	19 49,52	29 32,17	28 20,90	3439 4196,76
18	1082 1003,72	664 652,00	370 423,54	134 275,13	239 178,73	140 116,11	31 75,43	31 49,00	44 31,83	3 20,68	3770 3907,42
19	1093 897,43	691 582,96	262 378,69	132 246,00	85 159,80	17 103,81	4 67,44	103 43,81	79 28,46	16 18,49	3561 3432,55
20	778 724,56	397 470,67	243 305,74	94 198,61	64 129,02	6 83,82	104 54,45	121 35,37	36 22,98	5 14,93	2726 2863,57
21	807 534,47	381 347,19	142 225,53	25 146,51	8 95,17	129 61,83	71 40,17	11 26,09	2 16,95	3 11,01	2438 2287,18
22	537 367,11	195 238,47	29 154,91	26 100,63	59 65,37	73 42,47	45 27,59	10 17,92	9 11,64	2 7,56	1838 1763,93
23	536 239,54	56 155,60	22 101,08	64 65,66	151 42,66	33 27,71	10 18,00	8 11,70	22 7,60	5 4,94	1579 1323,92
24	171 150,99	37 98,08	100 63,72	55 41,39	21 26,89	19 17,47	1 11,35	1 7,37	1 4,79	1 3,11	920 973,52
25	77 93,08	163 60,47	156 39,28	22 25,52	9 16,58	4 10,77	2 7,00	2 4,54	4 2,95	6 1,92	718 705,13
26	172 56,60	92 36,77	31 23,88	17 15,51	2 10,08	31 6,55	3 4,25	3 2,76	4 1,80	4 1,17	482 505,16
27	77 34,12	95 22,17	53 14,40	4 9,35	27 6,08	35 3,95	17 2,56	3 1,67	1 1,08	0 0,70	454 359,06
28	56 20,47	60 13,30	19 8,64	19 5,61	27 3,65	1 2,37	1 1,54	0 1,00	2 0,65	1 0,42	259 253,80
29	58 12,24	33 7,95	3 5,17	8 3,36	5 2,18	1 1,42	1 0,92	1 0,60	1 0,39	1 0,25	183 178,68
30	39 7,31	22 4,75	10 3,08	1 2,00	6 1,30	5 0,85	2 0,55	3 0,36	0 0,23	0 0,15	158 125,45
31	60 4,36	44 2,83	1 1,84	5 1,19	3 0,78	2 0,50	1 0,33	1 0,21	0 0,14	0 0,09	189 87,91
32	47 2,60	3 1,69	3 1,10	27 0,71	0 0,46	1 0,30	1 0,20	5 0,13	1 0,08	1 0,05	121 61,52
33	29 1,55	4 1,00	21 0,65	3 0,42	5 0,28	2 0,18	12 0,12	1 0,08	0 0,05	0 0,03	116 43,01
34	25 0,92	31 0,60	12 0,39	5 0,25	2 0,16	7 0,11	0 0,07	1 0,04	2 0,03	0 0,02	126 30,05
35	41 0,55	11 0,36	10 0,23	6 0,15	8 0,10	3 0,06	3 0,04	0 0,03	0 0,02	0 0,01	112 20,98

Valor real acima, média a posteriori abaixo.

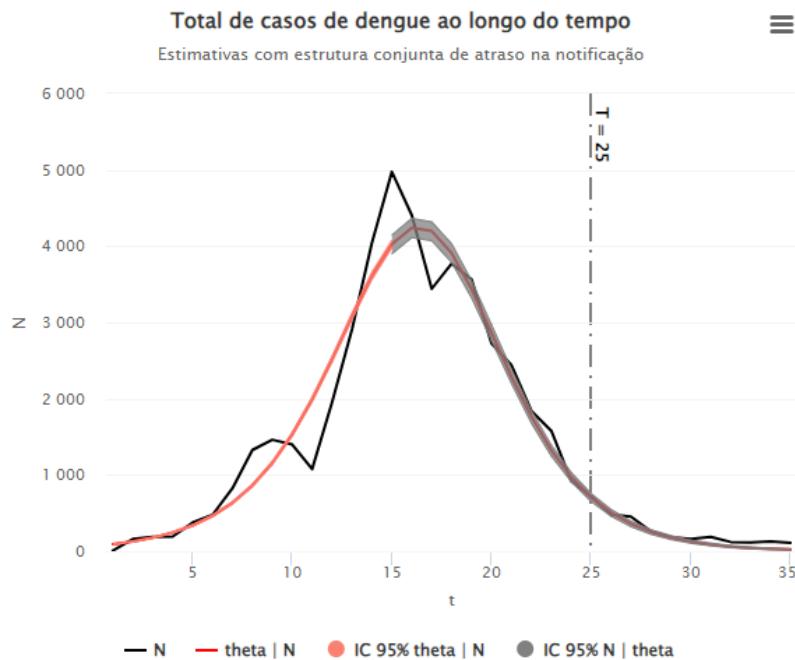
Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO

Tabela 6 - Intervalos de previsão dos casos de dengue com 95% de credibilidade – MC

t/d	1	2	3	4	5	6	7	8	9	10	N
16										11 [10; 27]	4404 [4111; 4363]
17									29 [20; 42]	28 [12; 29]	3439 [4069; 4317]
18								31 [35; 62]	44 [20; 42]	3 [12; 29]	3770 [3785; 4030]
19							4 [50; 82]	103 [31; 56]	79 [19; 39]	16 [10; 26]	3561 [3319; 3545]
20					6 [65; 100]	104 [40; 68]	121 [23; 46]	36 [14; 32]	5 [8; 22]		2726 [2761; 2969]
21				8 [76; 113]	129 [46; 76]	71 [28; 52]	11 [16; 35]	2 [8; 24]	3 [5; 17]		2438 [2198; 2382]
22			26 [82; 120]	59 [48; 79]	73 [29; 54]	45 [17; 37]	10 [9; 25]	9 [5; 18]	2 [3; 13]		1838 [1678; 1843]
23		22 [80; 119]	64 [50; 81]	151 [29; 54]	33 [17; 37]	10 [10; 26]	8 [5; 18]	22 [3; 13]	5 [1; 9]		1579 [1247; 1390]
24	37 [78; 116]	100 [49; 79]	55 [28; 53]	21 [16; 36]	19 [9; 25]	1 [4; 17]	1 [2; 12]	1 [1; 9]	1 [0; 6]		920 [912; 1032]
25	77 [73; 111]	163 [45; 75]	156 [27; 51]	22 [16; 35]	9 [8; 24]	4 [5; 17]	2 [2; 12]	2 [0; 8]	4 [0; 6]	6 [0; 4]	718 [651; 755]
26	172 [41; 70]	92 [25; 48]	31 [13; 32]	17 [7; 22]	2 [4; 16]	2 [2; 11]	31 [1; 8]	3 [0; 6]	4 [0; 4]	4 [0; 3]	482 [458; 547]
27	77 [23; 45]	95 [12; 30]	53 [7; 21]	4 [4; 15]	27 [1; 10]	35 [0; 7]	17 [0; 5]	3 [0; 4]	1 [0; 3]	0 [0; 2]	454 [320; 394]
28	56 [11; 28]	60 [7; 20]	19 [3; 14]	19 [1; 10]	27 [0; 7]	1 [0; 5]	1 [0; 4]	0 [0; 3]	2 [0; 2]	1 [0; 2]	259 [221; 282]
29	58 [6; 19]	33 [3; 13]	3 [1; 9]	8 [0; 7]	5 [0; 5]	1 [0; 4]	1 [0; 3]	1 [0; 2]	1 [0; 2]	0 [0; 1]	183 [153; 205]
30	39 [2; 12]	22 [1; 9]	10 [0; 6]	1 [0; 5]	6 [0; 3]	5 [0; 3]	2 [0; 2]	3 [0; 1]	0 [0; 1]	0 [0; 1]	158 [103; 146]
31	60 [1; 8]	44 [0; 6]	1 [0; 4]	5 [0; 3]	3 [0; 2]	2 [0; 2]	1 [0; 1]	0 [0; 1]	3 [0; 1]	0 [0; 1]	189 [70; 106]
32	47 [0; 6]	3 [0; 4]	3 [0; 3]	27 [0; 2]	0 [0; 2]	1 [0; 1]	1 [0; 1]	5 [0; 1]	1 [0; 1]	1 [0; 1]	121 [45; 75]
33	29 [0; 4]	4 [0; 3]	21 [0; 2]	3 [0; 2]	5 [0; 1]	2 [0; 1]	12 [0; 1]	1 [0; 1]	0 [0; 0]	0 [0; 0]	116 [30; 55]
34	25 [0; 3]	31 [0; 2]	12 [0; 2]	5 [0; 1]	2 [0; 1]	7 [0; 1]	0 [0; 1]	1 [0; 0]	2 [0; 0]	0 [0; 0]	126 [19; 40]
35	41 [0; 2]	11 [0; 1]	10 [0; 1]	6 [0; 1]	8 [0; 1]	3 [0; 1]	3 [0; 0]	0 [0; 0]	0 [0; 0]	0 [0; 0]	112 [13; 30]

