

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Modelo de espaço de estados para simulação do conteúdo de
matéria orgânica do solo**

Gilberto Volpe Neto

Dissertação de Mestrado do Programa de Mestrado Profissional em
Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Gilberto Volpe Neto

Modelo de espaço de estados para simulação do conteúdo de matéria orgânica do solo

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria.
VERSÃO REVISADA

Área de Concentração: Matemática, Estatística e Computação

Orientador: Prof. Dr. Paulino Ribeiro Villas Boas

USP – São Carlos
Junho de 2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

V931m Volpe, Gilberto
Modelo de espaço de estados para simulação do
conteúdo de matéria orgânica do solo / Gilberto
Volpe; orientador Paulino Ribeiro Villas Boas. --
São Carlos, 2024.
108 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Mestrado Profissional em Matemática, Estatística
e Computação Aplicadas à Indústria) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2024.

1. Modelo. 2. Inferência Bayesiana. 3. Nimble.
I. Ribeiro Villas Boas, Paulino , orient. II.
Título.

Gilberto Volpe Neto

State space model for simulating soil organic matter content

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master – Professional Masters in Mathematics, Statistics and Computing Applied to Industry. *FINAL VERSION*

Concentration Area: Mathematics, Statistics and Computing

Advisor: Prof. Dr. Paulino Ribeiro Villas Boas

USP – São Carlos
June 2024

*Este trabalho é dedicado a todos que buscam a construção de uma sociedade justa,
buscando um futuro sustentável, onde o capital não seja somente um criador de desigualdades.*

AGRADECIMENTOS

Inicialmente, gostaria de expressar minha gratidão à minha família, que foi construída ao longo do meu mestrado. Um agradecimento especial à minha esposa Manuela, por todo o amor e apoio incondicionais. Um abraço carinhoso para a minha primogênita, Olga, e a partilha de muitos momentos especiais com a minha segunda filha, Elis. Desejo muitas aventuras e corridas divertidas com minha caçula, Iza. Amo todas vocês e todo este trabalho é dedicado a garantir um futuro melhor para todas.

Não posso deixar de agradecer também às pessoas com as quais convivi desde o início do mestrado, pois cada uma delas contribuiu para moldar quem eu sou neste momento. Um grande abraço e carinho especial para Bruno Bessa, Sandro Rodrigues, Daniel Rodrigues, Gilberto Sarian, Andrea Aun, Eduardo Pavan, Anderson Forato, Jorge Jamil, Alessandro Laudares, Nelson Watanabe, Daniel Wada, Celina Moraes, Gustavo Machado, Elis Miada, Livia Parra e Adriano Neves.

Por fim, gostaria de agradecer ao professor doutor Paulino Villas-Boas por toda a orientação, paciência e conversas que tivemos ao longo deste trabalho. Sua orientação foi fundamental e, além disso, você se tornou um amigo. Sem você, esse trabalho não teria sido possível.

“Palavra puxa palavra, uma ideia traz outra, e assim se faz um livro, um governo, ou uma revolução, alguns dizem que assim é que a natureza compôs as suas espécies.”
(Machado de Assis)

RESUMO

NETO, G. V. **Modelo de espaço de estados para simulação do conteúdo de matéria orgânica do solo**. 2024. 108 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

A evolução dos diversos efeitos causados pelo aumento da temperatura da Terra está se tornando cada vez mais frequente, tornando necessário implementar novas formas de trabalho para mitigar ao máximo o efeito estufa. Uma maneira de reduzir a emissão de gases do efeito estufa é por meio do crédito de carbono, onde empresas emissoras podem compensar suas emissões por meio de projetos de absorção de gases. A agricultura é uma das principais emissoras, mas através do cultivo adequado das plantas e do manejo correto do solo, ela também pode se tornar uma grande captadora de carbono. No entanto, os métodos atuais para medir a quantidade de carbono no solo são complexos e custosos. Portanto, este estudo tem como objetivo desenvolver um modelo de espaço de estados com dois compartimentos, um de decaimento lento e outro de decaimento rápido, para modelar a quantidade de carbono armazenado no solo. O modelo proposto foi testado em dois conjuntos de dados: um conjunto simulado e outro com medições de três terrenos no Canadá. No caso dos dados simulados, o modelo demonstrou alta eficácia na simulação das cadeias de Markov e na precisão preditiva. O índice de Gelman-Rubin ficou próximo de 1,03, indicando uma boa convergência das cadeias de Markov. Em termos de precisão do modelo, o *MAPE* - *Mean Absolute Percentage Error* foi de apenas 0,61%. No conjunto de dados dos terrenos canadenses, as cadeias de Markov também convergiram com qualidade, com um valor de \hat{R} igual a 1 para os três terrenos. A precisão do modelo, avaliada pelo *MAPE*, foi de 5,29%, 0,017% e 0,021% para os terrenos analisados. Esses resultados evidenciam a eficiência do modelo na simulação e previsão do carbono orgânico no solo, tanto para dados simulados quanto para dados reais, fornecendo uma ferramenta confiável para entender as mudanças e os efeitos das práticas de manejo do solo ao longo do tempo.

Palavras-chave: Matéria orgânica no solo, Modelo de espaço de estados, Método de Monte Carlo com Cadeias de Markov (MCMC), NIMBLE.

ABSTRACT

NETO, G. V. **State space model for simulating soil organic matter content.** 2024. 108 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

The evolution of various effects caused by the increase in Earth's temperature is becoming more frequent, making it necessary to implement new working methods to mitigate the greenhouse effect as much as possible. One way to reduce greenhouse gas emissions is through carbon credits, where emitting companies can offset their emissions through gas absorption projects. Agriculture is one of the main contributors, but through proper plant cultivation and soil management, it can also become a significant carbon sink. However, current methods for measuring soil carbon content are complex and costly. Therefore, this study aims to develop a state-space model with two compartments, one for slow decay and another for fast decay, to model the amount of carbon stored in the soil. The proposed model was tested on two datasets: a simulated dataset and another with measurements from three Canadian terrains. In the case of the simulated data, the model demonstrated high effectiveness in simulating Markov chains and predictive accuracy. The Gelman-Rubin index was close to 1.03, indicating good convergence of the Markov chains. In terms of model accuracy, the mean absolute percentage error (MAPE) was only 0.61%. In the dataset of Canadian terrains, the Markov chains also converged well, with a \hat{R} value of 1 for all three terrains. The model's accuracy, evaluated by MAPE, was 5.29%, 0.017%, and 0.021% for the analyzed terrains. These results highlight the efficiency of the model in simulating and predicting soil organic carbon, both for simulated and real data, providing a reliable tool for understanding changes and the effects of soil management practices over time.

Keywords: Organic matter in soil, State-space model, Markov Chain Monte Carlo (MCMC), NIMBLE.

LISTA DE ILUSTRAÇÕES

Figura 1 – Distribuição dos emissores de gases	26
Figura 2 – Modelo de crédito de carbono	26
Figura 3 – Estrutura de modelo proposto no presente trabalho	29
Figura 4 – Situação de estacionaridade do algoritmos de MCMC	43
Figura 5 – Diagrama do gerador de matéria orgânica no solo	51
Figura 6 – Evolução da quantidade de matéria orgânica total no sistema sem erro de medição versus com erro de medição	52
Figura 7 – Comportamento da evolução do carbono no solo 244003 do Canadá	54
Figura 8 – Comportamento da evolução do carbono no solo 244007 do Canadá	54
Figura 9 – Comportamento da evolução do carbono no solo 244014 do Canadá	55
Figura 10 – Evolução da simulação de MCMC dos compartimentos de matéria orgânica do sistema no início do ciclo	60
Figura 11 – Evolução da simulação de MCMC para entrada de carbono	61
Figura 12 – Evolução da simulação de MCMC para λ_{rapido}	62
Figura 13 – Evolução da simulação de MCMC para λ_{lento}	62
Figura 14 – Evolução da simulação de MCMC para taxa de conversão da materia de compartimento rápido para o lento	62
Figura 15 – Evolução da simulação de MCMC para momento de conversão	63
Figura 16 – Evolução da simulação de MCMC para taxa erro na observação do carbono total	63
Figura 17 – Evolução da simulação de MCMC para o erro na transição entre os estados rápidos	64
Figura 18 – Evolução da simulação de MCMC para o erro na transição entre os estados lentos	64
Figura 19 – Comparação visual entre o total de carbono previsto pelo modelo e o total de entrada	65
Figura 20 – Diferença entre o total de carbono previsto pelo modelo versus a quantidade de entrada	66
Figura 21 – Normal Q-Qplot	66
Figura 22 – Gráfico mostrando o compartimento de ACF e PACF do resíduo do modelo	67
Figura 23 – Relação entre MPG vs Dart	74
Figura 24 – Convergência do modelo de regressão linear entre MPG e Dart	78

Figura 25 – Evolução da simulação de MCMC dos compartimentos de matéria orgânica do sistema para o solo canadense 244003 no início do ciclo	89
Figura 26 – Evolução da simulação de MCMC para entrada de carbono do sistema para o solo canadense 244003	90
Figura 27 – Evolução da simulação de MCMC para λ_{rapido} para o sistema de solo canadense 244003	90
Figura 28 – Evolução da simulação de MCMC para λ_{lento} para o sistema de solo canadense 244003	90
Figura 29 – Evolução da simulação de MCMC para a taxa de conversão para o sistema de solo canadense 244003	90
Figura 30 – Evolução da simulação de MCMC para taxa erro na observação do carbono total para o sistema de solos canadense 244003	91
Figura 31 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados rápidos para o sistema de solos canadense 244003	91
Figura 32 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados lentos para o sistema de solos canadense 244003	91
Figura 33 – Evolução da simulação de MCMC dos compartimentos de matéria orgânica do sistema para o solo canadense 244007 no início do ciclo	92
Figura 34 – Evolução da simulação de MCMC para entrada de carbono do sistema para o solo canadense 244007	92
Figura 35 – Evolução da simulação de MCMC para λ_{rapido} para o sistema de solo canadense 244007	92
Figura 36 – Evolução da simulação de MCMC para λ_{lento} para o sistema de solo canadense 244007	93
Figura 37 – Evolução da simulação de MCMC para a taxa de conversão para o sistema de solo canadense 244007	93
Figura 38 – Evolução da simulação de MCMC para taxa erro na observação do carbono total para o sistema de solos canadense 244007	93
Figura 39 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados rápidos para o sistema de solos canadense 244007	93
Figura 40 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados lentos para o sistema de solos canadense 244007	94
Figura 41 – Evolução da simulação de MCMC para o momento de conversão do carbono do compartimento rápido para o lento no sistema de solo canadense 244007	94
Figura 42 – Evolução da simulação de MCMC dos compartimentos de matéria orgânica do sistema para o solo canadense 244014 no início do ciclo	95
Figura 43 – Evolução da simulação de MCMC para entrada de carbono do sistema para o solo canadense 244014	95

Figura 44 – Evolução da simulação de MCMC para λ_{rapido} para o sistema de solo canadense 244014	95
Figura 45 – Evolução da simulação de MCMC para λ_{lento} para o sistema de solo canadense 244014	96
Figura 46 – Evolução da simulação de MCMC para a taxa de conversão para o sistema de solo canadense 244014	96
Figura 47 – Evolução da simulação de MCMC para taxa erro na observação do carbono total para o sistema de solos canadense 244014	96
Figura 48 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados rápidos para o sistema de solos canadense 244014	96
Figura 49 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados lentos para o sistema de solos canadense 244014	97
Figura 50 – Evolução da simulação de MCMC para o momento de conversão do carbono do compartimento rápido para o lento no sistema de solo canadense 244014	97
Figura 51 – Comparação visual entre o total de carbono previsto pelo modelo e o total de entrada no sistema do solo canadense 244003	98
Figura 52 – Diferença entre o total de carbono previsto pelo modelo versus a quantidade de entrada no sistema do solo canadense 244003	99
Figura 53 – Normal Q-Qplot do sistema de solo canadense 244003	99
Figura 54 – Gráfico mostrando o compartimento de ACF e PACF do resíduo do modelo no sistema de solo canadense 244003	100
Figura 55 – Comparação visual entre o total de carbono previsto pelo modelo e o total de entrada no sistema do solo canadense 244007	101
Figura 56 – Diferença entre o total de carbono previsto pelo modelo versus a quantidade de entrada no sistema do solo canadense 244007	102
Figura 57 – Normal Q-Qplot do sistema de solo canadense 244007	102
Figura 58 – Gráfico mostrando o compartimento de ACF e PACF do resíduo do modelo no sistema de solo canadense 244007	103
Figura 59 – Comparação visual entre o total de carbono previsto pelo modelo e o total de entrada no sistema do solo canadense 244014	104
Figura 60 – Diferença entre o total de carbono previsto pelo modelo versus a quantidade de entrada no sistema do solo canadense 244014	105
Figura 61 – Normal Q-Qplot do sistema de solo canadense 244014	105
Figura 62 – Gráfico mostrando o compartimento de ACF e PACF do resíduo do modelo no sistema de solo canadense 244014	106

LISTA DE ALGORITMOS

Algoritmo 1 – Filtro de Kalman	37
Algoritmo 2 – Metropolis-Hastings	41
Algoritmo 3 – Amostrador de Gibbs	41
Algoritmo 4 – Filtro de partículas Bootstrap	43

LISTA DE ABREVIATURAS E SIGLAS

APF	filtro de partículas auxiliar
EDO	Equação Diferencial Ordinária
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
IF2	filtro de iteração 2
IPCC	Painel Intergovernamental sobre Mudanças Climáticas
MAPE	<i>Mean Absolute Percentage Error</i>
NIMBLE	Numerical Inference for Statistical Models in a Bayesian and Likelihood Framework
ONU	Organização das Nações Unidas
SMC	Sequential Monte Carlo

SUMÁRIO

1	INTRODUÇÃO	25
2	REVISÃO TEÓRICA	31
2.1	Modelos de dinâmica de carbono	31
2.1.1	<i>Balanço de massa</i>	31
2.1.2	<i>Dependência do substrato com a decomposição</i>	32
2.1.3	<i>Taxa de decaimento</i>	33
2.1.4	<i>Transformações da matéria orgânica</i>	33
2.1.5	<i>Efeitos do meio ambiente</i>	34
2.1.6	<i>Interações da matéria orgânica</i>	35
2.2	Modelo de Espaço de Estados	35
2.2.1	<i>Representação do espaço de estado</i>	35
2.2.2	<i>Métodos de Kalman</i>	36
2.3	Inferência Bayesiana	37
2.3.1	<i>Métodos computacionais para Inferência Bayesiana</i>	39
2.3.2	<i>Método de Monte Carlo</i>	39
2.3.2.1	<i>Cadeias de Markov</i>	40
2.3.2.2	<i>Cadeias de Markov de tempo discreto</i>	40
2.3.3	<i>Método de Monte Carlos com Cadeias de Markov - MCMC</i>	40
2.3.3.1	<i>Metropolis-Hastings</i>	41
2.3.3.2	<i>Amostrador de Gibbs</i>	41
2.3.3.3	<i>Filtro de partícula Bootstrap</i>	42
2.3.4	<i>Convergência e avaliação dos métodos MCMC</i>	43
2.4	Modelos para estimação de incertezas	45
3	METODOLOGIA	49
3.1	Derivação da equação do modelo	49
3.2	Base de Dados	50
3.2.1	<i>Base de dados artificial</i>	50
3.2.2	<i>Base de dados Censo Agrícola Canadense</i>	52
3.3	Configuração dos modelos de inferência bayesiana	56
4	RESULTADO	59
4.1	Convergência do algoritmo de MCMC	59

4.1.1	<i>Matéria orgânica de solo simulado</i>	59
4.2	Precisão do modelo	65
4.2.1	<i>Matéria orgânica de solo simulado</i>	65
5	CONCLUSÃO	69
APÊNDICE A	NIMBLE	73
A.0.1	<i>Utilizando o NIMBLE</i>	74
A.0.2	<i>Definir o modelo</i>	75
A.0.3	<i>Definindo o objeto o NIMBLE</i>	75
A.0.4	<i>Compilar o modelo</i>	76
A.0.5	<i>Configurar o MCMC - Monte Carlo Markov Chain</i>	76
A.0.6	<i>Executando o NIMBLE</i>	77
A.0.7	<i>Inferência do modelo</i>	78
APÊNDICE B	CÓDIGOS	81
B.1	Implementação base de dados artificial	81
B.2	Implementação do modelo proposto	83
APÊNDICE C	GRÁFICOS	89
C.1	Convergência do algoritmo de MCMC	89
C.1.1	<i>Matéria orgânica de solo canadense 244003</i>	89
C.1.2	<i>Matéria orgânica de solo canadense 244007</i>	92
C.1.3	<i>Matéria orgânica de solo canadense 244014</i>	95
C.2	Precisão	98
C.2.1	<i>Precisão do modelo do solo canadense 244003</i>	98
C.2.2	<i>Precisão do modelo do solo canadense 244007</i>	101
C.2.3	<i>Precisão do modelo do solo canadense 2440014</i>	104
REFERÊNCIAS		107

INTRODUÇÃO

Em 2021, a Organização das Nações Unidas (ONU) divulgou a versão mais recente do relatório do Painel Intergovernamental sobre Mudanças Climáticas (IPCC), fornecendo evidências de que a atividade humana é diretamente responsável pelo aquecimento da atmosfera, dos oceanos e da terra.

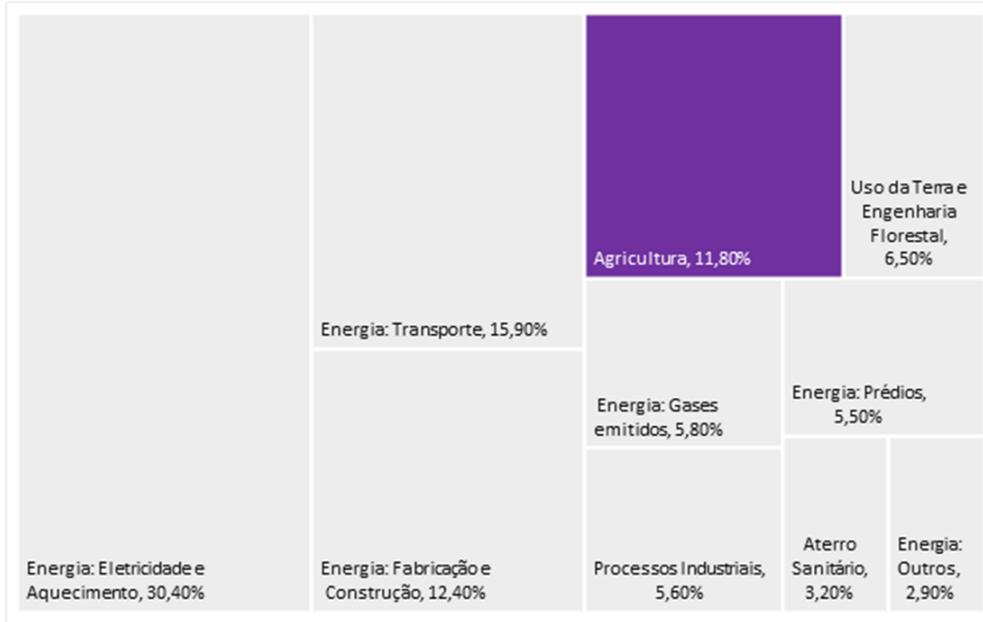
O aquecimento global do planeta Terra acarreta uma variedade de consequências alarmantes, dentre as quais se destaca o incremento na frequência de eventos climáticos extremos, tais como ondas de calor, precipitações intensas, períodos de seca e ciclones tropicais, como apresentados nos artigos de [Konisky, Hughes e Kaylor \(2015\)](#), [Wang, Jiang e Lang \(2017\)](#). Esses eventos representam desafios significativos tanto para a população em geral quanto para os governos, os quais frequentemente carecem de preparo para enfrentar a magnitude e a rapidez com que tais ocorrências se manifestam, como pôde ser constatado com as intensas chuvas ocorridas em Petrópolis, no Rio de Janeiro, em 2022.

A principal causa desse aumento de temperatura, discutido em [WRI-Brasil \(2020\)](#), é a queima de combustíveis fósseis para geração de energia, responsável por 77% das emissões de gases de efeito estufa. A agricultura também desempenha um papel significativo, contribuindo com 12% das emissões, enquanto os demais setores respondem por 11%, conforme ilustrado na [Figura 1](#).

Essas evidências ressaltam a necessidade urgente de ações e políticas que visem reduzir as emissões de gases de efeito estufa, mitigar os impactos das mudanças climáticas e promover a sustentabilidade ambiental. A conscientização e o engajamento global são fundamentais para enfrentar esse desafio e garantir um futuro mais seguro e sustentável para as gerações presentes e futuras.

Para mitigar as emissões de gases na atmosfera, a ONU desenvolveu o Protocolo de Kyoto, que introduziu o mecanismo de crédito de carbono. Esse mecanismo cria um mercado no qual são negociados créditos que representam a redução ou remoção de emissões de dióxido de

Figura 1 – Distribuição dos emissores de gases

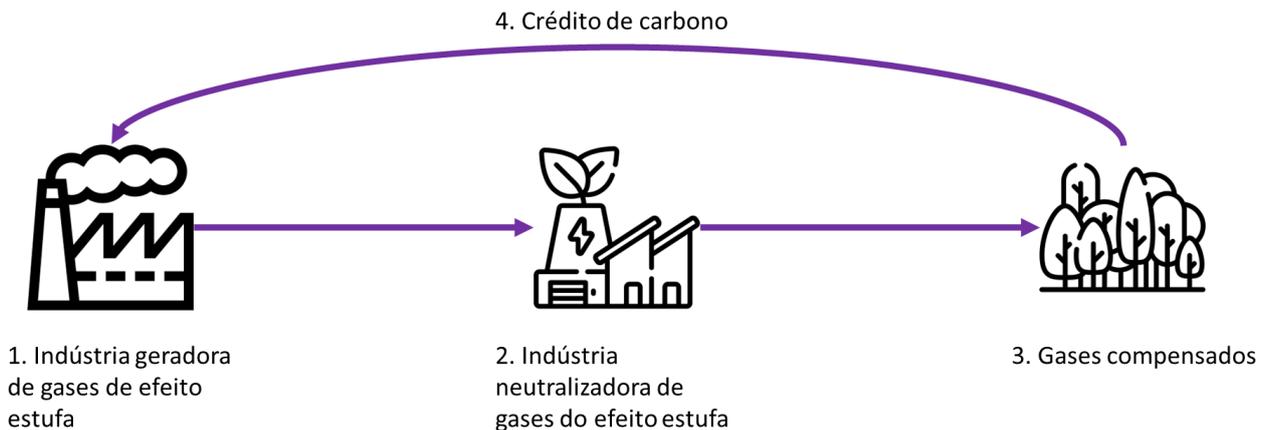
Principais fontes de gases do efeito estufa

carbono (CO₂).

O funcionamento do mercado de crédito de carbono é ilustrado na Figura 2. Nele, empresas que possuem limitações na redução direta de suas emissões de CO₂ têm a opção de adquirir créditos de carbono. Esses créditos são gerados por projetos ou iniciativas que promovem a redução ou remoção de CO₂ da atmosfera, como o reflorestamento ou a preservação de áreas florestais.

Figura 2 – Modelo de crédito de carbono

**Diagrama de crédito de carbono:
compra de sumidouros de gases do efeito estufa**



Ao adquirir esses créditos, as empresas estão compensando suas emissões ao investir em atividades que absorvem ou reduzem o carbono da atmosfera. Isso permite que elas cumpram metas de redução de emissões e contribuam para a mitigação das mudanças climáticas.

Um exemplo prático desse mecanismo é quando uma empresa adquire créditos de carbono para compensar parte de suas emissões. Esses créditos são provenientes de um projeto de reflorestamento, onde árvores são plantadas para absorver CO₂. Dessa forma, a empresa está financiando ações que combatem o desmatamento e contribuem para a redução líquida de emissões de gases de efeito estufa.

O mercado de crédito de carbono desempenha um papel importante no incentivo à adoção de práticas sustentáveis e na promoção de investimentos em projetos ambientais. Ele visa conciliar o desenvolvimento econômico com a preservação do meio ambiente, estimulando a transição para uma economia de baixo carbono.

A agricultura, como o segundo maior emissor de gases de efeito estufa, desempenha um papel importante na captura desses gases, uma vez que o manejo adequado do solo pode reduzir a quantidade de emissões ao longo do tempo. Pesquisas realizadas nos últimos anos pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) têm demonstrado tal fato.

No entanto, a mensuração precisa da quantidade de carbono armazenado no solo ainda é um desafio complexo e limitado em termos de capacidade de área. Atualmente, é necessário instalar vários cilindros para capturar amostras de carbono em diferentes pontos da área de cultivo, que são então medidas periodicamente após a instalação.

Diante deste cenário, surge a necessidade de desenvolver metodologias menos invasivas, baseadas em variáveis de fácil captura, como temperatura média, precipitação, quantidade de toneladas colhidas, número de plantações realizadas na área, tipo de cultivo, entre outros. Isso permitiria estimar de forma mais acessível e abrangente a quantidade de carbono armazenado no solo.

Essa abordagem é importante para promover a popularização do uso de créditos de carbono entre os agricultores, proporcionando-lhes meios mais acessíveis para mensurar e valorizar os benefícios ambientais gerados pelo manejo sustentável do solo. Além disso, a utilização de variáveis mais facilmente obtidas também pode contribuir para a adoção generalizada de práticas agrícolas mais sustentáveis, permitindo uma avaliação mais ampla do impacto ambiental das atividades agrícolas.

Sierra e Müller (2015) desenvolveram um framework baseado em seis princípios fundamentais que incluem o balanço de massa, a dependência do substrato com a decomposição, a taxa de decaimento, as transformações da matéria orgânica, os efeitos do meio ambiente e as interações da matéria orgânica. Essa abordagem visa avaliar diferentes tipos de modelos matemáticos que descrevem a evolução da matéria orgânica ao longo do tempo. O resultado

desse trabalho é expresso pela equação

$$\frac{d\mathbf{C}}{dt} = \mathbf{I}(t) + \mathbf{T}(\mathbf{C}, t) \times \mathbf{N}(\mathbf{C}, t) \times \mathbf{C}(t), \quad (1.1)$$

que sintetiza as interações complexas e dinâmicas envolvidas no ciclo da matéria orgânica do solo.

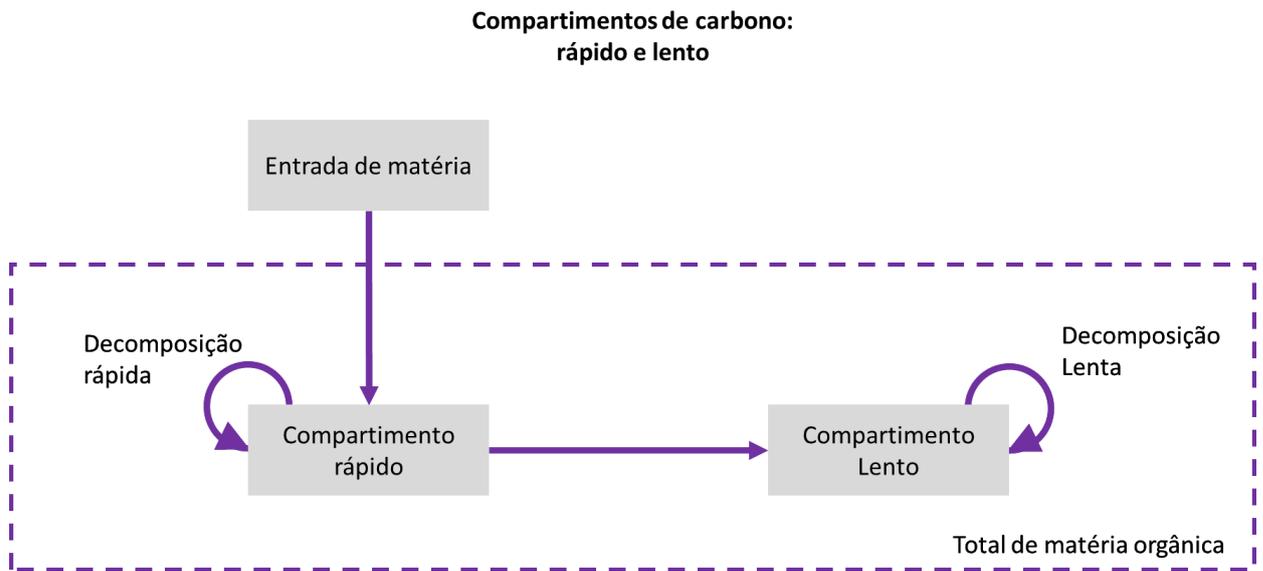
A partir da Equação Diferencial Ordinária (EDO) 1.1, podemos explorar novas abordagens para avaliar a evolução do carbono total no sistema, incluindo uma abordagem de inferência bayesiana. O processo envolve a construção de um espaço de estados que representa diversos compartimentos de carbono e a análise de sua evolução com base em medidas observadas.

O presente trabalho introduz uma nova abordagem na modelagem da absorção de carbono no solo, utilizando inferência bayesiana e modelos de espaço de estados para descrever a dinâmica do carbono ao longo do tempo, conforme ilustrado na Figura 3 e produzindo o conjunto de equações 3.5, onde C_{rapido} e C_{lento} significam a quantidade de carbono no compartimento rápido e lento respectivamente; λ_{rapido} e λ_{lento} significam decaimento em seus respectivos compartimentos e por último, I_{ext} a quantidade de carbono que entra no sistema via fonte externa e I_{int} o que é convertido do carbono rápido para o lento.

$$\begin{aligned} C_{rapido}(t) &= C_{rapido}(t-1) \times e^{-\lambda_{rapido}} + I_{ext} + C_{rapido}(t-1) + \epsilon^* \\ C_{lento}(t) &= C_{lento}(t-1) \times e^{-\lambda_{lento}} + I_{int} + C_{lento}(t-1) + \epsilon^* \\ C_{total}(t) &= C_{lento}(t) + C_{rapido}(t) + \mathbf{v}^* \end{aligned},$$

Para alcançar esse objetivo, esta dissertação está estruturada da seguinte forma: o Capítulo 2 oferece uma revisão teórica dos principais tópicos necessários para o desenvolvimento da dissertação, como a modelagem do carbono no solo e modelo de incertezas. O Capítulo 3 apresenta a metodologia desenvolvida, incluindo a modelagem matemática do modelo proposto, a implementação da simulação do carbono no terreno e, por fim, o uso do *NIMBLE* que é um *framework* em linguagem R para modelagem Bayesiana, para a implementação do modelo. No Capítulo 4, os resultados são apresentados e discutidos. Por fim, o Capítulo 5 traz as conclusões do trabalho, enfatizando as contribuições e possíveis direções futuras de pesquisa.

Figura 3 – Estrutura de modelo proposto no presente trabalho



REVISÃO TEÓRICA

Nas próximas seções, os fundamentos teóricos e conceituais subjacentes ao presente trabalho serão estabelecidos com base em uma seleção cuidadosa de artigos científicos. Essa abordagem, fundamentada em uma base acadêmica sólida, proporcionará um embasamento teórico consistente para a compreensão dos temas abordados, visando sustentar as análises e conclusões apresentadas neste trabalho.

2.1 Modelos de dinâmica de carbono

No artigo *A general mathematical framework for representing soil organic matter dynamics*, Sierra e Müller (2015) desenvolveram um *framework* baseado em seis princípios fundamentais que incluem o balanço de massa, a dependência do substrato com a decomposição, a taxa de decaimento, as transformações da matéria orgânica, os efeitos do meio ambiente e as interações da matéria orgânica para representar a evolução da matéria orgânica no solo.

2.1.1 Balanço de massa

O primeiro ponto fundamental do *framework* é o balanço de massas. Nesse conceito, reconhecemos a conservação da massa total do sistema. Considerando que o solo é um sistema aberto, podemos presumir que toda a massa que entra é direcionada para diversos “reservatórios”, que, no nosso caso, podem ser o ar ou podem se acumular dentro do próprio solo. Essa abordagem permite analisar e compreender a transferência e distribuição da matéria orgânica ao longo do sistema solo-atmosfera, considerando os diferentes fluxos e processos envolvidos. Ao utilizar o balanço de massas como base, podemos desenvolver modelos e estratégias de manejo que buscam otimizar a entrada, saída e armazenamento de matéria orgânica no solo, contribuindo para a sustentabilidade e saúde do ecossistema.

Matematicamente, temos a equação do balanço de massa de carbono:

$$\frac{dC}{dt} = I - O, \quad (2.1)$$

onde:

- $\frac{dC}{dt}$ representa a taxa de variação da massa de carbono no solo ao longo do tempo;
- I são as entradas de carbono no solo, que geralmente incluem raízes mortas, exsudatos radiculares (compostos orgânicos liberados pelas raízes das plantas no solo) e a queda/decomposição de matéria orgânica, como folhas, frutas, flores, restos de plantas e animais, no solo;
- O são as saídas de carbono do solo, que incluem o dióxido de carbono (CO_2) liberado para a atmosfera, o carbono erodido (transportado pela erosão do solo) e a dissolução do carbono da matéria orgânica.

Ao considerar o balanço de massas, podemos avaliar como as entradas e saídas de carbono influenciam o estoque total de carbono no solo, bem como sua dinâmica ao longo do tempo. Isso nos permite compreender os processos que afetam a quantidade de carbono armazenada no solo e desenvolver estratégias para promover a entrada e retenção desse elemento no solo, contribuindo para mitigar as mudanças climáticas e melhorar a saúde dos ecossistemas.

2.1.2 Dependência do substrato com a decomposição

Simplificando a dependência do substrato em uma frase, podemos dizer que não há decomposição sem um substrato adequado. Isso implica na existência de uma função de saída que depende da quantidade de carbono no solo. Podemos expressar essa relação por meio da seguinte equação:

$$O = f(C) \rightarrow O = e^{-k}C = kC, \quad (2.2)$$

onde O representa a taxa de saída de carbono do solo e C é a quantidade de carbono presente no solo.

Essa equação mostra que a quantidade de carbono liberada do solo está relacionada à quantidade de carbono disponível como substrato para a decomposição. A função $f(C)$ descreve a dependência dessa taxa de saída em relação ao estoque de carbono no solo. Essa relação permite entender como a quantidade de carbono no solo influencia a taxa de decomposição e, conseqüentemente, a dinâmica da matéria orgânica no ecossistema do solo. Atualizando a equação 2.1 para contemplar a decomposição da matéria orgânica, obtemos:

$$\frac{dC}{dt} = I - kC. \quad (2.3)$$

2.1.3 Taxa de decaimento

Os componentes do solo não formam uma mistura homogênea em que todas as etapas se comportam quimicamente, biologicamente e fisicamente da mesma forma. Isso leva à consideração de diferentes taxas de decaimento para cada componente, com suas entradas e saídas específicas. Podemos imaginar que o solo pode ser dividido em m compartimentos, cada um com comportamento distinto. Nesse caso, podemos descrever a dinâmica do sistema por meio da seguinte equação:

$$\begin{aligned} \frac{dC_1}{dt} &= I_1 - k_1 C_1 \\ \frac{dC_2}{dt} &= I_2 - k_2 C_2 \\ &\vdots \\ \frac{dC_m}{dt} &= I_m - k_m C_m \end{aligned} \quad (2.4)$$

usando uma notação matricial, podemos expressar a equação 2.4 de forma mais compacta. Definimos a matriz \mathbf{K} como uma matriz diagonal $m \times m$, na qual as taxas de decomposição k_i são dispostas ao longo da diagonal principal. Essa abordagem permite relacionar o solo a um conjunto de reservatórios de carbono interconectados. Nesse caso, podemos escrever a equação como:

$$\frac{d\mathbf{C}}{dt} = \mathbf{I} - \mathbf{K} \cdot \mathbf{C} \quad (2.5)$$

Esta forma matricial, equação 2.5, permite uma representação mais eficiente e generalizada do sistema, facilitando a modelagem e a análise da dinâmica do carbono no solo, considerando as diferentes taxas de decomposição para cada compartimento.

2.1.4 Transformações da matéria orgânica

A equação 2.5 mostra que temos diversos compartimentos com seus decaimentos, entretanto, eles não estão conectados entre si, impedindo que o mesmo carbono passe por diferentes taxas de decaimento, portanto, podemos reconsiderar o balanço de massa para levar tal fato em conta, seja i o i -ésimo compartimento:

$$\frac{dC_i}{dt} = I_{ext,i} + I_{int,i} - O_{ext,i} - O_{int,i} \quad (2.6)$$

os elementos “ext”, que estão fora do sistema, e “int”, que estão dentro do sistema, permitem interpretar a equação 2.6 da seguinte maneira: a taxa de decaimento do carbono no sistema é determinada pelas entradas provenientes de outros reservatórios, representadas por I_{int} , e pelas entradas de massa provenientes de fora do sistema, representadas por I_{ext} .

As saídas do sistema incluem a respiração com a liberação de CO_2 , representada por O_{ext} , e também a migração de carbono entre os reservatórios, representada por O_{int} . Essa equação descreve o balanço de massa do sistema, levando em consideração as transferências de carbono

entre diferentes componentes, tanto internos quanto externos ao sistema, permitindo uma análise mais abrangente da dinâmica do carbono no solo.

Precisamos reconsiderar o balanço de massa de todo o sistema que é representado pela equação 2.7:

$$\sum_i^m \frac{dC_i}{dt} = \sum_i^m I_i - \sum_i^m O_i, \quad (2.7)$$

além disso, precisamos considerar que para cada reservatório i as entradas internas são iguais às saídas de todo reservatório j , tal que $\forall j \neq i$. Portanto, temos:

$$I_{int,i} = \sum_{j,j \neq i}^m D_{i,j} \cdot O_{int,j}. \quad (2.8)$$

Na forma matricial temos a equação 2.9, onde D é uma matriz de uns indicando a direção de movimento do carbono, por exemplo, se $D_{2,3} = 1$, indica o deslocamento da massa do reservatório 2 para o reservatório 3,

$$I_{int} = D \times O_{int}, \quad (2.9)$$

podemos melhorar ainda mais a transferência de massa entre os reservatórios se construirmos uma matriz com diagonal igual a -1 e todos os demais coeficientes representando as proporções de finalização da transferência de carbono. Por exemplo, se o carbono sair do reservatório 2, em seguida for para o sétimo e, por fim, for para o terceiro, então a matriz T considerará apenas o início e o fim da transferência. Portanto, ao final, teremos a seguinte equação matricial:

$$\frac{d\mathbf{C}}{dt} = \mathbf{I} + \mathbf{T} \cdot \mathbf{K} \cdot \mathbf{C}. \quad (2.10)$$

2.1.5 Efeitos do meio ambiente

Uma das formas de considerar os efeitos do meio ambiente com relação a taxa de decomposição é construir uma função, $\xi(t)$, que dependa de diversas covariáveis, X_i como temperatura, umidade, presença de animais, incidência de chuvas entre outros e aplicar como efeito multiplicador nas taxas de decomposição, como na equação:

$$\frac{d\mathbf{C}}{dt} = \mathbf{I} + \xi(t) \cdot \mathbf{T} \cdot \mathbf{K} \cdot \mathbf{C}, \quad (2.11)$$

entretanto, para que isso seja verdade, estamos considerando que toda a transformação é constante e igual para todos os substratos, o que pode não ser verdadeiro. Portanto, uma forma de aprimorar a equação 2.11 é considerar que todos os elementos evoluem no tempo, não sendo mais constante, como na equação:

$$\frac{d\mathbf{C}(t)}{dt} = \mathbf{I}(t) + \mathbf{T}(t) \cdot \mathbf{K}(t) \cdot \mathbf{C}(t). \quad (2.12)$$

2.1.6 Interações da matéria orgânica

Diversos estudos recentes têm demonstrado que existem diferentes formas de interação entre a matéria orgânica do solo, o que resulta em efeitos variados nas taxas de decomposição e no substrato da matéria. Uma maneira de modelar esses efeitos é por meio da equação 2.13, em que a função de transferência T depende da matéria orgânica presente. Além disso, a equação das taxas de decomposição, representada por K , foi convertida em uma função N , indicando que não se trata apenas do processo de decaimento simples, sendo também uma função da quantidade de matéria orgânica no sistema,

$$\frac{d\mathbf{C}(t)}{dt} = \mathbf{I}(t) + \mathbf{T}(\mathbf{C}, t) \cdot \mathbf{N}(\mathbf{C}, t) \cdot \mathbf{C}(t). \quad (2.13)$$

2.2 Modelo de Espaço de Estados

Um modelo de espaço de estados, conforme livros de Brockwell e Davis (2016), Triantafyllopoulos (2021), é um modelo matemático-estatístico amplamente utilizado em diversas áreas, incluindo engenharia de controle. Esse modelo é especialmente útil para inferir informações que não podem ser medidas diretamente. Um exemplo prático é a estimação da posição de um caminhão com base em medidas realizadas por um satélite.

No contexto de um modelo de espaço de estados, consideramos que o sistema em estudo pode ser descrito por duas componentes principais: as variáveis de estado e as variáveis de observação. As variáveis de estado representam as quantidades desconhecidas que descrevem o sistema, enquanto as variáveis de observação são as medidas ou observações disponíveis que temos.

No caso do exemplo do caminhão, as variáveis de estado podem incluir a posição atual, a velocidade e a aceleração do veículo. No entanto, essas variáveis de estado não podem ser medidas diretamente por um satélite. Em vez disso, o satélite fornece observações indiretas, como a distância entre o satélite e o caminhão ou a direção em relação a uma referência.

O objetivo do modelo de espaço de estados é utilizar as observações disponíveis para inferir as informações desconhecidas do sistema, ou seja, as variáveis de estado. Assim, o modelo de espaço de estados permite estimar quantidades não diretamente observáveis com base em informações disponíveis. Ele é amplamente utilizado na engenharia de controle, mas também encontra aplicações em várias outras áreas, como economia, biologia e ciência da computação. Essa abordagem estatística oferece uma maneira poderosa de lidar com a incerteza e obter estimativas confiáveis de variáveis de interesse em sistemas complexos.

2.2.1 Representação do espaço de estado

A formulação matemática do modelo de espaço de estados consiste em duas equações fundamentais: a equação de observação e a equação de estados. Essas equações descrevem a

evolução do sistema e a relação entre as variáveis observadas e as variáveis não observadas.

A equação de observação relaciona as variáveis observadas, denotadas por Y_t , com as variáveis não observadas, também conhecidas como estados latentes ou ocultos, representadas por X_t . Essa equação estabelece como as observações são geradas a partir dos estados latentes, levando em consideração o possível ruído presente nas medições indiretas. Em sua forma geral, a equação de observação pode ser expressa da seguinte maneira:

$$Y_t = C_t X_t + v_t^*$$

Nessa equação, C_t é a matriz de transformação que mapeia os estados latentes para as observações, representando o modelo que descreve a relação entre o sistema real e as medições. $v_t^* \sim \mathcal{N}(0, R^*)$ é o termo de ruído que incorpora as incertezas associadas à medição indireta.

Por sua vez, a equação de estados descreve a atualização dos estados latentes ao longo do tempo. Ela relaciona o estado no tempo $t + 1$, denotado por X_{t+1} , com o estado no tempo t , X_t , considerando a possível presença de ruído na evolução do sistema. Essa equação pode ser escrita de forma geral como:

$$X_{t+1} = A_t X_t + B_t u_t + \varepsilon_t^*$$

Nessa equação, A_t é a matriz de atualização que determina como os estados evoluem ao longo do tempo, incorporando o modelo dinâmico do sistema. $\varepsilon_t^* \sim \mathcal{N}(0, Q^*)$ é o termo de ruído que captura as incertezas inerentes à evolução do sistema e que não podem ser adequadamente modeladas e B_t é o vetor de controle de entradas e u_t são as entradas do sistema.

2.2.2 Métodos de Kalman

Quando lidamos com modelos de espaço de estados, existem três possibilidades de inferência que podem ser aplicadas, dependendo do contexto e dos objetivos da análise:

1. **Predição:** Com base nos valores observados de Y_0, Y_1, \dots, Y_{t-1} é possível realizar uma estimativa dos próximos valores de $Y_t, Y_{t+1}, \dots, Y_{t+h}$, onde h representa o número de passos futuros a serem previstos. Essa abordagem permite projetar e antecipar as tendências e padrões futuros com base nas informações disponíveis até o momento $t - 1$.
2. **Filtragem:** Através dos valores observados Y_0, Y_1, \dots, Y_t é possível inferir o comportamento do espaço latente, que representa as variáveis não observáveis do sistema. Essa inferência nos permite acompanhar a evolução e o comportamento do sistema ao longo do tempo, mesmo que as variáveis latentes não sejam diretamente observáveis
3. **Suavização:** Com base nos valores observados de Y_0, Y_1, \dots, Y_n , onde $n > t$, é possível aplicar técnicas de suavização para aprimorar a estimação das medidas obtidas, resultando em um ajuste mais preciso. Essa abordagem é particularmente relevante para melhorar a estimação da posição do caminhão mencionado no início deste capítulo, utilizando as

observações disponíveis ao longo do tempo para retroativamente refinar e suavizar as estimativas da trajetória do veículo

Para resolver esses diversos problemas, utilizamos o algoritmo do filtro de Kalman quando o sistema é linearmente separável. No entanto, quando o sistema apresenta não linearidades, outras técnicas são necessárias, como discutido no capítulo sobre inferência bayesiana e filtro de partículas.

O algoritmo de Kalman é uma abordagem eficaz para a resolução das equações de espaço de estado em sistemas lineares. No Algoritmo 1, é apresentado o pseudocódigo do filtro de Kalman e ele possui duas etapas: predição e atualização.

Etapa de predição, onde A_t^T é a matriz transposta de A_t

$$\begin{aligned} X_t^{t-1} &= A_t * \hat{X}_{t-1} + Bu_t \\ P_t^{t-1} &= A_t \hat{P}_{t-1} A_t^T + Q^* \end{aligned}$$

Etapa de atualização:

$$\begin{aligned} K_t &= P_t^{t-1} C' (C P_t^{t-1} C' + R^*)^{-1} \\ \hat{X}_t &= X_t^{t-1} + K_t (Y_t - C X_t^{t-1}) \\ \hat{P}_t &= (I - K_t C) P_t^{t-1} \end{aligned}$$

O elemento K é conhecido como ganho de Kalman. R^* e \hat{P}_t são as matrizes de covariâncias do ruído e do processo, respectivamente. \hat{X}_t é o nosso melhor estimador do espaço latente X_t , e I é a matriz identidade.

Algoritmo 1 – Filtro de Kalman

- 1: inicialize o valor de $\hat{X}_0 = x_0, \hat{P}_0 = Q^*$
 - 2: **para** $t = 1, 2, \dots$ **faça**
 - 3: $X_t^{t-1} = A_t * \hat{X}_{t-1} + Bu_t$
 - 4: $P_t^{t-1} = A_t \hat{P}_{t-1} A_t^T + Q^*$
 - 5: $K_t = P_t^{t-1} C' (C P_t^{t-1} C' + R^*)^{-1}$
 - 6: $\hat{X}_t = X_t^{t-1} + K_t (Y_t - C X_t^{t-1})$
 - 7: $\hat{P}_t = (I - K_t C) P_t^{t-1}$
 - 8: **fim para**
-

2.3 Inferência Bayesiana

No modelo Bayesiano de inferência, conforme material desenvolvido por [Blevins \(2023\)](#), [Triantafyllopoulos \(2021\)](#), [Stata \(2016\)](#), todas as quantidades desconhecidas são tratadas como

variáveis aleatórias, diferentemente da abordagem Frequentista que considera os parâmetros desconhecidos como constantes. Para trabalharmos com o método bayesiano, precisamos considerar três etapas em nosso processo, entendendo que θ é o parâmetro desconhecido.

1. Ao selecionar a distribuição de $f(\theta)$, também conhecida como distribuição a priori, estamos levando em consideração nossos conhecimentos a priori sobre o parâmetro θ antes de interagir com os dados. Um exemplo prático disso é a distribuição de uma moeda, onde, a priori, consideramos que a moeda é honesta e, portanto, as probabilidades de obter cara (H) ou coroa (T) são iguais, ou seja, $P(H) = P(T) = \frac{1}{2}$. Nesse contexto, a escolha da distribuição a priori reflete nossas suposições e conhecimentos prévios sobre o parâmetro em questão.
2. Escolhemos um modelo estatístico, $f(x|\theta)$, que reflete nossas crenças sobre como os dados x impactam nosso parâmetro θ ;
3. Após observarmos os dados X_1, X_2, \dots, X_n , atualizamos nossas crenças e calculamos a distribuição a posteriori dada por $f(\theta|X_1, X_2, \dots, X_n)$.

A terceira etapa de cálculo apresentada pode ser expandida considerando o Teorema de Bayes, transformando a descrição em uma equação mais pragmática:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}$$

Somado ao Teorema de Bayes, se considerarmos também que as diversas observações são independentes entre si e igualmente distribuídas, temos que fazer a seguinte modificação na equação acima:

$$f(X_1, X_2, \dots, X_n|\theta) = \prod f(X_i|\theta) = \mathcal{L}_n(\theta)$$

$$f(\theta|X_1, X_2, \dots, X_n) = \frac{\prod f(X_i|\theta)f(\theta)}{\int \prod f(X_i|\theta)f(\theta)d\theta} = \frac{\mathcal{L}_n(\theta)f(\theta)}{c_n} \propto \mathcal{L}_n(\theta)f(\theta)$$

Portanto, a distribuição a posteriori é proporcional ao produto da verossimilhança pelos valores da distribuição a priori. Em alguns casos, é possível encontrar a distribuição a posteriori analiticamente. No entanto, na maioria dos casos práticos do dia-a-dia, a solução analítica não é viável, tornando necessário o uso de métodos computacionais para obter a distribuição a posteriori e realizar as inferências necessárias. Esses métodos computacionais, como o MCMC (Markov Chain Monte Carlo), permitem aproximar a distribuição a posteriori e fornecer uma base sólida para realizar análises estatísticas e inferências sobre o modelo.

2.3.1 Métodos computacionais para Inferência Bayesiana

Uma das principais técnicas para estimação da posteriori são os métodos de Monte Carlo e Cadeias de Markov, que são duas técnicas combinadas para a resolução de equações que não são viáveis de serem resolvidas analiticamente.