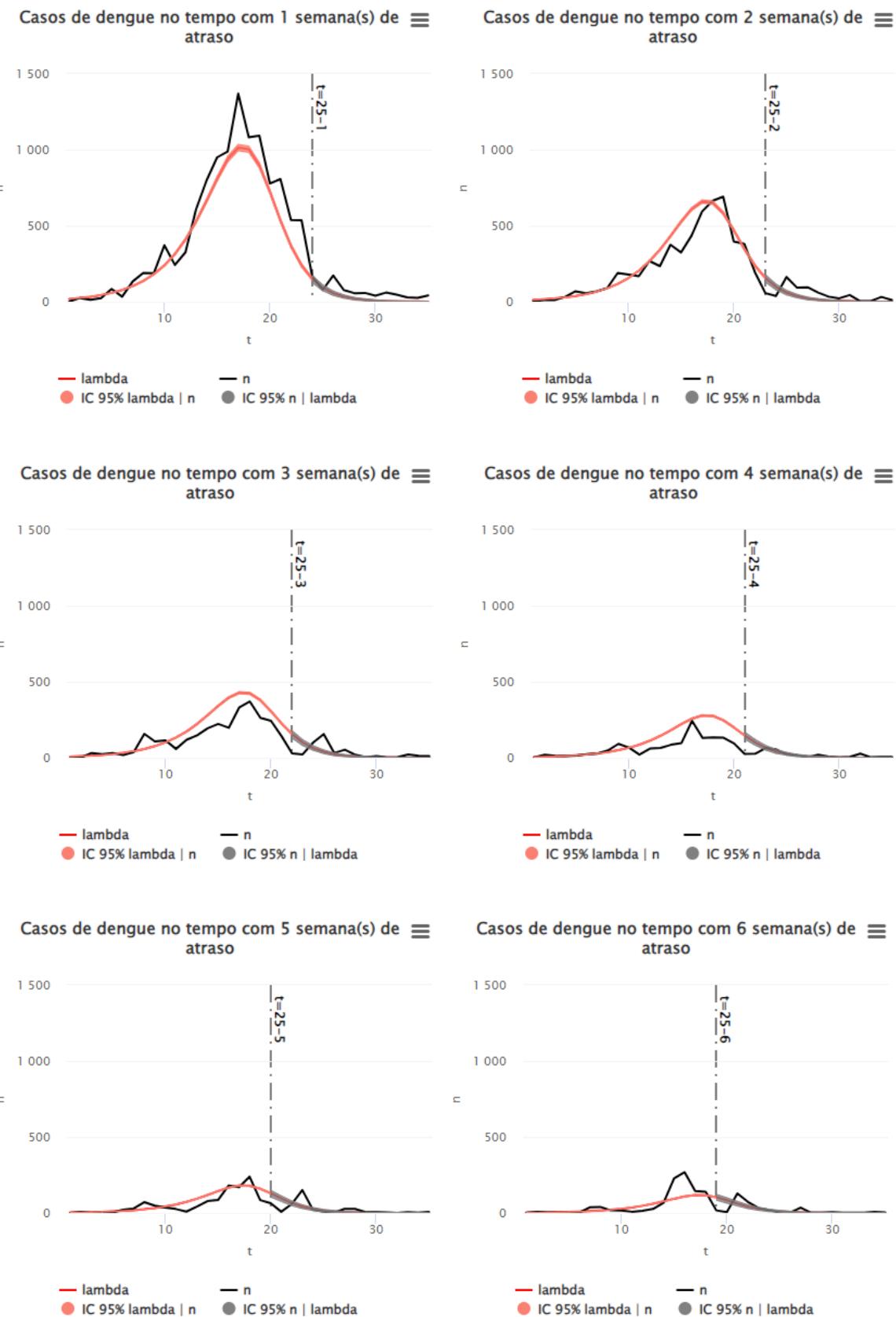
Valor real acima, intervalo de credibilidade abaixo.

Figura 8 - Estimativas e previsões do total de casos de dengue – MC

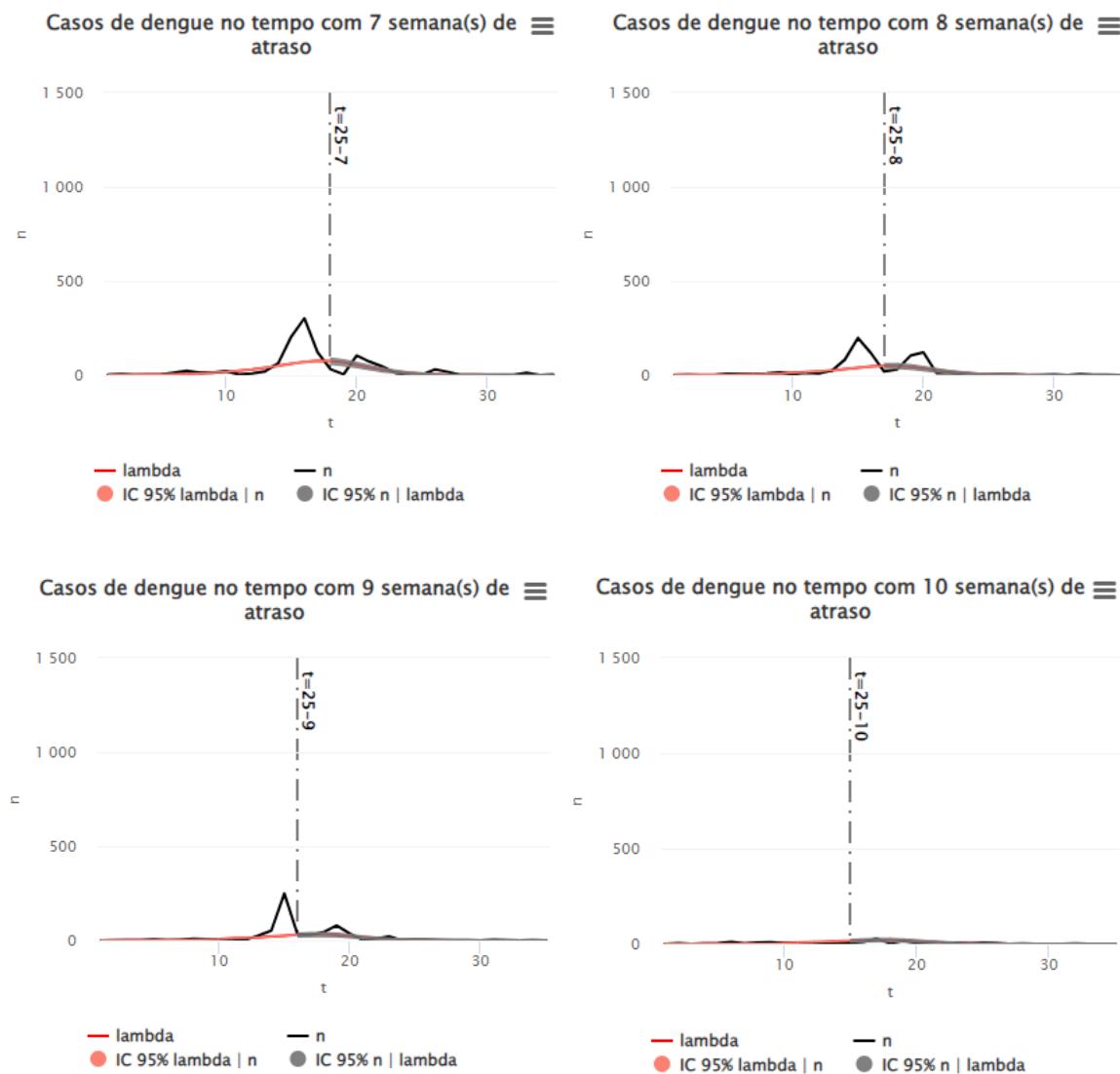


A Figura 9 exibe as estimativas, as previsões e os intervalos de credibilidade da média das distribuições a posteriori e dos casos de dengue para cada semana de atraso na notificação considerando os resultados do ajuste do modelo com estrutura conjunta de atraso na notificação. É importante salientar que à medida que o atraso na notificação aumenta, o impacto da diferença entre as estimativas ou previsões e os valores reais diminui. Portanto, as diferenças resultantes para as primeiras defasagens são mais graves do que as diferenças nos últimos atrasos, já que a medida que o atraso avança tem-se menos casos de dengue, resultando em contagens mais baixas.

Figura 9 - Estimativas e previsões para os casos de dengue por semana de atraso



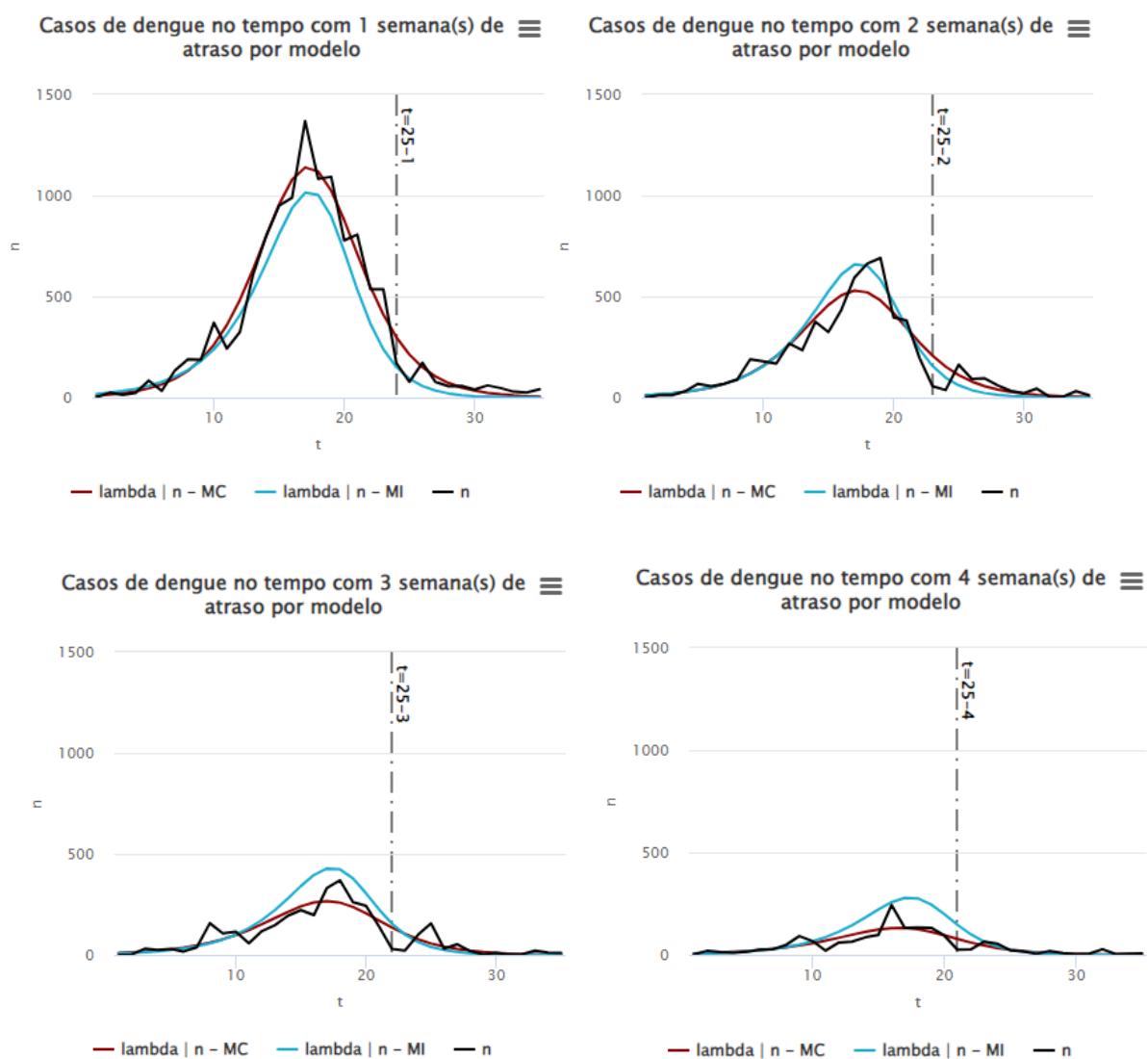
Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO



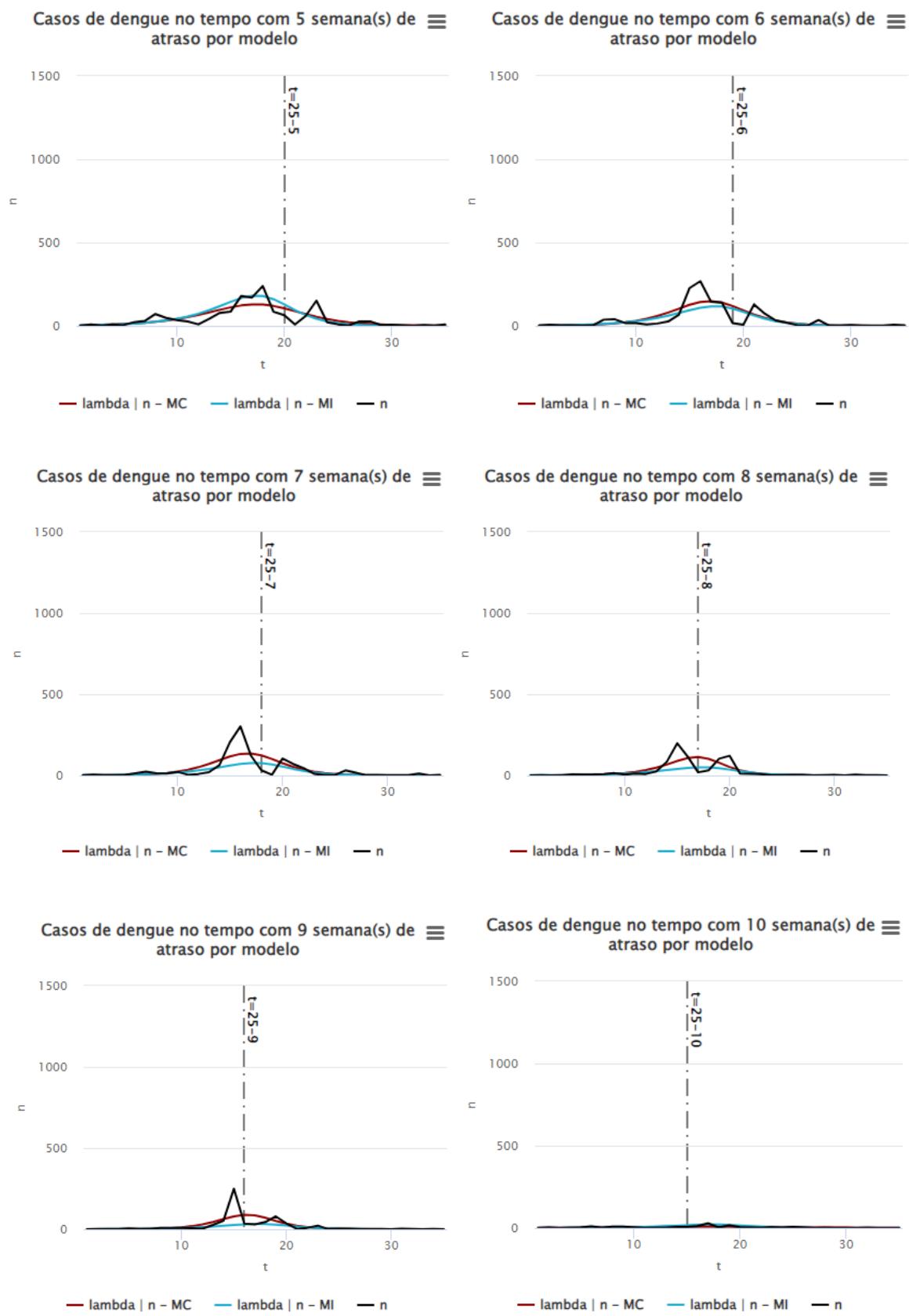
Comparação dos resultados dos modelos

Na Figura 10 são apresentadas de maneira visual as comparações das médias das distribuições a posteriori das contagens de casos de incidência de dengue em cada modelo por defasagem. Através das visualizações percebe-se que não existe uma conclusão clara de qual modelo melhor se ajustou aos casos de dengue, já que hora o modelo com estrutura conjunta de atraso na notificação (MC) teve uma performance melhor, e outras vezes o modelo com estrutura de atraso na notificação independente (MI) performou melhor.

Figura 10 - Médias das distribuições a posteriori para cada semana de atraso por modelo

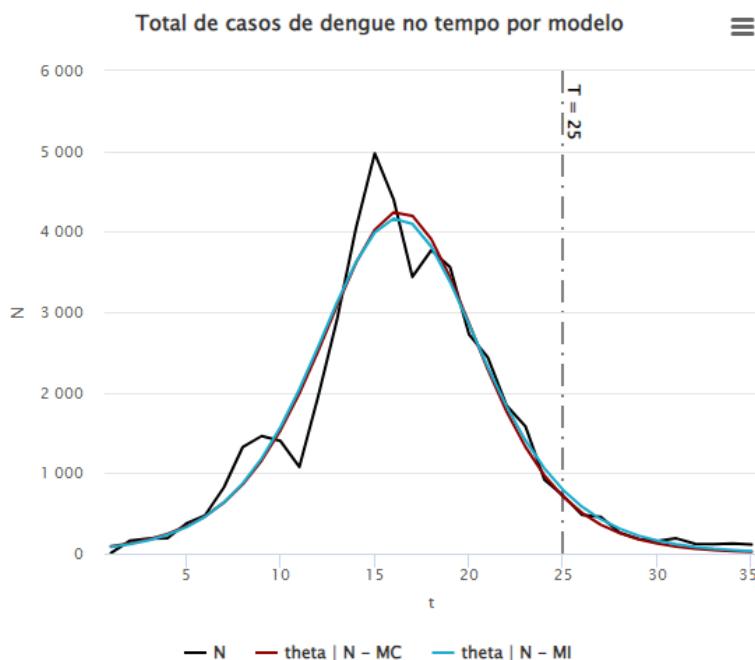


Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO



A Figura 11 exibe a média da aproximação das distribuições a posteriori do total de casos de incidência de dengue ao longo do tempo por modelo. Com essa imagem pode-se notar que, mesmo não havendo um padrão para avaliar o modelo com melhor performance por defasagem, há uma similaridade do desempenho dos modelos para o total de casos de dengue.

Figura 11 - Média da distribuição a posteriori do total de casos de dengue por modelo



A Tabela 7 apresenta as métricas MAE e RMSE como uma forma de comparar os erros de estimação e erros de previsão entre os modelos. Ressalta-se que parte das contagens de casos de dengue foram omitidas, sendo que, os erros de estimação foram calculados para as contagens consideradas no processo de ajuste dos modelos. Os erros de previsão *Nowcasting* foram calculados para os eventos ocorridos, mas não notificados na simulação de uma situação real. Por fim, os erros de previsão *Forecasting* foram calculados para as contagens de casos futuros da simulação de aplicação dos modelos. Sendo assim, nota-se que os erros de estimação do total de casos de dengue (N_t) foi menor para o modelo com estrutura conjunta de atraso na notificação. Os erros de previsão *Nowcasting* foram maiores que os erros de previsão *Forecasting* em ambos os modelos, sendo que, o modelo

com estrutura de atraso na notificação independente apresentou os menores erros de previsão.

Tabela 7 - MAE e RMSE por modelo

Estrutura de atraso	Dados	Erro de estimação		Erro de previsão			
		MAE	RMSE	MAE	RMSE	MAE	RMSE
Independente	n	34,79	56,22	18,44	29,36	9,81	20,55
	N	303,30	437,95	178,3	247,91	55,00	61,68
Conjunta	n	43,94	74,02	21,89	37,13	13,02	25,77
	N	290,40	420,14	187,95	273,86	58,5	69,17

A partir dos intervalos HPD com 95% de credibilidade para as previsões foram calculadas as probabilidades de cobertura para cada modelo. A probabilidade de cobertura para o modelo com estrutura de atraso na notificação foi 17,14%, enquanto para o modelo com estrutura conjunta de atraso na notificação a cobertura foi de 22,86%. Esses valores indicam que o modelo com estrutura conjunta de atraso na notificação teve um desempenho melhor para capturar as quantidades reais do total de casos de dengue. No entanto, ainda assim a porcentagem de cobertura foi muito baixa considerando um intervalo com 95% de credibilidade. As porcentagens de cobertura foram baixas porque os intervalos obtidos foram muito estreitos e não capturaram adequadamente a variabilidade das contagens. Uma possibilidade para solucionar essa questão poderia ser assumir uma distribuição de probabilidade Binomial Negativa para as contagens ao invés de usar a distribuição Poisson e esse levantamento é feito nas conclusões do trabalho.