2.3.2 Método de Monte Carlo

Os métodos de Monte Carlo são um conjunto de técnicas estatísticas que possibilitam o cálculo de estruturas complexas através da amostragem de valores. Eles são especialmente úteis quando enfrentamos problemas nos quais a solução analítica não é viável, como a resolução de integrais complicadas. Nesses casos, os métodos de Monte Carlo permitem aproximar esses cálculos complexos por meio da geração de amostras aleatórias, permitindo estimar valores numéricos ou realizar inferências estatísticas. Essas técnicas são amplamente utilizadas em diversas áreas, incluindo ciências, engenharia, finanças e pesquisa operacional, entre outras, onde a natureza estocástica ou a complexidade dos problemas inviabiliza abordagens analíticas diretas.

Vamos considerar um exemplo para ilustrar o método de Monte Carlo. Suponha que desejamos estimar o valor esperado θ de uma variável aleatória X , ou seja, $\theta = E(X)$. Se conseguirmos simular várias realizações independentes de X , em uma sequência de n simulações, teremos os valores X_1, X_2, \dots, X_n .

Podemos utilizar a média amostral $\bar{X} = \frac{1}{n} \sum X_i$ como um estimador para θ . De acordo com a Lei dos Grandes Números, à medida que o número de simulações aumenta, a média amostral se aproxima do valor esperado θ . Além disso, o valor esperado e a variância da média amostral são dados por:

$$E(\bar{X}) = \frac{1}{n} \sum X_i = \theta$$
$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Considerando o Teorema Central do Limite, para valores grandes de n , a média amostral \bar{X} segue uma distribuição aproximadamente normal, com a variabilidade dada por $\frac{\sigma}{\sqrt{n}}$. Isso significa que, à medida que o tamanho da amostra aumenta, a média amostral tenderá a se concentrar em torno do valor verdadeiro θ . Dessa forma, o método de Monte Carlo nos permite estimar o valor esperado θ usando a média amostral, aproveitando a convergência da média amostral para o valor verdadeiro à medida que o tamanho da amostra aumenta.

2.3.2.1 Cadeias de Markov

Processos estocásticos são uma classe de modelos nos quais estamos interessados em modelar uma variável aleatória ao longo do tempo. E no presente trabalho, estamos interessados nos processos estocásticos do tipo Cadeias de Markov que têm como principal propriedade a *memorylessness*. Trazendo essa palavra para a realidade do dia a dia, o estado atual do sistema contém toda a informação do passado. Além disso, podemos classificar as cadeias de Markov em tempo discreto, onde a mudança de estados é bem definida, e tempo contínuo, onde as mudanças de estados são dadas em saltos.

2.3.2.2 Cadeias de Markov de tempo discreto

As cadeias de Markov do tipo de tempo discreto são caracterizadas por terem um espaço de estados $X = S_1, S_2, \dots, S_k$, com uma distribuição inicial μ e uma matriz estocástica $P(S_j, S_k) = P(X_n = S_k | X_{n+1} = S_j)$. A principal característica das cadeias de Markov, no caso discreto é dada por:

$$Pr(X_{n+1} = x | X_1 = x_1, \dots, X_n = x_n) = Pr(X_{n+1} = x | X_n = x_n)$$

Adicionalmente a essa definição, temos algumas propriedades das cadeias de Markov em tempo discreto, tais como:

- **Estrutura de classe:** Se de um estado S_k chega-se a um outro estado S_j e vice-versa, temos então um conjunto chamado de classe e como veremos, todas as propriedades de Markov são válidas para a classe;
- **Recorrência e Transitoriedade:** uma cadeia de markov é recorrente quando temos que um estado é sempre visitado e transitoriedade quando temos um estado que é visitado somente um única vez e depois disso nunca mais. Ambas propriedades são de classes, portanto, definindo para um unico estado, temos para a cadeia toda
- **Medidas invariantes:** uma medida será invariante na cadeia de Markov se $\mu P = \mu$, se μ for uma probabilidade, então, a distribuição será invariante
- **Convergência:** em alguns casos temos que cadeias que rodem até um tempo suficientemente grande entram em equilíbrio e cadeia terá uma distribuição invariante que não depende do estado inicial.

2.3.3 Método de Monte Carlos com Cadeias de Markov - MCMC

Para obter as distribuições a posteriori necessárias para inferências bayesianas, é essencial recorrer a métodos computacionais que combinem cadeias de Markov e métodos de Monte Carlo.

Nessa categoria, destacam-se os algoritmos Metropolis-Hastings, amostrador de Gibbs e filtro de partículas Bootstrap. Essas abordagens computacionais são fundamentais para a resolução dessas distribuições, permitindo a obtenção de estimativas e análises mais precisas dentro do contexto bayesiano.

2.3.3.1 Metropolis-Hastings

O algoritmo de Metropolis-Hastings é uma técnica usada para gerar sequências de valores de uma distribuição alvo π . Em cada etapa da execução, um valor y_n é gerado com base no valor anterior y_{n-1} , usando uma probabilidade de transição denominada *alpha*. O novo valor proposto é então aceito ou rejeitado com base em critérios definidos pelo algoritmo. Ao final da execução, o algoritmo produz uma cadeia de Markov reversível em relação à distribuição estacionária desejada π . Essa distribuição π representa a distribuição alvo que pretendemos estimar, e o algoritmo garante que a cadeia de Markov converge para essa distribuição ao longo do tempo. Abaixo está o pseudo-código do algoritmo de Metropolis-Hastings 2:

Algoritmo 2 – Metropolis-Hastings

- 1: inicialize o valor de y_0
 - 2: **para** $iteration = 1, 2, \dots$ **faça**
 - 3: $y' \sim q(y'|y_{n-1})$
 - 4: $r = \frac{\pi(y')q(y_{n-1}|y')}{\pi(y_{n-1})q(y'|y_{n-1})}$
 - 5: $u \sim U[0, 1]$
 - 6: **se** $u \leq r$ **então** $y_n = y'$
 - 7: **senão** $y_n = y_{n-1}$
 - 8: **fim se**
 - 9: **fim para**
-

2.3.3.2 Amostrador de Gibbs

O amostrador de Gibbs é um método de amostragem que se baseia no algoritmo de Metropolis-Hastings, mas é especialmente adequado para lidar com distribuições conjuntas multivariadas. Ele permite obter amostras de uma distribuição conjunta conhecida, utilizando as distribuições condicionais de cada variável dada as outras. O pseudocódigo do amostrador, 3 de Gibbs, mostrando como o algoritmo funciona é:

Algoritmo 3 – Amostrador de Gibbs

- 1: inicialize os valores de $Y_0 = (y_1, \dots, y_k)$
 - 2: **para** $iteration = 1, 2, \dots$ **faça**
 - 3: Escolha um índice i aleatoriamente da sequência $1, \dots, k$
 - 4: Amostre uma variável onde $y'_i = q(y'_i|y_{-i})$ onde y_{-i} são todas as variáveis em Y exceto pela variável y_i
 - 5: $Y_N = (y_1, \dots, y_k)$
 - 6: **fim para**
-

O amostrador de Gibbs é eficaz para obter amostras de distribuições conjuntas multivariadas, desde que as distribuições condicionais sejam conhecidas ou possam ser calculadas. É importante ressaltar que a convergência da cadeia de Markov gerada pelo amostrador de Gibbs para a distribuição alvo pode ser lenta, e técnicas adicionais podem ser necessárias para melhorar a eficiência do algoritmo, como o descarte de amostras iniciais (*burn-in*) e o *thinning*.

2.3.3.3 Filtro de partícula Bootstrap

Em diversas situações do dia a dia, as observações podem depender do momento em que foram coletadas. Por exemplo, o valor de uma ação ao longo do dia, a temperatura de resfriamento de um metal após a usinagem, a contagem de veículos que passam ao longo de uma avenida, o crescimento de células não saudáveis ao longo do mês em pacientes com câncer, entre outros casos. Nessas circunstâncias, a ordem das observações é importante, o que requer considerar essa natureza ao construir métodos de inferência bayesiana. É aí que entram os métodos de filtro de partículas, também conhecidos como SMC (Sequential Monte Carlo).

O filtro de partículas é um método utilizado para estimar a distribuição a posteriori em um modelo sequencial, no qual as observações são recebidas em uma ordem específica. Ele aborda o desafio de atualizar a distribuição posterior de forma recursiva à medida que novas observações são disponibilizadas. O filtro de partículas emprega uma abordagem Monte Carlo, utilizando uma coleção de partículas para representar a distribuição a posteriori.

No início, as partículas são geradas a partir da distribuição a priori. À medida que novas observações são recebidas, as partículas são atualizadas por meio de um processo de amostragem e ponderação. As partículas que estão mais de acordo com as observações têm maiores pesos, enquanto as menos consistentes possuem pesos menores. O filtro de partículas realiza uma amostragem sequencial, em que as partículas são ajustadas e propagadas de acordo com o modelo do sistema e as observações recebidas.

Um dos principais algoritmos utilizados na implementação do filtro de partículas é o filtro de Bootstrap, algoritmo 4, que incorpora uma etapa de reamostragem. Essa etapa tem como objetivo eliminar partículas com baixa importância ponderada.

Após a atualização das partículas com base nas observações, cada partícula é atribuída a um peso que reflete sua importância relativa. As partículas que se ajustam melhor às observações têm pesos maiores, enquanto as menos consistentes têm pesos menores. No entanto, é possível que algumas partículas tenham pesos muito pequenos, o que implica em sua contribuição negligenciável para a representação da distribuição a posteriori.

A etapa de reamostragem do filtro de Bootstrap resolve esse problema. Ela consiste em selecionar, com base nos pesos, um conjunto de partículas para formar uma nova geração de partículas. Essa seleção é realizada de forma aleatória, mas ponderada pelos pesos das partículas, de modo que aquelas com pesos maiores têm mais chances de serem selecionadas várias vezes,

enquanto as de pesos menores são descartadas. Dessa forma, a nova geração de partículas é composta principalmente por aquelas que melhor representam a distribuição a posteriori.

É importante ressaltar que a reamostragem pode introduzir um certo grau de ruído no filtro de partículas, uma vez que a nova geração de partículas é formada por uma amostra finita das partículas antigas. No entanto, em muitos casos, esse efeito é mitigado por um número suficiente de partículas e por uma etapa de atualização adequada.

Algoritmo 4 – Filtro de partículas Bootstrap

```

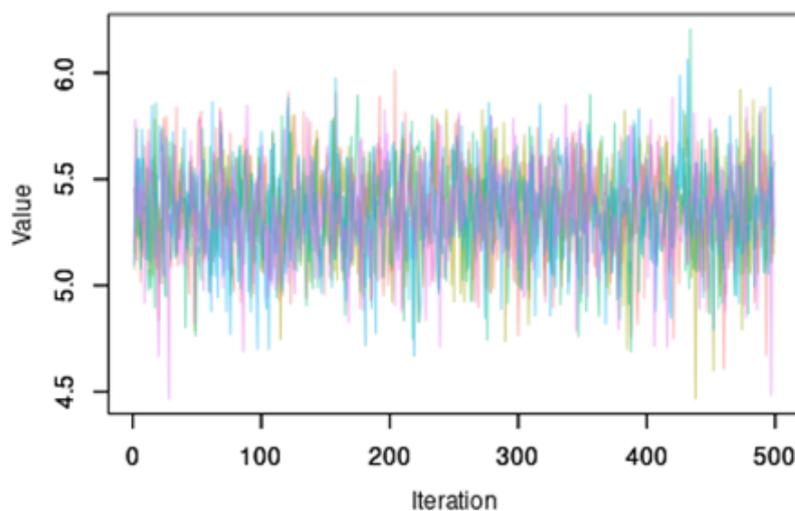
1: para  $i = 1, 2, \dots$  faça
2:   inicialize os valores de  $y_0^i \sim p(y_0)$ 
3: fim para
4: para  $t = 1, 2, \dots$  faça
5:   para  $i = 1, 2, \dots$  faça
6:      $\tilde{y}_t^i \sim p(y_t | t_{t-1}^i)$ 
7:      $\tilde{y}_{0:t}^i \leftarrow (y_{0:t-1}^i, \tilde{y}_t^i)$ 
8:   fim para
9:   para  $i = 1, 2, \dots$  faça
10:     $\tilde{w}_t^i \sim p(z_t | \tilde{y}_t^i)$ 
11:  fim para
12:  para  $i = 1, 2, \dots$  faça
13:    Amostre  $N$  elementos com reposição do conjunto  $\{\tilde{y}_{0:t}^i\}$  com os pesos  $\{\tilde{w}_t^i\}$ 
14:  fim para
15: fim para

```

2.3.4 Convergência e avaliação dos métodos MCMC

Para avaliar a qualidade de uma simulação, é importante que a sequência gerada seja estacionária. Isso significa que, ao longo do tempo, a distribuição dos valores amostrados não muda significativamente e segue uma distribuição de equilíbrio. A figura 4 ilustra esse conceito.

Figura 4 – Situação de estacionaridade do algoritmos de MCMC



A estacionariedade é crucial para garantir as propriedades desejadas das cadeias de Markov utilizadas nas inferências bayesianas. Quando a sequência está em modo "caminho aleatório", as amostras geradas são independentes e refletem adequadamente a distribuição alvo.

Além da inspeção visual da sequência gerada, existem métodos estatísticos para avaliar a convergência das cadeias de Markov simuladas. Um desses métodos é o teste de Gelman-Rubin, que compara a variabilidade entre diferentes cadeias e dentro da mesma cadeia.

No teste de Gelman-Rubin, várias cadeias independentes são simuladas a partir de distribuições iniciais diferentes. Em seguida, é calculada a proporção de variância entre as cadeias em relação à variância dentro de cada cadeia. Se essa proporção for próxima de 1, significa que as cadeias estão convergindo para a mesma distribuição.

Quando a diferença entre os valores amostrados entre as diferentes cadeias e dentro da mesma cadeia é pequena, podemos considerar que a convergência foi alcançada. Isso indica que a simulação produziu resultados consistentes e confiáveis.

O cálculo do teste de convergência de Gelman-Rubin pode ser realizado da seguinte forma. Suponha que temos M cadeias de Markov, cada uma com um comprimento de N passos, e estamos estimando o parâmetro θ . Seja $\hat{\theta}_m$ a média e $\hat{\sigma}_m^2$ a variância da m -ésima cadeia, e $\hat{\theta}$ a média geral. Podemos calcular as variâncias dentro das cadeias e entre as cadeias da seguinte maneira:

Dentro das cadeias:

$$W = \frac{1}{M} \sum \hat{\sigma}_m^2$$

Entre as cadeias:

$$B = \frac{N}{M-1} \sum (\hat{\theta}_i - \hat{\theta})^2$$

A variância combinada é dada por um estimador não viesado da variância a posteriori de θ :

$$\hat{V} = \frac{N-1}{N} W + \frac{M+1}{MN} B$$

Para avaliar a convergência das M cadeias de Markov, utilizamos a razão entre \hat{V} e W , que deve ser próxima de 1. Isso é expresso pela equação do Potencial de Redução de Fator de Escala (PSRF), onde \hat{d} é o número de graus de liberdade estimado por uma distribuição t :

$$R_c = \sqrt{\frac{\hat{d} + 3}{\hat{d} + 1} \frac{\hat{V}}{W}}$$

A interpretação desse valor é que, se R_c é próximo de um, podemos considerar que as cadeias convergiram para a distribuição a posteriori de forma adequada. Esse teste de convergência de Gelman-Rubin é uma ferramenta útil para garantir que as cadeias de Markov tenham atingido a convergência e que os resultados das inferências sejam confiáveis.

Portanto, a combinação do método de inspeção visual e do teste de Gelman-Rubin oferece uma abordagem robusta para avaliar a convergência das cadeias de Markov simuladas. Essa verificação é fundamental para garantir a confiabilidade das inferências realizadas no contexto da inferência bayesiana.

2.4 Modelos para estimação de incertezas

Nos estudos de solo conduzidos por [Davoudabadi Daniel Pagendam e Baldock \(2020\)](#) e [Clifford *et al.* \(2014\)](#), foi adotada uma abordagem bayesiana para investigar o decaimento do solo, empregando um conjunto de equações que modelam um único *pool* de carbono para estimar a quantidade desse elemento ao longo do tempo. Para atingir esse objetivo, os pesquisadores formularam uma equação recursiva que delinea o comportamento do carbono em função do tempo, considerando transformações internas e novas entradas de massa orgânica. Essa abordagem estabeleceu uma relação entre as observações reais e as projeções do modelo.

De forma simplificada, podemos reescrever a equação 2.4 para descrever o comportamento de um único *pool* de carbono e considerando η_C uma medida de erro, como:

$$C(t) = e^{-K\Delta t}C(t-1) + I_{C(t)} + \eta_C, \quad (2.14)$$

adicionalmente, a entrada de carbono $I_{C(t)}$ é uma função que varia em função de diversos fatores, tais como o manejo do solo $M(t)$ e $\rho C(t-1)$, em que ρ representa a taxa de matéria orgânica remanescente, incluindo raízes depositadas, material orgânico proveniente de animais e plantas mortas, bem como o tipo de solo $S(t)$. Essa relação pode ser expressa como:

$$I_{C(t)} = M(t) + \rho C(t-1) + S(t). \quad (2.15)$$

Finalmente, é crucial estabelecer a ligação entre o modelo de comportamento teórico e as observações derivadas das medidas obtidas durante os experimentos realizados. Essa relação é expressa pela equação:

$$Y_{C(t)}|C(t) \sim N(c(t), \sigma_{\epsilon C(t)}^2). \quad (2.16)$$

No estudo do artigo previamente citado, os autores escolheram para análise o solo de Tarlee, localizado no sul da Austrália que leva em consideração quatro variáveis latentes: X_C , X_G , X_W e X_P , em que as letras correspondem aos seguintes significados: C para carbono, W para trigo (*wheat*), G para grão e P para pasto seco. E, as observações são dados pelas variáveis Y , Y_C , Y_G , Y_W e Y_P . Gerando o seguinte, modelo de processo de evolução no tempo.

$$\begin{aligned}
X_{C(t)} &\sim LN(X_{C(t-1)}e^{-K\Delta t} + I_{C(t)}, \eta_t^i) \\
X_{G(t)} &\sim LN(\mu_G + \rho_G(\log(X_{W(t-1)}) - \mu_G), \sigma_G^2) \\
X_{W(t)} &\sim LN(\log(h) + \rho_W(\log(X_{W(t-1)})), \sigma_W^2) \\
X_{P(t)} &\sim LN(\mu_P + \rho_P(\log(X_{P(t-1)}) - \mu_P), \sigma_P^2)
\end{aligned}$$

Na equação fornecida, as variáveis $\eta_t^i(t) \sim N(0, \sigma_\eta^2)$ representam termos de erro seguindo uma distribuição normal com média zero e variância σ_η^2 . Além disso, ρ_G e ρ_P são parâmetros autoregressivos. Por último, a função $I_{C(t)}$ modela o efeito do manejo do solo, determinando a quantidade de carbono presente no solo.

$$I_{C(t)} = \begin{cases} c(X_{G(t)} - X_{W(t)}) + cr_w X_{W(t)} & \text{Trigo para Grão} \\ cpX_{W(t)} + cr_w X_{W(t)} & \text{Trigo para Feno} \\ cX_{P(t)} + cr_p X_{P(t)} & \text{Pasto} \\ cpX_{P(t)} + cr_p X_{P(t)} & \text{Pasto para Feno} \\ 0 & \text{Pousio} \end{cases},$$

com os parâmetros p representando a proporção da colheita acima do solo, r_P e r_W denotando as taxas de conversão do pasto e da colheita de trigo, e finalmente, c sendo a proporção de carbono no solo, obtemos as seguintes equações de de atualização com base nas observações e na tabela 1 o espaço paramétrico do modelo:

$$\begin{aligned}
Y_{C(t)}|X_{C(t)} &\sim LN(\log(x_{C(t)}), \sigma_{\varepsilon C(t)}^2) \\
Y_{G(t)}|X_{G(t)} &\sim LN(\log(x_{G(t)}), \sigma_{\varepsilon G(t)}^2) \\
Y_{W(t)}|X_{W(t)} &\sim LN(\log(x_{W(t)}), \sigma_{\varepsilon W(t)}^2) \\
Y_{P(t)}|X_{P(t)} &\sim LN(\log(x_{P(t)}), \sigma_{\varepsilon P(t)}^2)
\end{aligned}$$

Parâmetro	Priori	Tipo
$X_{C(t)}$	<i>NormalTruncada</i> (40, 10^2 , <i>minimo</i> = 0)	Não informativa
K	$LN(-2.71, (0.127)^2)$	Informativa
C	$N(0.45, (0.001)^2)$	Informativa
r_w	$N(0.5, (0.067)^2)$	Informativa
r_p	$N(1, (0.125)^2)$	Informativa
P	<i>Beta</i> (89.9, 809.1)	Informativa
H	$LN(0.825, (0.36)^2)$	Pouco informativa
μ_G	$N(0.42, (1.18)^2)$	Pouco informativa
μ_W	$N(1.24, (1.12)^2)$	Pouco informativa
μ_P	$N(1.41, (1.81)^2)$	Pouco informativa
ρ_G	<i>Uniforme</i> (-1, 1)	Não informativa
ρ_P	<i>Uniforme</i> (-1, 1)	Não informativa
σ_N^2	<i>Inv - Gamma</i> (0.001, 0.001)	Não informativa
σ_G^2	<i>Inv - Gamma</i> (0.001, 0.001)	Não informativa
σ_W^2	<i>Inv - Gamma</i> (0.001, 0.001)	Não informativa
σ_P^2	<i>Inv - Gamma</i> (0.001, 0.001)	Não informativa
$\sigma_{\varepsilon C}^2$	0.025	Fixo
$\sigma_{\varepsilon G}^2$	0.023	Fixo
$\sigma_{\varepsilon W}^2$	0.133	Fixo
$\sigma_{\varepsilon P}^2$	0.067	Fixo

Tabela 1 – Distribuição a Priori dos parâmetros do modelo

Com a base teórica estabelecida neste capítulo, complementada pela revisão de artigos. Podemos explorar como o trabalho atual foi concebido, delineando a metodologia adotada, as etapas do processo de pesquisa e as análises realizadas.

METODOLOGIA

3.1 Derivação da equação do modelo

Para formular a equação proposta neste trabalho, precisamos considerar a equação 2.10 e desmembrá-la para os dois compartimentos existentes: rápido e lento. Em seguida, convertemos a equação para o formato de espaço de estados, como descrito na equação 3.1,

$$\begin{aligned} \dot{\mathbf{X}} &= \mathbf{A}\mathbf{X} + \mathbf{B}u + \boldsymbol{\varepsilon}^* \\ \mathbf{Y} &= \mathbf{C}\mathbf{X} + \mathbf{v}^* \end{aligned} \quad (3.1)$$

Inicialmente, temos a equação 3.2,

$$\begin{bmatrix} \frac{dC_{rapido}(t)}{dt} \\ \frac{dC_{lento}(t)}{dt} \end{bmatrix} = \begin{bmatrix} I_{ext}(t) \\ I_{int}(t) \end{bmatrix} + \begin{bmatrix} e^{-k_{rapido} \times t} \times C_{rapido}(t) \\ e^{-k_{lento} \times t} \times C_{lento}(t) \end{bmatrix}, \quad (3.2)$$

na qual há apenas entrada no compartimento rápido, $I_{ext}(t)$, indicando que a matéria orgânica entra apenas nesse compartimento. Em seguida, ocorre a conversão do compartimento rápido para o lento, evidenciando que a entrada $I_{int}(t)$ permite apenas a conversão em uma direção, conforme descrito anteriormente.

Com um rearranjo da equação 3.2, podemos chegar ao formato necessário para a equação de espaço de estado, 3.1, apresentado na equação 3.3,

$$\begin{bmatrix} \frac{dC_{rapido}(t)}{dt} \\ \frac{dC_{lento}(t)}{dt} \end{bmatrix} = \begin{bmatrix} e^{-k_{rapido} \times t} & 0 \\ 0 & e^{-k_{lento} \times t} \end{bmatrix} \times \begin{bmatrix} C_{rapido}(t) \\ C_{lento}(t) \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} I_{ext}(t) \\ I_{int}(t) \end{bmatrix} + \boldsymbol{\varepsilon}^* \quad (3.3)$$

que considera os erros ou incertezas do modelo e considerando que observamos somente a soma dos componentes, temos a equação 3.4:

$$C_{total} = \begin{bmatrix} 1 & 1 \end{bmatrix} \times \begin{bmatrix} C_{rapido}(t) \\ C_{lento}(t) \end{bmatrix} + \mathbf{v}^* \quad (3.4)$$

A partir da equação 3.3, se considerarmos um delta pequeno, o mesmo da simulação computacional, podemos escrever nossa equação do modelo de forma mais eficiente para a execução das simulações propostas, chegando na equação 3.5,

$$\begin{aligned} C_{rapido}(t) &= C_{rapido}(t-1) \times e^{-k_{rapido}} + I_{ext} + C_{rapido}(t-1) + \varepsilon^* \\ C_{lento}(t) &= C_{lento}(t-1) \times e^{-k_{lento}} + I_{int} + C_{lento}(t-1) + \varepsilon^* \\ C_{total}(t) &= C_{lento}(t) + C_{rapido}(t) + v^* \end{aligned} \quad , \quad (3.5)$$

considerando que o carbono do tempo anterior, $C_{reservatorio}(t-1)$, é uma entrada automática no sistema, podemos simplificar ainda mais a equação, escrevendo da seguinte forma. 3.6,

$$\begin{aligned} C_{rapido}(t) &= C_{rapido}(t-1) \times e^{-k_{rapido}} + I'_{rapido}(t-1) + \varepsilon^* \\ C_{lento}(t) &= C_{lento}(t-1) \times e^{-k_{lento}} + I'_{lento}(t-1) + \varepsilon^* \\ C_{total}(t) &= C_{lento}(t) + C_{rapido}(t) + v^* \end{aligned} \quad , \quad (3.6)$$

portanto, a partir da equação 3.5, podemos construir as distribuições iniciais para compreender o comportamento da matéria orgânica no solo ao longo do tempo, servindo de base para uma metodologia menos invasiva do que temos atualmente para medir a quantidade de carbono no solo.