Conclusões

Neste trabalho foram propostos modelos com duas estruturas diferentes para incorporar o atraso na notificação de séries epidemiológicas, gerar estimativas e previsões para as quantidades de interesse. Um dos modelos incorpora uma estrutura de atraso na notificação independente, enquanto o outro é mais restrito e específica uma estrutura conjunta de atraso na notificação.

Os modelos foram especificados através de modelos hierárquicos Bayesianos e foi realizada a aplicação para casos semanais de incidência de dengue no Rio de Janeiro ocorridos entre janeiro de 2011 e abril de 2012. Foram construídas estimativas e previsões pontuais e intervalares dos casos de dengue simulando uma situação real onde parte das contagens não havia sido observada devido ao atraso na notificação. Além disso, o método MCMC foi utilizado para obter amostras das distribuições de interesse dos modelos.

Os resultados da aplicação foram satisfatórios para as estimativas das médias das distribuições a posteriori. Os modelos apresentaram um desempenho similar para o total de casos de dengue, sendo que, o modelo com estrutura conjunta de atraso na notificação obteve menores erros MAE e RMSE de estimativa para o total de contagem de incidência de dengue. Por outro lado, os erros MAE e RMSE de previsão foram menores para o modelo com estrutura de atraso na notificação independente. Ressalta-se que MAE e o RMSE possuem algumas limitações como sensibilidade a *outliers* e não fornecem uma medida direta de incerteza ou probabilidade associada às estimativas ou previsões dos modelos. Portanto, em trabalhos futuros é importante avaliar outras métricas para comparar os resultados dos modelos.

Ao avaliar os resultados das previsões dos casos de dengue, conclui-se que os intervalos de credibilidade foram muito estreitos e não foram capazes de incorporar adequadamente a variabilidade das contagens de dengue nessa aplicação. Uma das formas possíveis que pode ser avaliada futuramente para contornar esse impasse é adotar outra distribuição de probabilidade para os dados de contagem, como a Binomial Negativa ao invés da Poisson. Isso porque, a distribuição de probabilidade Poisson assume que a média e a variância dos dados são iguais, propriedade

conhecida como equidispersão. Já Binomial Negativa tem maior flexibilidade, pois permite que a média e a variância dos dados sejam diferentes, podendo se ajustar melhor aos dados quando a variância é maior do que o valor esperado pela distribuição Poisson.

No contexto de séries epidemiológicas é considerável incorporar elementos de efeito espacial nos modelos e comparar o desempenho dos modelos em diferentes épocas para uma mesma região geográfica. Também pode ser interessante avaliar outras formulações para o efeito das defasagens (β_d) e efeito do tempo (α_t) no modelo com estrutura conjunta de atraso na notificação.

Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO

Anexo

As Tabelas 8 e 9 apresentam os intervalos HPD com 95% credibilidade para as médias das distribuições a posteriori ou casos de dengue de cada modelo.

Tabela 8 - Intervalos de credibilidade do MI

t/d	1	2	3	4	5	6	7	8	9	10	N
1	2 [8.83; 11.91]	1 [9.23; 12.9]	0 [8.01; 11.93]	2 [5.29; 8.54]	0 [2.67; 4.69]	1 [0.65; 1.5]	1 [0.29; 0.8]	0 [0.15; 0.56]	0 [0.04; 0.24]	0 [0.32; 1.9]	[77.04; 87.9]
2	24 [13.05; 17.05]	11 [12.74; 17.2]	2 [10.69; 15.25]	19 [6.88; 10.59]	8 [3.61; 6.13]	7 [0.98; 2.11]	5 [0.48; 1.19]	2 [0.24; 0.82]	1 [0.08; 0.39]	3 [0.57; 2.28]	159 [110.11; 123.58]
3	13 [19.23; 24.35]	11 [17.63; 23]	31 [14.33; 19.54]	13 [9.05; 13.24]	2 [4.85; 7.82]	3 [1.51; 3]	2 [0.75; 1.75]	0 [0.41; 1.21]	2 [0.14; 0.61]	0 [0.82; 2.65]	186 [156.94; 173.47]
4	23 [28.13; 34.61]	31 [24.17; 30.56]	24 [19.04; 24.96]	10 [11.97; 16.67]	8 [6.68; 10.14]	4 [2.3; 4.25]	1 [1.26; 2.62]	0 [0.69; 1.78]	0 [0.26; 0.96]	1 [1.23; 3.16]	192 [223.16; 243.14]
5	84 [41.11; 49.22]	68 [33.22; 40.75]	30 [25.11; 31.76]	14 [15.45; 20.66]	4 [8.88; 12.98]	1 [3.52; 6.03]	4 [2.02; 3.87]	7 [1.08; 2.56]	5 [0.52; 1.54]	3 [1.64; 3.65]	374 [316.41; 339.92]
6	33 [59.68; 69.66]	56 [45.51; 54.24]	17 [33.13; 40.52]	24 [20.11; 25.86]	22 [12.07; 16.65]	3 [5.3; 8.47]	13 [3.25; 5.74]	5 [1.79; 3.76]	2 [0.93; 2.39]	10 [2.15; 4.28]	476 [445.07; 472.29]
7	133 [86.54; 98.59]	68 [62.06; 72.14]	38 [43.24; 51.47]	27 [25.7; 31.98]	29 [16.24; 21.43]	37 [8.05; 12.02]	23 [5.19; 8.48]	6 [2.84; 5.41]	2 [1.69; 3.76]	1 [2.73; 5.01]	824 [620.78; 652.04]
8	189 [125.21; 139.48]	89 [83.94; 95.46]	157 [56.18; 65.43]	48 [32.79; 39.8]	71 [21.64; 27.53]	40 [12.12; 16.97]	13 [8.26; 12.5]	9 [4.86; 7.96]	9 [3.06; 5.91]	8 [3.35; 5.83]	1326 [856.27; 892.49]
9	187 [178.58; 195.52]	189 [112.73; 125.92]	107 [72.69; 83.04]	91 [41.31; 49.26]	48 [28.53; 35.28]	17 [18.01; 23.87]	12 [12.91; 18.3]	14 [7.5; 11.63]	8 [5.4; 9.22]	9 [3.89; 6.59]	1461 [1162.33; 1205.5]
10	371 [250.59; 270.98]	179 [149.03; 164.53]	115 [92.47; 104.51]	67 [51.73; 60.87]	35 [37.27; 44.99]	16 [26.65; 33.66]	22 [20.15; 26.85]	5 [11.95; 17.02]	8 [9.3; 14.32]	4 [4.44; 7.37]	1399 [1547.16; 1600.26]
11	242 [347.57; 372.27]	168 [195.7; 214.09]	58 [115.97; 130.24]	20 [63.66; 74.37]	26 [47.82; 56.84]	8 [38.25; 46.68]	4 [30.69; 38.94]	12 [18.83; 25.1]	6 [15.8; 22.26]	0 [4.96; 8.1]	1077 [201.22; 2076.63]
12	325 [468.82; 499.22]	268 [250.45; 272.45]	118 [143.21; 160.05]	60 [76.73; 89.16]	9 [60.22; 70.73]	14 [53.88; 64.11]	9 [45.32; 55.51]	3 [28.39; 36.17]	3 [25.84; 34.21]	2 [5.47; 8.8]	1954 [2534.9; 2612.4]
13	607 [615.27; 652.26]	234 [313.31; 339.22]	147 [173.16; 192.63]	64 [90.97; 104.96]	42 [74.53; 86.76]	26 [73.3; 85.96]	20 [64.5; 77.03]	24 [42.27; 52.58]	26 [39.3; 50.07]	1 [5.77; 9.18]	2907 [3080.31; 3167.86]
14	798 [775.41; 818.83]	375 [379.67; 409.39]	194 [202.68; 224.24]	85 [104.2; 119.66]	78 [89.32; 103.36]	67 [95.22; 110.62]	63 [86.54; 101.92]	83 [60; 73.71]	52 [56.04; 69.58]	6 [6.08; 9.52]	4041 [3576.91; 3670.46]
15	950 [929.32; 977.76]	324 [441.86; 474.66]	222 [229.2; 252.26]	97 [115.12; 131.86]	85 [103.35; 119.09]	227 [116.95; 135.26]	205 [107.35; 125.88]	198 [80.09; 97.42]	249 [71.12; 87.24]	6 [6.28; 9.7]	4973 [3945.01; 4041.07]
16	899 [1054.3; 1105.83]	435 [489.05; 524.26]	197 [247.75; 272.04]	243 [121.89; 139.7]	180 [114.14; 131.19]	268 [132.62; 153.67]	302 [121.63; 142.87]	116 [97.46; 117.41]	34 [79.64; 98.06]	11 [3; 13]	4404 [4032; 4282]
17	1369 [1112.61; 1166.99]	594 [510.76; 547.98]	332 [253.75; 279.31]	131 [121.64; 140.02]	170 [119.7; 137.96]	144 [137.68; 160.18]	122 [123.7; 146.03]	19 [102.02; 124.47]	29 [66; 102]	28 [3; 13]	3439 [3970; 4216]
18	1082 [1092.43; 1148.03]	664 [501.43; 539.71]	370 [246.91; 271.9]	134 [114.72; 133.16]	239 [118.3; 137.09]	140 [129.11; 150.65]	31 [112.81; 133.72]	31 [82; 121]	44 [55; 87]	3 [3; 13]	3770 [3695; 3936]
19	1093 [998.28; 1051.03]	691 [462.13; 498.88]	262 [225.2; 250.42]	132 [103.15; 120.15]	85 [110.09; 128.06]	17 [110.21; 128.62]	4 [80; 119]	103 [61; 94]	79 [37; 64]	16 [2; 12]	3561 [3260; 3484]
20	778 [855.16; 901.4]	397 [401.31; 433.19]	243 [194.93; 217.49]	94 [87.27; 102.01]	64 [97.44; 113.47]	6 [75; 111]	104 [58; 92]	121 [37; 65]	36 [23; 45]	5 [2; 12]	2726 [2750; 2957]
21	807 [692.1; 730.38]	381 [331.73; 358.22]	142 [161.03; 179.89]	25 [70.7; 83.02]	8 [69; 105]	129 [51; 83]	71 [38; 66]	11 [21; 42]	2 [13; 30]	3 [2; 11]	2438 [2220; 2406]
22	537 [535.04; 566.28]	195 [260.85; 282.94]	29 [127.08; 142.77]	26 [44; 74]	59 [53; 85]	73 [34; 60]	45 [23; 46]	10 [9; 25]	9 [6; 19]	2 [1; 10]	1838 [1747; 1915]
23	536 [398.11; 424.4]	56 [197.5; 216.42]	22 [84; 123]	64 [32; 58]	151 [40; 68]	33 [21; 42]	104 [13; 31]	121 [3; 15]	36 [2; 12]	5 [1; 10]	1579 [1328; 1475]
24	171 [287.89; 310.65]	37 [129; 176]	100 [59; 93]	55 [22; 44]	21 [29; 53]	19 [12; 29]	1 [7; 21]	1 [2; 10]	1 [1; 8]	1 [1; 9]	920 [998; 1123]
25	77 [183; 240]	163 [61; 96]	156 [27; 52]	22 [9; 25]	9 [13; 30]	4 [2; 13]	2 [1; 10]	2 [0; 4]	4 [0; 4]	6 [1; 8]	718 [733; 843]
26	172 [124; 172]	92 [61; 96]	31 [27; 52]	17 [9; 25]	2 [13; 30]	35 [1; 10]	3 [0; 4]	1 [0; 4]	4 [1; 8]	4 [53; 628]	482
27	77 [83; 123]	95 [40; 69]	53 [18; 39]	4 [5; 18]	27 [7; 22]	35 [2; 10]	17 [0; 7]	3 [0; 3]	1 [0; 2]	0 [0; 7]	454 [382; 463]
28	56 [56; 88]	60 [26; 50]	19 [12; 29]	19 [3; 14]	27 [5; 17]	1 [0; 7]	1 [0; 5]	0 [0; 2]	2 [0; 2]	1 [0; 6]	259 [272; 340]
29	58 [35; 62]	33 [17; 37]	3 [6; 21]	8 [1; 10]	5 [3; 13]	1 [0; 6]	1 [0; 3]	1 [0; 1]	1 [0; 1]	0 [0; 6]	183 [192; 251]
30	39 [22; 44]	22 [11; 27]	10 [4; 16]	1 [1; 8]	6 [1; 9]	5 [0; 3]	2 [0; 2]	3 [0; 1]	0 [0; 1]	0 [0; 5]	158 [135; 184]
31	60 [13; 32]	44 [6; 20]	1 [2; 12]	5 [0; 6]	3 [0; 7]	2 [0; 3]	1 [0; 2]	0 [0; 1]	3 [0; 1]	0 [0; 5]	189 [94; 136]
32	47 [8; 23]	3 [4; 15]	3 [1; 9]	27 [0; 5]	0 [0; 5]	1 [0; 2]	1 [0; 1]	5 [0; 1]	1 [0; 0]	1 [0; 4]	121 [65; 100]
33	29 [5; 17]	4 [2; 11]	21 [0; 7]	3 [0; 4]	5 [0; 1]	2 [0; 1]	12 [0; 1]	1 [0; 0]	0 [0; 0]	0 [0; 4]	116 [46; 75]
34	25 [2; 12]	31 [1; 8]	12 [0; 5]	5 [0; 3]	2 [0; 1]	7 [0; 1]	0 [0; 0]	1 [0; 0]	2 [0; 0]	0 [0; 4]	126 [30; 55]
35	41 [1; 9]	11 [0; 6]	10 [0; 4]	6 [0; 2]	8 [0; 3]	3 [0; 1]	3 [0; 1]	0 [0; 1]	0 [0; 0]	0 [0; 3]	112 [20; 41]