3.2 Base de Dados

O presente trabalho utilizou quatro bases de dados, sendo uma delas gerada artificialmente devido à dificuldade de encontrar conjuntos de dados disponíveis, e as outras três foram obtidas a partir do censo agropecuário canadense.

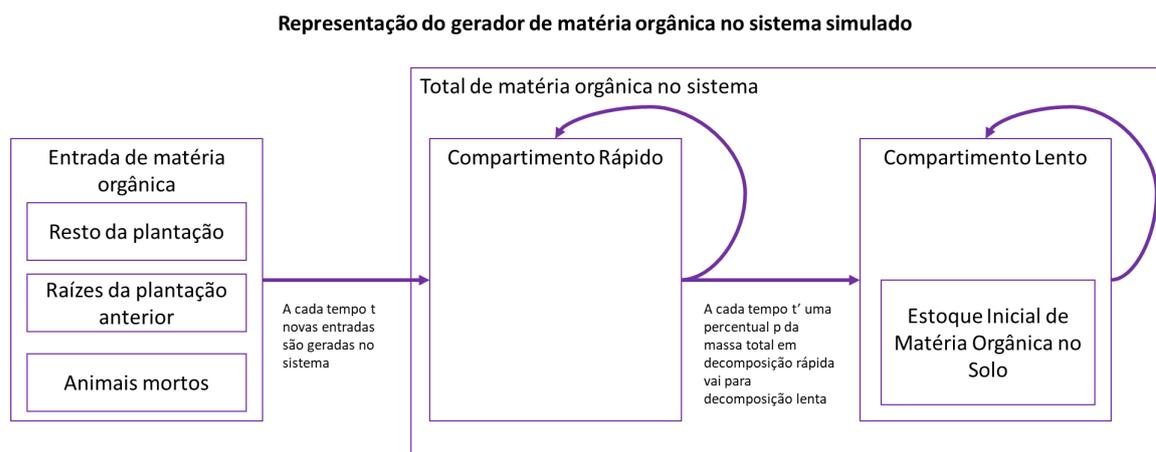
3.2.1 Base de dados artificial

Uma das principais desafios na modelagem de carbono no solo reside na obtenção de conjuntos de dados adequados para testar os modelos propostos. Dada essa situação, o presente trabalho desenvolveu um simulador de decaimento de carbono, que serve como base para o modelo proposto. Esse simulador permite gerar dados sintéticos que representam a dinâmica de carbono no solo, possibilitando assim a avaliação e o aprimoramento do modelo proposto. Essa abordagem é essencial para lidar com a escassez de dados observacionais precisos e disponíveis, além de permitir uma avaliação mais detalhada do desempenho do modelo em diferentes cenários e condições.

Para esse propósito, considerou-se que o solo possui dois compartimentos para o decréscimo da matéria orgânica. O primeiro compartimento é responsável pela decomposição inicial de resíduos da colheita, restos de plantas, animais mortos e outros materiais orgânicos, os quais começam a se decompor. Com o tempo, esses materiais se tornam mais estáveis e cristalizados, resultando em uma decomposição mais lenta. Além disso, o solo também contém

carbono orgânico já existente, proveniente de processos anteriores de decomposição. Esses dois compartimentos de decaimento representam as principais fontes de matéria orgânica no solo e são considerados no modelo proposto. Na Figura 5, temos um esquemático do modelo de solo proposto e no apêndice B temos os códigos utilizados no presente trabalho.

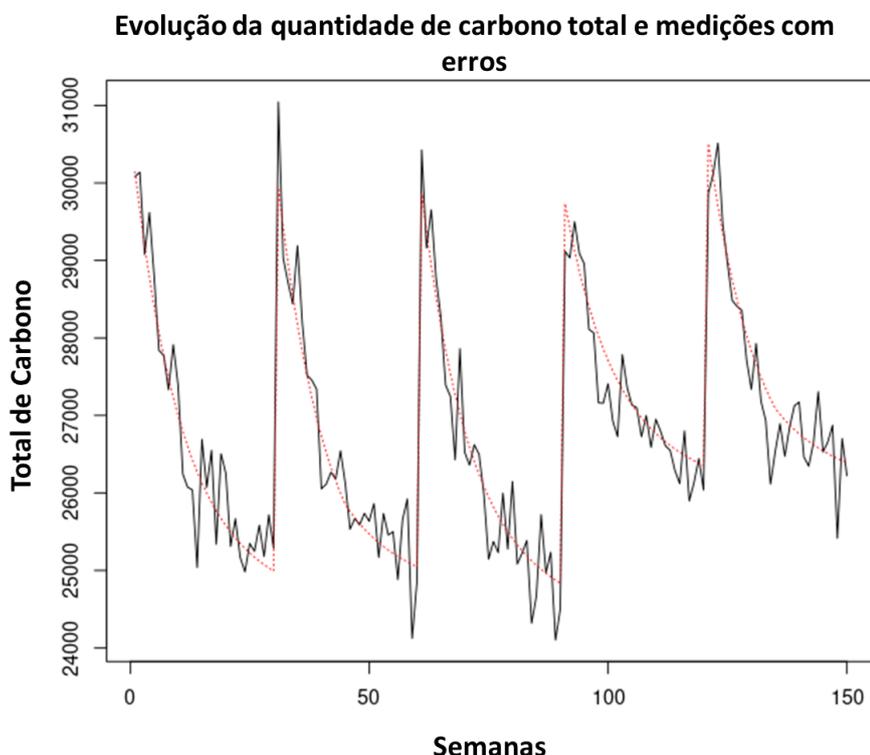
Figura 5 – Diagrama do gerador de matéria orgânica no solo



Na geração da simulação do solo, considerou-se um período total de 150 semanas, representado pela variável *Total*. Além disso, foi considerado que a cada 30 semanas há entrada de novas matérias orgânicas no solo, e a cada 45 semanas ocorre a conversão da massa do compartimento de decaimento rápido para o de decaimento lento. Abaixo está o código para a geração do carbono no solo, onde teremos quatro variáveis de Carbono:

- **Carbono lento:** a variável $X_{carbono_slow}$ representa a quantidade de carbono que foi transferida para o compartimento lento do sistema, conforme ilustrado na Figura ??.
- **Carbono rápido:** a variável $X_{carbono_fast}$ representa a quantidade de carbono que foi adicionada e armazenada no compartimento rápido, conforme ilustrado na Figura ??.
- **Carbono total irrealista:** a variável $X_{carbono_smooth}$ representa a quantidade total de carbono no sistema, sem erros de medição, conforme ilustrado na Figura 6, onde é representada pela linha vermelha tracejada.
- **Carbono total realista:** a variável $X_{carbono}$ representa a quantidade total de carbono no sistema, considerando um erro no momento da medição descrito por $\varepsilon \sim \mathcal{N}(0, 500)$, conforme ilustrado na Figura 6, onde é representada pela linha preta contínua.

Figura 6 – Evolução da quantidade de matéria orgânica total no sistema sem erro de medição versus com erro de medição



3.2.2 Base de dados Censo Agrícola Canadense

O governo canadense realiza a quantificação do carbono no solo a cada cinco anos, por meio do censo agrícola, e quando não ocorre, o valor de carbono é construído através de uma interpolação linear. Portanto, o presente conjunto de dados apresenta 3408 regiões com medição de carbono, com as seguintes colunas de descrição em seus dados e pode ser acessado em: <"<https://open.canada.ca/data/en/dataset/7a31d717-529f-45c2-9273-e7ddfb767331/resource/fb04b9d1-cc25-4723-9f7f-d6673f4be94c#resource-list>">:

- **SOIL_LANDSCAPE_ID:** Identificador de polígono de paisagem do solo do Canadá;
- **PROVINCE:** O nome da província onde os dados são encontrados;
- **YEAR:** Ano do dado;
- **POLYGON_WEIGHT:** O peso atribuído a este registro (polígono) em comparação com outros registros (polígonos) ao agregar os dados para unidades de mapa mais amplas, por

exemplo, para escalas de EcoRegião ou EcoDistrito;

- **PROV_PCT:** A ponderação percentual atribuída à porção do polígono da Paisagem de Solo do Canadá (identificado pelo SOIL_LANDSCAPE_ID) que se encontra na província identificada. Se um polígono da Paisagem de Solo do Canadá for dividido por uma fronteira provincial, esse valor será inferior a 100;
- **SOCC_VAL:** Uma faixa de valores que descrevem a mudança no SOC em um determinado ano ($\text{kg ha}^{-1} \text{ ano}^{-1}$);
- **SOCC_CLASS:** Uma classificação que descreve a mudança no carbono orgânico do solo em um determinado ano;
- **SOCC_CLASS_EN:** Descrição textual da classe em Inglês;
- **SOCC_CLASS_FR:** Descrição textual da classe em Francês;
- **RSOC_VAL:** Uma faixa de valores expressa como uma proporção do nível atual de carbono orgânico do solo em relação a um valor de referência modelado de carbono orgânico do solo para uma pastagem permanente de gramíneas com manejo extensivo;
- **RSOC_CLASS:** Uma classificação que descreve o nível atual de carbono orgânico do solo em relação a um valor de referência modelado de carbono orgânico do solo para uma pastagem permanente de gramíneas com manejo extensivo;
- **RSOC_CLASS_EN:** Descrição textual da classe em Inglês;
- **SOC_CLASS_FR:** Descrição textual da classe em Francês;
- **SDR_CLASS:** Uma classificação que descreve o risco geral de degradação do carbono orgânico do solo;
- **SDR_CLASS_EN:** Descrição textual da classe em Inglês;

- **SDR_CLASS_FR**: Descrição textual da classe em Francês.

Para o presente trabalho seguimos somente com as colunas SOIL_LANDSCAPE_ID, YEAR e SOCC_VAL e selecionamos os terrenos com código 244003, 244007 e 244014. Nas Figuras 7, 8 e 9 temos a evolução do carbono em anos.

Figura 7 – Comportamento da evolução do carbono no solo 244003 do Canadá

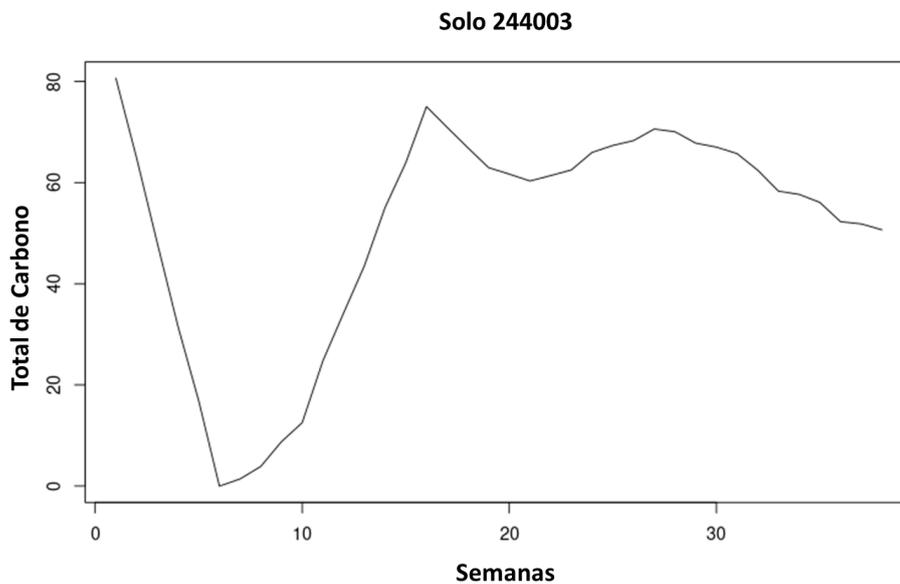


Figura 8 – Comportamento da evolução do carbono no solo 244007 do Canadá

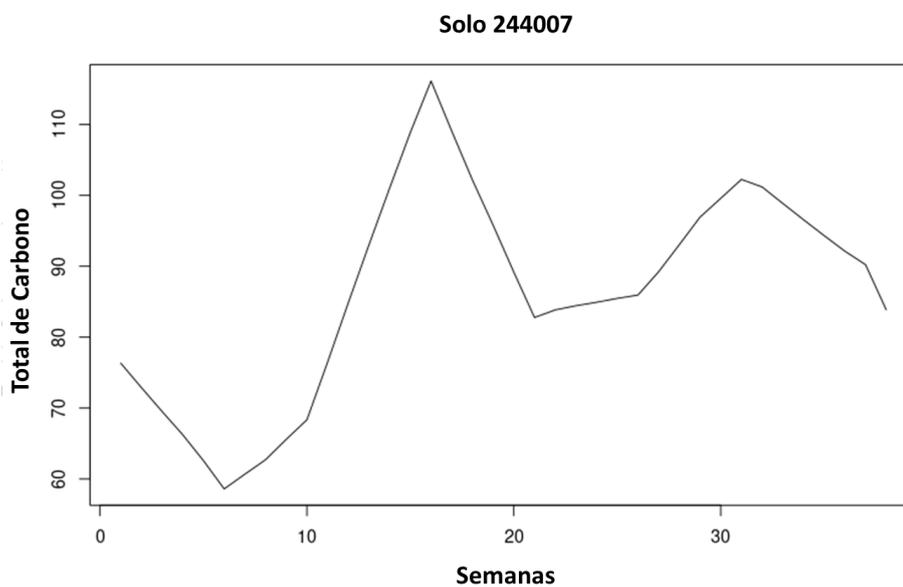
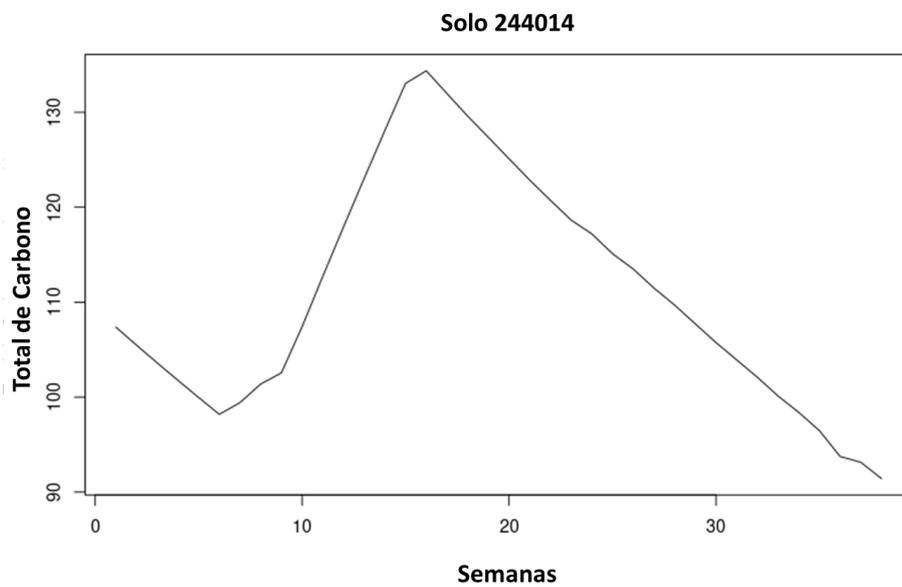


Figura 9 – Comportamento da evolução do carbono no solo 244014 do Canadá



3.3 Configuração dos modelos de inferência bayesiana

Ambas bases de dados consideraram um sistema com duas variáveis latentes: carbono de decomposição rápida, C_{rapido} , e carbono de decomposição lenta, C_{lento} . Nesse sistema, ocorre sempre a transição da massa de carbono de decomposição rápida para a lenta e a entrada ocorre apenas no compartimento rápido.

As principais diferenças entre os dois modelos são a quantidade de entrada de carbono. No caso artificial, é em torno de 5 mil, enquanto no caso real é em torno de 5. Além disso, os momentos de conversão também são diferentes: no caso artificial, é em torno de 45, e no caso real, é em torno de 10. Por último, a taxa de conversão da massa que sai do rápido e vai para o lento também difere, sendo em torno de 34% no caso artificial e cerca de 20% no caso real.

Modelo para solo artificial

$$\lambda_{lento} \sim \mathcal{N}(400, 50)$$

$$\lambda_{rapido} \sim \mathcal{N}(1000, 50)$$

$$\rho_{lento} \sim \mathcal{N}(10, 1)$$

$$\rho_{rapido} \sim \mathcal{N}(10, 1)$$

$$\rho_{observacao} \sim \mathcal{N}(10, 1)$$

$$momento \sim \mathcal{N}(44, 1)$$

$Entrada_{carbono}(t) \sim \acute{E} \mathcal{N}(5000, 10)$ para todo instante t que é múltiplo de 45

$Tx_{conversao}(t) \sim \acute{E} \mathcal{N}(0.34, 1)$ para todo t que é múltiplo da variavel momento

$$C_{convertido}(t) = C_{rapido}(t-1) \times Tx_{conversao}(t)$$

$$C_{lento}(t) \sim N((C_{lento}(t-1) + C_{convertido}(t)) \times e^{-\lambda_{lento}}, \rho_{lento})$$

$$C_{rapido}(t) \sim N((C_{rapido}(t-1) - C_{convertido}(t) + Entrada_{carbono}(t)) \times e^{-\lambda_{rapido}}, \rho_{rapido})$$

$$C_{total}(t) = C_{lento}(t) + C_{rapido}(t)$$

$$C_{observado}(t) \sim \mathcal{N}(C_{total}(t), \rho_{observacao})$$

Modelo para solo do censo agrícola canadense

$$\lambda_{lento} \sim \mathcal{N}(400, 50)$$

$$\lambda_{rapido} \sim \mathcal{N}(1000, 50)$$

$$\rho_{lento} \sim \mathcal{N}(10, 1)$$

$$\rho_{rapido} \sim \mathcal{N}(10, 1)$$

$$\rho_{observacao} \sim \mathcal{N}(10, 1)$$

$$momento \sim \mathcal{N}(10, 1)$$

$Entrada_{carbono}(t) \sim \acute{E} \mathcal{N}(5, 10)$ para todo instante t que é múltiplo de 45

$Tx_{conversao}(t) \sim \acute{E} \mathcal{N}(0.20, 1)$ para todo t que é múltiplo da variavel momento

$$C_{convertido}(t) = C_{rapido}(t-1) \times Tx_{conversao}(t)$$

$$C_{lento}(t) \sim N((C_{lento}(t-1) + C_{convertido}(t)) \times e^{-\lambda_{lento}}, \rho_{lento})$$

$$C_{rapido}(t) \sim N((C_{rapido}(t-1) - C_{convertido}(t) + Entrada_{carbono}(t)) \times e^{-\lambda_{rapido}}, \rho_{rapido})$$

$$C_{total}(t) = C_{lento}(t) + C_{rapido}(t)$$

$$C_{observado}(t) \sim \mathcal{N}(C_{total}(t), \rho_{observacao})$$

Com a metodologia detalhada neste capítulo, estamos preparados para avançar para a análise dos resultados no próximo capítulo. Essa metodologia fornece o arcabouço necessário para uma investigação abrangente e precisa, permitindo a reprodutibilidade dos resultados encontrados.

RESULTADO

4.1 Convergência do algoritmo de MCMC

No presente capítulo são apresentados os resultados das simulações realizadas, acompanhados de uma discussão detalhada sobre os principais aspectos e informações obtidas para o solo simulado e no apêndice C os resultados dos demais solos canadenses. Os tempos computacionais foram medidos para ambos os solos e colocado na tabela 2, a configuração utilizada foi uma máquina virtual na plataforma Posit, <<https://posit.cloud/>>, com 16 GB de memória RAM e 4 cpus dedicadas.

Modelo	Tempo médio (min)	Desvio Padrão (min)
Modelo simulado	85.20	3.74
Modelo solo canadense	9.47	1.14

Tabela 2 – Tempo computacional para execução do modelo de solo simulado e dos solos canadenses

4.1.1 *Matéria orgânica de solo simulado*

Para avaliar a convergência de um algoritmo de MCMC e verificar se os resultados são confiáveis, é importante considerar dois fatores: o valor do teste de Gelman-Rubin e a aparência das simulações em relação ao formato de ruído ou caminho aleatório.

O teste de Gelman-Rubin, também conhecido como diagnóstico R-hat, é uma medida estatística que compara a variabilidade dentro das cadeias de Markov com a variabilidade entre as cadeias. Com valor próximo de 1 indica que as cadeias convergiram e fornecem estimativas consistentes. Valores maiores que 1 podem indicar que as cadeias não convergiram completamente ou que há alguma dependência entre elas.

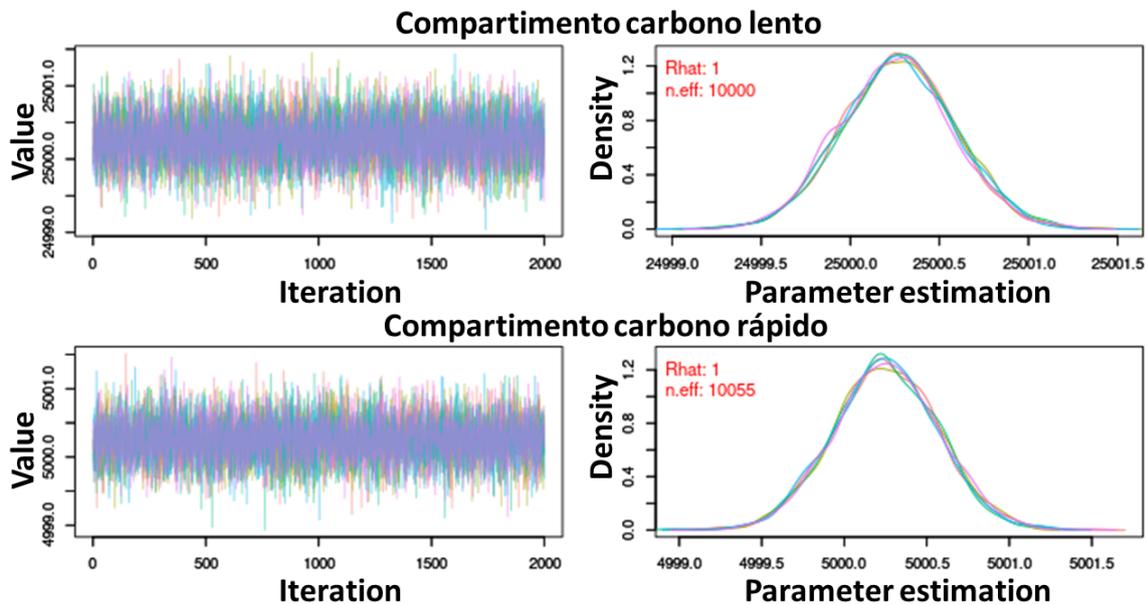
Além disso, é importante examinar visualmente as simulações geradas pelo algoritmo. Se os valores parecerem um ruído aleatório ou seguirem um caminho aleatório, isso sugere que

o algoritmo está explorando de forma adequada o espaço de parâmetros e produzindo resultados confiáveis. Por outro lado, se houver algum padrão, tendência ou dependência aparente nos valores simulados, isso pode indicar problemas de convergência ou amostragem insuficiente.

Portanto, ao analisar os resultados de um algoritmo de MCMC, é recomendado verificar se o valor do teste de Gelman-Rubin está próximo de 1 e se as simulações exibem um comportamento semelhante ao ruído aleatório ou caminho aleatório. Essas avaliações ajudam a determinar se os resultados são confiáveis e se o algoritmo convergiu adequadamente.

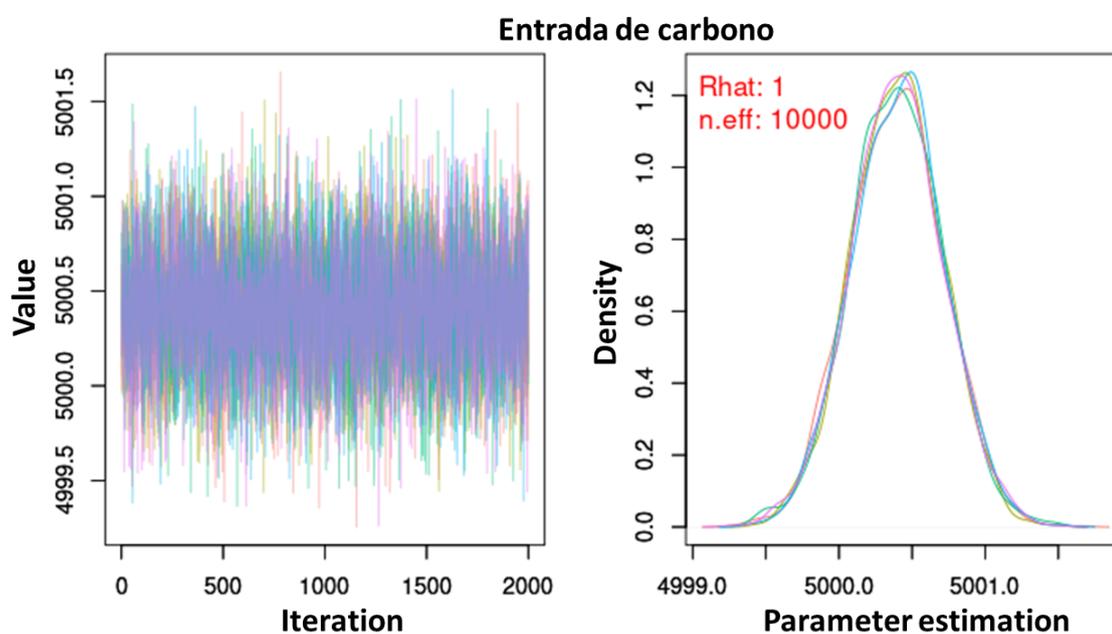
Inicialmente, explorando a situação dos compartimentos de matéria orgânica, tanto o lento ($x_{\text{carbono_lento_rapido}}[1,t]$) quanto o rápido ($x_{\text{carbono_lento_rapido}}[2,t]$), temos uma boa convergência do MCMC, conforme Figura 10.

Figura 10 – Evolução da simulação de MCMC dos compartimentos de matéria orgânica do sistema no início do ciclo



Uma segunda métrica que estamos avaliando é a quantidade de carbono que entra no sistema a cada 45 semanas, conforme mostrado na Figura 11. Nossa avaliação indica que o algoritmo de MCMC foi capaz de simular adequadamente essa variável do modelo. Os valores simulados estão em concordância com os dados de entrada reais, o que sugere que o modelo está capturando de forma precisa a dinâmica do carbono no sistema. Essa é uma indicação positiva da capacidade do modelo em reproduzir os padrões observados e em fornecer estimativas confiáveis para a quantidade de carbono que entra no sistema ao longo do tempo.

Figura 11 – Evolução da simulação de MCMC para entrada de carbono



O próximo conjunto de variáveis a ser avaliado é composto por λ_{rapido} , λ_{lento} e a taxa de conversão da matéria do compartimento rápido para o lento, que em nosso modelo de dinâmica do solo simulado foi definida como 30%. Conforme podemos observar nas Figuras 12 e 13, os valores simulados para os lambdas estão próximos do valor teórico esperado. No entanto, a taxa de conversão, figura 14, apresenta um valor abaixo do esperado, em torno de 5%. Essa diferença pode ser atribuída às incertezas dentro do modelo, que diminuem a importância dessa variável e afetam sua estimativa. É importante ressaltar que a taxa de conversão é uma variável chave para a dinâmica do sistema, e sua estimativa precisa é fundamental para uma modelagem adequada. Portanto, a avaliação e o aprimoramento dessa variável são áreas de interesse para futuras pesquisas e refinamentos do modelo.

Figura 12 – Evolução da simulação de MCMC para λ_{rapido}

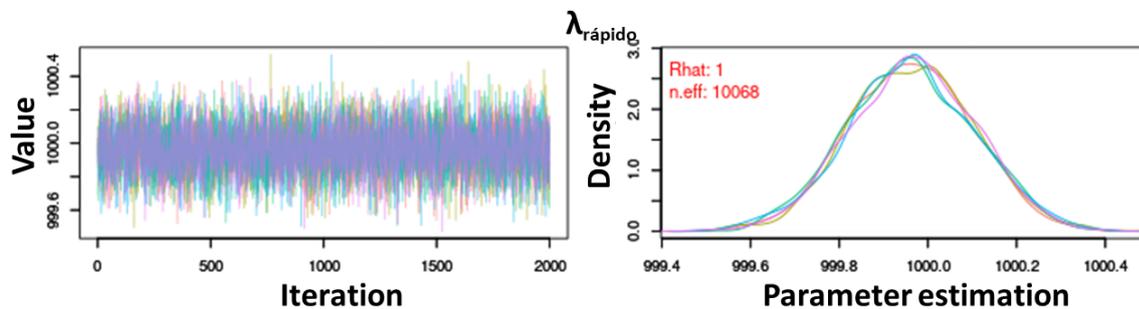


Figura 13 – Evolução da simulação de MCMC para λ_{lento}

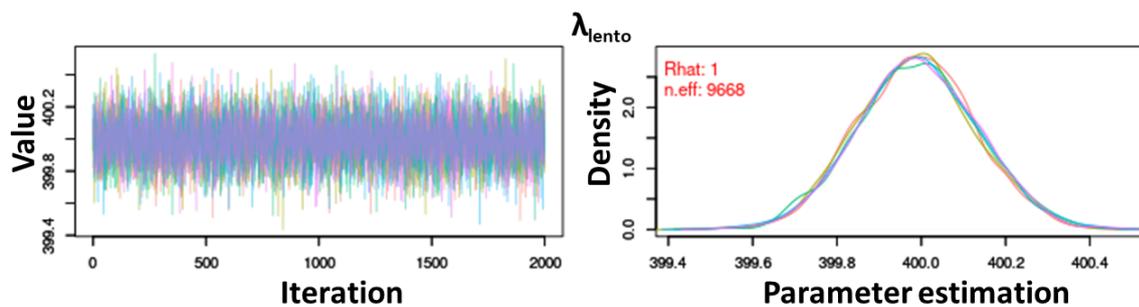
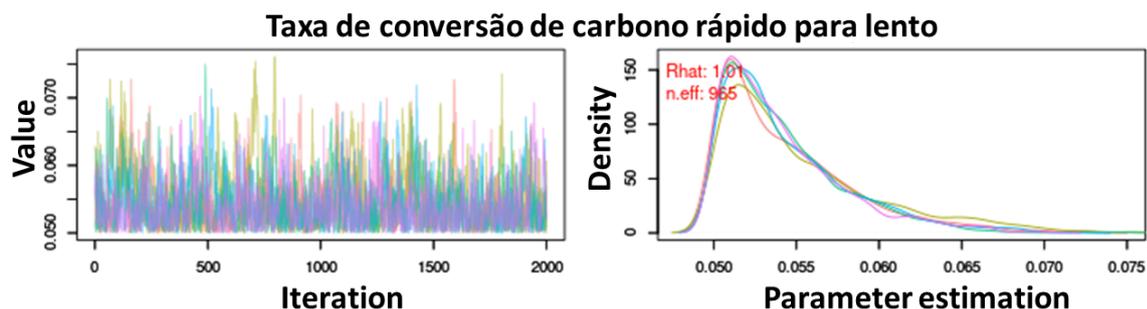


Figura 14 – Evolução da simulação de MCMC para taxa de conversão da matéria de compartimento rápido para o lento



Na última etapa, avaliamos quatro variáveis finais: o tempo de conversão de rápido para lento, cujo valor simulado foi de 45 semanas, enquanto o modelo proposto encontrou 42 semanas, conforme figura 15; e os erros associados à observação, figura 16, e à evolução dos estados rápidos e lentos, figuras 17 e 18.

Figura 15 – Evolução da simulação de MCMC para momento de conversão

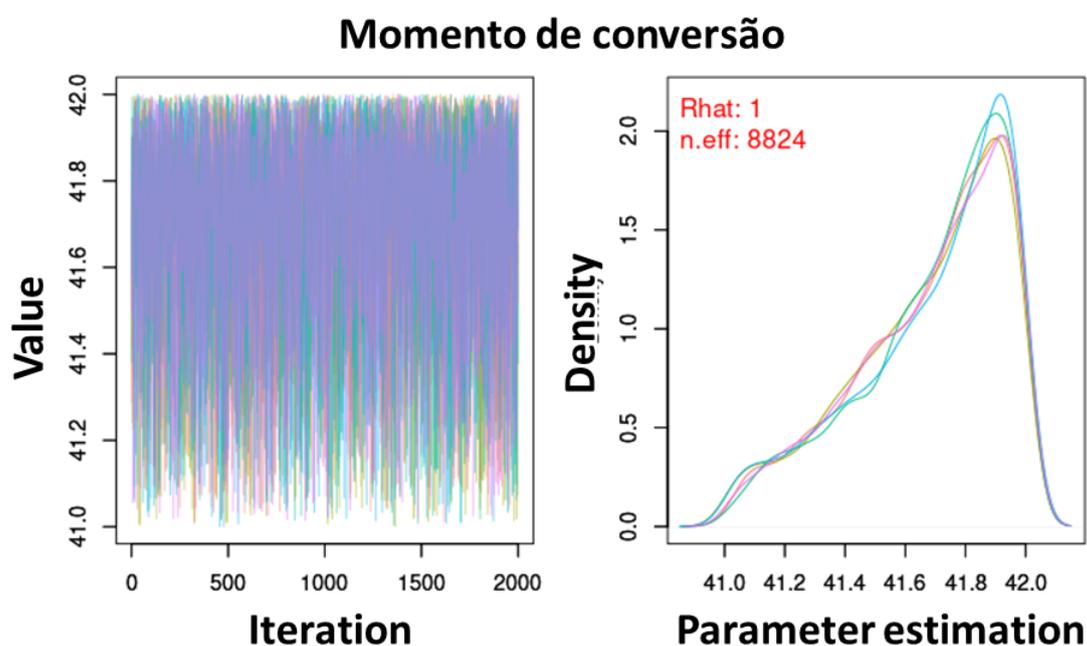


Figura 16 – Evolução da simulação de MCMC para taxa erro na observação do carbono total

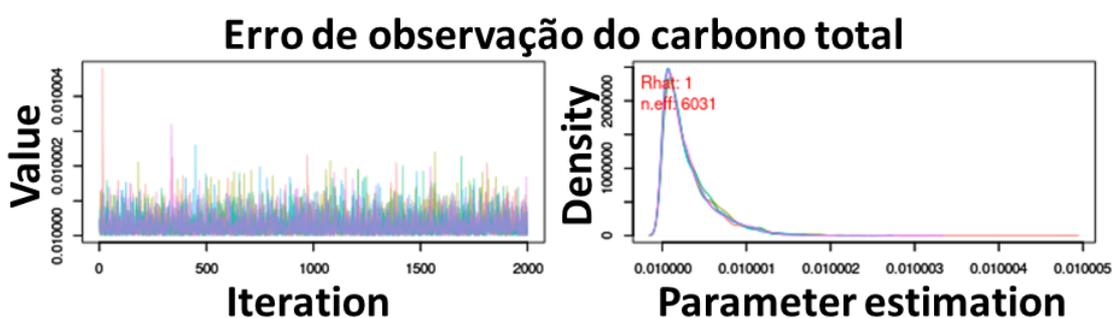


Figura 17 – Evolução da simulação de MCMC para o erro na transição entre os estados rápidos

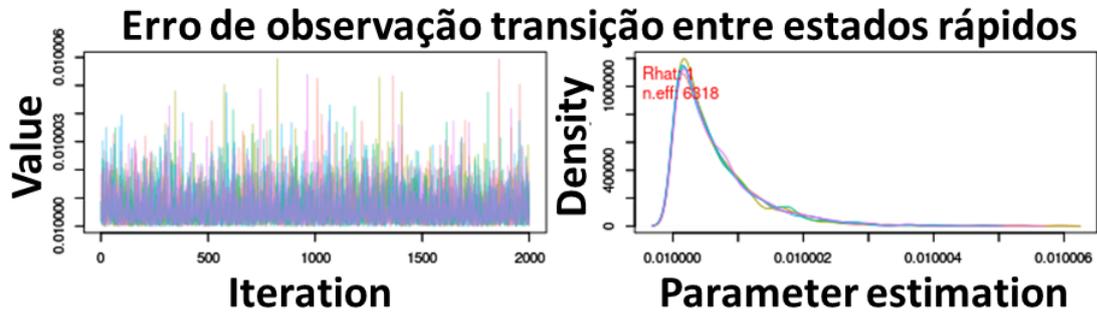
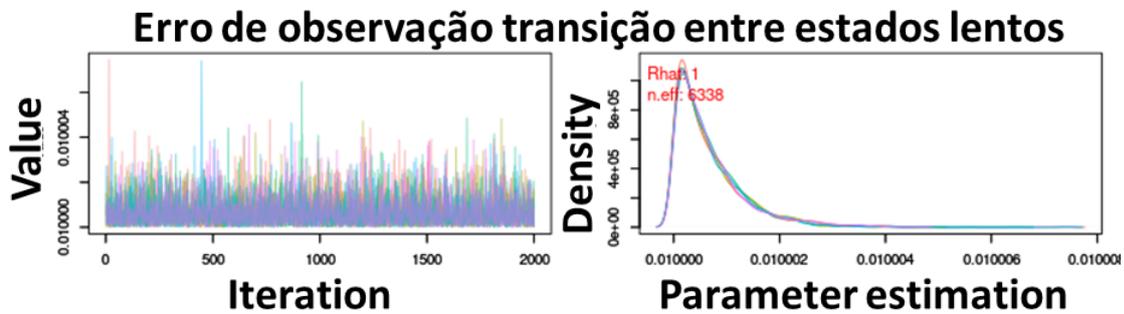


Figura 18 – Evolução da simulação de MCMC para o erro na transição entre os estados lentos



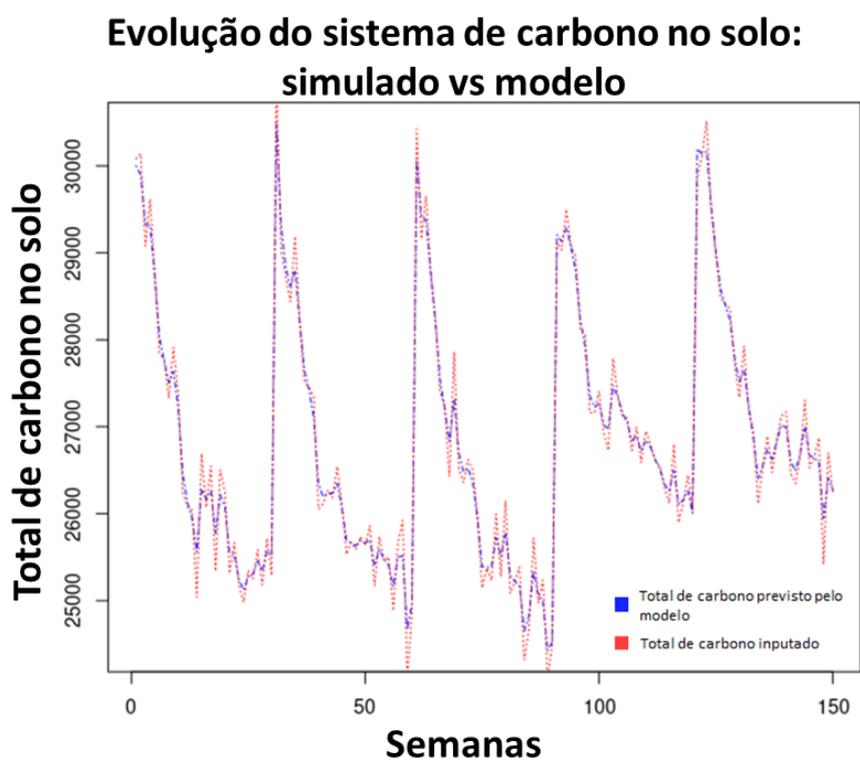
Conforme podemos observar nas Figuras 16, 17 e 18, todas as três variáveis apresentaram uma ótima convergência, com um valor de $\hat{R} = 1$ para todas elas. Isso indica que o algoritmo de MCMC convergiu bem e os resultados são consistentes. No entanto, é importante notar que houve uma pequena discrepância entre o valor simulado e o valor estimado para o tempo de conversão, o que sugere a possibilidade de ajustar esse parâmetro no modelo para obter uma melhor correspondência com o valor esperado. No geral, esses resultados demonstram a eficácia do modelo proposto em capturar as características e dinâmicas do sistema de carbono no solo.

4.2 Precisão do modelo

4.2.1 Matéria orgânica de solo simulado

No ambiente totalmente simulado, temos um maior controle sobre os principais fatores que influenciam a evolução total da matéria orgânica no solo. Um dos primeiros resultados apresentados é a comparação entre o total de carbono simulado e as previsões do modelo, conforme mostrado na Figura 19. Visualmente, temos a percepção de que o modelo teve um bom desempenho, possuindo *Mean Absolute Percentage Error* (MAPE) igual a 0,61%, sendo capaz de prever com qualidade a evolução da matéria orgânica no solo.

Figura 19 – Comparação visual entre o total de carbono previsto pelo modelo e o total de entrada



Um segundo gráfico de interesse é verificar se os resíduos do modelo proposto seguem uma distribuição normal com média zero. Para realizar essa análise, realizamos uma inspeção visual por meio do gráfico de resíduos, apresentado na Figura 20, e também utilizamos um gráfico Q-QPlot, mostrado na Figura 21.

Figura 20 – Diferença entre o total de carbono previsto pelo modelo versus a quantidade de entrada

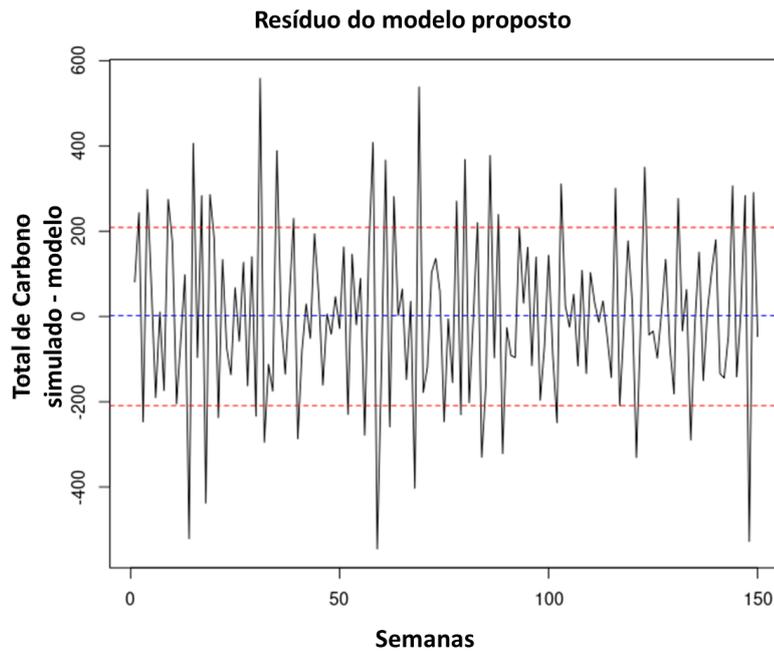
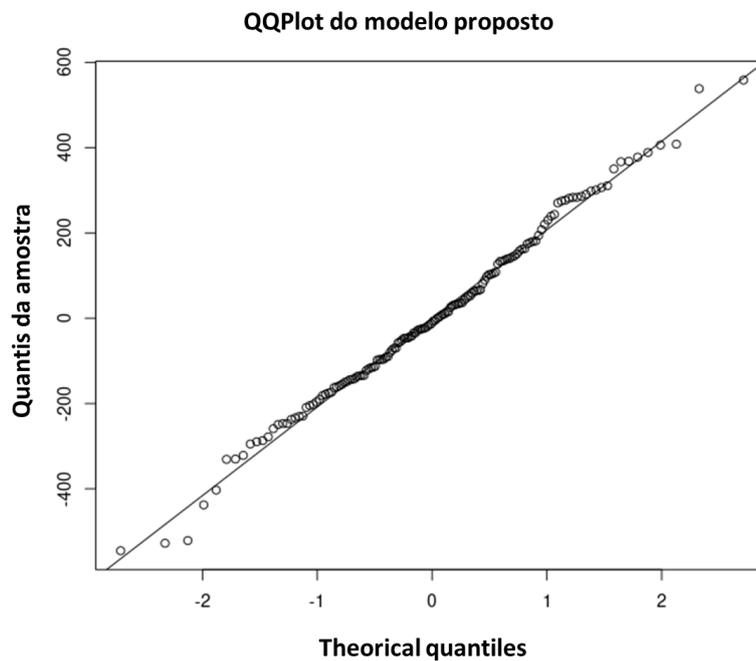
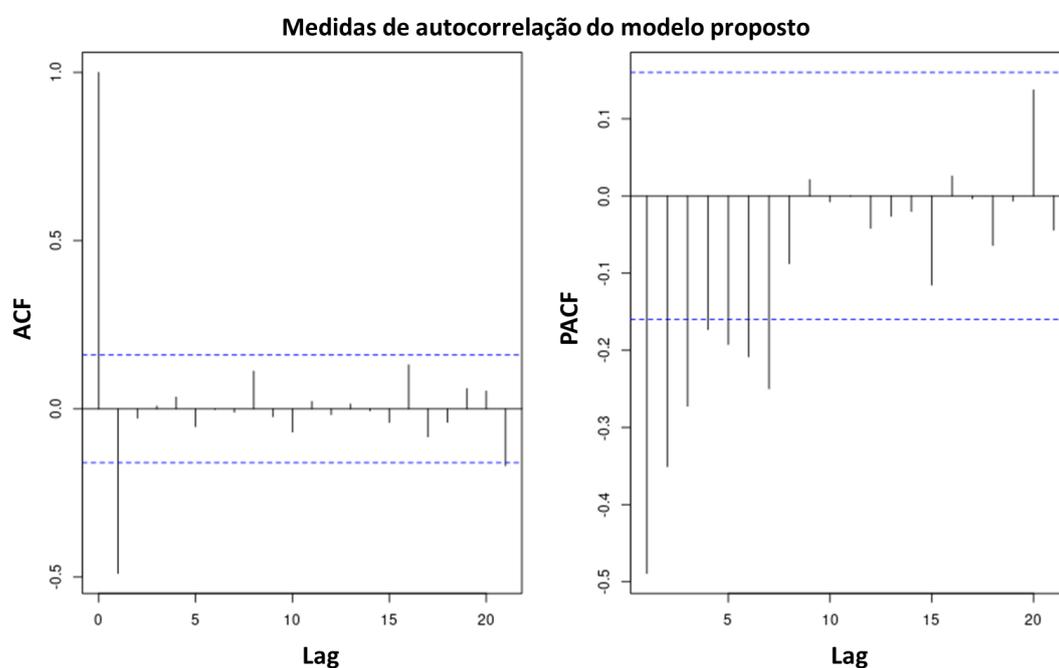


Figura 21 – Normal Q-Qplot



Portanto, com base nas Figuras 20 e 21, podemos inferir que o modelo se ajustou bem aos dados e forneceu uma boa explicação para o conjunto de dados simulados propostos. No entanto, ao considerar medidas tradicionais de séries temporais, como a Função de Auto-Correlação (ACF) e a Função de Auto-Correlação Parcial (PACF), conforme mostrado na Figura 22, observamos que o modelo ainda exibe autocorrelações que não foram totalmente eliminadas ou corrigidas. Essas autocorrelações indicam uma dependência do termo anterior, sugerindo a presença de um modelo de média móvel com ordem igual a 1. Além disso, há uma autocorrelação significativa entre diferentes momentos de t até $t + 7$, o que sugere a presença de um modelo autorregressivo de ordem 7.

Figura 22 – Gráfico mostrando o compartimento de ACF e PACF do resíduo do modelo



CONCLUSÃO

A frequência de eventos climáticos tem aumentado ano após ano devido à influência do efeito estufa causado por nós. Nossa sociedade possui uma grande capacidade de solucionar diversos desafios que enfrentamos ao longo de nossa jornada evolutiva, e uma dessas soluções é o crédito de carbono, que auxilia as empresas a equilibrarem suas emissões. Um dos setores que emite grande quantidade de dióxido de carbono (CO₂) é a agricultura, porém, também é um dos principais capturadores desse gás. A EMBRAPA desempenhou um papel fundamental no avanço da agricultura brasileira e tem demonstrado que o manejo adequado do solo é importante não apenas para o futuro do planeta, mas também para a qualidade das safras.

Entretanto, enfrentamos uma grande dificuldade em medir de forma barata e abrangente a quantidade de carbono armazenada no solo, pois o método atual é altamente invasivo, exigindo a instalação de cilindros no solo e medições frequentes. No entanto, o presente trabalho conseguiu demonstrar a eficácia do uso de modelos de solo para estimar a quantidade de matéria orgânica armazenada ao longo do tempo. Isso é válido tanto para períodos de semanas, como no caso dos dados simulados, quanto para períodos de anos, como nos solos canadenses. Esse avanço representa uma alternativa viável e menos onerosa para monitorar as mudanças no carbono do solo em larga escala.

O presente trabalho, em termos de acurácia, alcançou um MAPE de 0,61% para o solo simulado e de 5,29%, 0,017% e 0,021% para os solos canadenses. Isso indica que o modelo proposto é capaz de aprender de forma significativa a evolução do armazenamento de carbono no solo. No entanto, quando se métodos de Monte Carlos, não podemos avaliar apenas a precisão do modelo, mas também verificar se ele convergiu de forma adequada. Nesse sentido, o presente trabalho foi satisfatório, pois todas as simulações convergiram adequadamente, apresentando valores de \hat{R} sempre próximos de um. Isso indica que o modelo foi capaz de capturar a variabilidade dos dados e fornecer estimativas confiáveis.