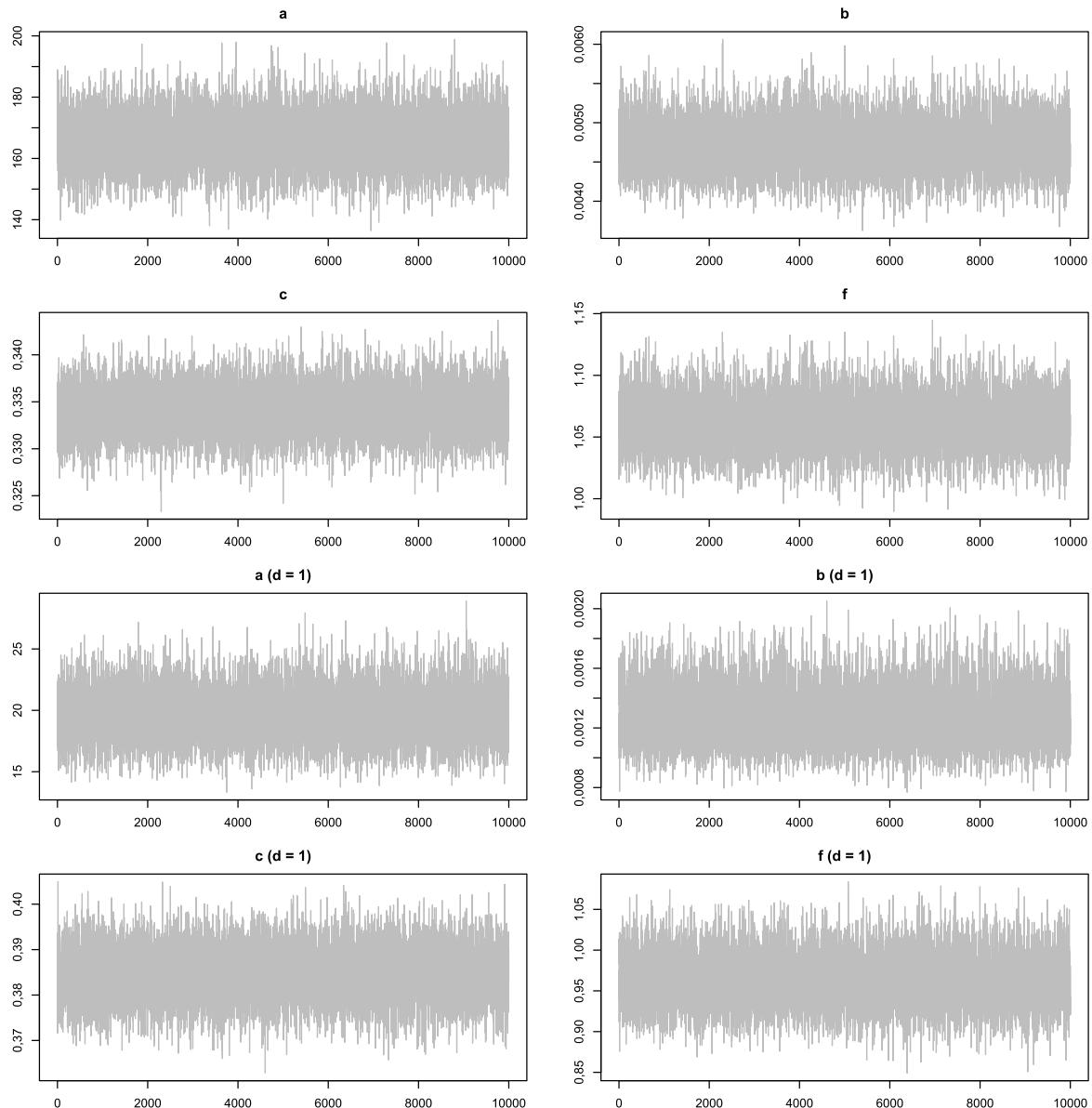
Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO

Tabela 9 - Intervalos de credibilidade do MC

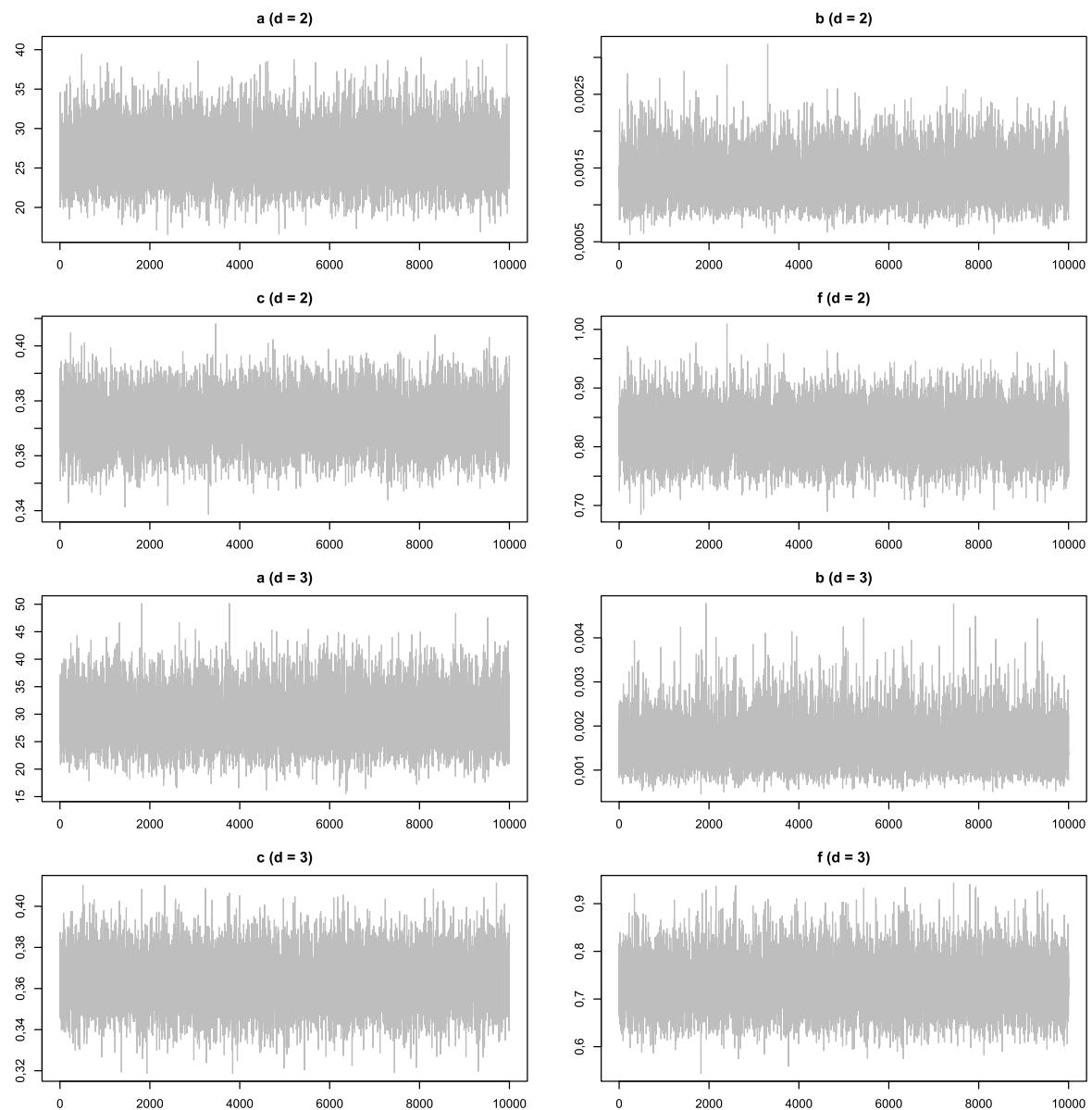
/d	1	2	3	4	5	6	7	8	9	10	N
1	2	1	0	2	0	1	1	0	0	0	7
	[17,16; 19,85]	[11,14; 12,86]	[7,24; 8,36]	[4,7; 5,43]	[3,05; 3,53]	[1,98; 2,3]	[1,28; 1,49]	[0,83; 0,98]	[0,53; 0,63]	[0,35; 0,41]	[85,7; 97,59]
2	24	11	2	19	8	7	5	2	1	3	159
	[23,01; 26,28]	[14,92; 17,01]	[9,69; 11,05]	[6,29; 7,17]	[4,08; 4,67]	[2,65; 3,05]	[1,71; 1,98]	[1,11; 1,29]	[0,72; 0,84]	[0,46; 0,55]	[119,91; 134,41]
3	13	11	31	13	2	3	2	0	2	0	186
	[30,75; 34,71]	[19,98; 22,5]	[12,98; 14,61]	[8,45; 9,52]	[5,47; 6,19]	[3,55; 4,03]	[2,29; 2,62]	[1,49; 1,72]	[0,97; 1,12]	[0,62; 0,73]	[167,85; 185,23]
4	23	31	24	10	8	4	1	0	1	1	192
	[41,2; 45,92]	[26,8; 29,81]	[17,39; 19,34]	[11,31; 12,59]	[7,32; 8,18]	[4,76; 5,34]	[3,08; 3,49]	[2; 2,27]	[1,29; 1,48]	[0,83; 0,96]	[234,09; 254,64]
5	84	68	30	14	4	1	4	7	5	3	374
	[55,17; 60,79]	[35,91; 39,46]	[23,29; 25,58]	[15,15; 16,67]	[9,81; 10,83]	[6,36; 7,07]	[4,11; 4,61]	[2,66; 3,01]	[1,72; 1,95]	[1,11; 1,28]	[325,85; 349,72]
6	33	56	17	24	22	3	13	5	2	10	476
	[74,02; 80,63]	[48,07; 52,21]	[31,21; 33,89]	[20,26; 22,04]	[13,12; 14,33]	[8,5; 9,35]	[5,5; 6,1]	[3,56; 3,98]	[2,3; 2,6]	[1,48; 1,69]	[450,11; 477,43]
7	133	68	38	27	29	37	23	6	2	1	824
	[98,85; 106,57]	[64,13; 68,91]	[41,7; 44,79]	[27,09; 29,15]	[17,57; 19]	[11,38; 12,4]	[7,35; 8,07]	[4,75; 5,26]	[3,07; 3,44]	[1,98; 2,25]	[618,34; 649,76]
8	189	89	157	48	71	40	13	9	9	8	1326
	[131,64; 140,62]	[85,67; 91,14]	[55,67; 59,19]	[36,16; 38,55]	[23,42; 25,1]	[15,15; 16,37]	[9,78; 10,67]	[6,34; 6,98]	[4,08; 4,55]	[2,64; 2,98]	[843,3; 879,93]
9	187	189	107	91	48	17	12	14	8	9	1461
	[175,46; 185,92]	[114,3; 120,55]	[74,14; 78,16]	[48,17; 50,94]	[31,21; 33,21]	[20,15; 21,62]	[13,04; 14,13]	[8,43; 9,24]	[5,43; 6,03]	[3,51; 3,94]	[1135,11; 1178,95]
10	371	179	115	67	35	16	22	5	8	4	1399
	[232,1; 244,55]	[151,4; 158,63]	[98,46; 103,03]	[63,86; 67,09]	[41,27; 43,65]	[26,72; 28,52]	[17,23; 18,6]	[11,13; 12,16]	[7,2; 7,97]	[4,64; 5,21]	[1505,6; 1558,73]
11	242	168	58	20	26	8	4	12	6	0	1077
	[306,39; 321,37]	[199,79; 208,34]	[129,8; 135,17]	[84,15; 88,02]	[54,43; 57,36]	[35,18; 37,45]	[22,75; 24,5]	[14,68; 16]	[9,46; 10,45]	[6,1; 6,83]	[1956,93; 2021,92]
12	325	268	118	60	9	14	9	9	3	2	1954
	[400,13; 418,6]	[260,75; 271,18]	[169,55; 176,07]	[109,91; 114,64]	[71,06; 74,74]	[45,89; 48,78]	[29,64; 31,87]	[19,15; 20,84]	[12,35; 13,62]	[7,96; 8,9]	[2478,37; 2555,85]
13	607	234	147	64	42	26	20	24	26	1	2907
	[515,47; 538,52]	[335,8; 348,63]	[218,32; 226,36]	[141,5; 147,37]	[91,49; 96,11]	[59,1; 62,77]	[38,17; 41,01]	[24,64; 26,82]	[15,9; 17,53]	[10,27; 11,47]	[3035,49; 3122,67]
14	798	375	194	85	78	67	63	83	52	6	4041
	[649,69; 678,16]	[423,4; 439,09]	[274,9; 284,83]	[178,47; 185,66]	[115,33; 121,02]	[74,5; 79,04]	[48,07; 51,63]	[31,03; 33,75]	[20,04; 22,09]	[12,94; 14,45]	[3562,53; 3657,07]
15	950	324	222	97	85	227	205	198	249	6	4973
	[791,55; 825,39]	[515,92; 534,56]	[335,52; 347,24]	[217,46; 226]	[140,72; 147,52]	[90,77; 96,24]	[58,69; 63]	[37,86; 41,16]	[24,43; 26,9]	[15,77; 17,6]	[3976,46; 4076,46]
16	989	435	197	243	180	268	302	116	34	11	4404
	[918,26; 956,95]	[598,98; 620,02]	[389,09; 402,31]	[252,23; 262,06]	[163,03; 170,91]	[105,44; 111,73]	[68,03; 73,01]	[43,82; 47,65]	[28,26; 31,14]	[10; 27]	[4111; 4363]
17	1369	594	332	131	170	144	122	19	29	28	3439
	[993,71; 1035,42]	[647,21; 670,1]	[420,85; 435,64]	[272,43; 283,44]	[176,07; 184,82]	[113,94; 120,93]	[73,52; 78,98]	[47,4; 51,59]	[20; 42]	[12; 29]	[4069; 4317]
18	1082	664	370	134	239	140	31	31	44	3	3770
	[982,34; 1023,89]	[640,47; 663,71]	[415,6; 430,8]	[269,2; 280,61]	[174,19; 183,19]	[112,59; 119,71]	[72,75; 78,29]	[35; 62]	[20; 42]	[12; 29]	[3785; 4030]
19	1093	691	262	132	85	17	4	103	79	16	3561
	[879,61; 916,28]	[572,45; 593,73]	[371,74; 385,71]	[240,69; 251,26]	[156,64; 164,01]	[100,7; 107,32]	[50; 82]	[31; 56]	[19; 39]	[10; 26]	[3319; 3545]
20	778	397	243	94	64	6	104	121	36	5	2726
	[708,89; 739,69]	[461,49; 480,07]	[299,58; 312,05]	[193,86; 203,2]	[125,38; 132,7]	[65; 100]	[40; 68]	[23; 46]	[14; 32]	[8; 22]	[2761; 2969]
21	807	381	142	25	8	129	71	11	2	3	2438
	[520,58; 548,94]	[338,46; 356,18]	[219,53; 231,53]	[142,12; 150,68]	[76; 113]	[46; 76]	[28; 52]	[16; 35]	[8; 24]	[5; 17]	[2198; 2382]
22	537	195	29	26	59	73	45	10	9	2	1838
	[353,86; 381,13]	[229,66; 247,22]	[148,94; 160,73]	[82; 120]	[48; 79]	[29; 54]	[17; 37]	[9; 25]	[5; 18]	[3; 13]	[1678; 1843]
23	536	56	22	64	151	33	10	8	22	5	1579
	[227,01; 251,7]	[147,55; 163,56]	[80; 119]	[50; 81]	[29; 54]	[17; 37]	[10; 26]	[5; 18]	[3; 13]	[1; 9]	[1247; 1390]
24	171	37	100	55	21	19	1	1	1	1	920
	[140,48; 160,74]	[78; 116]	[49; 79]	[28; 53]	[16; 36]	[9; 25]	[4; 17]	[2; 12]	[1; 9]	[0; 6]	[912; 1032]
25	77	163	156	22	9	4	2	2	4	6	718
	[73; 111]	[48; 75]	[27; 51]	[16; 35]	[8; 24]	[5; 17]	[2; 12]	[0; 8]	[0; 6]	[0; 4]	[651; 755]
26	172	92	31	17	2	2	31	3	4	4	482
	[41; 70]	[25; 48]	[13; 32]	[7; 22]	[4; 16]	[2; 11]	[1; 8]	[0; 6]	[0; 4]	[0; 3]	[458; 547]
27	77	95	53	4	27	35	17	3	1	0	454
	[23; 45]	[12; 30]	[7; 21]	[4; 15]	[1; 10]	[0; 7]	[0; 5]	[0; 4]	[0; 3]	[0; 2]	[320; 394]
28	56	60	19	19	27	1	1	0	2	1	259
	[11; 28]	[7; 20]	[3; 14]	[1; 10]	[0; 7]	[0; 5]	[0; 4]	[0; 3]	[0; 2]	[0; 2]	[221; 282]
29	58	33	3	8	5	1	1	1	1	0	183
	[6; 19]	[3; 13]	[1; 9]	[0; 7]	[0; 5]	[0; 4]	[0; 3]	[0; 2]	[0; 2]	[0; 1]	[153; 205]
30	39	22	10	1	6	5	2	3	0	0	158
	[2; 12]	[1; 9]	[0; 6]	[0; 5]	[0; 3]	[0; 2]	[0; 1]	[0; 1]	[0; 1]	[0; 1]	[103; 146]
31	60	44	1	5	3	2	1	0	3	0	189
	[1; 8]	[0; 6]	[0; 4]	[0; 3]	[0; 2]	[0; 2]	[0; 1]	[0; 1]	[0; 1]	[0; 1]	[70; 106]
32	47	3	3	27	0	1	1	5	1	1	121
	[0; 6]	[0; 4]	[0; 3]	[0; 2]	[0; 2]	[0; 1]	[0; 1]	[0; 1]	[0; 1]	[0; 1]	[45; 75]
33	29	4	21	3	5	2	12	1	0	0	116
	[0; 4]	[0; 3]	[0; 2]	[0; 2]	[0; 1]	[0; 1]	[0; 1]	[0; 1]	[0; 0]	[0; 0]	[30; 55]
34	25	31	12	5	2	7	0	1	2	0	126
	[0; 3]	[0; 2]	[0; 2]	[0; 1]	[0; 1]	[0; 1]	[0; 1]	[0; 0]	[0; 0]	[0; 0]	[19; 40]
35	41	11	10	6	8	3	3	0	0	0	112
	[0; 2]	[0; 1]	[0; 1]	[0; 1]	[0; 1]	[0; 1]	[0; 0]	[0; 0]	[0; 0]	[0; 0]	[13; 30]