Do ponto de vista qualitativo, apresentamos um modelo simples para simular a dinâmica

do solo, permitindo que outros pesquisadores que enfrentam dificuldades de acesso a conjuntos de dados possam criar suas próprias dinâmicas de solo. Além disso, conseguimos desmistificar a biblioteca “NIMBLE” na linguagem R, fornecendo material em língua portuguesa para que estudantes de graduação, mestrado ou doutorado possam utilizá-la no desenvolvimento de modelos de espaço de estados e aplicar inferência em diversos outros problemas. Isso amplia o acesso a ferramentas poderosas e promove a disseminação do conhecimento nessa área de pesquisa.

Como proposta para trabalhos futuros, sugere-se:

- Aprimorar o modelo para considerar momentos sem observação é uma área de desenvolvimento importante. Isso permitiria acompanhar as medidas de solos reais em intervalos de tempo mais espaçados, por exemplo, a cada 6 semanas, e modelar considerando a ausência de observações em alguns momentos. Esse aprimoramento seria especialmente útil quando lidamos com dados parciais, nos quais só temos informações em determinados momentos. Ele poderia melhorar a capacidade do modelo em estimar a dinâmica do carbono no solo, mesmo quando há lacunas nos dados observados. Essa abordagem mais flexível e adaptável seria valiosa para fornecer *insights* sobre as mudanças no armazenamento de carbono e orientar práticas de manejo do solo de forma mais precisa.
- Uma possível melhoria seria explorar novas formas de entrada de matéria orgânica no solo, levando em consideração entradas mais aleatórias ao longo do período analisado. Isso poderia refletir a variabilidade natural dos processos de decomposição e entrada de matéria orgânica no solo. Ao introduzir uma maior aleatoriedade nas entradas, o modelo seria capaz de capturar melhor as flutuações e as não-linearidades dos processos biogeoquímicos. Isso permitiria uma representação mais realista da dinâmica do carbono no solo e proporcionaria uma melhor compreensão dos padrões temporais de acumulação e degradação da matéria orgânica. Além disso, ao considerar entradas mais aleatórias, seria possível explorar diferentes cenários e condições ambientais, o que poderia fornecer *insights* valiosos sobre a resposta do solo a diferentes condições de manejo e mudanças climáticas.
- Além do filtro de partículas tradicional, como o *bootstrap*, é recomendável explorar a aplicação de outros métodos, como o filtro de partículas auxiliar (APF), o filtro de Liu e Wester e o filtro de iteração 2 (IF2). Esses filtros podem fornecer abordagens alternativas e complementares para melhorar a eficiência e a precisão do modelo. O filtro de partículas auxiliar, por exemplo, utiliza informações auxiliares para reponderar as partículas e melhorar a representação do estado estimado. O filtro de Liu e Wester é conhecido por sua capacidade de lidar com distribuições não lineares e não gaussianas, sendo especialmente útil em cenários complexos. Já o filtro de iteração 2 é uma técnica avançada que combina o uso de vários filtros de partículas em um processo iterativo para

obter estimativas mais precisas. A avaliação desses filtros adicionais pode contribuir para aprimorar a inferência e proporcionar uma análise mais robusta dos dados do modelo de espaço de estados.

NIMBLE

Numerical Inference for Statistical Models in a Bayesian and Likelihood Framework (NIMBLE) é um *framework* escrito em R, de [Valpine et al. \(2023\)](#), codificando a linguagem BUGS, o qual permite que pessoas que conheçam essa linguagem possam criar de forma declarativa modelos hierárquicos complexos, tornando a inferência bayesiana mais fácil de ser executada. Suas principais características são apresentadas abaixo:

- **Especificação Declarativa do Modelo:** Com o *NIMBLE*, você especifica seu modelo estatístico de forma declarativa usando uma sintaxe que se assemelha à notação matemática padrão. Isso facilita a expressão de modelos hierárquicos complexos e a especificação de dependências entre os componentes do modelo.
- **Linguagem Integrada:** O *NIMBLE* combina as capacidades de modelagem do BUGS e do JAGS com o poder de programação do R. Podendo usar o rico ecossistema de bibliotecas e funções do R diretamente dentro da especificação do modelo *NIMBLE*.
- **Cadeias de Markov Monte Carlo (MCMC):** O *NIMBLE* suporta algoritmos MCMC para amostragem de distribuições a posteriori dos parâmetros do modelo. Ele fornece algoritmos MCMC eficientes, como o No-U-Turn Sampler (NUTS) e o Hamiltonian Monte Carlo (HMC), que lidam com modelos de alta dimensão e complexidade.
- **Compilação Eficiente:** O *NIMBLE* compila dinamicamente a especificação do modelo em código C++, o que leva a uma execução eficiente e melhora o desempenho computacional em comparação com implementações puras em R.
- **Modelos Hierárquicos:** O *NIMBLE* é especialmente útil para a construção de modelos hierárquicos, nos quais os dados estão organizados em níveis aninhados. Ele permite a especificação fácil de efeitos aleatórios, parâmetros compartilhados e outras estruturas complexas comumente encontradas em modelagem hierárquica.

- **Funções e Distribuições Personalizadas:** O *NIMBLE* suporta a definição de funções e distribuições personalizadas, permitindo incorporar componentes de modelagem especializados ou distribuições de probabilidade não padrão em seus modelos.
- **Computação Paralela:** O *NIMBLE* fornece suporte integrado para computação paralela, permitindo distribuir a carga computacional entre vários núcleos ou até mesmo entre várias máquinas.

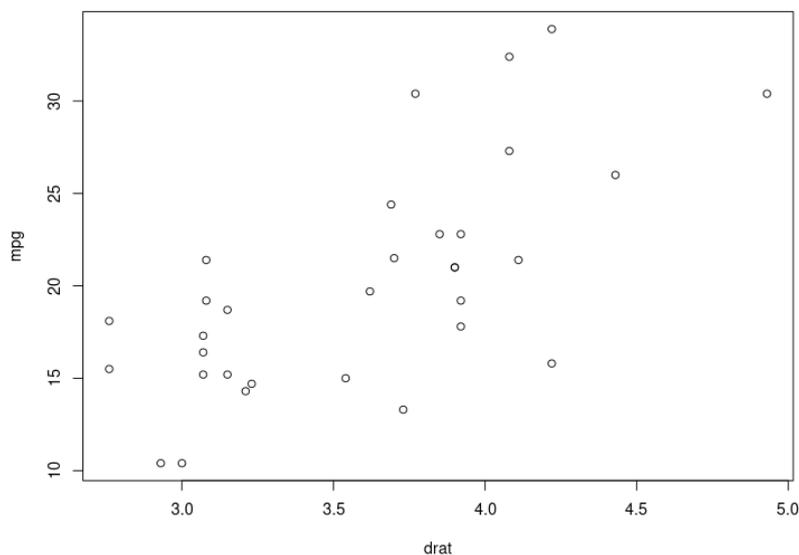
Para escrever um programa utilizando o *NIMBLE*, é necessário seguir as seguintes etapas:

- 1) Definir o modelo;
- 2) Definir objeto *NIMBLE*;
- 3) Compilar o modelo;
- 4) Configurar o MCMC;
- 5) Executar a simulação; e, por fim,
- 6) Realizar as inferências desejadas.

A.0.1 Utilizando o NIMBLE

Um dos *datasets* mais famosos da linguagem R é o *mtcars*, composto por onze diferentes características de 32 veículos diferentes do início dos anos 1970. Ele será a base para a construção do nosso modelo de regressão linear e demonstração do uso do *NIMBLE*. Em particular modelaremos o consumo de milhas por galão através da relação do eixo traseiro, *drat*, figura 23. Nosso modelo bayesiano será descrito através de uma normal com média $\beta \times \text{drat}$ e desvio padrão de σ . Portanto, teremos a seguinte equação em nosso modelo, que será implementada no *NIMBLE*:

Figura 23 – Relação entre MPG vs Dart



$$\beta \sim \mathcal{N}(0, 1)$$

$$\sigma \sim \Gamma(1, 1)$$

$$mpg_i \sim \mathcal{N}(\beta \times dart_i, \sigma)$$

A.0.2 Definir o modelo

Na primeira etapa da utilização do *NIMBLE* é necessário declarar as variáveis, especificando suas distribuições e relações conforme descritas no modelo. Para exemplificar a relação entre "mpg" e "dart", apresentamos a seguinte declaração:

```

1 library(nimble)
2
3 data(mtcars)
4
5 x <- mtcars$drat
6 y <- mtcars$mpg
7
8 code <- nimbleCode({
9   # declarando as variaveis
10  # distribuição a priori do coeficiente linear da equação
11  beta ~ dnorm(0, 1)
12
13  # distribuição a priori do desvio padrão
14  sigma ~ dgamma(1, 1)
15
16  for (i in 1:N) {
17    #equação de verosimilhança dos dados observados
18    y[i] ~ dnorm(beta * x[i], sigma)
19  }
20 })
21
22 constants <- list(N = length(x))
23 data <- list(y = y, x = x)
24 inits <- list(beta = 0, sigma = 1)

```

A.0.3 Definindo o objeto o NIMBLE

Após definir o modelo, os dados e as constantes, é possível criar um objeto *NIMBLE*, que servirá como base para os cálculos subsequentes.

```

1 model <- nimbleModel(code, constants = constants, data = data, inits = inits)

```

A.0.4 Compilar o modelo

Após definir o objeto, é necessário compilá-lo para a linguagem C++, que proporciona maior eficiência nos cálculos computacionais. Para realizar essa etapa, utiliza-se o comando `compileNimble`.

```
1 Cmodel <- compileNimble(model, showCompilerOutput = T)
```

A.0.5 Configurar o MCMC - Monte Carlo Markov Chain

Na terceira etapa, configuramos o MCMC para calcular as estatísticas do modelo proposto. Primeiro, utilizamos a função `configureMCMC` para habilitar o cálculo do WAIC (Watanabe-Akaike Information Criteria), uma medida que ajuda na seleção de modelos. Em seguida, especificamos as variáveis que desejamos monitorar, que são β , σ , e y . Essas variáveis são atribuídas ao nosso objeto de configuração. Em seguida, construímos o modelo MCMC utilizando o método `buildMCMC`, que compila o modelo para otimizar a velocidade de execução. Dessa forma, obtemos resultados mais eficientes.

```
1 MCMCConf <- configureMCMC(Cmodel, enableWAIC = TRUE)
2 monitors <- c('beta'
3             , 'sigma'
4             , 'logProb_y'
5             )
6 MCMCConf$addMonitors(monitors)
7 exampleModelMCMC <- buildMCMC(MCMCConf, print=T)
8 cMCMC <- compileNimble(exampleModelMCMC,
9                       project=Cmodel, resetFunctions=T, showCompilerOutput = TRUE)
```

O código acima funciona para modelos não sequenciais. Em casos sequenciais, que são Sequential Monte Carlo (SMC), é necessário um conjunto maior de comandos para especificar corretamente a construção do simulador de MCMC, permitindo assim explorarmos outros *samplers* e formas de avaliar o modelo, conforme a sequência de código abaixo:

```
1 # CexampleModel é o modelo nimble já compilado
2 exampleModelMCMCConf <- configureMCMC(CexampleModel)
3 exampleModelMCMCConf$resetMonitors()
4 monitors <- c('observado'
5             , 'x_carbono_lento_rapido'
6             )
7 exampleModelMCMCConf$addMonitors(monitors)
8
9 # variavel considerada latente, ou seja, não observamos ela diretamente
```

```

10 # ela tem influência direta na variavel que observamos
11 latentVariables <- 'x_carbono_lento_rapido'
12 # é a classe de filtro que queremos utilizar, podendo ser outros tipos como também
13 # o Kalman Filter
14 bootstrapFilter <- buildBootstrapFilter(cabornModel,
15                                       latentVariables,
16                                       control = list(saveAll=T,
17                                                     smoothing=T,
18                                                     initModel=T,
19                                                     thresh=0.70))
20 # lista de variaveis que queremos impactar com a mudança da variavel latent
21 targets <- c( 'lambda_lento'
22              , 'lambda_rapido'
23             )
24
25 # número de partículas para samplear e estimar o valor da distribuição a posteriori
26 numParticles <- 100
27 exampleModelMCMCConf$addSampler(target = targets,
28                                 type= 'RW_PF_block',
29                                 print=T,
30                                 control= list(pf=bootstrapFilter,
31                                               adaptive=T,
32                                               latents=latentVariables,
33                                               pfNparticles=numParticles))
34
35
36 exampleModelMCMC <- buildMCMC(exampleModelMCMCConf, print=T)
37 cMCMC <- compileNimble(exampleModelMCMC, project=cabornModel,
38                       resetFunctions=T, showCompilerOutput = TRUE)

```

A.0.6 Executando o NIMBLE

Para obter as inferências desejadas do modelo *NIMBLE*, executamos o comando `runMCMC` com os seguintes parâmetros:

- **niter:** Indica o número total de iterações que desejamos executar no código. No nosso caso, definimos o valor como 5000, o que significa que o modelo será executado cinco mil vezes.
- **nburnin:** Especifica quantas iterações iniciais devem ser descartadas antes de começarmos a considerar as amostras. Essa etapa é útil para garantir que o algoritmo tenha tempo suficiente para atingir a convergência.
- **thin:** Define o espaçamento entre as amostras que queremos capturar. Em alguns casos, as amostras podem ser autocorrelacionadas, portanto, desejamos capturar os dados apenas a cada 'n' unidades de tempo para reduzir a autocorrelação.

- **nchains:** Indica quantas cadeias de Markov desejamos executar. Ter pelo menos duas cadeias é necessário para avaliar a convergência do modelo.
- **setSeed:** Permite definir um valor de semente para reproduzir os experimentos. Isso garante que os resultados sejam consistentes e replicáveis.
- **summary:** Salva o resultado sumarizado das simulações, fornecendo informações resumidas sobre as estatísticas do modelo, como média, mediana, intervalo de confiança, entre outros.

Esses parâmetros permitem controlar e ajustar o processo de execução do modelo MCMC para obter as inferências desejadas.

```

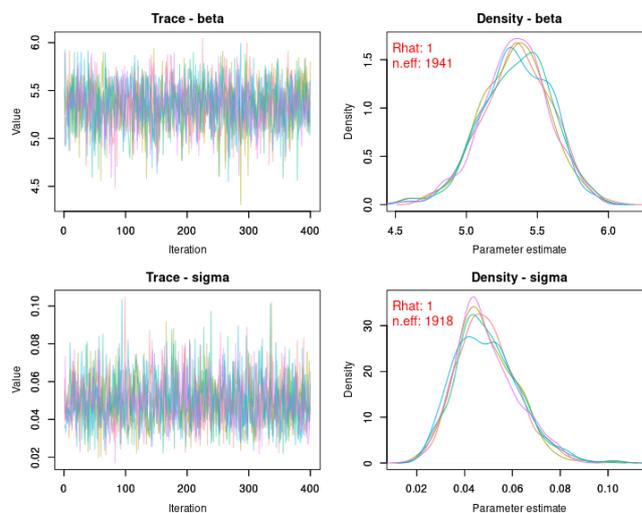
1 cMCMCout <- runMCMC(cMCMC,niter = 5000, nburnin = 1000, thin = 10,
2                   setSeed = F,
3                   summary = T, nchains = 5)

```

A.0.7 Inferência do modelo

A inferência do modelo é obtida através da execução do comando **runMCMC**. Primeiro, compilamos os resultados de todas as cadeias de Markov simuladas. Em seguida, utilizamos os comandos do pacote **MCMCvis** para visualizar a convergência e os resultados do modelo de forma mais clara e compreensível. A Figura 24 ilustra um exemplo dessa visualização.

Figura 24 – Convergência do modelo de regressão linear entre MPG e Dart



Essa abordagem nos permite analisar e interpretar os resultados da inferência de forma mais eficaz, facilitando a compreensão dos resultados do modelo bayesiano.

```
1 mcmc.output <- mcmc.list( as.mcmc(cMCMCout$samples$chain1)
2                          ,as.mcmc(cMCMCout$samples$chain2)
3                          ,as.mcmc(cMCMCout$samples$chain3)
4                          ,as.mcmc(cMCMCout$samples$chain4)
5                          ,as.mcmc(cMCMCout$samples$chain5)
6 )
7
8 MCMCsummary(object = mcmc.output, round = 2)
9 MCMCplot(object = mcmc.output, params = 'beta')
10 MCMCplot(object = mcmc.output, params = 'sigma')
11
12
13 MCMCtrace(object = mcmc.output,
14           pdf = FALSE, # no export to PDF
15           ind = TRUE, # separate density lines per chain
16           Rhat = TRUE,
17           n.eff = TRUE,
18           params = c("beta", "sigma"))
```

CÓDIGOS

B.1 Implementação base de dados artificial

```
1 Total <- 150
2
3 total_massa_carbono <- 30000
4 total_massa_carbono_slow <- 25000
5 total_massa_entrada <- 5000
6 reducao <- 1
7
8 razao <- 2.04E-06/5.05E-04
9 slow_rate <- (razao/reducao)/10
10 fast_rate <- 1/10
11
12 X_carbono_sample <- c()
13 X_carbono <- rep(0,Total)
14 X_carbono_smooth <- rep(0,Total)
15 X_carbono_slow <- rep(0,Total)
16 X_carbono_slow_only <- rep(0,Total)
17 X_carbono_fast <- rep(0,Total)
18
19 sd_entrada <- 300
20 sd_medida <- 500
21
22 list_carbon_fast <- list()
23 list_start_carbon_fast <- c()
24 idx_carbon_fast <- 1
25 indice_conversao_fast_slow <- 1
26
27 erro_medida <- 100
28 tempo_metrica <- 30
```

```
29 tempo_momento_conversao <- 45
30 entrada_de_carbono_sim <- rep(0,Total)
31
32 for(t in 1:Total){
33
34   tempo_entrada <- t %% tempo_metrica
35   if(tempo_entrada == 1){
36     entrada_de_carbono_sim[t] <- rnorm(1, mean=total_massa_entrada,sd=sd_entrada)
37   }
38
39   if( t > 1){
40
41     if(tempo_entrada == 1){
42       X_carbono_fast[t-1] <- X_carbono_fast[t-1] + entrada_de_carbono_sim[t]
43     }
44
45     tmp <- 0
46     tempo_conversao <- t %% tempo_momento_conversao
47     if(tempo_conversao == 1){
48       tmp <- X_carbono_fast[t-1]*0.35
49       X_carbono_fast[t-1] <- X_carbono_fast[t-1] - tmp
50     }
51
52     X_carbono_slow_only[t] <- X_carbono_slow_only[t-1]*exp(-slow_rate)
53
54     X_carbono_slow[t] <- (X_carbono_slow[t-1] + tmp)*exp(-slow_rate)
55     X_carbono_fast[t] <- X_carbono_fast[t-1]*exp(-fast_rate)
56
57
58     X_carbono[t] <- X_carbono_slow[t] + X_carbono_fast[t] + rnorm(1,0,sd_medida)
59
60     X_carbono_smooth[t] <- X_carbono_slow[t] + X_carbono_fast[t]
61   }else{
62
63     X_carbono_slow[1] <- total_massa_carbono_slow
64     X_carbono_slow_only[1] <- total_massa_carbono_slow
65     X_carbono_fast[1] <- entrada_de_carbono_sim[1]
66
67     X_carbono[1] <- X_carbono_slow[1] + X_carbono_fast[1] + rnorm(1,0,sd_medida)
68
69     X_carbono_smooth[1] <- X_carbono_slow[1] + X_carbono_fast[1]
70   }
71 }
```

B.2 Implementação do modelo proposto

O modelo matemático representado pela Figura 3 foi implementado considerando as etapas descritas no apêndice de revisão do *NIMBLE*. Inicialmente, foram importadas as bibliotecas necessárias e, em seguida, foi feita a construção do modelo conforme descrito a seguir.

```

1  output <- require(nimble)
2  if(!output){
3    instal.package('nimble')
4  }
5  output <- require(nimbleSMC)
6  if(!output){
7    instal.package('nimbleSMC')
8  }
9  output <- require(coda)
10 if(!output){
11   instal.package('codal')
12 }
13 output <- require(MCMCvis)
14 if(!output){
15   instal.package('MCMCvis')
16 }
17
18 bayesCarbonStockOverTime <- nimbleCode({
19
20   lambda_lento   ~ T(dnorm(400,50),100,1000) #dunif(300,600)
21   lambda_rapido  ~ T(dnorm(1000,50),100,2000) #dunif(900,1100)
22
23   erro_observacao ~ T(dnorm(10,1),0.01,1000)
24   erro_transicao_rapido ~ T(dnorm(10,1),0.01,1000)
25   erro_transicao_lento ~ T(dnorm(10,1),0.01,1000)
26
27   x_carbono_lento_rapido[1,1] ~ T(dnorm(25000,10),0,1000000)
28   x_carbono_lento_rapido[2,1] ~ T(dnorm(5000,10),0,1000000)
29
30   entrada_de_carbono ~ T(dnorm(5000,10),1,100000)
31
32   estrutura_dinamica[1] <- x_carbono_lento_rapido[1,1] + x_carbono_lento_rapido[2,1]
33
34   observado[1] ~ dnorm(estrutura_dinamica[1], erro_observacao)
35
36   conversao_rapido_lento[1] <- 0
37
38   taxa_conversao ~ T(dnorm(0.34,1),0.05,1) #dunif(0.20,0.50)
39   tempo_momento_conversao_rapido_devagar_poisson ~ T(dnorm(44,1),15,65)
40
41   for(t in 2:Tempo_Total){
42

```

```

43   conversao_rapido_lento[t] <- x_carbono_lento_rapido[2,t-1]*(
44     calculate_return_distribution(t,
45     tempo_momento_conversao_rapido_devagar_poisson,taxa_conversao) )
46
47   x_carbono_lento_rapido[1,t] ~ dnorm( ( x_carbono_lento_rapido[1,t-1]
48     + conversao_rapido_lento[t]
49     )*exp(-lambda_lento/1000000)
50     ,erro_transicao_lento)
51
52   x_carbono_lento_rapido[2,t] ~ dnorm( ( x_carbono_lento_rapido[2,t-1]
53     - conversao_rapido_lento[t]
54     + entrada_de_carbono*ifelse( t %% Tempo_entrada_carbono) == 1,1,0)
55     )*exp(-lambda_rapido/10000) #
56     ,erro_transicao_rapido)
57
58   estrutura_dinamica[t] <- x_carbono_lento_rapido[1,t] + x_carbono_lento_rapido[2,t]
59
60   observado[t] ~ dnorm(estrutura_dinamica[t], erro_observacao)
61 }
62 })
63
64 calculate_return_distribution <- nimbleFunction(
65   run = function(t = integer(), Tempo_conversao_rapido_devagar = integer(),
66     taxa_conversao = double()){
67     returnType(double(0))
68     bollean_logic <- t %% Tempo_conversao_rapido_devagar
69     if (is.na(bollean_logic)) {
70       print("Number is :", t, Tempo_conversao_rapido_devagar)
71       values <- 0
72     }else {
73       if( bollean_logic == 1){
74         values <- taxa_conversao
75       }else {
76         values <- 0
77       } }
78
79     return(values)
80   }
81 )

```

Em seguida, preparamos a base de dados como *input* para as simulações e também construímos o objeto do modelo *NIMBLE* para compilação.

Realizamos a configuração do algoritmo de MCMC, adicionando as variáveis que serão monitoradas. No nosso caso, configuramos os compartimentos de carbono como variáveis latentes. Em seguida, configuramos o filtro de Bootstrap, definindo os seus alvos, ou seja, as variáveis que ele irá modificar. Também definimos o amostrador do método.

Por fim, criamos o objeto MCMC do *NIMBLE*, compilamos nosso modelo e configuramos o método de amostragem.

```

1 cabornModel <- nimbleModel(code=bayesCarbonStockOverTime,
2                           constants = list(
3                               Tempo_Total=length_data_to_use
4                               ,Tempo_entrada_carbono = tempo_metrica
5                               ),
6                           data = simulatedData)
7
8 CexampleModel <- compileNimble(cabornModel,showCompilerOutput = T)

```

```

1 exampleModelMCMCConf <- configureMCMC(CexampleModel)
2 exampleModelMCMCConf$resetMonitors()
3 monitors <- c('observado'
4               , 'logProb_observado'
5               , 'conversao_rapido_lento'
6               , 'erro_observacao'
7               , 'erro_transicao_rapido'
8               , 'erro_transicao_lento'
9               , 'lambda_lento'
10              , 'lambda_rapido'
11              , 'estrutura_dinamica'
12              , 'entrada_de_carbono'
13              , 'x_carbono_lento_rapido'
14              , 'taxa_conversao'
15              , 'tempo_momento_conversao_rapido_devagar_poisson'
16             )
17 exampleModelMCMCConf$addMonitors(monitors)
18
19 latentVariables <- 'x_carbono_lento_rapido'
20 bootstrapFilter <- buildBootstrapFilter(cabornModel,
21                                         latentVariables,
22                                         control = list(saveAll=T,
23                                                         smoothing=T,
24                                                         initModel=T,
25                                                         thresh=0.70))
26 targets <- c( 'erro_observacao'
27               , 'erro_transicao_rapido'
28               , 'erro_transicao_lento'
29               , 'lambda_lento'
30               , 'lambda_rapido'
31               , 'entrada_de_carbono'
32               , 'taxa_conversao'
33               , 'tempo_momento_conversao_rapido_devagar_poisson'
34             )

```

```

35
36 numParticles <- 100
37 exampleModelMCMCConf$addSampler(target = targets,
38                               type= 'RW_PF_block',
39                               print=T,
40                               control= list(pf=bootstrapFilter,
41                                             adaptive=T,
42                                             latents=latentVariables,
43                                             pfNparticles=numParticles))
44
45 exampleModelMCMC <- buildMCMC(exampleModelMCMCConf, print=T)
46 cMCMC <- compileNimble(exampleModelMCMC,project=cabornModel,
47                       resetFunctions=T,showCompilerOutput = TRUE)

```

Em nossa última etapa, executamos o algoritmo de MCMC com as seguintes configurações: um total de 30 mil iterações, um período de *burn-in* de 10 mil iterações, capturando os dados a cada 10 iterações. Não fixamos uma semente geradora e utilizamos cinco cadeias de Markov.

```

1 cMCMCout <- runMCMC(cMCMC,niter = 30000, nburnin = 10000, thin = 10,
2                   setSeed = F,
3                   summary = T, nchains = 5)
4
5 posteriorSamples <- as.mcmc(as.matrix(cMCMCout$samples$chain2))
6 chainsList <- mcmc.list( as.mcmc(cMCMCout$samples$chain1)
7                       ,as.mcmc(cMCMCout$samples$chain2)
8                       ,as.mcmc(cMCMCout$samples$chain3)
9                       ,as.mcmc(cMCMCout$samples$chain4)
10                      ,as.mcmc(cMCMCout$samples$chain5)
11 )

```

Para avaliação do modelo, foi executado os seguintes comandos que são provenientes da biblioteca MCMCVis.

```

1 MCMCtrace(object = chainsList,
2           pdf = FALSE,
3           ind = TRUE,
4           Rhat = TRUE, # add Rhat
5           n.eff = TRUE, # add eff sample size
6           params = c('x_carbono_lento_rapido'))
7
8 MCMCtrace(object = chainsList,
9           pdf = FALSE,
10          ind = TRUE,
11          Rhat = TRUE, # add Rhat

```

```
12     n.eff = TRUE, # add eff sample size
13     params = c('entrada_de_carbono'))
14
15
16 MCMCtrace(object = chainsList,
17           pdf = FALSE,
18           ind = TRUE,
19           Rhat = TRUE, # add Rhat
20           n.eff = TRUE, # add eff sample size
21           params = c('lambda_rapido', 'lambda_lento', 'taxa_conversao'))
22
23 MCMCtrace(object = chainsList,
24           pdf = FALSE,
25           ind = TRUE,
26           Rhat = TRUE, # add Rhat
27           n.eff = TRUE, # add eff sample size
28           params = c('tempo_momento_conversao_rapido_devagar_poisson'))
29
30 MCMCtrace(object = chainsList,
31           pdf = FALSE,
32           ind = TRUE,
33           Rhat = TRUE, # add Rhat
34           n.eff = TRUE, # add eff sample size
35           params = c('erro_observacao', 'erro_transicao_rapido', 'erro_transicao_lento'))
```

GRÁFICOS

C.1 Convergência do algoritmo de MCMC

C.1.1 *Matéria orgânica de solo canadense 244003*

Para as diferentes variáveis propostas para o modelo de solo canadense 244003, em todas as execuções foram encontrados \hat{R} próximo de 1, garantindo a convergência da cadeia de Markov.

Figura 25 – Evolução da simulação de MCMC dos compartimentos de matéria orgânica do sistema para o solo canadense 244003 no início do ciclo

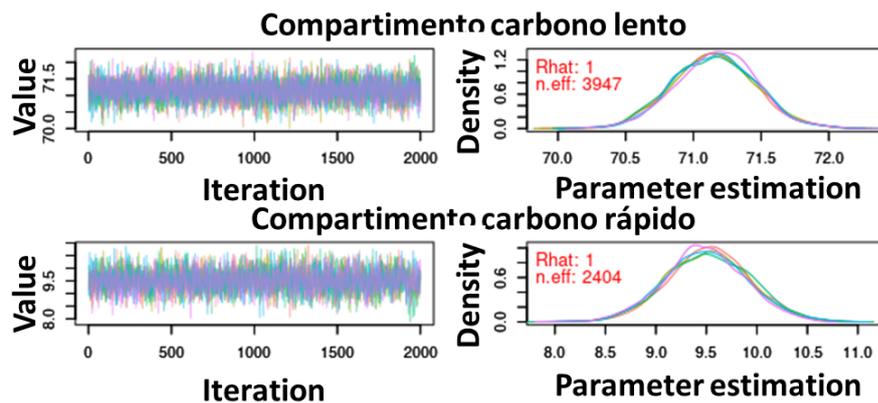


Figura 26 – Evolução da simulação de MCMC para entrada de carbono do sistema para o solo canadense 244003

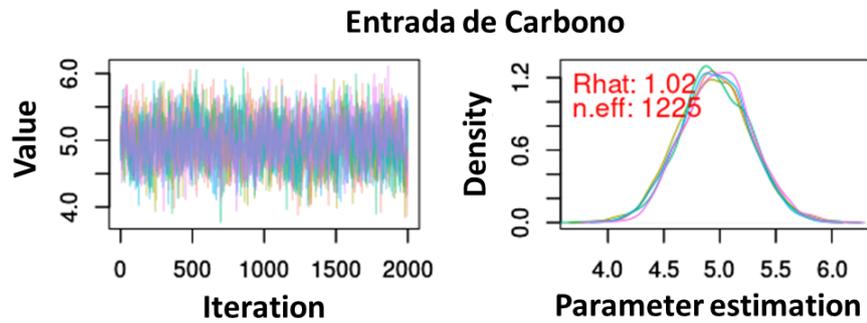


Figura 27 – Evolução da simulação de MCMC para $\lambda_{rápido}$ para o sistema de solo canadense 244003

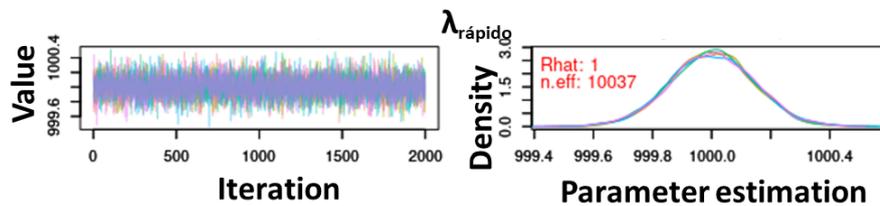


Figura 28 – Evolução da simulação de MCMC para λ_{lento} para o sistema de solo canadense 244003

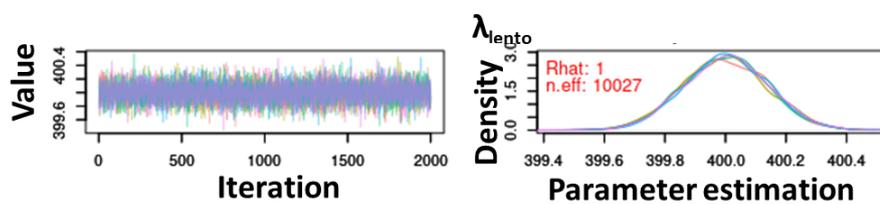


Figura 29 – Evolução da simulação de MCMC para a taxa de conversão para o sistema de solo canadense 244003

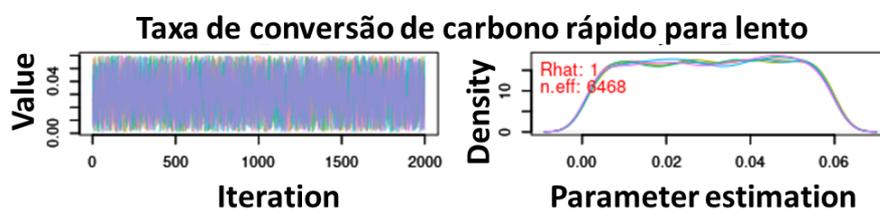


Figura 30 – Evolução da simulação de MCMC para taxa erro na observação do carbono total para o sistema de solos canadense 244003

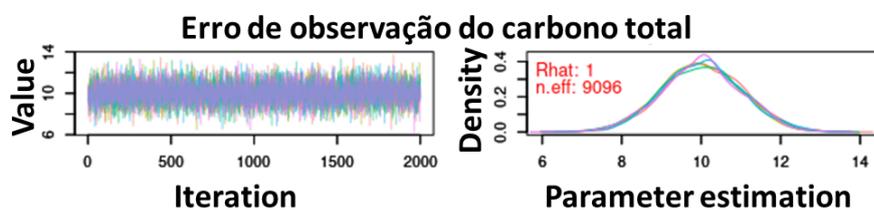


Figura 31 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados rápidos para o sistema de solos canadense 244003

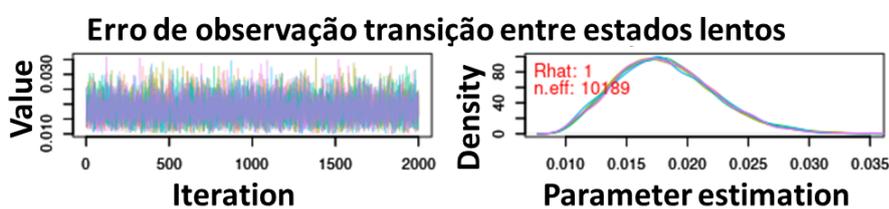
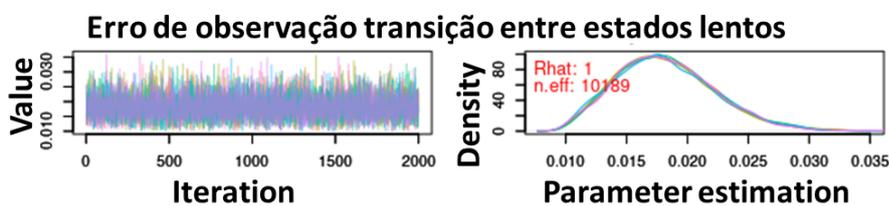


Figura 32 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados lentos para o sistema de solos canadense 244003



C.1.2 Matéria orgânica de solo canadense 244007

Para as diferentes variáveis propostas para o modelo de solo canadense 244007, em todas as execuções foram encontrados \hat{R} próximo de 1, garantindo a convergência da cadeia de Markov.

Figura 33 – Evolução da simulação de MCMC dos compartimentos de matéria orgânica do sistema para o solo canadense 244007 no início do ciclo

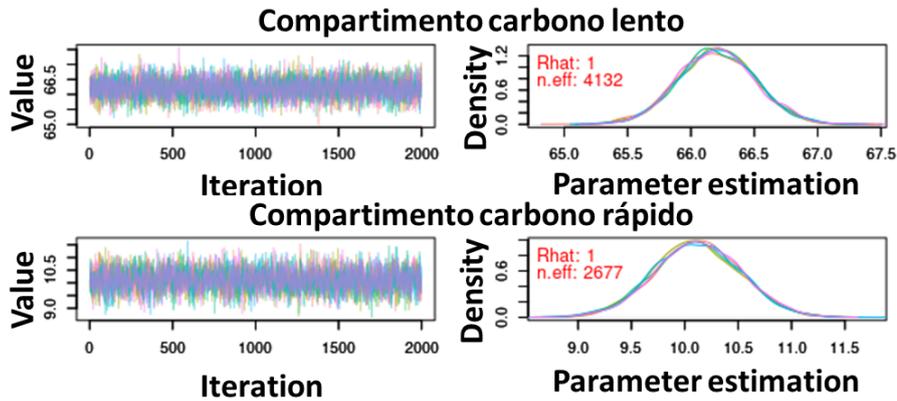


Figura 34 – Evolução da simulação de MCMC para entrada de carbono do sistema para o solo canadense 244007

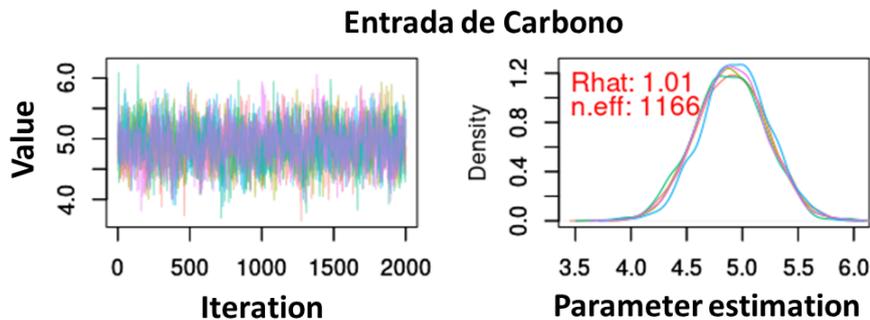


Figura 35 – Evolução da simulação de MCMC para $\lambda_{\text{rápido}}$ para o sistema de solo canadense 244007

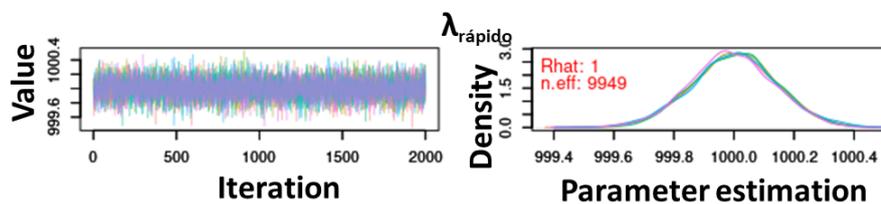


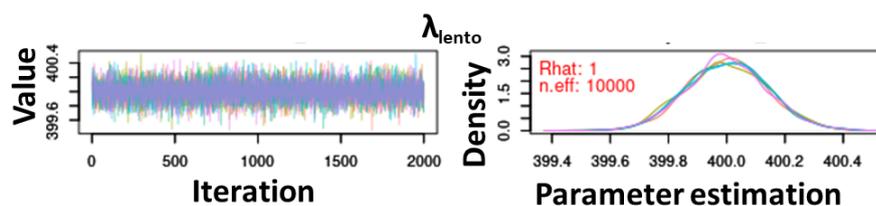
Figura 36 – Evolução da simulação de MCMC para λ_{lento} para o sistema de solo canadense 244007

Figura 37 – Evolução da simulação de MCMC para a taxa de conversão para o sistema de solo canadense 244007

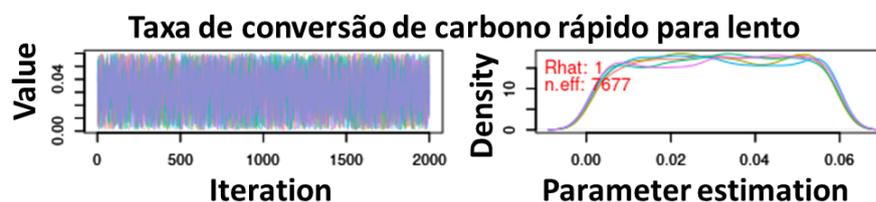


Figura 38 – Evolução da simulação de MCMC para taxa erro na observação do carbono total para o sistema de solos canadense 244007

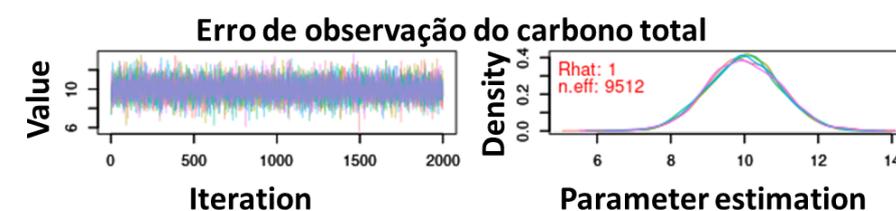


Figura 39 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados rápidos para o sistema de solos canadense 244007

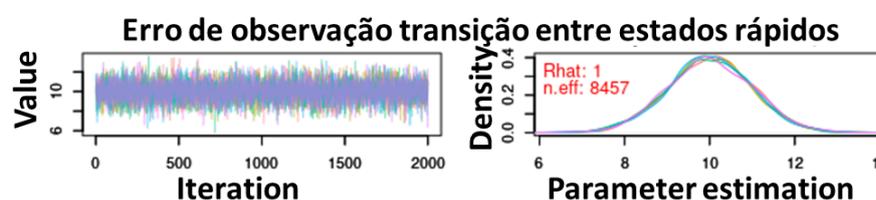


Figura 40 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados lentos para o sistema de solos canadense 244007

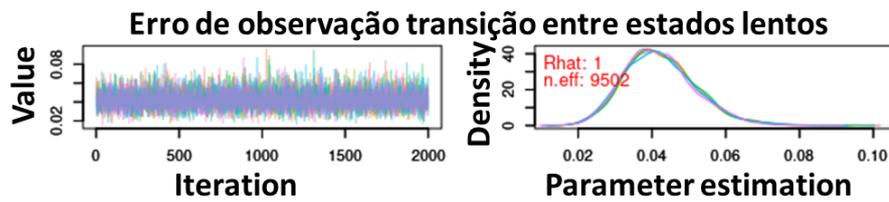
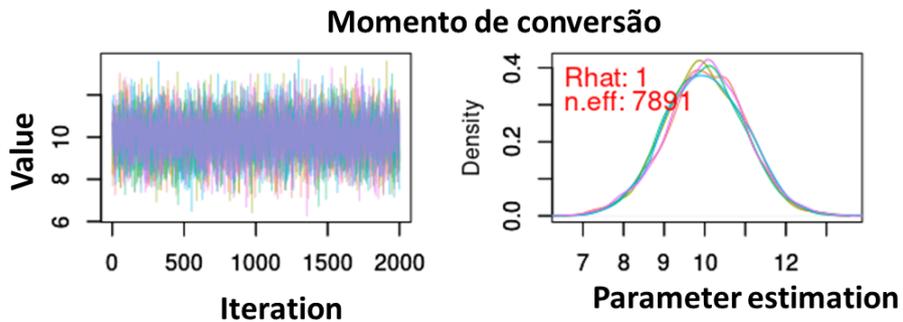


Figura 41 – Evolução da simulação de MCMC para o momento de conversão do carbono do compartimento rápido para o lento no sistema de solo canadense 244007



C.1.3 Matéria orgânica de solo canadense 244014

Para as diferentes variáveis propostas para o modelo de solo canadense 244014, em todas as execuções foram encontrados \hat{R} próximo de 1, garantindo a convergência da cadeia de Markov.

Figura 42 – Evolução da simulação de MCMC dos compartimentos de matéria orgânica do sistema para o solo canadense 244014 no início do ciclo

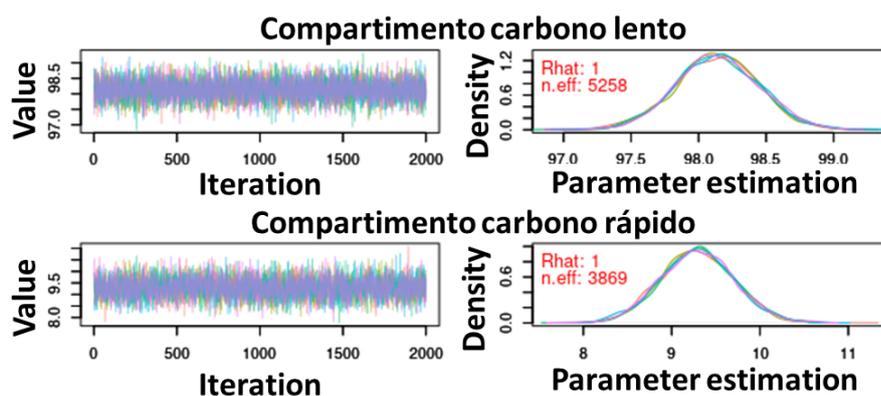


Figura 43 – Evolução da simulação de MCMC para entrada de carbono do sistema para o solo canadense 244014

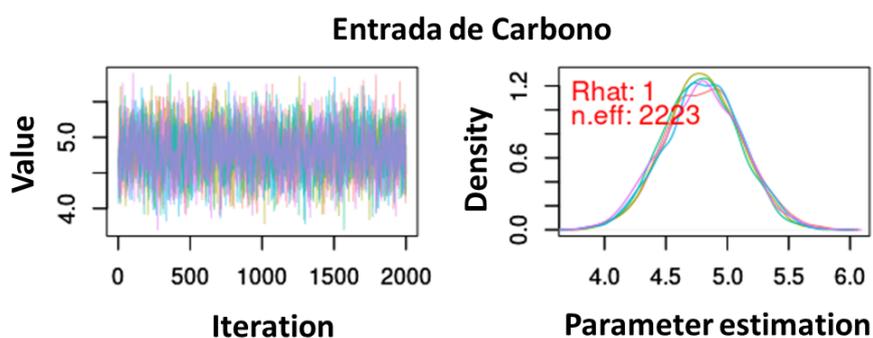


Figura 44 – Evolução da simulação de MCMC para $\lambda_{rápido}$ para o sistema de solo canadense 244014

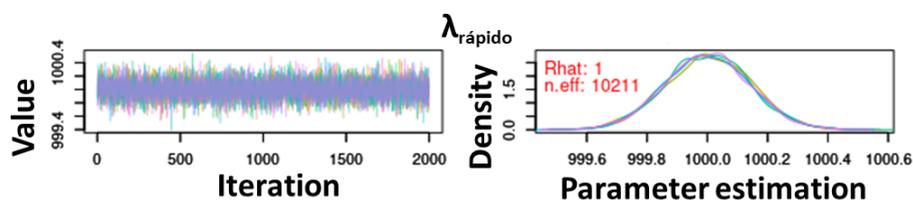


Figura 45 – Evolução da simulação de MCMC para λ_{lento} para o sistema de solo canadense 244014

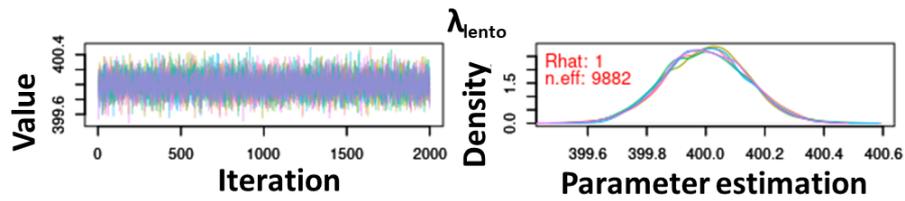


Figura 46 – Evolução da simulação de MCMC para a taxa de conversão para o sistema de solo canadense 244014

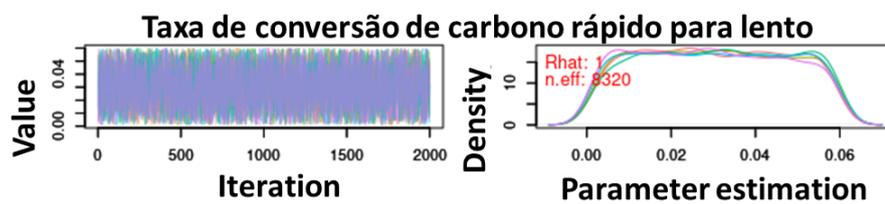


Figura 47 – Evolução da simulação de MCMC para taxa erro na observação do carbono total para o sistema de solos canadense 244014

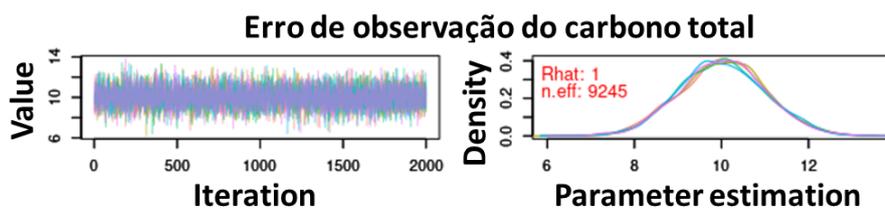


Figura 48 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados rápidos para o sistema de solos canadense 244014

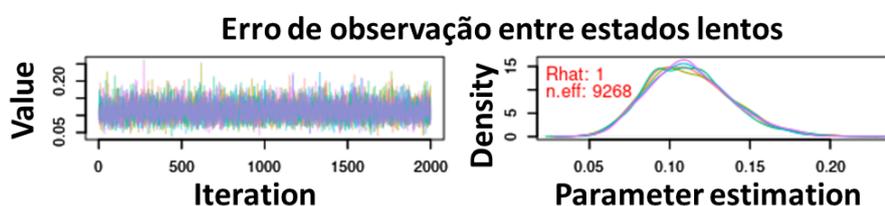


Figura 49 – Evolução da simulação de MCMC para taxa erro na observação da evolução dos estados lentos para o sistema de solos canadense 244014

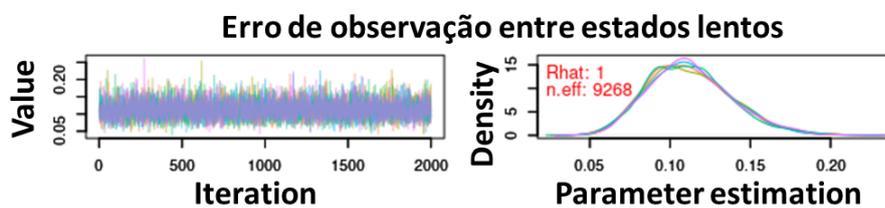
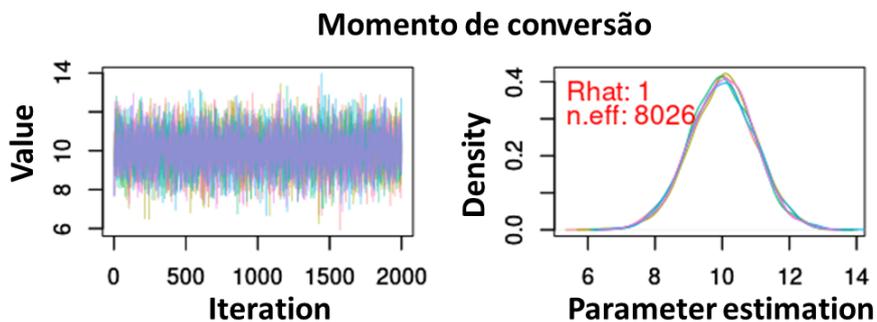


Figura 50 – Evolução da simulação de MCMC para o momento de conversão do carbono do compartimento rápido para o lento no sistema de solo canadense 244014

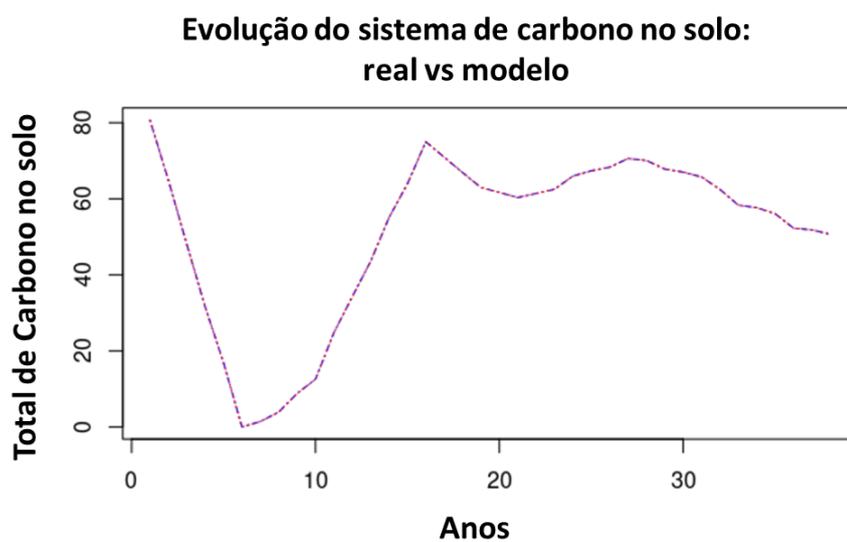


C.2 Precisão

C.2.1 Precisão do modelo do solo canadense 244003

Visualmente, temos a percepção de que o modelo teve um bom desempenho, possuindo MAPE igual a 5,29%, sendo capaz de prever com qualidade a evolução da matéria orgânica no solo.