As Figuras 12 e 13 apresentadas a seguir ilustram os traços das cadeias de Markov dos hiper parâmetros dos modelos ajustados aos casos reais de dengue.

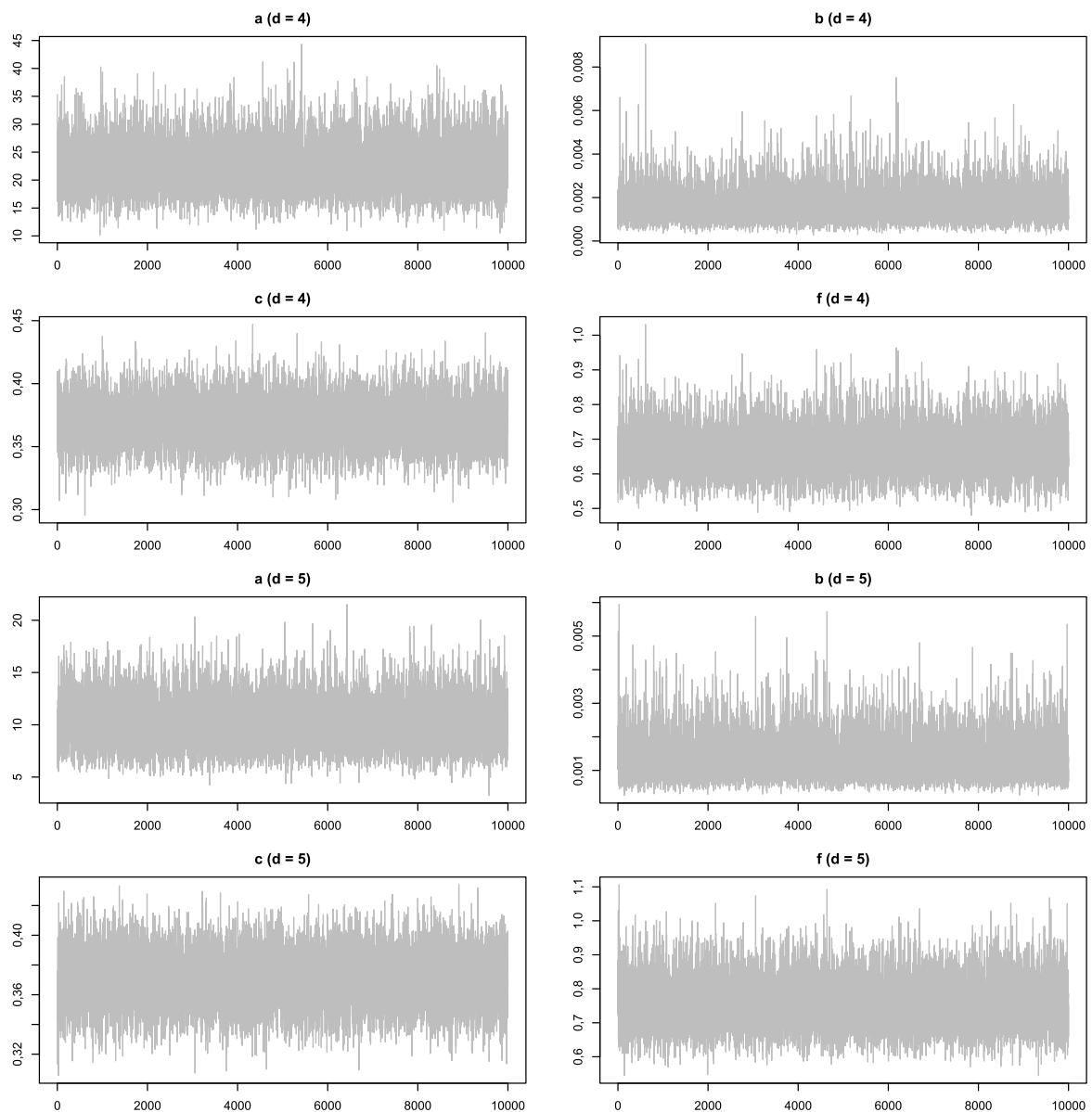
Figura 12 - Traços das cadeias de Markov para os hiper parâmetros do MI



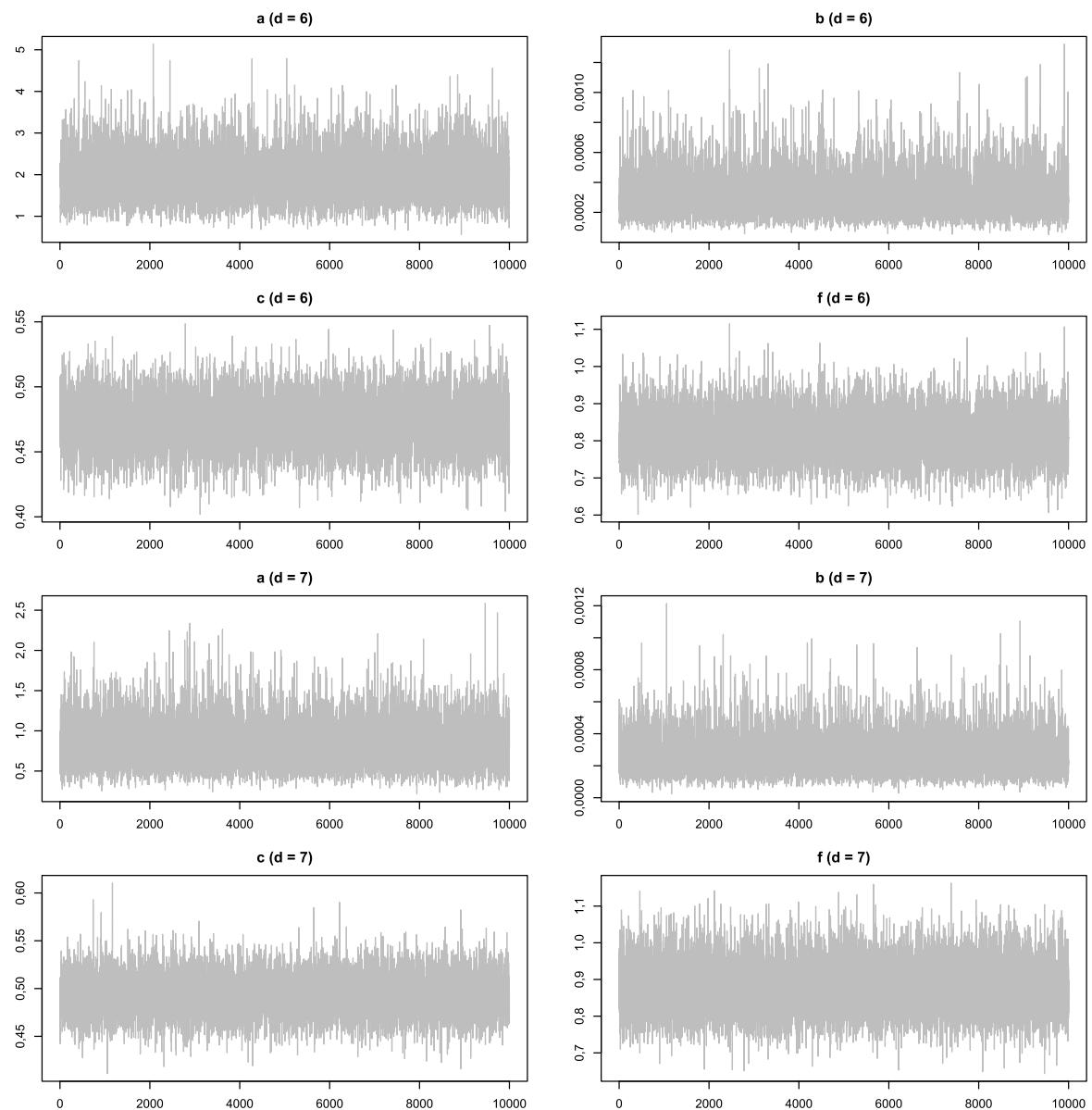
Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO



Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO



Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO



Universidade Federal de Minas Gerais
TRABALHO FINAL DE CONCLUSÃO DE CURSO

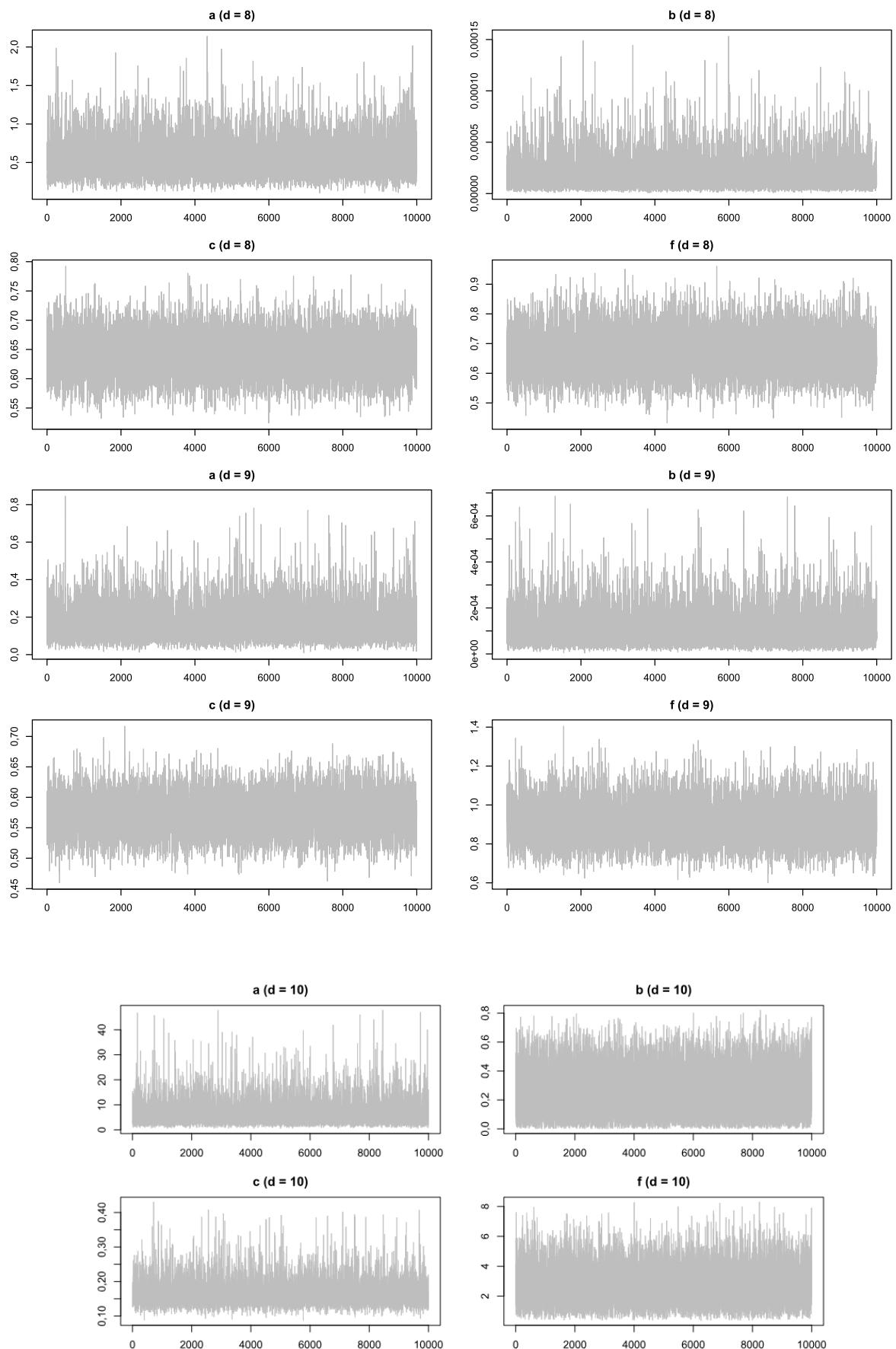
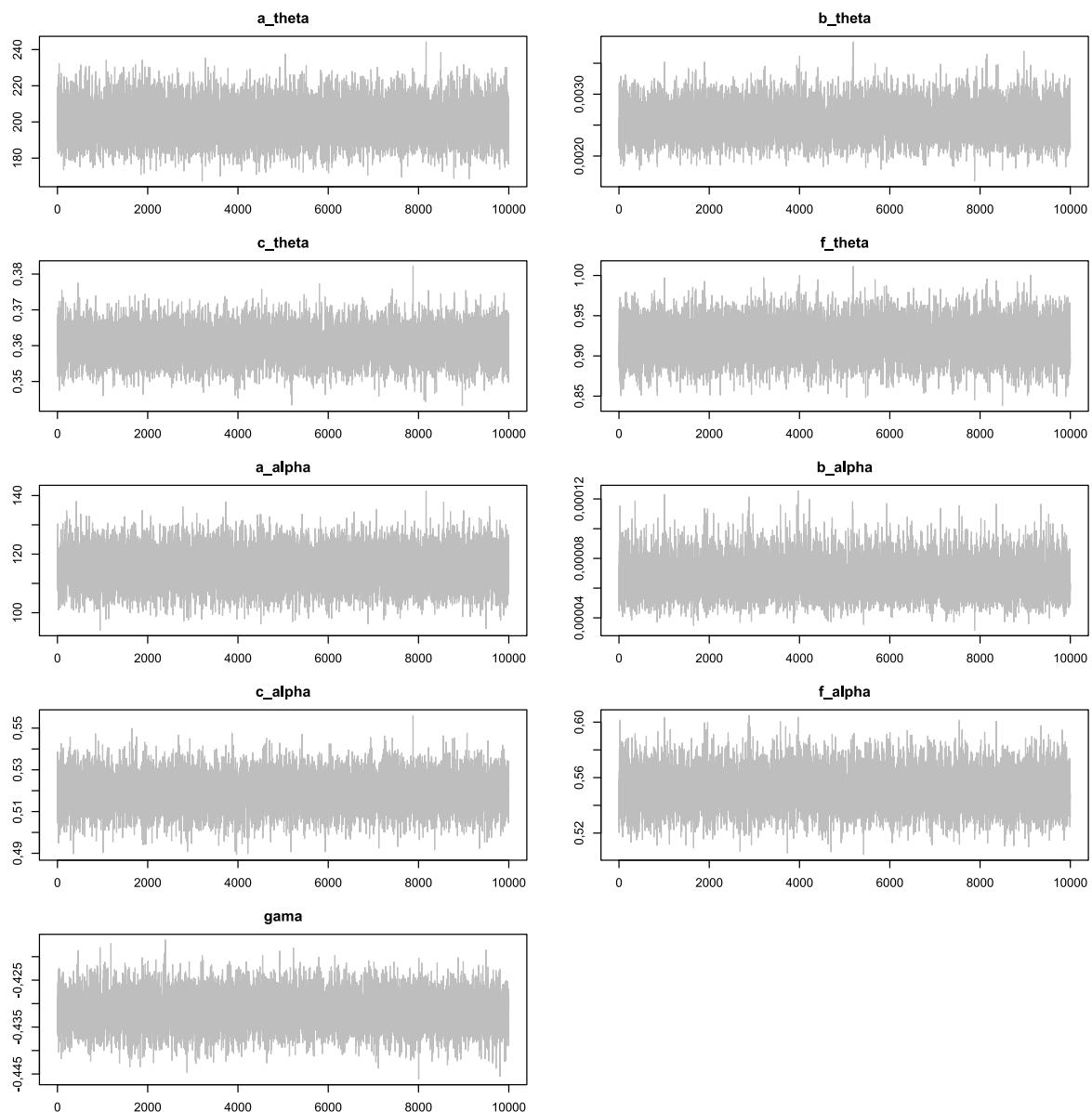


Figura 13 - Traços das cadeias de Markov para os hiper parâmetros do MC



Referências

BASTOS, Leonardo S., et al. **A modelling approach for correcting reporting delays in disease surveillance data.** Statistics in Medicine, 2019.

CHAI, T., DRAXLER, R.R. **Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature.** Geosci. Model Dev., 2014.

GAMERMAN, D. e LOPES, H. F. **Markov Chain Monte Carlo: stochastic simulation for Bayesian inference.** CRC press, 2006.

GAMERMAN, D., et al. **Building a Platform for Data-Driven Pandemic Prediction from Data Modelling to Visualisation - The CovidLP Project.** CRC press, 2022.

GELMAN, A., et al. **Bayesian Data Analysis.** 2013.

MING-HUI C. and QUI-MAN S. **Monte Carlo Estimation of Bayesian Credible and HPD Intervals,** 1998.

NOLAU, Izabel, GAMERMAN, Dani, BASTOS, Leonardo S. **A modeling approach to forecasting data with reporting delay.** Simpósio Nacional de Probabilidade e Estatística, 2022.

STEVEN M. and R. **Elliot Churchill. Principals and Practice of Public Health Surveillance,** 2000.