Figura 51 – Comparação visual entre o total de carbono previsto pelo modelo e o total de entrada no sistema do solo canadense 244003

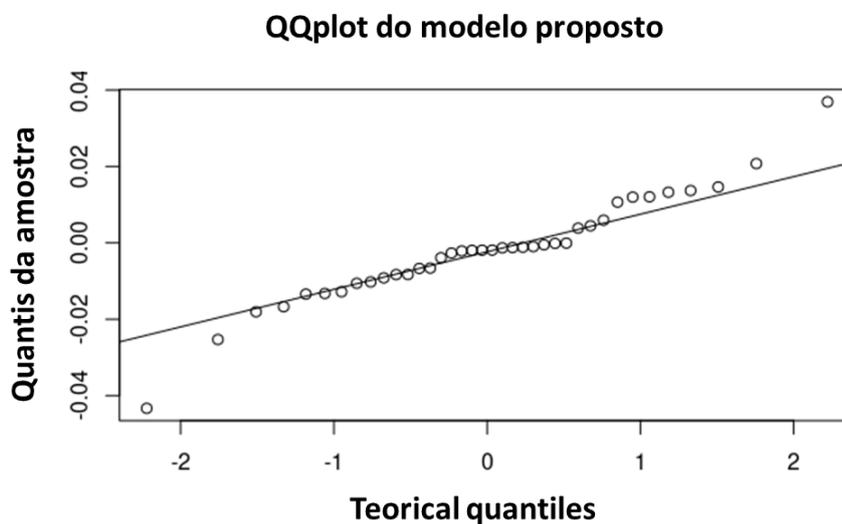


Podemos inferir que o modelo se ajustou bem aos dados, com base nas figuras 52 e 53, forneceu uma boa explicação para o conjunto de dados simulados propostos.

Figura 52 – Diferença entre o total de carbono previsto pelo modelo versus a quantidade de entrada no sistema do solo canadense 244003

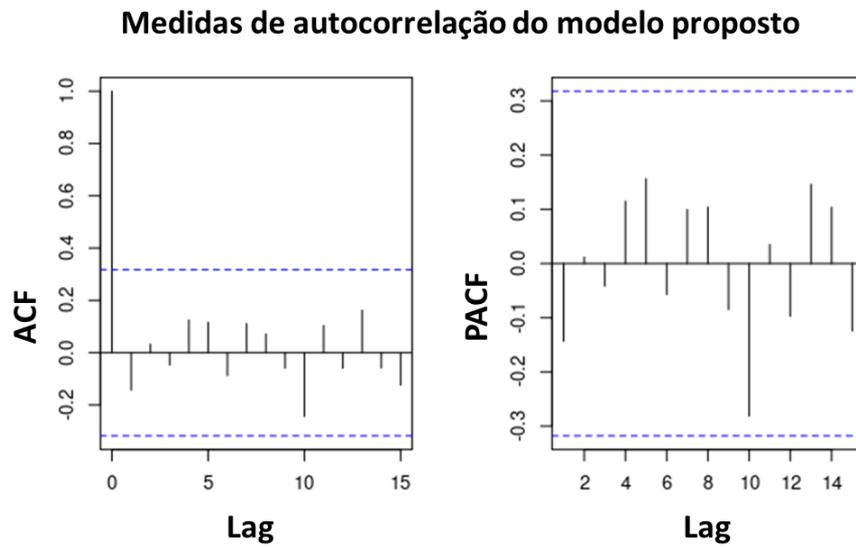


Figura 53 – Normal Q-Qplot do sistema de solo canadense 244003



Na última figura, figura 54 é possível notar que o modelo para o solo 244003 conseguiu ajustar corretamente aos dados.

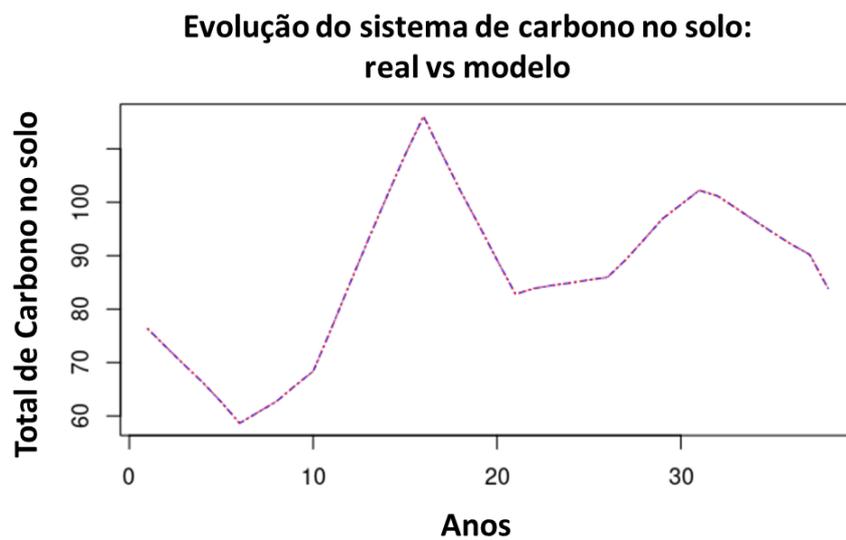
Figura 54 – Gráfico mostrando o compartimento de ACF e PACF do resíduo do modelo no sistema de solo canadense 244003



C.2.2 Precisão do modelo do solo canadense 244007

Visualmente, temos a percepção de que o modelo teve um bom desempenho, possuindo MAPE igual a 0,017%, sendo capaz de prever com qualidade a evolução da matéria orgânica no solo.

Figura 55 – Comparação visual entre o total de carbono previsto pelo modelo e o total de entrada no sistema do solo canadense 244007



Podemos inferir que o modelo se ajustou bem aos dados, com base nas figuras 56 e 57, forneceu uma boa explicação para o conjunto de dados simulados propostos, salve um pequeno outlier.

Figura 56 – Diferença entre o total de carbono previsto pelo modelo versus a quantidade de entrada no sistema do solo canadense 244007

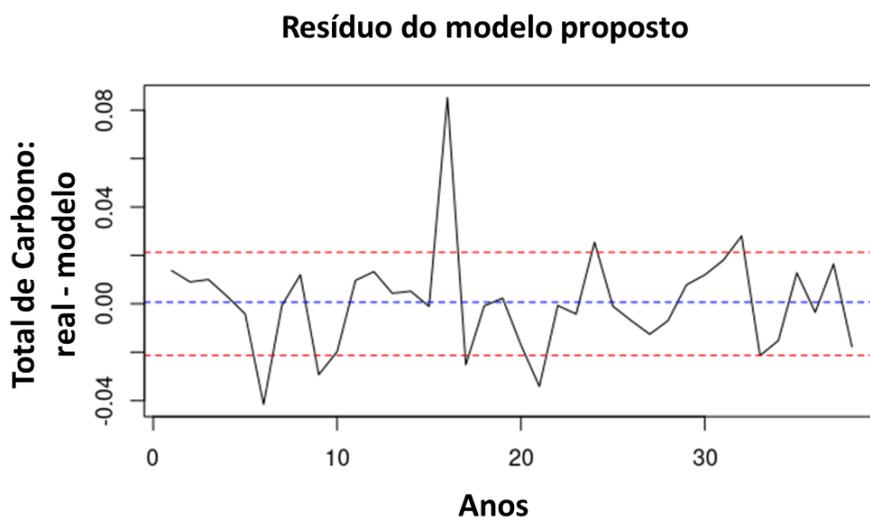
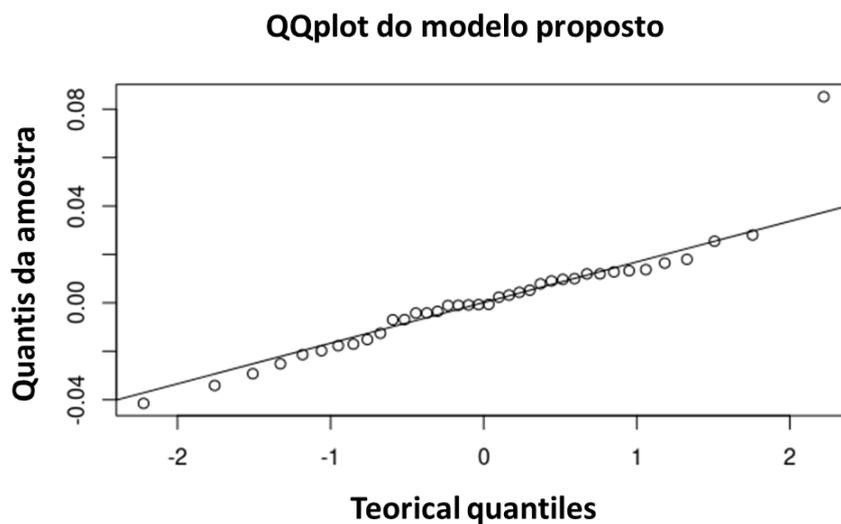
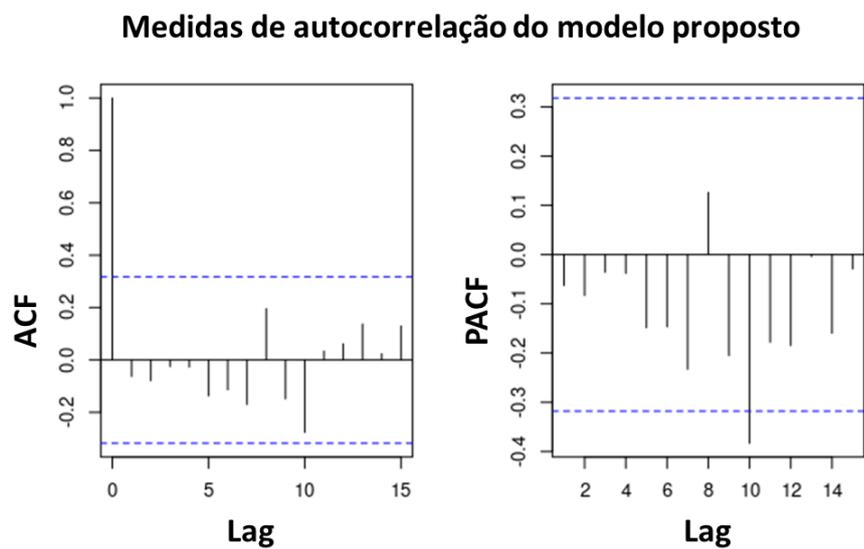


Figura 57 – Normal Q-Qplot do sistema de solo canadense 244007



Na última figura, figura 58 é possível notar que o modelo para o solo 244007 exceto por uma pequena correlação parcial no *lag* 10.

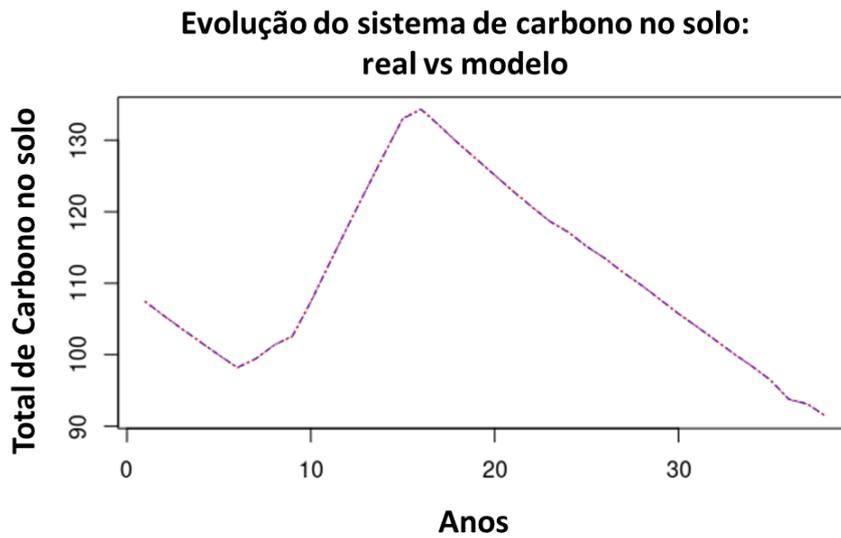
Figura 58 – Gráfico mostrando o compartimento de ACF e PACF do resíduo do modelo no sistema de solo canadense 244007



C.2.3 Precisão do modelo do solo canadense 2440014

Visualmente, temos a percepção de que o modelo teve um bom desempenho, possuindo MAPE igual a 0,021%, sendo capaz de prever com qualidade a evolução da matéria orgânica no solo.

Figura 59 – Comparação visual entre o total de carbono previsto pelo modelo e o total de entrada no sistema do solo canadense 244014

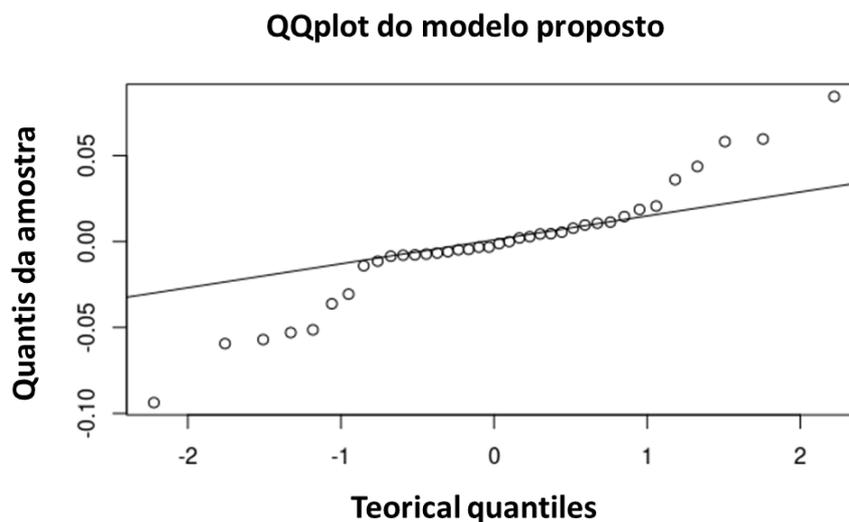


Podemos inferir que o modelo se ajustou bem aos dados, com base nas figuras 60 e 61, forneceu uma boa explicação para o conjunto de dados simulados propostos, salve um pequeno conjunto de outlier simétrico.

Figura 60 – Diferença entre o total de carbono previsto pelo modelo versus a quantidade de entrada no sistema do solo canadense 244014

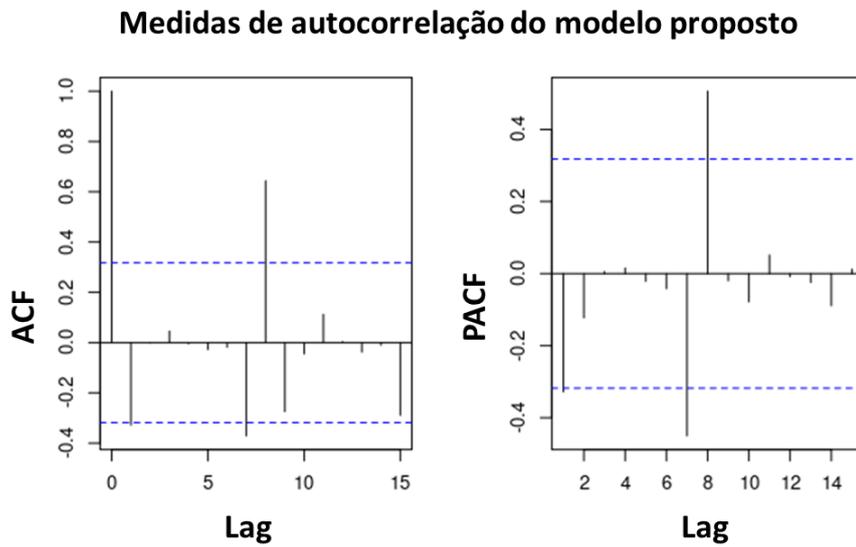


Figura 61 – Normal Q-Qplot do sistema de solo canadense 244014



Na última figura, figura 62 é possível notar que o modelo para o solo 244014 exceto por uma pequena correlação parcial e total nos *lags* 7 e 9.

Figura 62 – Gráfico mostrando o compartimento de ACF e PACF do resíduo do modelo no sistema de solo canadense 244014



REFERÊNCIAS

BATLLE-AGUILAR, J.; BROVELLI, A.; PORPORATO, A.; BARRY, D. A. Modelling soil carbon and nitrogen cycles during land use change. a review. **Agronomy for Sustainable Development** volume, 2010. Nenhuma citação no texto.

BLEVINS, J. **Introduction to Sequential Monte Carlo Methods**. 2023. Disponível em: <<https://jblevins.org/notes/smc-intro>>. Citado na página 37.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to Time Series and Forecasting**. [S.l.]: Springer, 2016. Citado na página 35.

BRONICK, C. J.; LAL, R. Soil structure and management: a review. **GEODERMA**, 2014. Nenhuma citação no texto.

CLIFFORD, D.; PAGENDAMA, D.; BALDOCK, J.; CRESSIEC, N.; FARQUHARSON, R.; FARRELL, L. M. M.; MURRAY, L. Rethinking soil carbon modelling: a stochastic approach to quantify uncertainties. **Environmetrics**, 2014. Citado na página 45.

DAVOUDABADI DANIEL PAGENDAM, C. D. M. J.; BALDOCK, G. W. J. Advanced bayesian approaches for state-space models with a case study on soil carbon sequestration. **Environmental Modelling and Software**, 2020. Citado na página 45.

de Valpine, P.; PACIOREK, C.; TUREK, D.; MICHAUD, N.; ANDERSON-BERGMAN, C.; OBERMEYER, F.; Wehrhahn Cortes, C.; RODRIGUEZ, A.; Temple Lang, D.; PAGANIN, S. **NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling**. [S.l.], 2023. R package version 1.0.0. Disponível em: <<https://cran.r-project.org/package=nimble>>. Citado na página 73.

FAN, X.; GAO, D.; ZHAO, C.; WANG, C.; QU, Y.; ZHANG, J.; BAI, E. Improved model simulation of soil carbon cycling by representing the microbially derived organic carbon pool. **The ISME Journal**, 2021. Nenhuma citação no texto.

KONISKY, D.; HUGHES, L.; KAYLOR, C. Extreme weather events and climate change concern. **Climatic Change** 134, 2015. Citado na página 25.

LIANG, Y.; GOLLANY, H. T.; RICKMAN, R. W.; ALBRECHT, S. L.; FOLLETT, R. F.; WILHELM, W. W.; NOVAK, J. M.; DOUGLAS, J. C. L. Cqestr simulation of management practice effects on long-term soil organic carbon. **Soil & Water Management & Conservation**, 2008. Nenhuma citação no texto.

LUO, Y.; AHLSTRÖM, A.; ALLISON, S. D.; BATJES, N. H.; BROVKIN, V.; CARVALHAIS, N.; CHAPPELL, A.; CIAIS, P.; DAVIDSON, E. A.; FINZI, A.; GEORGIU, K.; GUENET, B.; HARARUK, O.; HARDEN, J. W.; HE, Y.; HOPKINS, F.; JIANG, L.; KOVEN, C.; JACKSON, R. B.; JONES, C. D.; LARA, M. J.; LIANG, J.; MCGUIRE, A. D.; PARTON, W.; PENG, C.; RANDERSON, J. T.; SALAZAR, A.; SIERRA, C. A.; SMITH, M. J.; TIAN, H.; TODD-BROWN, K. E. O.; TORN, M.; GROENIGEN, K. J. van; WANG, Y. P.; WEST, T. O.; WEI, Y.;

- WIEDER, W. R.; XIA, J.; XU, X.; XU, X. feng; ZHOU, T. Toward more realistic projections of soil carbon dynamics by earth system models. **Global Biogeochemical Cycles**, 2015. Nenhuma citação no texto.
- SIERRA, C. A.; HOYT, A. M.; HE, Y.; TRUMBORE, S. E. Soil organic matter persistence as a stochastic process: Age and transit time distributions of carbon in soils. **Global Biogeochemical Cycles**, 2018. Nenhuma citação no texto.
- SIERRA, C. A.; MALGHANI, S.; MÜLLER, M. Model structure and parameter identification of soil organic matter models. **Soil Biology & Biochemistry**, 2015. Nenhuma citação no texto.
- SIERRA, C. A.; MÜLLER, M. A general mathematical framework for representing soil organic matter dynamics. **Concepts & synthesis**, 2015. Citado nas páginas 27 e 31.
- SIERRA, C. A.; MÜLLER, M.; TRUMBORE, S. E. Models of soil organic matter decomposition: the soilr package, version 1.0. **Geoscientific Model Development**, 2012. Nenhuma citação no texto.
- SIERRA, C. A.; TRUMBORE, S. E.; DAVIDSON, E. A.; VICCA, S.; JANSSENS, I. Sensitivity of decomposition rates of soil organic matter with respect to simultaneous changes in temperature and moisture. **Journal of Advances in Modeling Earth Systems**, 2015. Nenhuma citação no texto.
- SIX, J.; CONANT, R. T.; PAUL, E. A.; PAUSTIAN, K. Stabilization mechanisms of soil organic matter: Implications for c-saturation of soils. **Plant and Soil**, 2002. Nenhuma citação no texto.
- SOLLINS, P.; HOMANN, P.; CALDWELL, B. A. Stabilization and destabilization of soil organic matter: mechanisms and controls. **GEODERMA**, 1996. Nenhuma citação no texto.
- STATA. **Gelman–Rubin convergence diagnostic using multiple chains**. 2016. Disponível em: <<https://blog.stata.com/2016/05/26/gelman-rubin-convergence-diagnostic-using-multiple-chains/#:~:text=TheGelman\T1\textendashRubinconvergenceidiagnostic&text=Theconvergenceisassessedby,betweenthesevariancesindicatenonconvergence.>> Citado na página 37.
- TRIANTAFYLLOPOULOS, K. **Bayesian Inference of State Space Models**. [S.l.]: Springer, 2021. Citado nas páginas 35 e 37.
- WANG, X.; JIANG, D.; LANG, X. Future extreme climate changes linked to global warming intensity. **Science Bulletin**, 2017. Citado na página 25.
- WRI-BRASIL. **4 gráficos para entender as emissões de gases de efeito estufa por país e por setor**. 2020. Disponível em: <<https://www.wribrasil.org.br/noticias/4-graficos-para-entender-emissoes-de-gases-de-efeito-estufa-por-pais-e-por-setor>>. Citado na página 25.
- YU, L.; AHRENS, B.; WUTZLER, T.; SCHRUMPF, M.; ZAEHLE, S. Jena soil model (jsm v1.0; revision 1934): a microbial soil organic carbon model integrated with nitrogen and phosphorus processes. **Geoscientific Model Development**, 2019. Nenhuma citação no texto.

